

Chen, Ke; Poon, Ser-Huang

Working Paper

Variance swap premium under stochastic volatility and self-exciting jumps

Manchester Business School Working Paper, No. 634

Provided in Cooperation with:

Manchester Business School, The University of Manchester

Suggested Citation: Chen, Ke; Poon, Ser-Huang (2013) : Variance swap premium under stochastic volatility and self-exciting jumps, Manchester Business School Working Paper, No. 634, The University of Manchester, Manchester Business School, Manchester

This Version is available at:

<https://hdl.handle.net/10419/102387>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Working Paper Series

Variance Swap Premium under Stochastic Volatility and Self-Exciting Jumps

Ke Chen
Ser-Huang Poon

Manchester Business School Working Paper No 634

Manchester Business School

Copyright © 2013, AUTHOR CHEN & POON All rights reserved.
Do not quote or cite without permission from the author.

Manchester Business School
The University of Manchester
Booth Street West
Manchester M15 6PB

+44(0)161 306 1320
<http://www.mbs.ac.uk/cgi/apps/research/working-papers/>

ISSN 0954-7401

The working papers are produced by The University of Manchester - Manchester Business School and are to be circulated for discussion purposes only. Their contents should be considered to be preliminary. The papers are expected to be published in due course, in a revised form and should not be quoted without the authors' permission.

Variance Swap Premium under Stochastic Volatility and Self-Exciting Jumps

Ke Chen and Ser-Huang Poon*

January 13, 2013

*Ke Chen (ke.chen@postgrad.mbs.ac.uk) and Ser-Huang Poon (ser-huang.poon@mbs.ac.uk) are both at the Manchester Business School, University of Manchester. We would like to thank Simon Acomb, Michael Brennan, Peter Carr, Riccardo Rebonato, Stephan Taylor, and participants at Quant Congress London for many helpful comments and suggestions.

Variance Swap Premium under Stochastic Volatility and Self-Exciting Jumps

Abstract

We introduce a stochastic volatility model with self-exciting jump intensity to capture the change in pricing dynamic triggered by big negative stock returns. The stochastic variance and jump intensity, and their risk premium are estimated jointly from daily stock returns and option data over 2007-2010. The model is calibrated to cumulants implied from option prices instead of option prices directly. We find evidence that the time varying jump intensity plays a very important role in the sub-prime crisis and explained most of the risk premium, while in other calmer periods, stochastic variance accounts for most of the risk premium.

JEL Classification: G12, G13

Keywords: Hawkes process, Volatility Surface, Volatility Risk Premium, Jump Risk Premium, Skew Premium

Variance Swap Premium under Stochastic Volatility and Self-Exciting Jumps

1 Introduction

It is well known in the finance literature that volatility of asset return is not constant over time.¹ Models with time varying volatility, e.g. ARCH/GARCH model and Heston model, show some success in fitting both the historical time series of stock returns and option prices (see Engle (1982), Duan (1995), Eraker (1998), and Heston (1993)). More recent papers add jumps to the price and volatility dynamics, and achieve a significantly improved calibration (see Duffie, Pan and Singleton (2000) and Eraker (2004)). In this change of paradigm, from a simple stock returns uncertainty to a risk framework with stochastic volatility and jump processes, one question that immediately followed is “how much premium is associated with each of these new risk factors?” A major concern of investors and traders of vanilla and volatility derivatives is the amount of risk premium they pay for these securities. The objective of this paper is to show the different components of risk premium in implied variance swap prices.

A variance swap is a forward contract on future “realized” variance. Demeterfi, Derman, Kamal and Zou (1999) show that variance swap can be replicated, under the risk neutral measure, by a weighted sum of OTM options across all strikes. Technically, the expected future variance is equal to a weighted integral of implied volatility of all strikes. Carr and Wu (2009) calculate the difference between the realized variance and the variance expected under the risk neutral measure; they find variance risk premium, as reflected by this difference, is economically significant. Earlier, Pan (2002) shows that in a stochastic volatility framework, the volatility risk premium also changes over time.

More recently, Bergomi (2004) shows that any single factor volatility risk model cannot

¹In this paper, unless otherwise specified, we use ‘volatility’ and ‘variance’ interchangeably.

produce the volatility term structure implied by option prices observed in the market; for the minimum, a second volatility risk factor is needed. Subsequently, Bergomi (2008) proposes a model with two volatility risk factors of different timescales, as reflected in their different speeds of mean reversion, and a stock price dynamic that follows a pure diffusion process. This two factor model, when calibrated to the same market data, was shown to produce very different prices for more complex exotic volatility derivatives. By using high frequency stock return data, Todorov (2010) isolates realized jumps from diffusion process, and finds variance risk premium to be abnormally high after a realized jump. The findings in Todorov (2010) suggest that realized jumps in the stock price might be another source of risk premium.

Based on these previous studies, we propose a two-factor stochastic volatility and self exciting jump model. We keep the jump process in the stock price dynamic since it appears to capture the market's price reaction during the crisis period well under both the physical and the risk neutral measures. Unlike the studies mentioned above, we set the jump intensity to follow a self exciting process, as in Hawkes (1971), to allow direct feedback from stock return jump to future jumps. The idea of a time varying jump intensity is not new. The GARJI (Generalised Autoregressive Conditional Jump Intensity) model proposed by Maheu and McCurdy (2004), as shown below, has the jump intensity, λ_t , following an autoregressive process,

$$\begin{aligned}\lambda_t &= \lambda_0 + \rho\lambda_{t-1} + \gamma\xi_{t-1} \\ \xi_{t-1} &= \mathbb{E}[n_{t-1}|r_{t-1}] - \lambda_{t-1}\end{aligned}$$

where $\mathbb{E}[n_{t-1}|r_{t-1}]$ is the expected number of jumps conditional on the observed stock return, ξ_{t-1} is the unexpected jump intensity residual that has an impact on the jump intensity in the next period through the parameter γ , and ρ controls the speed of decay to the baseline level λ_0 . Maheu and McCurdy (2004) fit the GARJI and the GARCH models to a group of individual stock returns and find a much faster reverting speed in the jump intensity above

compared with the volatility decay speed in GARCH.

A more recent paper by Ait-Sahalia, Cacho-Diaz and Laeven (2010) propose a model with both stochastic volatility and mutually exciting jumps for the extremal dependence across international markets, and find strong evidence of self excitement and contagion effect. In contrast to Ait-Sahalia et al. (2010), we focus on univariate feedback effect and we have slightly modified the Hawkes process to allow feedback from only *negative* stock return jump in our proposed Stochastic Volatility Self-Exciting Jump (SVSEJ) model. Two models similar to ours were studied in Carr and Wu (2008) and Fulop, Li and Yu (2012). The difference between these two papers and our study lies in the definition of jump and the impact of jump on the subsequent stock price dynamic. The jump in our model is defined as a rare event that triggers a big reaction in the stock price dynamic. In the calm period, the jump intensity in our model is small; most of the return variations and risk premium will be explained by the stochastic variance. When there is a large negative stock return shock, the market becomes chaotic and the jump intensity in our model is then excited and triggers further large jumps. Our empirical investigation indicates that such a sudden switch in model dynamic is important, especially for explaining the large skewness observed in the option market during the financial crisis. High jump intensity in the crisis period represents bigger uncertainty under the physical measure, which in turns attracts a higher risk premium for the big uncertainty, under the risk neutral measure, if there will be more future big losses.

To estimate the model parameters and the associated risk premium, a joint estimation is carried out using information under both \mathbb{P} and \mathbb{Q} measures, i.e. by calibrating to both historical stock prices and option prices. Instead of using option prices directly, we use the implied cumulants under the \mathbb{Q} measure. Cumulants, similar to moments, are summary statistics of the distributions implied by the volatility surface, and can be used to distinguish the intrinsic properties of the different volatility risk factors. Furthermore, we derive the analytical pricing formulae for the log contract and the variance swap, and use it to assess how the risk premium in implied variance swap price changes in response to a negative stock

return jump. The biggest challenge in the calibration of our model is that it is strongly nonlinear and non-Gaussian; the conventional Kalman filter and its variants are no longer appropriate tools for estimation. So we resort to using particle filter in the implementation, which then leads to another problem that the likelihood is not continuous with respect to the parameters to be estimated. To overcome this last problem, we adopt the Expectation Maximisation (EM) algorithm since it requires the least amount of computational effort and it converges quickly.

The rest of this paper is organized as follows. We start with the introduction of our Stochastic Volatility with Self Exciting Jump model in Section 2. We explain the joint estimation method in Section 3. Section 4 discusses the empirical finding and the time varying risk premium of variance swap. Section 5 concludes. All the technical details of the derivations are included in the Appendix.

2 Stochastic Volatility Self-Exciting Jump Model

While stochastic volatility model is well established in the finance literature, strong empirical evidence from market observed stock and option prices suggest there are jumps in both stock and volatility dynamics. Inspired by the findings in Todorov (2010) that the variance risk premium is abnormally high after a realized jump, we introduce another risk factor that is associated with a large negative shock in stock returns as follows:

$$\begin{aligned}
\frac{dS_t}{S_t} &= udt + \sqrt{v_t}dW_{1,t} + \int_{\mathbb{R} \times \mathbb{R}^+} (e^{J_x} - 1) [\mu(dJ_x, dJ_\lambda, dt) - \pi_x(J_x)dJ_x\lambda_t dt] \\
dv_t &= \kappa_v(\theta_v - v_t)dt + \sigma_v\sqrt{v_t}dW_{2,t} \\
d\lambda_t &= \kappa_\lambda(\theta_\lambda - \lambda_t)dt + \sigma_\lambda\sqrt{\lambda_t}dW_{3,t} + \int_{\mathbb{R} \times \mathbb{R}^+} 1_{J_x < 0} J_\lambda \mu(dJ_x, dJ_\lambda, dt) \\
\langle dW_{1,t}, dW_{2,t} \rangle &= \rho dt
\end{aligned} \tag{1}$$

where $\mu(dJ_x, dJ_\lambda, dt)$ denotes the random measures of a two dimensional jump in price and jump intensity; there is no correlation between $W_{2,t}$ and $W_{3,t}$, and between $W_{1,t}$ and $W_{3,t}$. We assume that jumps in the stock price, J_x , and jumps in the jump intensity, J_λ , are compound Poisson jumps, where $J_x \sim \pi_x(J_x)$ is Double Exponential(η_u, η_d, p) and $J_\lambda \sim \pi_\lambda(J_\lambda)$ is Exponential(η). We further assume that jump size and jump timing are independent, and that the jump sizes in these two processes are independent. Therefore, the compensator of the random measures can be written as

$$\begin{aligned}
\pi(dJ_x, dJ_\lambda)dJ_x dJ_\lambda \lambda_t dt &= \pi_x(J_x)\pi_\lambda(J_\lambda)dJ_x dJ_\lambda \lambda_t dt \\
\int_{\mathbb{R} \times \mathbb{R}^+} (e^{J_x} - 1)\pi(J_x, J_\lambda)dJ_\lambda dJ_x \lambda_t dt &= \int \left(\int_{\mathbb{R} \times \mathbb{R}^+} (e^{J_x} - 1)\pi(J_x)dJ_x \right) \pi(J_\lambda)dJ_\lambda \lambda_t dt \\
&= \left(\int_{\mathbb{R} \times \mathbb{R}^+} (e^{J_x} - 1)\pi(J_x)dJ_x \right) \lambda_t dt \int \pi(J_\lambda)dJ_\lambda \\
&= \int_{\mathbb{R} \times \mathbb{R}^+} (e^{J_x} - 1)\pi(J_x)dJ_x \lambda_t dt
\end{aligned}$$

as shown in equation (1). It should be noted that κ_λ is not the true mean reverting speed, because the innovation term is not compensated and hence $d\lambda_t$ is not a martingale. If we adjust the random measure by its compensator, we have $\varkappa_\lambda = \kappa_\lambda - (1 - p)\eta$ and $\vartheta_\lambda = \frac{\kappa_\lambda \theta_\lambda}{\varkappa_\lambda}$, which are the mean reversion speed and the long run mean of the jump intensity respectively.

We have explained in the introduction that our SVSEJ model has features resembling, but not the same as, the jump models in e.g. Maheu and McCurdy (2004), Ait-Sahalia et al. (2010) and Todorov (2010). The most important unique feature of our SVSEJ model is that the jump in the jump intensity is triggered by a negative jump in stock return. A positive jump in the stock return will not cause the jump intensity to jump. This choice is empirically motivated. During our pilot studies, we have estimated a model similar to (1) using a long period of daily stock returns but with a constant jump intensity. The estimation is done via MCMC (Markov Chain Monte Carlo). The results show that the standard error of upward jump size is very big compared with the negative jump counterpart. Based on the samples drawn from the MCMC, we calculated probability of subsequent stock return jumps

following a stock return jump. We find that the probability of jump is almost constant, and below the overall average, after a positive stock return jump. In sharp contrast, after a negative stock return jump, the probability of jump doubled and decayed monotonically to the overall average after about 10 days. We interpreted this as a clear evidence of time varying jump intensity and self-excitement triggered by a negative stock return jump, and hence created the model in (1).

2.1 Dynamics under \mathbb{Q} Measure

2.1.1 Measure Change

Since our objective is to identify the risk premium, we need the dynamics under the \mathbb{Q} measure. The measure change from \mathbb{P} to \mathbb{Q} is,

$$\begin{aligned} \frac{d\mathbb{Q}}{d\mathbb{P}} \Big|_{\mathcal{F}_t} &= \mathcal{E} \left(- \int_0^t \gamma_v \sqrt{v_s} dW_{2,s} \right) \times \mathcal{E} \left(- \int_0^t \gamma_\lambda \sqrt{\lambda_s} dW_{3,s} \right) \\ &\times \mathcal{E} \left(\int_0^t (e^{-\gamma_{J_x} J_x - \gamma_{J_\lambda} J_\lambda} - 1) (\mu(dJ_x, dJ_\lambda, dt) - \nu(dJ_x, dJ_\lambda) \lambda_t dt) \right) \end{aligned}$$

where \mathcal{E} is the stochastic exponential.

For the Brownian motion term, we shift the mean and maintain the affine form of the model. Therefore, $dW_{2,t}^{\mathbb{Q}} = \gamma_v \sqrt{v_t} dt + dW_{2,t}^{\mathbb{P}}$ and $dW_{3,t}^{\mathbb{Q}} = \gamma_\lambda \sqrt{\lambda_t} dt + dW_{3,t}^{\mathbb{P}}$ are Brownian motions under \mathbb{Q} measure. For the compound Poisson jump, we change the distribution of jump size by a constant shift. The distribution for jump size in the stock return becomes $\pi_x^{\mathbb{Q}}(J_x) = \text{Double Exponential}(\eta_u^{\mathbb{Q}}, \eta_j^{\mathbb{Q}}, p^{\mathbb{Q}})$, where $\eta_u^{\mathbb{Q}} = \frac{1}{1+\gamma_{J_x} \eta_u^{\mathbb{P}}}$, $\eta_d^{\mathbb{Q}} = \frac{1}{1-\gamma_{J_x} \eta_d^{\mathbb{P}}}$, and $p^{\mathbb{Q}} = p^{\mathbb{P}} \left(p^{\mathbb{P}} + (1-p^{\mathbb{P}}) \frac{1+\gamma_{J_x} \eta_u^{\mathbb{P}}}{1-\gamma_{J_x} \eta_d^{\mathbb{P}}} \right)^{-1}$. The distribution of jump size in the jump intensity becomes $\pi_\lambda^{\mathbb{Q}}(J_\lambda) = \text{Exponential}(\eta^{\mathbb{Q}})$, where $\eta^{\mathbb{Q}} = \frac{1}{1+\gamma_{J_\lambda} \eta^{\mathbb{P}}}$.² Hence the dynamics of our SVSEJ model under \mathbb{Q} measure is as follows

²For the measure change of the exponential distribution, we have

$$\begin{aligned}
\frac{dF_t}{F_t} &= \sqrt{v_t}dW_{1,t}^{\mathbb{Q}} + \int_{\mathbb{R} \times \mathbb{R}^+} (e^{J_x} - 1)(\mu^{\mathbb{Q}}(dJ_x, dJ_\lambda, dt) - \pi_x^{\mathbb{Q}}(J_x)dJ_x\lambda_t dt) \\
dv_t &= \kappa_v^{\mathbb{Q}}(\theta_v^{\mathbb{Q}} - v_t)dt + \sigma_v\sqrt{v_t}dW_{2,t}^{\mathbb{Q}} \\
d\lambda_t &= +\kappa_\lambda^{\mathbb{Q}}(\theta_\lambda^{\mathbb{Q}} - \lambda_t)dt + \sigma_\lambda\sqrt{\lambda_t}dW_{3,t}^{\mathbb{Q}} + \int_{\mathbb{R} \times \mathbb{R}^+} 1_{J_x < 0} J_\lambda \mu^{\mathbb{Q}}(dJ_x, dJ_\lambda, dt) \\
\langle dW_{1,t}, dW_{2,t} \rangle &= \rho dt
\end{aligned} \tag{2}$$

where $\kappa_v^{\mathbb{Q}} = \kappa_v + \gamma_v\sigma_v$, $\theta_v^{\mathbb{Q}} = \frac{\kappa_v\theta_v}{\kappa_v^{\mathbb{Q}}}$, $\kappa_\lambda^{\mathbb{Q}} = \kappa_\lambda + \gamma_\lambda\sigma_\lambda$, and $\theta_\lambda^{\mathbb{Q}} = \frac{\kappa_\lambda\theta_\lambda}{\kappa_\lambda^{\mathbb{Q}}}$. Since the random measure in the jump intensity equation is not compensated, so the risk premium related to γ_{J_λ} does not appear in equation (2). But for the true mean reverting speed and the long run mean of jump intensity, we have $\varkappa_\lambda^{\mathbb{Q}} = \kappa_\lambda + \gamma_\lambda\sigma_\lambda - \gamma_{J_\lambda}(1-p)\eta$ and $\vartheta_\lambda^{\mathbb{Q}} = \frac{\kappa_\lambda\theta_\lambda}{\varkappa_\lambda^{\mathbb{Q}}}$, in which γ_{J_λ} has a direct impact on the model parameters under the risk neutral measure.

2.1.2 Characteristic Function of Log Price

Our proposed SVSEJ model is designed to be affine, so we are able to use the transform approach from Heston (1993) to price any claim contingent on the final state of the underlying. Therefore, we can calibrate our model to fit European option prices listed in the market.

$$\begin{aligned}
\frac{1}{\eta^{\mathbb{Q}}}e^{-x/\eta^{\mathbb{Q}}} &= A\frac{1}{\eta^P}e^{-x/\eta^P}e^{-\gamma x} = \frac{A}{\eta^P}e^{-x(\frac{1}{\eta^P} + \gamma)} \\
\eta^{\mathbb{Q}} &= \frac{1}{\frac{1}{\eta^P} + \gamma} = \eta^P \frac{1}{1 + \gamma\eta^P}
\end{aligned}$$

Given $\eta^{\mathbb{Q}}$, we can back out A

$$A = \frac{\eta^P}{\eta^{\mathbb{Q}}} = 1 + \gamma\eta^P$$

First, write the moment generating function (MGF), $G(\cdot)$, as

$$\begin{aligned} G(\omega, x_t, v_t, \lambda_t, t, T) &= \phi(-i\omega, x_t, v_t, \lambda_t, t, T) \\ \phi(\omega, x_t, v_t, \lambda_t, t, T) &= \mathbb{E}[e^{i\omega x_T} | x_t, v_t, \lambda_t] = \int_{-\infty}^{\infty} e^{i\omega x_T} f(x_T | x_t, v_t, \lambda_t) dx_T \end{aligned}$$

where $\phi(\cdot)$ is the characteristic function, $x_T = \ln F_T$ is the log price, which is driven by the stochastic process in (2).

By Feynman-Kac theorem, the expectation of any function of x_T , can be determined by a PIDE (Partial Integral Differential Equation). Following the approach by Pan (2002) and Sepp (2003), we derive the PIDE for the moment generating function (MGF) of x_T for our SVSEJ model,

$$\begin{aligned} 0 &= G_t - \left(\frac{1}{2}v + \lambda_t \mathbb{E}[e^{J_x} - 1]\right)G_x + \frac{1}{2}v_t G_{xx} + \kappa(\theta - v_t)G_v \\ &\quad + \frac{1}{2}\sigma^2 v_t G_{vv} + \rho\sigma v_t G_{xv} + \kappa_\lambda(\theta_\lambda - \lambda_t)G_\lambda \\ &\quad + \lambda_t \int [G(\omega, x + J_x, v_t, \lambda_t + 1_{J < 0} J_\lambda, t, T) - G(\omega, x, v_t, \lambda_t, t, T)] \nu(dJ_x, dJ_\lambda) \quad (3) \end{aligned}$$

with initial condition at time T , $G_T = e^{x_T \omega}$. As shown in Appendix A, there is no analytical solution for the MGF G in (3) because of the self exciting jump part.

2.2 Model Discussion

Our proposed SVSEJ model is a two-factor model. The single factor model is well studied in the literature and some single factor models provide good fit for both stock price movement and option volatility surface (see Pan (2002) and Eraker (2004)). However, there are a few fundamental problems with the single factor models. First, the volatility and skewness generated by the single factor models are one-to-one correspondence for the same maturity,³ which

³The volatility surface is a function of the state variables. For the single factor model, there is only one state variable. If the value at any point on the volatility surface is known, we can back out the state of the

is not the case in the market data. Figure 1 plots the 3rd cumulant (unnormalised skewness) against the 2nd cumulant (variance) implied by options with two months to maturity. It is clear that implied variance and skewness are not one-to-one correspondence especially when the variance is big. In another words, there are more than one factor controlling the market price dynamics. Second, the term structure of implied variance is fixed in the single factor models. This means that given the price of a variance swap maturing in e.g. one month, prices for variance swaps of other maturities will be also known. The implication is that the correlation term structure of variance derivatives will always be one. But this is not true with the real market data.

Bergomi (2005) shows that more than one volatility risk factors are needed to fit the variance dynamics. Cont and Da Fonseca (2002) also claim that a multi-factor model is necessary to capture the volatility surface dynamics. Todorov (2010) finds risk premium for variance swap increased dramatically after a realized jump has occurred. This suggests that one of the risk factors could be strongly associated with stock return jumps. To accommodate all the empirically observed market features, we construct the SVSEJ model and link the stock return jump to future jump intensity in a self exciting manner. In our SVSEJ model, a jump in the jump intensity not only causes jump in variance, but also introduces a big skewness in the distribution implied by prices of short maturity options. The large implied skewness, which we witnessed during the crisis, means that there was a huge premium for deep OTM put options for protection from further big future losses. Mathematically, it is very difficult to generate such a large skewness from the continuous diffusion part of the stock price dynamic, because skewness generated by the continuous part is in the order of $(T - t)^2$. When time to maturity, $(T - t)$, is small, the diffusion component cannot provide enough skewness to match the observed skewness in the implied volatility surface. On the other hand, the skewness generated by jumps is proportional to $(T - t)$. If we allow the jump intensity itself to jump, we effectively introduced jumps in both variance and skewness.

factor, and use it to calculate other values of the volatility surface such as skewness.

Since jump in the jump intensity also causes a jump in the total variance, the dynamics of stochastic variance in our model is kept simple. We introduce another Gaussian term, $\sigma_\lambda \sqrt{\lambda_t} dW_{3,t}$, in the dynamic of the jump intensity. The contribution of this term is small when the market is calm. The more important point is that when a jump occurs, the jump intensity is excited; this Brownian motion term can help to explain the big fluctuations of the jump intensity afterwards.

In similar work, Carr and Wu (2008) and Fulop et al. (2012) use Gamma process to model the jump dynamic and restrict the jump size in stock return to be proportional to the jump size in the jump intensity. Compared with the compound Poisson process used in our model, Gamma process has jumps in any time interval and the number of jumps is infinite. So the jump intensity will jump proportionally to the jump size in stock returns everyday in their models. However, we believe that market responds to past returns is strongly nonlinear depending on the size of the loss. To make a clear distinction between large and small stock returns, and between positive and negative stock returns, we therefore chose the compound Poisson process, which has a finite number of jumps, for modelling rare jumps. In our SVSEJ model, the jump size in stock return and the jump size in the jump intensity are independent. We believe that changes in stock returns and jump intensity may be related, but not of any simple functional form due to many reasons such as liquidity and trading rules. When there is a big fluctuation in the option market, it is not necessary that we can observe a corresponding movement in daily or weekly stock returns. One interesting example is the flash crash on May 6, 2010, when the Dow Jones Industrial Average suffered the biggest one day loss in history, but recovered within minutes. The option market responded to the crash, and VIX increased by 27.5% on that day and continued to increase by another 22% on the following day. In contrast, the daily stock return was just -3.3% on that day and -1.3% on the following day.

Compared with the GARJI model in Maheu and McCurdy (2004), the jump intensity in our model is driven by stock return jumps, and not by the expected jump number. Moreover,

the jump intensity in our model only responds to negative jumps in stock returns and the size of the response is random. In our model, both jump and volatility processes are stochastic, while in the GARJI model they are both deterministic.

3 Joint Estimation under Two Measure

3.1 Cumulants of Log Return

A distribution can be fully represented by its cumulants (or moments). The result of Edgeworth expansion suggests that the lower cumulants are the most important in explaining the shape of the distribution. Here, we choose to calibrate to the second and the third cumulants implied by prices of options with two and three months to maturity every day.⁴ To use cumulants in the estimation, we need to derive the cumulant implied by the model parameters and by the market option prices. Since our mode is affine, the cumulant generating function, Υ , of log returns is an affine function of the state variable,⁵

$$\begin{aligned}\Upsilon(\omega, v_t, \lambda_t, t, T) &= \ln G(\omega, x_t, v_t, \lambda_t, t, T) - \ln e^{x_t\omega} \\ &= A(\omega, t, T) + B(\omega, t, T)v_t + C(\omega, t, T)\lambda_t\end{aligned}$$

So the i th cumulant of log return can be calculated using,

$$\kappa_i = \left. \frac{\partial^i \Upsilon(\omega, v_t, \lambda_t, t, T)}{\partial \omega^i} \right|_{\omega=0}$$

⁴Under the risk neutral measure, the first cumulant turns out to be the same as the integrated variance scaled by $2T$, while the second cumulant is the total variance of S_T . The first cumulant is ignored in the estimation because the dynamics of the first cumulant is very similar to that of the second cumulant and does not carry any extra information. The fair price of the variance swap is actually a constant times the price of log contract, which is the first cumulant.

⁵ Υ is the cumulant generating function for the log return, not log price, so we need to subtract the term, $\ln e^{x_t\omega}$.

which gives,

$$\kappa_i = \left. \frac{\partial^i A(\omega, t, T)}{\partial \omega^i} \right|_{\omega=0} + \left. \frac{\partial^i B(\omega, t, T)}{\partial \omega^i} \right|_{\omega=0} v_t + \left. \frac{\partial^i C(\omega, t, T)}{\partial \omega^i} \right|_{\omega=0} \lambda_t \quad (4)$$

We show in Appendix B how to derive the first derivatives ($i = 1$) of A , B , and C with respect to ω for the first cumulant κ_1 .⁶ The higher order derivatives with respect to ω for the higher order cumulants can be derived similarly. The results in Appendix B are substituted into equation (4) to produce the second and third cumulants, κ_2 and κ_3 .

High skewness (or third cumulant) implied by short maturity options, such as that observed during the crisis, can only be explained by jumps. To understand this, apply Taylor expansion on κ_2 and κ_3 and let $(T - t) \rightarrow 0$ for short maturity options. Then

$$\begin{aligned} \kappa_2 &\approx v_t(T - t) + \mathbb{E}[J_x^2] \lambda_t (T - t) + O((T - t)^2) \\ \kappa_3 &\approx \mathbb{E}[J_x^3] \lambda_t (T - t) + O((T - t)^2) \end{aligned}$$

It is clear that, for short maturity options, the second cumulant κ_2 (or variance) is due to the diffusion and the jump parts, both in the order of $(T - t)$. However, the third cumulant κ_3 (unnormalised skewness) is completely dominated by the jump component only.

Equation (4) provides the model implied cumulants. The cumulants implied by market data can be constructed from OTM option prices as follows:

⁶Although we do not use the first cumulant in the estimation. The first cumulant derived in Appendix B is used later to determine the variance swap risk premium.

$$\begin{aligned}
\kappa_1^* &= M_1 \\
\kappa_2^* &= M_2 - M_1^2 \\
\kappa_3^* &= M_3 - 3M_2M_1 + 2M_1^3
\end{aligned}$$

where M_1 , M_2 and M_3 are the first three moments of $\ln\left(\frac{F_T}{F_t}\right)$ as follows:

$$\begin{aligned}
M_1 &= -\int_0^{F_t} \frac{1}{K^2} P(K) dK - \int_{F_t}^{\infty} \frac{1}{K^2} C(K) dK \\
M_2 &= \int_0^{F_t} \frac{2}{K^2} \left(1 - \ln \frac{K}{F_t}\right) P(K) dK + \int_{F_t}^{\infty} \left(1 - \ln \frac{K}{F_t}\right) C(K) dK \\
M_3 &= \int_0^{F_t} \frac{3}{K^2} \left(2 \ln \frac{K}{F_t} - \ln^2 \frac{K}{F_t}\right) P(K) dK + \int_{F_t}^{\infty} \frac{3}{K^2} \left(2 \ln \frac{K}{F_t} - \ln^2 \frac{K}{F_t}\right) C(K) dK
\end{aligned}$$

and $P(K)$ and $C(K)$ are the (interpolated and extrapolated) market put and call SPX option prices at strike K (see Section 4.1). Details of the derivations following Carr and Madan (2002) can be found in Appendix C.

3.2 Estimating Hidden State Model

The dynamics of our SVSEJ model has two latent processes, v_t and λ_t . We can rewrite the model into a state-space form and use filtering technique to back out the latent variables. For a sample period of length T , the observation on each day is $y_t = \{r_t, \kappa_t^j(\tau_j)\}$, where $r_t = \ln(S_{t+1}/S_t)$ is the log return and $\kappa_t^i(\tau_j)$ is the i th cumulant of log return implied from time t prices of j th maturity option as described in the previous section. Therefore, the *measurement equations* can be written as,

$$r_t = \left(u - \frac{v_{t-1}^2}{2} - \lambda_{t-1} \left(\frac{p}{1-\eta_u} + \frac{1-p}{1+\eta_d} - 1 \right) \right) \Delta t + \sqrt{v_{t-1}} Z_{1,t} + 1_{N_t > 0} J_{x,t} \quad (5)$$

$$\kappa_t^i(\tau_j) = A_{i,j} + B_{i,j} v_t + C_{i,j} \lambda_t + \sigma_{i,j} \epsilon_{i,j,t} \quad (6)$$

where jump timing $N_t \sim \text{Poisson}(\lambda_t \Delta t)$, $1_{N_t > 0}$ is an indicator function which takes value 1 when $N_t > 0$ or 0 otherwise, jump size $J_{x,t} \sim \pi_x(J_x)$ has a double exponential distribution, $Z_{1,t}$ is a standard normal random variable, and $\epsilon_{i,j,t}$ (for $i = 2, 3$, $j = 2, 3$) are four uncorrelated standard normal random variables.⁷ The cumulants used in the equation (6) are the second and the third cumulants of the 2-month and 3-month implied distributions. More detailed discussion regarding the choice of cumulants is in Section 4. As mentioned in the previous section, $A_{i,j}$, $B_{i,j}$, and $C_{i,j}$ can be derived in close form as shown in Appendix B.

Based on the Euler approximation, the *transition equations* that drive the set of latent processes, $x_t = \{v_t, \lambda_t\}$ are

$$v_t = v_{t-1} + \kappa_v (\theta_v - v_{t-1}) \Delta t + \sigma_v \sqrt{v_{t-1}} \left(\rho Z_{1,t} + \sqrt{1-\rho^2} Z_{2,t} \right) \quad (7)$$

$$\lambda_t = \lambda_{t-1} + \kappa_\lambda (\theta_\lambda - \lambda_{t-1}) \Delta t + \sigma_\lambda \sqrt{\lambda_{t-1}} Z_{3,t} + 1_{N_t > 0, J_{x,t} < 0} J_{\lambda,t} \quad (8)$$

where jump size $J_{\lambda,t} \sim \pi_\lambda(J_\lambda)$ has an exponential distribution, $Z_{2,t}$ and $Z_{3,t}$ are uncorrelated standard normal random variables, and $1_{N_t > 0, J_{x,t} < 0}$ is an indicator function which takes the value 1 when $N_t > 0$ and $J_{x,t} < 0$, or the value 0 otherwise.

⁷The integral for the jump term is

$$\int_{\mathbb{R} \times \mathbb{R}^+} (e^{J_x} - 1) (\mu(dJ_x, dJ_\lambda, dt) - \pi_x(J_x) dJ_x \lambda_t dt) = 1_{N_t > 0} J_{x,t} - \lambda_t dt \int_{\mathbb{R}} (e^{J_x} - 1) \pi_x(J_x) dJ_x,$$

and

$$\int_{\mathbb{R}} (e^{J_x} - 1) \pi_x(J_x) dJ_x = \left(\frac{p}{1-\eta_u} + \frac{1-p}{1+\eta_d} - 1 \right)$$

appears as a compensator in the Δt term in equation (5).

3.3 Filtering Problem

We cannot use Kalman filter or its variants to estimate the state-space model in Section 3.2 as the model is highly non-Gaussian and nonlinear. Instead, a particle filter is used to approximate the predictive and filtered distributions. There is an extensive literature on how to make particle filter efficient and accurate. The readers are referred to the book by Doucet et al. (2001) for further details.

Compared with the other applications of particle filtering, the biggest challenge in estimating the model in Section 3.2 is due to the fact that jumps in both the observation and the transition equations are defined to be rare but have a big influence on how the observed and latent variables behave. To make the filtering process more efficient in the standard Sequential Importance Resampling (SIR) framework (see Gordon et al. (1993)), the jump timing $N_t = \{0, 1\}$ is initially drawn from a Bernoulli distribution.⁸ The step-by-step procedure of the filtering process is described below:

1. Initialisation: with initial values $\Theta = \{\mu, \kappa_v, \theta_v, \sigma_v, \rho, \kappa_\lambda, \theta_\lambda, \sigma_\lambda, \eta, \eta_u, \eta_d, p, \gamma_v, \gamma_\lambda, \gamma_{J_x}, \gamma_{J_\lambda}, \sigma_{22}, \sigma_{23}, \sigma_{32}, \sigma_{33}\}$, set $t = 0$ and $x_0 = \{v_0, \lambda_0\}$, where $v_0 = \theta_v$ and $\lambda_0 = \theta_\lambda$.
2. Prediction: starting from $t = 1$, we have for $m = 1 : M$, where $M = 2000$ is the number of particles:
 - (a) Prediction of jump: Since there is maximum one jump per day, $N_t^m \in \{0, 1\}$; we draw M samples of N_t^m from a Bernoulli distribution with $\pi = 0.5$. For each $N_t^m = 1$, we draw $J_{x,t}^m$ from the double exponential distribution (η_u, η_d, p) , and for each $J_{x,t}^m < 0$, we draw $J_{\lambda,t}^m$ from the exponential distribution (η) .
 - (b) Prediction of latent process:
 - i. Calculate $Z_{1,t}^m$ from (5) using $J_{x,t}^m$ from Step (a) and the observed value of r_t ;

⁸Random variable from the Bernoulli distribution only takes value of 1 or 0. Here we ignore the possibility of multiple jumps on each day. The result obtained later shows that the jump intensity is small enough to justify this assumption.

- ii. Simulate M samples of $Z_{2,t}^m$ and use it, together with $Z_{1,t}^m$ from Step i, to calculate v_t^m from equations (7);
 - iii. Simulate M samples of $Z_{3,t}^m$ and use it, together with $J_{\lambda,t}^m$ from Step (a), to calculate λ_t^m from equation (8).
- (c) Calculate the importance weight for the jump samples:

$$\pi_t^m = \frac{\text{Poisson}(N_t^m, \lambda_t^m)}{\text{Bernoulli}(N_t^m, \pi = 0.5)} \quad (9)$$

In equation (9), the numerator is the probability of jump in the model, the denominator is the probability of jump used in the sampling process; π_t^m is equivalent to the likelihood ratio in the Monte Carlo method.⁹

3. Updating:

- (a) Calculate $\epsilon_{i,j,t}^m$ from observed values of $\kappa_t^i(\tau_j)$ from equation (6) for $i = 2, 3$ and $j = 2, 3$ using the parameter values, Θ , and particle values of the latent variables in Step 2.
- (b) Calculate the importance and normalised weights of the particles using

$$w_t^m = \pi_t^m \times \Phi(Z_{1,t}^m) \times \prod_{i,j} \Phi(\epsilon_{i,j}^m)$$

$$\bar{w}_t^m = \frac{w_t^m}{\sum_{m=1}^M w_t^m}$$

- (c) Resample the particles $\tilde{J}_{x,t}^m$ from $J_{x,t}^m$, $\tilde{J}_{\lambda,t}^m$ from $J_{\lambda,t}^m$, \tilde{v}_t^m from v_t^m , and $\tilde{\lambda}_t^m$ from λ_t^m according to the normalized weight, \bar{w}_t^m .

4. Let $t = t + 1$ and start again from Step 2 until $t = T$.

⁹The reader is reminded here that π_t^m and π denote probabilities, whereas π_x and π_λ denote distributions. In equation (9), π is set equal to 0.5 to represent a naive prior.

3.4 Estimation Method

The estimation is done by Maximum Likelihood Estimation (MLE). Given the filtered samples, the likelihood is given by,

$$\begin{aligned}
 L[\Theta|y_{1:T}] &= \prod_{t=1}^T p(y_t|y_{1:t-1}, \Theta) \\
 &= \prod_{t=1}^T \int p(y_t|x_t, \Theta)p(x_t|y_{1:t-1}, \Theta)dx_t \\
 &\approx \prod_{t=1}^T \sum_{m=1}^M w_t^m
 \end{aligned}$$

where w_t^m is the importance weight from the previous subsection. The availability of the importance weights has avoided the need to evaluate (conditional) likelihood of y_t and x_t of which closed forms are not available.

However, given the nature of Monte Carlo, the approximated likelihood function based on Filtered process is not continuous.¹⁰ Therefore, we cannot apply classical optimisation algorithm. Malik and Pitt (2011) proposed a method to approximate the likelihood for a state space model, which is continuous with respect to the change of the parameters values. However, such a method is difficult to use for multidimensional problem, which is our case here. Instead, we adopt the Expectation Maximisation (EM) together with smoothed particles for the estimation. Compared with direct MLE, EM method requires more computation, because it is an iterative procedure and it relies on the smoothed particle paths, which is computationally expensive to calculate. Despite these complications, other approaches would require even more computation efforts. Here we explained the EM procedure, which mainly follows Hurzeler and Kunsch (1998), for use with the particle filter and smoother.

1. Filtering: Using the steps described in Section 3.3 and initial parameter values Θ to

¹⁰The discontinuity mainly comes from the resampling step, in which one has to sample from the distribution represented by particles associated with different weights

generate M filtered and re-sampled particles $\tilde{x}_t^m = \{\tilde{v}_t^m, \tilde{\lambda}_t^m\}$ for $t = 1, \dots, T$.

2. Smoothing: produce the smoothed path $\hat{x}_{1:T}$ backward, for $t = T, \dots, 1$, based on the filtered samples from Step 1. We explained the step-by-step smoothing procedures below:

- (a) Start at $t = T$ with the special case, $p(\hat{x}_T|y_{1:T}, \Theta) \sim p(\tilde{x}_T|y_{1:T}, \Theta)$. For this special case, the smoothed particles are the same as the filtered particles, i.e. $\hat{x}_T^m = \tilde{x}_T^m$ for $m = 1 : M$, $M = 2000$.
- (b) Set $t = T - 1$ and $t + 1 = T$, sample \hat{x}_t from

$$p(\hat{x}_t|\hat{x}_{t+1}^m, y_{1:T}, \Theta) \propto p(\hat{x}_t|y_{1:t}, \Theta)p(\hat{x}_{t+1}^m|\tilde{x}_t, \Theta)$$

Here, we use the filtered particles $\{\tilde{x}_t^{\tilde{m}}\}_{\tilde{m}=1:M}$ as the candidates for \hat{x}_t . In this case, $p(\hat{x}_t|y_{1:t}, \Theta)$ becomes a constant, $\frac{1}{M}$, since the likelihood of the filtered particles after resampling are all equal. Then we only need to calculate

$$p(\hat{x}_t|\hat{x}_{t+1}^m, y_{1:T}, \Theta) \propto p(\hat{x}_{t+1}^m|\tilde{x}_t^{\tilde{m}}, \Theta)$$

where m is the counter for the smoothed particles, and \tilde{m} is the counter for the filtered particles. In other cases where there is no confusion over which particle is used, m is used generally as the counter.

- (c) For each m , the probability $p(\hat{x}_{t+1}^m|\tilde{x}_t^{\tilde{m}}, \Theta)$ for a given \tilde{m} is calculated from equations (7) and (8) as follows:
 - i. Calculate $Z_{1,t}^{\tilde{m}}$ from (5) using $\tilde{J}_{x,t}^{\tilde{m}}$ and the observed value of r_t ;
 - ii. Use $Z_{1,t}^{\tilde{m}}$ from Step i, $\tilde{v}_t^{\tilde{m}}$ and \hat{v}_{t+1}^m to calculate $Z_{2,t}^{\tilde{m}}$ from equations (7).
 - iii. Use $\tilde{J}_{\lambda,t}^{\tilde{m}}$, $\tilde{\lambda}_t^{\tilde{m}}$ and $\hat{\lambda}_{t+1}^m$ to calculate of $Z_{3,t}^{\tilde{m}}$ from equation (8).

iv. Calculate the likelihood values below for each \tilde{m} ,

$$p(\hat{x}_{t+1}^m | \tilde{x}_t^{\tilde{m}}, \Theta) = \Phi(Z_{2,t}^{\tilde{m}}) \Phi(Z_{3,t}^{\tilde{m}}) \Phi(\tilde{J}_{\lambda,t}^{\tilde{m}}) \Phi(\tilde{J}_{x,t}^{\tilde{m}}) \quad (10)$$

v. Repeat Step i to Step iv for $\tilde{m} = 1 : M$.

(d) The smoothed particle, \hat{x}_t^m , is re-sampled from $\{\tilde{x}_t^{\tilde{m}}\}_{\tilde{m}=1:M}$ using normalised probability

$$\frac{\pi^{\tilde{m}}}{\sum_{\tilde{m}=1}^M \pi^{\tilde{m}}}$$

where $\pi^{\tilde{m}} = p(\hat{x}_{t+1}^m | \tilde{x}_t^{\tilde{m}}, \Theta)$ from equation (10).

(e) Repeat step (c) and (d) for $m = 1 : M$.¹¹

(f) Set $t = t - 1$ and repeat from step (c) till $t = 1$.

3. Expectation: Calculate the expectation of log likelihood based on smoothed samples.

$$\begin{aligned} Q(\Theta^* | \Theta) &= \mathbb{E}_{x_{1:T}, y_{1:T}, \Theta} [\ln p(x_{1:T}, y_{1:T} | \Theta^*)] \\ &\approx \frac{1}{M} \sum_{t=1}^T \sum_{i=1}^M [\ln p(\hat{x}_t^m | \hat{x}_{t-1}^m, \Theta^*) + \ln p(y_t | \hat{x}_t^m, \Theta^*)] \end{aligned}$$

where $\Theta^* = \Theta$ and $\{\hat{x}_t^m\}_{t=1}^T$ are smoothed samples from Step 2. In the calculation of the log likelihood, $p(\hat{x}_t^m | \hat{x}_{t-1}^m, \Theta^*) = \Phi(\hat{Z}_{2,t}^m) \Phi(\hat{Z}_{3,t}^m) \Phi(\hat{J}_{\lambda,t}^m) \Phi(\hat{J}_{x,t}^m)$ and $p(y_t | \hat{x}_t^m, \Theta^*) = \Phi(\hat{Z}_{1,t}^m) \times \prod_{i,j} \Phi(\epsilon_{i,j}^m)$, both calculated using the smoothed samples.

4. Maximisation: Maximise $Q(\Theta^* | \Theta)$ with respect to Θ^* , and let $\Theta = \Theta^*$. Then use the new Θ to repeat from Step 1 until the parameters converged.

¹¹For the smoothing step, the computation is in the order of M^2 , whereas the computation for the filtering step previously is in the order of M .

4 Empirical Results

4.1 Data Description

The data used in this study include S&P 500 daily closing price and the prices of the corresponding European options with maturities up to six months. The sample period is from October 2007 to October 2010 which includes the important period of sub-prime crisis. We hope to explain how the market behaves in such a turbulent period. Vanilla option prices are used to construct the cumulants of log return. On the other hand, as shown in equation (6), to calculate cumulants, we need option prices with strike price ranges from 0 to infinity, which we approximate by interpolating and extrapolating the market option prices using the SVI function in Gatheral (2004). As shown in Appendix C, the higher the moments, the more weights are placed on the deep OTM options and the bigger the impact of extrapolation errors. Taking into account of these pitfalls, we chose to calibrate only the second and the third cumulants, which correspond to the second and third centered moments.

One important source of errors in fitting equation (6) is from the interpolation between maturities. In this study, we focus on the short maturities options, namely two and three months. This is because the stock market movement has the greatest impact on the short maturity options which are also the most liquid. In order to get the cumulants at the targeted maturities of interest, we linearly interpolate the two adjacent cumulants from the two nearest maturities. For one month maturity, one of the two cumulants used in the interpolation is less than one month. It is well known that options expiring within two weeks are subject to large trading noise which will severely impact on the quality of the cumulants calculated from the option prices. In this study, we exclude all the options that mature within two weeks. We also remove, on each day, all options with trading volume less than 100. At the final stage, we dropped the interpolated one month maturity options because the noise introduced by the interpolation is too big and it is not possible to have a continuous time series since half the time, the first maturity options are dropped because

the time to maturity is less than two weeks.

Table 1: Summary Statistics

	Mean	Standard Deviation
Log Return	-0.0003	0.0192
$\kappa_2(\tau_2)$	0.0170	0.0153
$\kappa_3(\tau_2)$	-0.0048	0.0072
$\kappa_2(\tau_3)$	0.0262	0.0213
$\kappa_3(\tau_3)$	-0.0090	0.0123

In total, the data set consists of 770 daily observations; one market log return and four cumulants for each day. We do not have any identification problem, since there are five measurement equations each day, which is more than the number of transition equations.¹² Table 1 reports the mean and standard deviation of the daily log return and cumulants. The mean value of the log return is negative due to the big loss during the sub-prime crisis. We can also see big standard deviation for all the four cumulants as a result of the crisis.

4.2 Empirical Findings and Analysis

Table 2 reports the parameters estimates and the associated standard errors. Due to the large amount of data used in the estimation, most of the parameters values show strong statistical significant except the mean return, μ , under the \mathbb{P} measure. The mean return is estimated from only the log return series. The cumulants implied by options, which is under the \mathbb{Q} measure, contains no information about the mean return.

For the ease of comparison, we reported the parameters values under the \mathbb{P} and the \mathbb{Q} measures in Table 3. The relationship of parameters values under the two measures are given

¹²There are two transition equations. But the jump timing is also a latent process which we need to take into account.

Table 2: Parameters Estimates

Mean (Standard Deviation)		Mean (Standard Deviation)	
u	-0.1106 (0.0574)	σ_λ	10.0698 (0.0026)
κ_v	2.3910 (0.0170)	η	5.5406 (0.0002)
θ_v	0.0949 (0.0003)	γ_v	-1.1549 (0.0096)
σ_v	0.7311 (0.0017)	γ_λ	3.7750 (0.0009)
ρ	-0.8699 (0.0025)	γ_{J_x}	19.1631 (0.0034)
η_u	0.0352 (0.0001)	γ_{J_λ}	-0.1579 (0.0000)
η_d	0.0320 (0.0000)	$\sigma_{2,2}$	0.0015 (0.0000)
p	0.4543 (0.0003)	$\sigma_{2,3}$	0.0010 (0.0000)
κ_λ	7.6358 (0.0090)	$\sigma_{3,2}$	0.0010 (0.0000)
θ_λ	0.0218 (0.0002)	$\sigma_{3,3}$	0.0008 (0.0000)

This table reports the parameters estimates from Monte Carlo-Expectation Maximisation with standard errors in parentheses.

below,

$$\begin{aligned}
\kappa_v^{\mathbb{Q}} &= \kappa_v^{\mathbb{P}} + \gamma_v \sigma_v \\
\theta_v^{\mathbb{Q}} &= \frac{\kappa_v^{\mathbb{P}} \theta_v^{\mathbb{P}}}{\kappa_v^{\mathbb{Q}}} \\
\eta_u^{\mathbb{Q}} &= \frac{\eta_u^{\mathbb{P}}}{1 + \eta_u^{\mathbb{P}} \gamma_{J_x}} \\
\eta_d^{\mathbb{Q}} &= \frac{\eta_d^{\mathbb{P}}}{1 - \eta_d^{\mathbb{P}} \gamma_{J_x}} \\
p^{\mathbb{Q}} &= \frac{p(1 - \gamma_{J_x} \eta_d^{\mathbb{P}})}{p(1 - \gamma_{J_x} \eta_d^{\mathbb{P}}) + (1 - p)(1 + \gamma_{J_x} \eta_u^{\mathbb{P}})} \\
\kappa_\lambda^{\mathbb{Q}} &= \kappa_\lambda^{\mathbb{P}} + \gamma_\lambda \sigma_\lambda \\
\theta_\lambda^{\mathbb{Q}} &= \frac{\kappa_\lambda^{\mathbb{P}} \theta_\lambda^{\mathbb{P}}}{\kappa_\lambda^{\mathbb{Q}}} \\
\eta^{\mathbb{Q}} &= \frac{\eta^{\mathbb{P}}}{1 + \eta^{\mathbb{P}} \gamma_{J_\lambda}}
\end{aligned}$$

The parameters values for ρ , σ_v , and σ_λ are the same under the \mathbb{P} and the \mathbb{Q} measures.

Correspond to previous studies; we also obtain a negative volatility risk premium γ_v , which results in a smaller mean reverting speed κ_v and a higher long run mean θ_v for stochastic variance under the \mathbb{Q} measure. The leverage effect between stochastic variance and log return is characterised by the correlation ρ . Studies that use only stock return series

Table 3: Parameters Comparison under Two Measure

	P Measure	Q Measure
κ_v	2.3910	1.5467
θ_v	0.0949	0.1467
η_u	0.0352	0.0210
η_d	0.0320	0.0827
p	0.4543	0.1613
κ_λ	7.6358	45.6493
\varkappa_λ	4.6123	8.5145
θ_λ	0.0218	0.0036
ϑ_λ	0.0361	0.0196
η	5.5406	44.2755

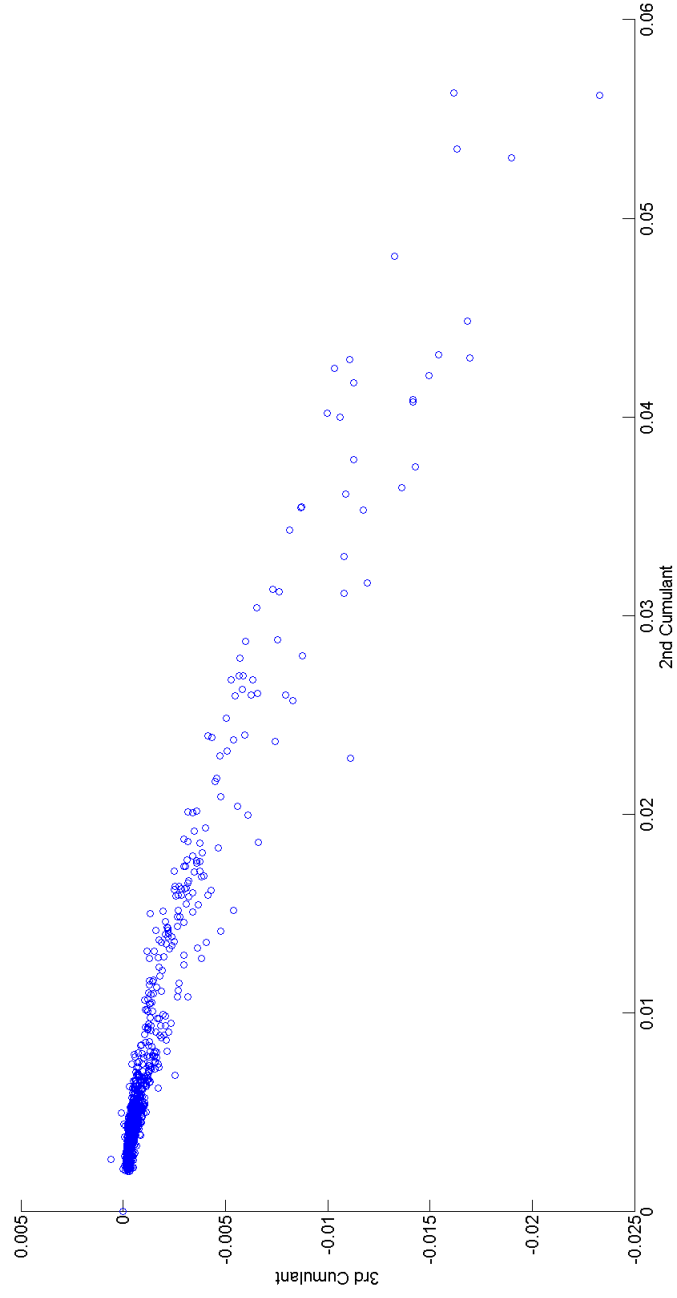
$\varkappa_\lambda = \kappa_\lambda - (1 - p)\eta$ and $\vartheta_\lambda = \frac{\kappa_\lambda\theta_\lambda}{\varkappa_\lambda}$ are the mean reverting speed and long run mean of the jump intensity after compensating the random measure.

for estimation (e.g., Fulop et al. (2012) and Eraker (2004)) typically produce a ρ around -0.6. Studies that use option data such as Bergomi (2004), Duffie et al. (2000), and Carr and Wu (2008) produce a much more negative ρ often closed to -1. Our $\rho=-0.8699$, produced from joint estimation using both option and stock returns data, falls between the two numbers reported by previous studies.

The most drastic differences between the two measures are related to the jump part. First, we observe a big negative shift in the distribution of jump size in log returns, producing a much higher risk neutral skewness implied by option prices. The premium paid for the skewness is further amplified by the huge increase of jump size in jump intensities under the Q measure. An interesting observation is in the change in the mean reversion speed $\varkappa_\lambda = \kappa_\lambda - (1 - p)\eta$ and the long run mean $\vartheta_\lambda = \frac{\kappa_\lambda\theta_\lambda}{\varkappa_\lambda}$ of the jump intensity. Unlike the stochastic variance, the jump intensity decays faster under the Q measure than under the P measure. However, the big jump size under the Q measure suggests a very high premium for short maturity option that decay quickly with option maturity. In terms of implied volatility surface, the results correspond to a high skewness at short maturities that flattens out quickly as maturity gets longer.

Figure 2 plots the log return series with the smoothed jump size. The estimation of

Figure 1: Scatter Plot of Skewness (3rd cumulant) against Variance (2nd cumulant)



This graph shows the scatter plot of 2nd and 3rd cumulants, which are variance and unnormalised skewness respectively, implied by SPX options maturing in two months.

jumps depends on the observed stock return as well as the estimated instantaneous variance. Therefore, the continuous big fluctuations during the sub-prime crisis are not all classified as jumps. Only the first few days of the crisis period are identified as jumps when the estimated instantaneous variance is low. Moreover, the estimation of jumps in this paper also depends on the information contained in the option prices. For example, there is a big loss in the stock market before October 2008, which was not identified as a jump because of the lack of implied skew in the option prices. Such an inconsistency between the two markets is part of the reason why we disentangle the dependency of the jump size in log return and the jump size in jump intensity.

In contrast, Carr and Wu (2008) and Fulop et al. (2012) assume that jump size in return is proportional to the jump size in the jump intensity. Such a specification works fine when the estimation uses only option prices or only stock return data but not both. It can be used to produce the risk premium for stock return jumps if the estimation uses only option data. The jump size in jump intensity is much harder to estimate using only stock return data. However, if the estimation uses both option data and stock return data, we can relax this constraint. For the sake of parsimony, we do not impose any other dependence structure.

In Figures 3 and 4, we plot the time series of the four cumulants calculated from market option prices together with the filtered counterparts from the estimation. The filtered processes track the market observed cumulants well in both calm and crisis periods. Compared with the jump estimates in Figure 2, we can see that the periods with higher skewness (third cumulant) during the subprime crisis started with negative jumps in stock returns.

Figure 5 plots the time series of filtered stochastic variance and filtered stochastic jump intensity. It is clear that stochastic variance explains most of the market pricing dynamics in the calm period when the stochastic jump intensity is small and negligible. During the crisis period, the stochastic jump intensity becomes very prominent. Compared with the previous figures, we notice that the start of the crisis period marked by negative stock return jumps is followed by big fluctuations in the stochastic jump intensity. The fluctuations of

Figure 2: SPX Log Return and the Mean of Jump Size

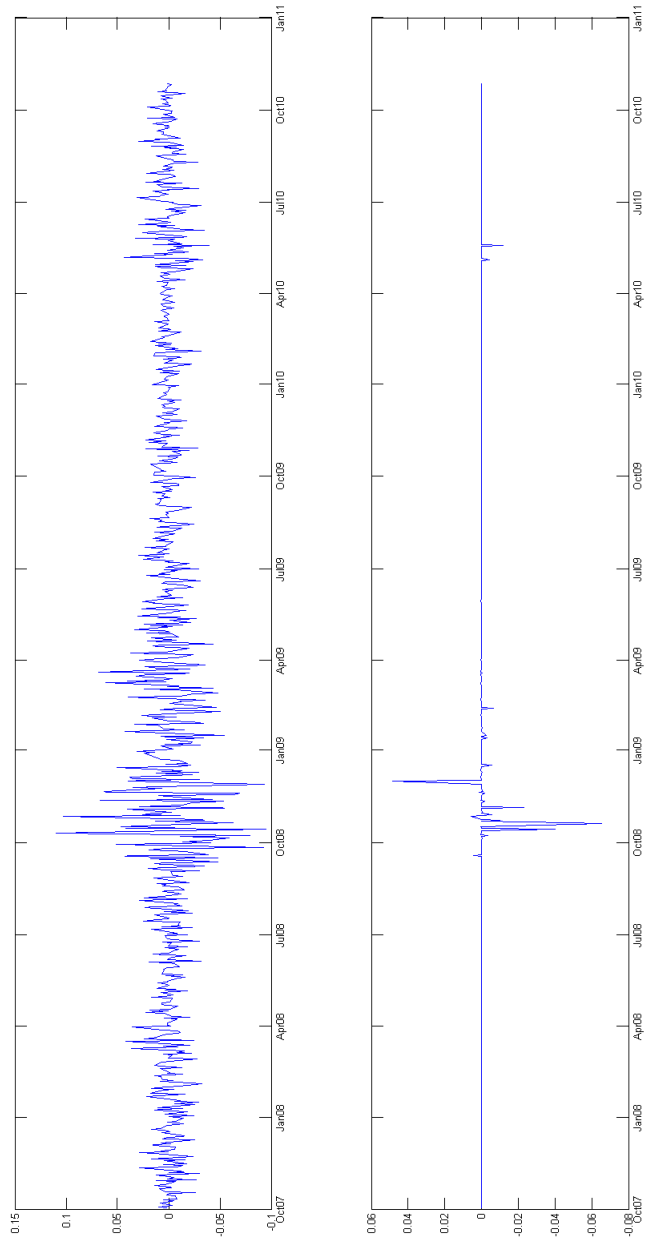
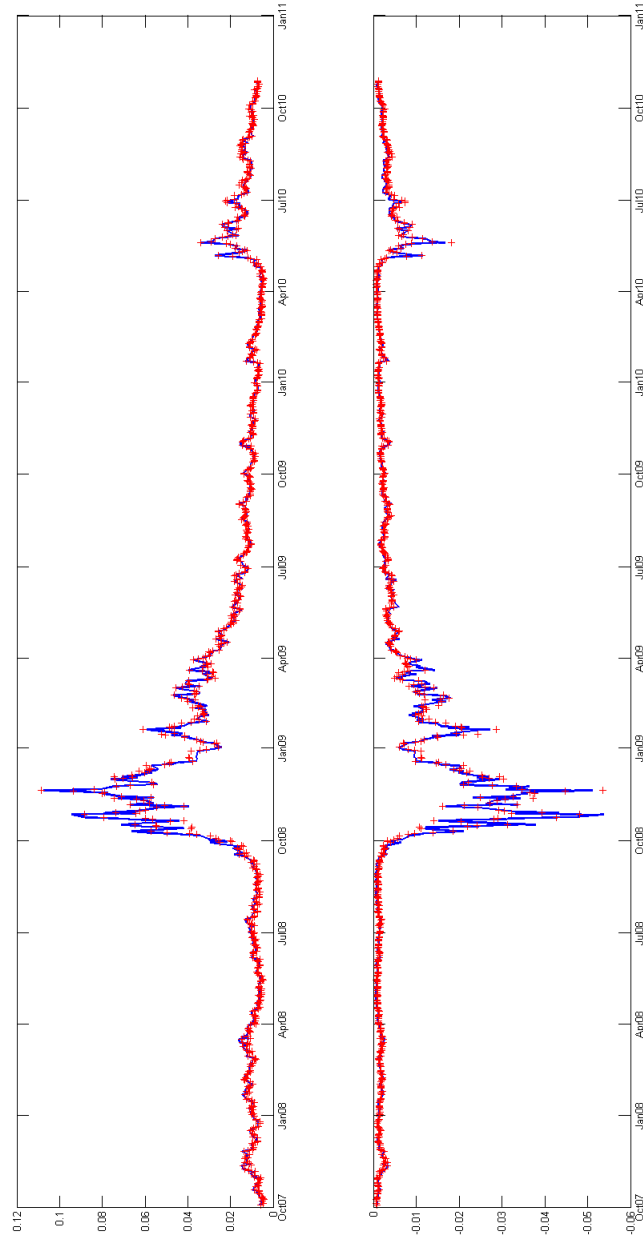
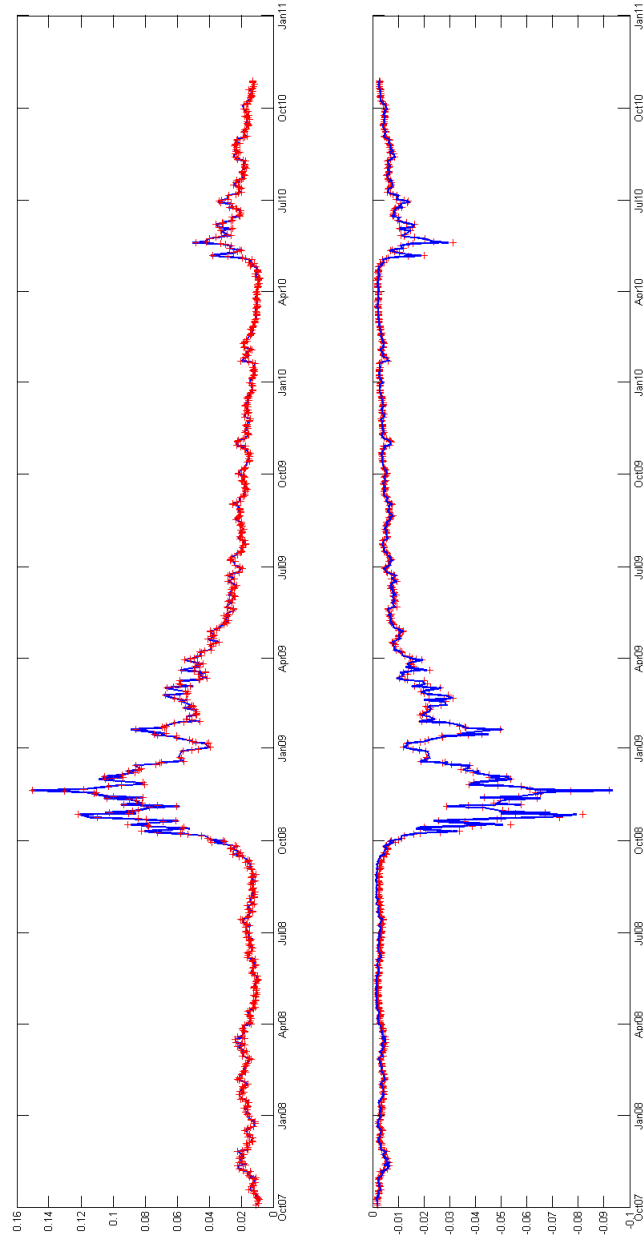


Figure 3: 2nd and 3rd Cumulants implied from SPX options with two months to maturity



The 2nd cumulant is in the top panel, whereas the 3rd cumulant is in the bottom panel. The blue solid line is the cumulant derived from option data and the red crosses are the filtered value from the model.

Figure 4: 2nd and 3rd Cumulants implied from SPX options with three months to maturity



The 2nd cumulant is in the top panel, whereas the 3rd cumulant is in the bottom panel. The blue solid line is the cumulant derived from option data and the red crosses are the filtered value from the model.

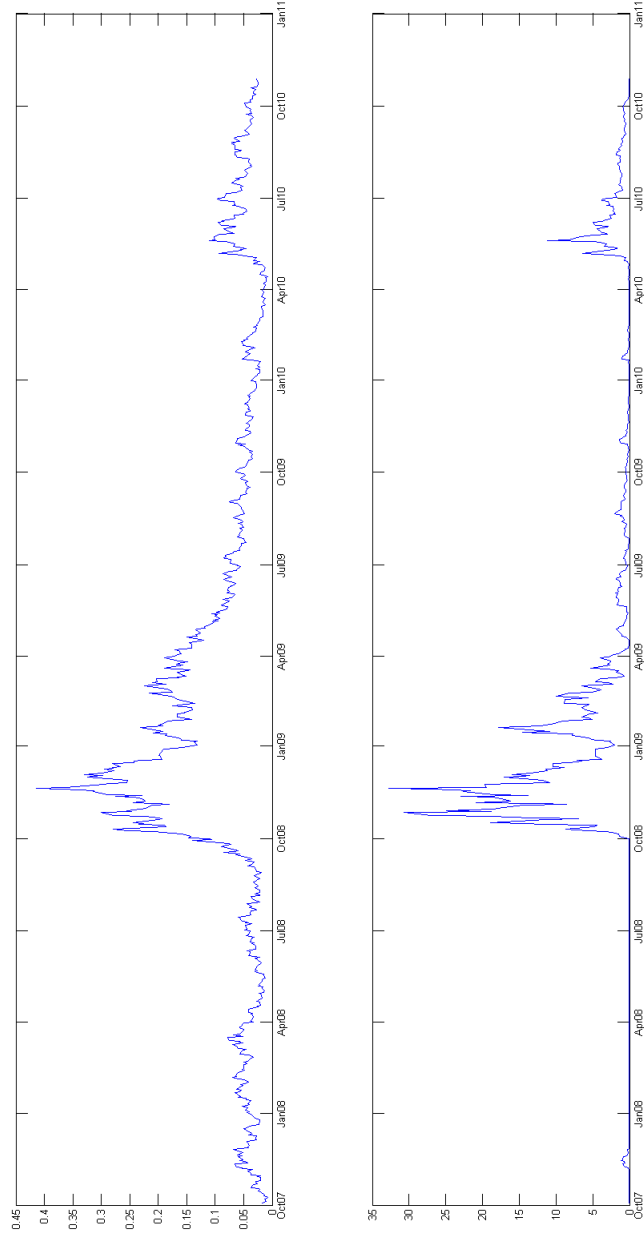
the jump intensity is controlled by the diffusion part of the jump intensity dynamics whose variance is proportional to $\sqrt{\lambda_t}$; the larger the jump intensity, the bigger its fluctuations. The fluctuations in the stochastic jump intensity are not necessarily associated with or driven by jumps in the stock return.

To see the different roles played by variance and jump intensity, we plot the contributions to the cumulants by these two stochastic components. Figures 6 and 7 show, respectively, the proportion of the implied cumulants of 2-month and 3-month maturity options that are explained by jump intensity. Before the sub-prime crisis, we can see both the second and third cumulants are mainly explained by stochastic variance. In the sub-prime crisis, jump intensity explains most of the skewness (third cumulant); the leverage effect between stochastic variance and stock return is not enough to account for the big skewness implied by prices of short maturity options. For the second cumulant, stochastic variance is always the most important factor. If we do not use the skewness information in the estimation, it might not be possible to identify the jump intensity process, since stochastic variance is capable of fully explain the second cumulant.

4.3 Risk Premium of Variance Swap

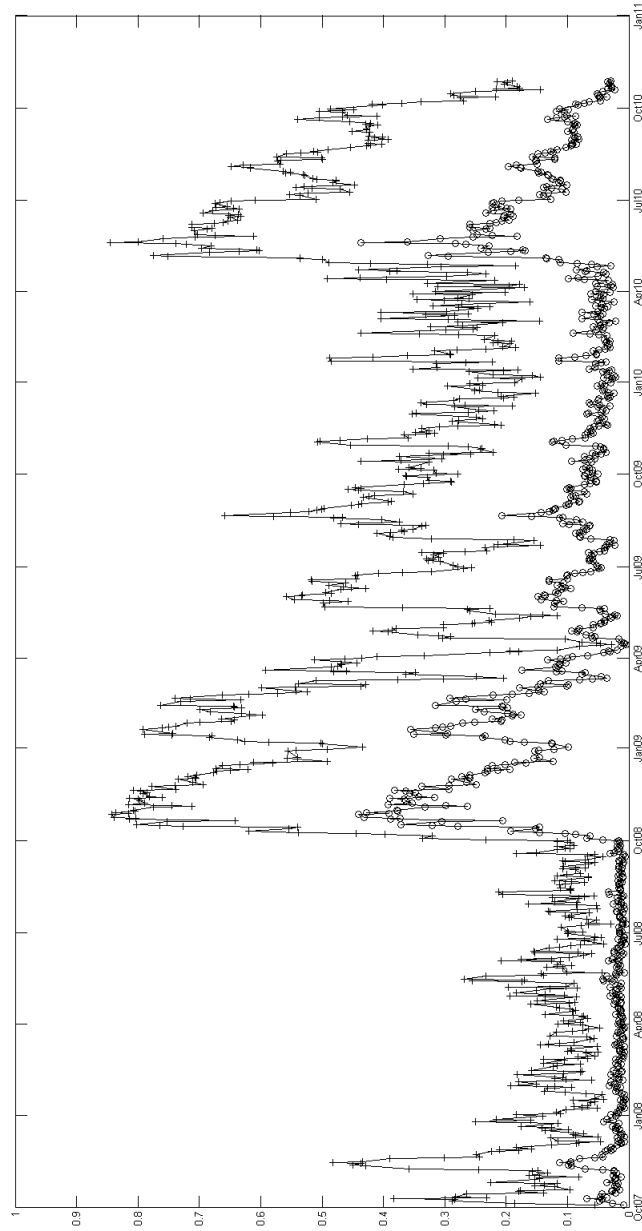
Since we have the estimates for the risk premium for both stochastic variance and jump intensity, we can revisit the question, “how the risk premium of variance swap reacts to negative stock return jumps?” Variance swap is a forward contract on realised variance. Demeterfi et al. (1999) show that the expectation of future variance, or the fair price of a variance swap, is a weighted integral of option prices across all strikes, assuming that there is no jump in equity price. This result strongly influenced the way in which the CBOE calculates the spot VIX index (see CBOE (2009)) defined as the square root of the fair price of variance swap. Carr and Wu (2009) establish the relationship between log contract and variance swap in the presence of jumps. If admitting jumps, the variance swap calculated by CBOE is,

Figure 5: Filtered Values of Stochastic Variance and Stochastic Jump Intensity



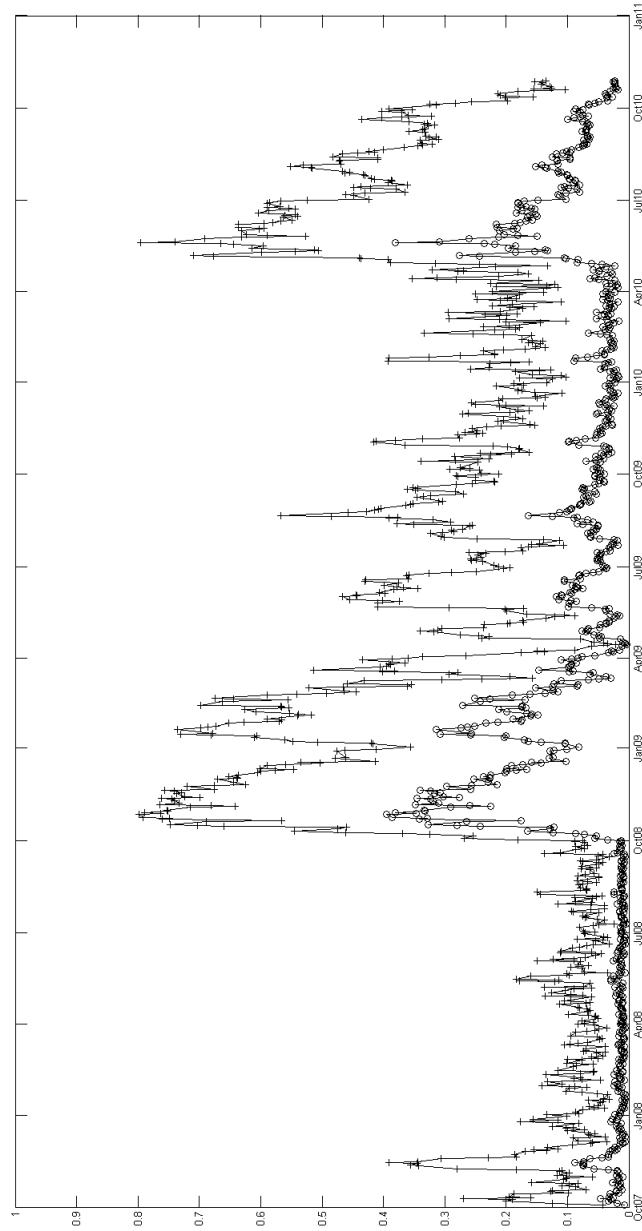
The filtered stochastic variance is in the top panel. The filtered stochastic jump intensity is in the bottom panel.

Figure 6: Percentage contribution by jump intensity to the implied cumulants of 2-month options



The line with circles is the percentage of contribution to the 2nd cumulant (variance) by jump intensity; the line with crosses is the percentage of contribution to 3rd cumulant (skewness) by jump intensity.

Figure 7: Percentage contribution by jump intensity to the implied cumulants of 3-month options



The line with circles is the percentage of contribution to the 2nd cumulant (variance) by jump intensity; the line with crosses is the percentage of contribution to 3rd cumulant (skewness) by jump intensity.

$$\begin{aligned}
\mathbb{E}[\text{VS}^{t,T}|\mathcal{F}_t] &= (T-t)(\theta_v + 2\mathbb{E}[e^{J_x} - 1 - x]\vartheta) + \frac{1 - e^{-\kappa_v(T-t)}}{\kappa_v}(v_t - \theta_v) \\
&\quad + 2\mathbb{E}[e^{J_x} - 1 - J_x]\frac{1 - e^{-\varkappa_\lambda(T-t)}}{\varkappa_\lambda}(\lambda_t - \vartheta_\lambda)
\end{aligned} \tag{11}$$

From the original definition of variance swap, $\mathbb{E}[\text{VS}^{t,T}|\mathcal{F}_t] = \int_t^T \langle d \ln S_t \rangle$, we have

$$\begin{aligned}
\mathbb{E}[\text{VS}^{t,T}|\mathcal{F}_t] &= (T-t)(\theta_v + \mathbb{E}[J_x^2]\vartheta) + \frac{1 - e^{-\kappa_v(T-t)}}{\kappa_v}(v_t - \theta_v) \\
&\quad + \mathbb{E}[J_x^2]\frac{1 - e^{-\varkappa_\lambda(T-t)}}{\varkappa_\lambda}(\lambda_t - \vartheta_\lambda)
\end{aligned} \tag{12}$$

The difference between these two definitions are of the third order of the jump size, i.e. $O(\mathbb{E}[J_x^3])$, which is very small and we shall omit it in this paper. In the remainder of this paper, we will use the second definition of variance swap in (12) to calculate the time varying risk premium for the variance swap. Following Todorov (2010), we define the variance swap risk premium as the difference between the expectations under the two measures,

$$VR_\tau(t) = \mathbb{E}^{\mathbb{P}}[\text{VS}^{t,T}|\mathcal{F}_t] - \mathbb{E}^{\mathbb{Q}}[\text{VS}^{t,T}|\mathcal{F}_t] \tag{13}$$

where $\tau = T - t$, and $\mathbb{E}^{\mathbb{P}}[\text{VS}^{t,T}|\mathcal{F}_t]$ and $\mathbb{E}^{\mathbb{Q}}[\text{VS}^{t,T}|\mathcal{F}_t]$ are calculated using the parameters values under the two measures and reported in Table 3 together with filtered values of v_t and λ_t . Figure 8 plots the risk premium associated with the one month variance swap. Since the risk premium is also an affine function of the state variables, the risk premium will jump whenever the state variables jump. Using high frequency data to identify realized stock return jumps, Todorov (2010) finds the decay speed of variance swap premium is abnormally high after realized jumps. In our two factor specification, we confirm this observation.

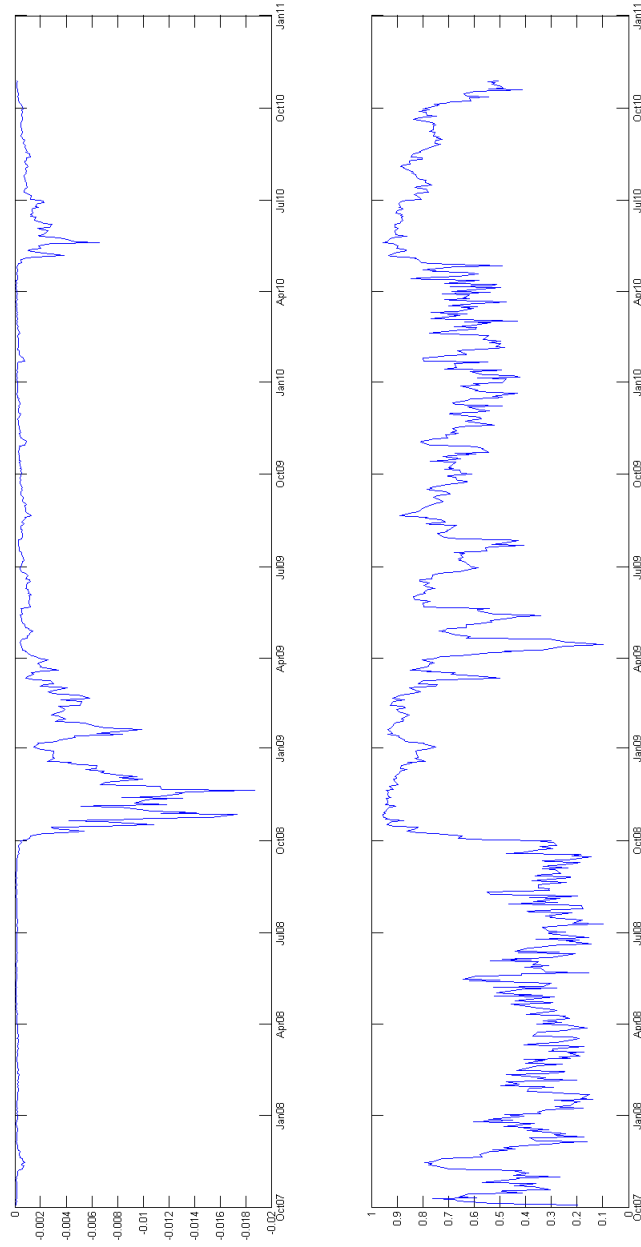
The risk premium and the percentage contribution from the jump intensity shot up at the beginning of the sub-prime crisis. The percentage contribution from the jump intensity to the risk premium decays slowly after this initiate shock till the Greek debt crisis in 2010. Since the percentage contribution from the jump intensity decays after jumps, we expect a faster decay speed for the jump intensity than that for the stochastic variance, $\varkappa_\lambda > \kappa_v$, which is consistent with the parameters estimates reported in Table 3.

An interesting observation to note by comparing Figure 8 with Figures 6 and 7 is that the proportion of risk premium due to jump is much bigger than jump's contribution to the total variance. This means investors pay a higher risk premium for the same amount of variance that produced by jumps. As shown by Bakshi and Madan (2006), risk premium for variance comes from higher moments. For each unit of variance produced from jump, the corresponding skewness and kurtosis associated with jumps are much higher than those associated with the diffusion process, and hence the risk premium paid for each unit of jump is much higher. In the sub-prime crisis, jumps only account for 40% of the variance, but in terms of risk premium for variance swap, over 90% of the premium is due to jumps.

5 Conclusions

We proposed a stochastic variance with self exciting jump (SVSEJ) model for explaining the time varying risk premium of variance swap based on premium embedded in short term maturity option prices. Instead of using option prices directly, we track the cumulants implied by options to distinguish the different roles played by stochastic variance and self exciting jump. It is shown that the stochastic variance is able to fit market dynamics very well in the normal period and accounts for most of the time varying risk premium. However, in the crisis period, the market behaves very differently. The high skewness under \mathbb{Q} measure implied in the prices of short term maturity options can only be explained by jumps with high jump intensity.

Figure 8: One Month Variance Swap Risk Premium and Contribution by Jump Intensity



Top panel plots the risk premium for variance swap in total; Bottom panel plots the contribution by jump intensity to the variance swap risk premium.

To cater for both calm or crisis periods, we use a Poisson process to capture rare jump events that act as the turning points from calm to crisis periods. Such an assumption is based on what we observe in the history of stock markets. Jumps is triggered by abrupt big movements in the stock return, and also characterised by the very skewed implied volatility surface. The self exciting jump in our model explains well the high skewness triggered by a negative stock return jump at the start of the crisis period. We introduce another diffusion term in the jump intensity to describe the big fluctuation of jump intensity in the aftermath reflecting investor’s uncertainty about future jumps. A mean reverting term (together with some other conditions)¹³ guarantees the stationarity of the process, ensuring the market pricing dynamics will revert back to the normal condition eventually. The estimation of the model shows that the reverting speeds of the stochastic variance and the self exciting jump are different under both measures, resulting in different risk premium between calm and crisis periods.

In this study, we only focus on the risk premium of options with two and three months to maturity because we do not have liquid data for long maturity options. It is possible that we may need extra factors to explain prices of options with medium and long maturities. For example, the long run mean of the stochastic variance at long maturity implied by our model is a constant, while we can see big changes day by day of the ATM implied volatility of long maturity options. A stochastic long run mean might be a potential third factor if one wants to include prices of long maturity options in the calibration. Other possible improvements concern the particle filter method, especially the sampling of jumps in the prediction step. One could also consider the use of parallel computing to speed up the EM algorithm.

¹³The technical condition required here is $\varkappa_\lambda > 0$ or $\kappa_\lambda > (1 - p)\eta$.

Appendix

A. Solution of PIDE for Moment Generating Function

We derive the solution form for the moment generating function in equation (3), which has the following form,

$$G(\omega, x_t, v_t, \lambda_t, t, T) = e^{\omega x_t + A(\omega, t, T) + B(\omega, t, T)v_t + C(\omega, t, T)\lambda_t} \quad (14)$$

Given the initial condition at time T that $G(\omega, x_T, V_T, \lambda_T, T, T) = e^{x_T \omega}$, we have $A(\omega, T, T) = 0$, $B(\omega, T, T) = 0$, and $C(\omega, T, T) = 0$. Substitute this solution form into equation (3), we have

$$\begin{aligned} & -\frac{1}{2}v_t\omega - \lambda_t\mathbb{E}[e^J - 1]\omega + \frac{1}{2}v_t\omega^2 + \kappa(\theta - v_t)B(\omega, t, T) + \frac{1}{2}\sigma^2v_tB(\omega, t, T)^2 \\ & + \rho\sigma v_t\omega B(\omega, t, T) + \kappa_\lambda(\theta_\lambda - \lambda_t)C(\omega, t, T) + \lambda_t \int_{\mathbb{R} \times \mathbb{R}^+} (e^{J\omega + 1_{J < 0}J_\lambda C(\omega, t, T)} - 1)\nu(dJ, dJ_\lambda) \\ = & -\frac{\partial A(\omega, t, T)}{\partial t} - \frac{\partial B(\omega, t, T)}{\partial t}v_t - \frac{\partial C(\omega, t, T)}{\partial t}\lambda_t \end{aligned}$$

Regroup the equation with respect to the state variables v_t and λ_t ,

$$\begin{aligned} 0 = & \left(-\frac{1}{2}\omega + \omega^2 - \kappa B(\omega, t, T) + \frac{1}{2}\sigma^2 B(\omega, t, T)^2 + \rho\sigma\omega B(\omega, t, T) + \frac{\partial B(\omega, t, T)}{\partial t}\right)v_t \\ & + \left(-\mathbb{E}[e^J - 1]\omega - \kappa_\lambda C(\omega, t, T) + \int_{\mathbb{R} \times \mathbb{R}^+} (e^{J\omega + 1_{J < 0}J_\lambda C(\omega, t, T)} - 1)\nu(dJ, dJ_\lambda)\right. \\ & \left. + \frac{\partial C(\omega, t, T)}{\partial t}\right)\lambda_t + \kappa_v\theta_v B(\omega, t, T) + \kappa_\lambda\theta_\lambda C(\omega, t, T) + \frac{\partial A(\omega, t, T)}{\partial t} \end{aligned}$$

Since v_t and λ_t are stochastic and not identical, to let the RHS always equals to zero, we have

$$\frac{\partial A(\omega, t, T)}{\partial t} = -\kappa_v \theta_v B(\omega, t, T) - \kappa_\lambda \theta_\lambda C(\omega, t, T) \quad (15)$$

$$\frac{\partial B(\omega, t, T)}{\partial t} = \frac{1}{2}(\omega - \omega^2) + (\kappa - \rho\sigma\omega)B(\omega, t, T) - \frac{1}{2}\sigma^2 B(\omega, t, T)^2 \quad (16)$$

$$\begin{aligned} \frac{\partial C(\omega, t, T)}{\partial t} &= \kappa_\lambda C(\omega, t, T) + \mathbb{E}[e^{J_x} - 1]\omega \\ &\quad - \int (e^{J_x \omega + 1_{J_x < 0} J_\lambda C(\omega, t, T)} - 1) \nu(dJ_x, dJ_\lambda) \end{aligned} \quad (17)$$

Therefore, the PIDE is decomposed as a set of ODEs. The second nonlinear ODE (16) is known as the Riccati differential equation, which has analytical solution, while the third ODE (17), which corresponds to the self exciting jump intensity, has no analytical solution and has to be solved numerically.¹⁴

B. Price of Log Contract and Variance Swap

The relationship between moment generating function and log contract is given by,

$$\begin{aligned} \mathbb{E}[\ln F_T | \mathcal{F}_t] &= \frac{\partial G(\omega, x_t, v_t, \lambda_t, t, T)}{\partial \omega} \Big|_{\omega=0} \\ &= \ln F_t + \frac{\partial A(\omega, t, T)}{\partial \omega} \Big|_{\omega=0} + \frac{\partial B(\omega, t, T)}{\partial \omega} \Big|_{\omega=0} v_t + \frac{\partial C(\omega, t, T)}{\partial \omega} \Big|_{\omega=0} \lambda_t \end{aligned} \quad (18)$$

Taking derivative with respect to ω on equations (15), (16), and (17), we have,

¹⁴The numerical method used here is the Runge-Kutta solver.

$$\begin{aligned}
\frac{\partial^2 A(\omega, t, T)}{\partial t \partial \omega} \Big|_{\omega=0} &= -\kappa_v \theta_v \frac{\partial B(\omega, t, T)}{\partial \omega} \Big|_{\omega=0} - \kappa_\lambda \theta_\lambda \frac{\partial C(\omega, t, T)}{\partial \omega} \Big|_{\omega=0} \\
\frac{\partial^2 B(\omega, t, T)}{\partial t \partial \omega} \Big|_{\omega=0} &= \frac{1}{2} - \rho \sigma B(0, t, T) + \kappa \frac{\partial B(\omega, t, T)}{\partial \omega} \Big|_{\omega=0} - \sigma^2 B(0, t, T) \frac{\partial B(\omega, t, T)}{\partial \omega} \Big|_{\omega=0} \\
\frac{\partial^2 C(\omega, t, T)}{\partial t \partial \omega} \Big|_{\omega=0} &= \kappa_\lambda \frac{\partial C(\omega, t, T)}{\partial \omega} \Big|_{\omega=0} + \mathbb{E}[e^{J_x} - 1] \\
&\quad - \int_{\mathbb{R} \times \mathbb{R}^+} (J_x + 1_{J_x < 0} J_\lambda) \frac{\partial C(\omega, t, T)}{\partial \omega} \Big|_{\omega=0} e^{1_{J_x < 0} J_\lambda C(0, t, T)} \nu(dJ_x, dJ_\lambda)
\end{aligned}$$

Given the initial conditions $A(\omega, T, T) = 0$, $B(\omega, T, T) = 0$, and $C(\omega, T, T) = 0$, we have $A(0, t, T)$, $B(0, t, T)$, and $C(0, t, T)$ are all equal to zero, which simplify the above ODE set as,

$$\begin{aligned}
\frac{\partial^2 A(\omega, t, T)}{\partial t \partial \omega} \Big|_{\omega=0} &= -\kappa_v \theta_v \frac{\partial B(\omega, t, T)}{\partial \omega} \Big|_{\omega=0} - \kappa_\lambda \theta_\lambda \frac{\partial C(\omega, t, T)}{\partial \omega} \Big|_{\omega=0} \\
\frac{\partial^2 B(\omega, t, T)}{\partial t \partial \omega} \Big|_{\omega=0} &= \frac{1}{2} + \kappa_v \frac{\partial B(\omega, t, T)}{\partial \omega} \Big|_{\omega=0} \\
\frac{\partial^2 C(\omega, t, T)}{\partial t \partial \omega} \Big|_{\omega=0} &= \kappa_\lambda \frac{\partial C(\omega, t, T)}{\partial \omega} \Big|_{\omega=0} + \mathbb{E}[e^{J_x} - 1] \\
&\quad - \int_{\mathbb{R} \times \mathbb{R}^+} (J_x + 1_{J_x < 0} J_\lambda) \frac{\partial C(\omega, t, T)}{\partial \omega} \Big|_{\omega=0} \nu(dJ_x, dJ_\lambda)
\end{aligned}$$

Taking integral with respect to t and using the initial conditions, we have,

$$\begin{aligned}
\frac{\partial A(\omega, t, T)}{\partial \omega} \Big|_{\omega=0} &= \frac{1 - (T-t)\kappa_v - e^{-\kappa_v(T-t)}}{2\kappa_v} \theta_v + \vartheta \frac{1 - (T-t)\varkappa_\lambda - e^{-\varkappa_\lambda(T-t)}}{\varkappa_\lambda^2} \kappa_\lambda \theta_\lambda \\
\frac{\partial B(\omega, t, T)}{\partial \omega} \Big|_{\omega=0} &= \frac{e^{-\kappa_v(T-t)} - 1}{2\kappa_v} \\
\frac{\partial C(\omega, t, T)}{\partial \omega} \Big|_{\omega=0} &= \xi \frac{e^{-\varkappa_\lambda(T-t)} - 1}{\varkappa_\lambda}
\end{aligned}$$

where $\varkappa_\lambda = \kappa_\lambda - P(J < 0)\eta$ and $\xi = \mathbb{E}[e^{J_x} - 1 - J_x]$. Substituting above results into equation (18), the price of log contract is given by,

$$\begin{aligned}
\mathbb{E}[\ln F_T | \mathcal{F}_t] &= \ln F_t - \frac{1}{2}(T-t)\theta_v - \frac{1 - e^{-\kappa_v(T-t)}}{2\kappa_v}(v_t - \theta_v) \\
&\quad - \xi(T-t)\vartheta_\lambda - \xi \frac{1 - e^{-\varkappa_\lambda(T-t)}}{\varkappa_\lambda}(\lambda_t - \vartheta_\lambda)
\end{aligned} \tag{19}$$

where $\vartheta_\lambda = \frac{\kappa_\lambda}{\varkappa_\lambda}\theta_\lambda$ is the long run mean of the jump intensity. Using the relationship between log contract and variance swap (see Carr and Wu (2009)),

$$\mathbb{E}[\ln F_T | \mathcal{F}_t] = \ln F_t - \frac{1}{2}\mathbb{E}[VS^{t,T} | \mathcal{F}_t] + \int_t^T \int_{\mathbb{R} \times \mathbb{R}^+} (e^{J_x} - 1 - J_x - \frac{J_x^2}{2})\nu(dJ_x, dt)$$

Noticing that $\xi = \mathbb{E}[e^{J_x} - 1 - J_x] = \mathbb{E}[\frac{J_x^2}{2}] + \mathbb{E}[e^{J_x} - 1 - J_x - \frac{J_x^2}{2}]$ and using the assumption that jump timing and size are independent, we substitute above relationship into equation (19) to have,

$$\begin{aligned}
\mathbb{E}[VS^{t,T} | \mathcal{F}_t] &= (T-t)(\theta_v + \mathbb{E}[J_x^2]\vartheta_\lambda) + \frac{1 - e^{-\kappa_v(T-t)}}{\kappa_v}(v_t - \theta_v) \\
&\quad + \mathbb{E}[J_x^2] \frac{1 - e^{-\varkappa_\lambda(T-t)}}{\varkappa_\lambda}(\lambda_t - \vartheta_\lambda)
\end{aligned} \tag{20}$$

We can see from equation (20) that over the time interval from t to T , as the contribution to the variance, diffusion and jump parts have the same functional form in terms of their mean reversion rate and long run mean, except that the variance from jumps is scaled by $\mathbb{E}[J_x^2]$, which is the variance for each jump. So using variance swap prices alone is not enough to separate the two state variables in our model.

For the second and third cumulants we tracked in the estimation, we follow the same procedure and derive the close form by using Wolfram Mathematica,

$$\frac{\partial^2 A(\omega, t, T)}{\partial \omega^2} \Big|_{\omega=0} = \frac{e^{-2(T-t)\kappa_v}}{8\kappa_v^3} \theta_v ((1 + 4e^{(T-t)\kappa_v} - 5e^{2(T-t)\kappa_v})\sigma^2) \quad (21)$$

$$\begin{aligned} &+ 2e^{(T-t)\kappa_v} (8(-1 + e^{(T-t)\kappa_v})\rho\sigma) \\ &+ (2 + e^{(T-t)\kappa_v})(T-t)\sigma^2 \kappa_v + 8e^{2(T-t)\kappa_v} (T-t)\kappa_v^3 \\ &- 8e^{(T-t)\kappa_v} (-1 + (T-t)\rho\sigma + e^{(T-t)\kappa_v} (1 + (T-t)\rho\sigma)) \kappa_v^2 \\ &+ \frac{e^{-2(T-t)\varkappa_\lambda}}{\varkappa_\lambda^3} \vartheta(e^{(T-t)\varkappa_\lambda} \varkappa_\lambda^2 (1 + e^{(T-t)\varkappa_\lambda} (-1 + (T-t)\varkappa_\lambda)) J_2 \\ &- (-1 + p)\eta\xi(\eta(1 + 4e^{(T-t)\varkappa_\lambda} (1 + (T-t)\varkappa_\lambda)) \\ &+ e^{2(T-t)\varkappa_\lambda} (-5 + 2(T-t)\varkappa_\lambda)) \xi \\ &+ 2e^{(T-t)\varkappa_\lambda} \varkappa_\lambda (2 + (T-t)\varkappa_\lambda + e^{(T-t)\varkappa_\lambda} (-2 + (T-t)\varkappa_\lambda)) \eta_d) \end{aligned}$$

$$\frac{\partial^2 B(\omega, t, T)}{\partial \omega^2} \Big|_{\omega=0} = \frac{e^{-2(T-t)\kappa_v}}{4\kappa_v^3} (-\sigma^2 + e^{2(T-t)\kappa_v} \sigma^2 + 4e^{(T-t)\kappa_v} \rho\sigma\kappa_v) \quad (22)$$

$$\begin{aligned} &- 2e^{(T-t)\kappa_v} (T-t)\sigma^2 \kappa_v - 4e^{(T-t)\kappa_v} \kappa_v^2 + 4e^{2(T-t)\kappa_v} \kappa_v^2 \\ &+ 4e^{(T-t)\kappa_v} (T-t)\rho\sigma\kappa_v^2 - 4e^{2(T-t)\kappa_v} \rho\sigma\kappa_v) \end{aligned}$$

$$\frac{\partial^2 C(\omega, t, T)}{\partial \omega^2} \Big|_{\omega=0} = \frac{e^{-2(T-t)\varkappa_\lambda}}{\varkappa_\lambda^3} (e^{(T-t)\varkappa_\lambda} (-1 + e^{(T-t)\varkappa_\lambda}) \varkappa_\lambda^2 J_2) \quad (23)$$

$$\begin{aligned} &- 2(-1 + p)\eta\xi(\eta(-1 + e^{2(T-t)\varkappa_\lambda} - 2e^{(T-t)\varkappa_\lambda} (T-t)\varkappa_\lambda) \xi \\ &+ e^{(T-t)\varkappa_\lambda} \varkappa_\lambda (-1 + e^{(T-t)\varkappa_\lambda} - (T-t)\varkappa_\lambda) \eta_d) \end{aligned}$$

$$\begin{aligned}
\frac{\partial^3 A(\omega, t, T)}{\partial \omega^3} \Big|_{\omega=0} &= -\frac{e^{-3(T-t)\kappa_v}}{16\kappa_v^5} \sigma \theta_v (-(-1 - 6e^{(T-t)\kappa_v} \\
&\quad -15e^{2(T-t)\kappa_v} + 22e^{3(T-t)\kappa_v}) \sigma^3 \\
&\quad + 6e^{(T-t)\kappa_v} \sigma ((-5 - 8e^{(T-t)\kappa_v} + 13e^{2(T-t)\kappa_v}) \rho \sigma \\
&\quad + \sigma ((1 - 8e^{(T-t)\kappa_v} + 7e^{2(T-t)\kappa_v}) \rho \\
&\quad + (1 + 3e^{(T-t)\kappa_v} + e^{2(T-t)\kappa_v})(T-t)\sigma)) \kappa_v \\
&\quad - 6e^{(T-t)\kappa_v} (2(-1 + (T-t)\rho\sigma)\sigma \\
&\quad + 2e^{2(T-t)\kappa_v} (12\rho\rho\sigma + 5\sigma + 2(T-t)\rho\sigma^2 \\
&\quad + (T-t)\rho\sigma^2) - e^{(T-t)\kappa_v} (24\rho\rho\sigma \\
&\quad + 8\sigma - 12t\rho\sigma\sigma - 4t\rho\sigma^2 + (T-t)^2\sigma^3)) \kappa_v^2 \\
&\quad + 12e^{2(T-t)\kappa_v} (-(T-t)(-4 - 2e^{(T-t)\kappa_v} + (T-t)\rho\sigma)\sigma \\
&\quad + \rho(-8 + 8t\rho\sigma + 4e^{(T-t)\kappa_v} (2 + (T-t)\rho\sigma) - (T-t)^2\sigma^2)) \kappa_v^3 \\
&\quad - 24e^{2(T-t)\kappa_v} (T-t)\rho(2 + 2e^{(T-t)\kappa_v} - (T-t)\rho\sigma) \kappa_v^4 \\
&\quad - \frac{e^{-3(T-t)\varkappa_\lambda}}{\varkappa_\lambda^5} \vartheta (-e^{2(T-t)\varkappa_\lambda} \varkappa_\lambda^4 (1 + e^{(T-t)\varkappa_\lambda} (-1 + (T-t)\varkappa_\lambda)) J_3 \\
&\quad - 3e^{(T-t)\varkappa_\lambda} (-1 + p)\eta\kappa^2 J_2 (\eta(1 + 4e^{(T-t)\varkappa_\lambda} (1 + (T-t)\varkappa_\lambda) \\
&\quad + e^{2(T-t)\varkappa_\lambda} (-5 + 2(T-t)\varkappa_\lambda)) \xi + e^{(T-t)\varkappa_\lambda} \varkappa_\lambda (2 + (T-t)\varkappa_\lambda \\
&\quad + e^{(T-t)\varkappa_\lambda} (-2 + (T-t)\varkappa_\lambda)) \eta_d \\
&\quad + (-1 + p)\eta\xi(\eta^2(-\varkappa_\lambda(-1 + 9e^{(T-t)\varkappa_\lambda} + 9e^{2(T-t)\varkappa_\lambda} (1 + 2(T-t)\varkappa_\lambda) \\
&\quad + e^{3(T-t)\varkappa_\lambda} (-17 + 6t\kappa)) + 2(-1 + p)\eta(1 + 6e^{(T-t)\varkappa_\lambda} (1 + (T-t)\varkappa_\lambda) \\
&\quad + e^{3(T-t)\varkappa_\lambda} (-22 + 6t\kappa) + 3e^{2(T-t)\varkappa_\lambda} (5 + 6t\kappa + 2(T-t)^2\varkappa_\lambda^2))) \xi^2 \\
&\quad - 3e^{(T-t)\varkappa_\lambda} \eta\kappa(\varkappa_\lambda(1 + 4e^{(T-t)\varkappa_\lambda} (1 + (T-t)\varkappa_\lambda) \\
&\quad + e^{2(T-t)\varkappa_\lambda} (-5 + 2(T-t)\varkappa_\lambda)) \\
&\quad + [\dots]
\end{aligned} \tag{24}$$

$$\begin{aligned}
[\dots] = & -2(-1+p)\eta(2+(T-t)\varkappa_\lambda + 2e^{(T-t)\varkappa_\lambda}(2+(T-t)\varkappa_\lambda)^2 \\
& + e^{2(T-t)\varkappa_\lambda}(-10+3(T-t)\varkappa_\lambda))\xi\eta_d \\
& -3e^{2(T-t)\varkappa_\lambda}\varkappa_\lambda^2(-(-1+p)\eta(6+4t\kappa+(T-t)^2\varkappa_\lambda^2 \\
& +2e^{(T-t)\varkappa_\lambda}(-3+(T-t)\varkappa_\lambda)) \\
& +2\kappa(2+(T-t)\varkappa_\lambda + e^{(T-t)\varkappa_\lambda}(-2+(T-t)\varkappa_\lambda))\eta_d^2))
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^3 B(\omega, t, T)}{\partial \omega^3} \Big|_{\omega=0} = & \frac{3e^{-3(T-t)\kappa_v}}{16\kappa_v^5} \sigma(-(-1-2e^{(T-t)\kappa_v} + e^{2(T-t)\kappa_v} \\
& +2e^{3(T-t)\kappa_v})\sigma^3 \\
& +2e^{(T-t)\kappa_v} \sigma(4(-2+e^{(T-t)\kappa_v} + e^{2(T-t)\kappa_v})\rho\sigma \\
& +\sigma(2(-1+e^{(T-t)\kappa_v})^2\rho + (2+e^{(T-t)\kappa_v})(T-t)\sigma))\kappa_v \\
& -2e^{(T-t)\kappa_v}(4(-1+(T-t)\rho\sigma)\sigma + 4e^{2(T-t)\kappa_v}(2\rho\rho\sigma + \sigma) \\
& -e^{(T-t)\kappa_v}(8\rho\rho\sigma - 8t\rho\sigma\sigma + (T-t)^2\sigma^3))\kappa_v^2 \\
& +4e^{2(T-t)\kappa_v}((T-t)(4-(T-t)\rho\sigma)\sigma \\
& +\rho(-4+4e^{(T-t)\kappa_v} + 4t\rho\sigma - (T-t)^2\sigma^2))\kappa_v^3 \\
& +8e^{2(T-t)\kappa_v}(T-t)\rho(-2+(T-t)\rho\sigma)\kappa_v^4
\end{aligned} \tag{25}$$

$$\begin{aligned}
\frac{\partial^3 C(\omega, t, T)}{\partial \omega^3} \Big|_{\omega=0} &= \frac{e^{-3(T-t)\varkappa_\lambda}}{\varkappa_\lambda^5} (e^{2(T-t)\varkappa_\lambda} (-1 + e^{(T-t)\varkappa_\lambda}) \varkappa_\lambda^4 J_3) \\
&+ 3e^{(T-t)\varkappa_\lambda} (-1 + p) \eta \kappa^2 J_2 (2\eta (-1 + e^{2(T-t)\varkappa_\lambda}) \\
&- 2e^{(T-t)\varkappa_\lambda} (T-t) \varkappa_\lambda) \xi \\
&+ e^{(T-t)\varkappa_\lambda} \varkappa_\lambda (-1 + e^{(T-t)\varkappa_\lambda} - (T-t) \varkappa_\lambda) \eta_d \\
&- 3(-1 + p) \eta \xi (\eta^2 (\varkappa_\lambda (-1 + 6e^{(T-t)\varkappa_\lambda} - 2e^{3(T-t)\varkappa_\lambda}) \\
&+ e^{2(T-t)\varkappa_\lambda} (-3 + 6(T-t) \kappa)) \\
&+ 2(-1 + p) \eta (-1 + 2e^{3(T-t)\varkappa_\lambda} - 2e^{(T-t)\varkappa_\lambda} (1 + 2(T-t) \varkappa_\lambda) \\
&+ e^{2(T-t)\varkappa_\lambda} (1 - 2(T-t) \varkappa_\lambda - 2(T-t)^2 \varkappa_\lambda^2)) \xi^2 \\
&- 2e^{(T-t)\varkappa_\lambda} \eta \kappa (-\varkappa_\lambda (1 - e^{2(T-t)\varkappa_\lambda} + 2e^{(T-t)\varkappa_\lambda} (T-t) \varkappa_\lambda) \\
&- (-1 + p) \eta (-3 + 3e^{2(T-t)\varkappa_\lambda} - 2(T-t) \varkappa_\lambda) \\
&- 2e^{(T-t)\varkappa_\lambda} (T-t) \varkappa_\lambda (2 + (T-t) \varkappa_\lambda)) \xi \eta_d \\
&- e^{2(T-t)\varkappa_\lambda} \varkappa_\lambda^2 (-2\kappa (1 - e^{(T-t)\varkappa_\lambda} + (T-t) \varkappa_\lambda) \\
&- (-1 + p) \eta (-2 + 2e^{(T-t)\varkappa_\lambda} - 2(T-t) \varkappa_\lambda - (T-t)^2 \varkappa_\lambda^2) \eta_d^2)
\end{aligned} \tag{26}$$

where $J_2 = \mathbb{E}[J_x^2]$ and $J_3 = \mathbb{E}[J_x^3]$.

C. Cumulants derived from Option Prices

As shown in Carr and Madan (2002), for any twice differentiable function $f(F_T)$ of the final future price F_T ,

$$\begin{aligned}
f(F_T) &= f(F_t) + f'(F_t)(F_T - F_t) \\
&+ \int_0^{F_t} f''(K)(K - F_T)^+ dK + \int_{F_t}^\infty f''(K)(F_T - K)^+ dK
\end{aligned}$$

Let $f(\cdot) = \left(\ln \frac{\cdot}{F_t}\right)^i$, where i denotes moment order, and taking expectation on both sides, we can calculate the i th moment of the log return,

$$\mathbb{E}[f(F_T)] + \int_0^{F_t} f''(K)P(K)dK + \int_{F_t}^{\infty} f''(K)C(K)dK \quad (27)$$

where $C(K)$ and $P(K)$ are option prices at strike price K without discounting. So, for the first three moments of $\left(\ln \frac{F_T}{F_t}\right)$, we have,

$$\begin{aligned} M_1 &= -\int_0^{F_t} \frac{1}{K^2}P(K)dK - \int_{F_t}^{\infty} \frac{1}{K^2}C(K)dK \\ M_2 &= \int_0^{F_t} \frac{2}{K^2} \left(1 - \ln \frac{K}{F_t}\right) P(K)dK + \int_{F_t}^{\infty} \left(1 - \ln \frac{K}{F_t}\right) C(K)dK \\ M_3 &= \int_0^{F_t} \frac{3}{K^2} \left(2 \ln \frac{K}{F_t} - \ln^2 \frac{K}{F_t}\right) P(K)dK + \int_{F_t}^{\infty} \frac{3}{K^2} \left(2 \ln \frac{K}{F_t} - \ln^2 \frac{K}{F_t}\right) C(K)dK \end{aligned}$$

To convert moments to cumulants, we have, option implied cumulants

$$\begin{aligned} \kappa_1^* &= M_1 \\ \kappa_2^* &= M_2 - M_1^2 \\ \kappa_3^* &= M_3 - 3M_2M_1 + 2M_1^3 \end{aligned}$$

References

- Ait-Sahalia, Y., Cacho-Diaz, J., and Laeven, R. J. (2010). Modeling financial contagion using mutually exciting jump processes. Technical report, National Bureau of Economic Research, Inc.
- Bakshi, G. and Madan, D. (2006). A theory of volatility spreads. *Management Science*, 52(12):1945–1956.
- Bergomi, L. (2004). Smile dynamics. *Risk*.
- Bergomi, L. (2005). Smile dynamics ii. *Risk*.
- Bergomi, L. (2008). Smile dynamics iii. *Risk*.
- Carr, P. and Madan, D. (2002). Towards a theory of volatility trading. pages 417–427.
- Carr, P. and Wu, L. (2008). Leverage effect, volatility feedback, and self-exciting market disruptions: Disentangling the multi-dimensional variations in s&p 500 index options. *SSRN eLibrary*.
- Carr, P. and Wu, L. (2009). Variance risk premiums. *Review of Financial Studies*, 22(3):1311–1341.
- CBOE (2009). The cboe volatility index - vix. Technical report.
- Cont, R. and Da Fonseca, J. (2002). Dynamics of implied volatility surfaces. *Quantitative Finance*, 2(1):45–60.
- Demeterfi, K., Derman, E., Kamal, M., and Zou, J. (1999). More than you ever wanted to know about volatility swaps. Quantitative strategies research notes, Goldman Sachs.
- Doucet, A., de Freitas, N., and Gordon, N., editors (2001). *Sequential Monte Carlo Methods in Practice (Statistics for Engineering and Information Science)*. Springer, 1 edition.

- Duan, J.-C. (1995). The garch option pricing model. *Mathematical Finance*, 5(1):13–32.
- Duffie, D., Pan, J., and Singleton, K. (2000). Transform analysis and asset pricing for affine jump-diffusions. *Econometrica*, 68:1343–1376.
- Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, 50(4):987–1007.
- Eraker, B. (1998). Mcmc analysis of diffusion models with application to finance. *Journal of Business and Economic Statistics*, 19:177–191.
- Eraker, B. (2004). Do stock prices and volatility jump? reconciling evidence from spot and option prices. *Journal of Finance*, 59(3):1367–1404.
- Fulop, A., Li, J., and Yu, J. (2012). Bayesian learning of impacts of self-exciting jumps in returns and volatility. Working Papers 03-2012, Singapore Management University, School of Economics.
- Gatheral, J. (2004). A parsimonious arbitrage-free implied volatility parameterization with application to the valuation of volatility derivatives. In *Global Derivatives & Risk Management*.
- Gordon, N. J., Salmond, D. J., and Smith, A. F. M. (1993). Novel approach to nonlinear/non-gaussian bayesian state estimation. *Iee Proceedings F Radar and Signal Processing*, 140.
- Hawkes, A. G. (1971). Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 58:83–90.
- Heston, S. L. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies*, 6:327–343.
- Hurzeler, M. and Kunsch, H. R. (1998). Monte carlo approximations for general state-space models. *Journal Of Computational And Graphical Statistics*, 7(2):175– 193.

- Maheu, J. M. and McCurdy, T. H. (2004). News arrival, jump dynamics and volatility components for individual stock returns. *Journal of Finance*, 59(2):755–793.
- Malik, S. and Pitt, M. K. (2011). Particle filters for continuous likelihood evaluation and maximisation. *Journal of Econometrics*, 165(2):190–209.
- Pan, J. (2002). The jump-risk premia implicit in options: Evidence from an integrated time-series study. *Journal of Financial Economics*, 63:3–50.
- Sepp, A. (2003). Fourier transform for option pricing under affine jump-diffusions: An overview. *SSRN eLibrary*.
- Todorov, V. (2010). Variance risk-premium dynamics: The role of jumps. *Review of Financial Studies*, 23(1):345–383.