Self-Exciting Jumps, Learning, and Asset Pricing Implications

Author(s): Andras Fulop, Junye Li and Jun Yu

Source: *The Review of Financial Studies*, March 2015, Vol. 28, No. 3 (March 2015), pp. 876-912

Published by: Oxford University Press. Sponsor: The Society for Financial Studies.

Stable URL: https://www.jstor.org/stable/24465729

Self-Exciting Jumps, Learning, and Asset Pricing Implications

Andras Fulop
ESSEC Business School, Paris-Singapore

Junye Li
ESSEC Business School, Paris-Singapore

Jun Yu
School of Economics and Lee Kong Chian School of Business, Singapore Management University

The paper proposes a self-exciting asset pricing model that takes into account co-jumps between prices and volatility and self-exciting jump clustering. We employ a Bayesian learning approach to implement real-time sequential analysis. We find evidence of self-exciting jump clustering since the 1987 market crash, and its importance becomes more obvious at the onset of the 2008 global financial crisis. We also find that learning affects the tail behaviors of the return distributions and has important implications for risk management, volatility forecasting, and option pricing. (*JEL* C11, C13, C32, G12)

The financial meltdown of 2008 and the European debt crisis of 2010 to 2012 have impacted financial markets worldwide and have had far-reaching consequences for the world economy. These market turmoils raise questions about how likely extreme events are and how they can be modeled. Recent empirical studies find that a big jump in asset prices tends to be associated with an abrupt move in asset volatility, a phenomenon labeled as co-jumps of prices and volatility (Eraker, Johannes, and Polson 2003; Eraker 2004; Jacod and Todorov 2010; Todorov and Tauchen 2011). A further intriguing empirical observation is that an extreme movement in markets tends to be followed by another extreme movement, resulting in self-exciting jump clustering (Yu 2004; Carr and Wu 2011; Aït-Sahalia, Cacho-Diaz, and Laeven Forthcoming).

In the current paper, we propose a self-exciting asset pricing model where both co-jumps of prices and volatility and self-exciting jump clustering are

We would like to thank Pietro Veronesi (the editor) and an anonymous referee for their constructive comments that helped us substantially improve the paper. We are also grateful for comments of Yacine Ait-Sahalia, Jin-Chuan Duan, Nicolas Chopin, Laurent Calvet, and seminar participants at National University of Singapore, Singapore Management University, Xiamen Univeristy (China), SMU-ESSEC Symposium on Empirical Finance and Financial Econometrics, 2012 China International Conference in Finance, and Princeton/QUT/SMU Tripartite Conference on Financial Econometrics. Send correspondence to Junye Li, 100 Victoria Street, NLB #13-02, Singapore 188064; telephone: +65 6413 9462. E-mail: li@essec.edu.

doi:10.1093/rfs/hhu078 Advance Access publication November 3, 2014

allowed. In our specification, negative jumps play a crucial role. In particular, whenever there is a negative jump in asset returns, it is simultaneously passed on to both diffusion variance and the jump intensity. Therefore, the likelihood of future extreme events can be enhanced by either jumps in diffusion volatility or increases in the jump intensity or both. The importance of negative jumps is consistent with the well-documented empirical regularity in financial markets that economic agents react more strongly to bad macroeconomic surprises than to good ones (Andersen, Bollerslev, and Diebold 2007). Our model is quite flexible, and it has closed-form conditional expectations of the volatility components, making it convenient to use in volatility forecasting and risk management.

Traditional asset pricing theories usually have a strong assumption that endows economic agents with more precise information of the model and parameters than that available to econometricians. While this approach simplifies model specification and inference, it ignores the need for updating the long-run components of uncertainty, and it may lead to underestimation of risks encountered by economic agents. L. Hansen (2007, pp. 1–2) makes the following argument:

> Should we put econometricians and economic agents on comparable footing, or should we endow economic agents with much more refined knowledge? ... As the statistical problem that agents confront in our models is made complex, rational expectations' presumed confidence in their knowledge of the probability specification becomes more tenuous.

Given that our model has a complex structure and contains multiple unobserved dynamic factors, we take a different route and consider a Bayesian economic agent who faces the same belief-updating problems as confronted by the econometrician. She takes parameters, latent states, and models as unknowns and uses Bayes' rule to update her beliefs sequentially over time as market information becomes available. This may lead to differences between ex ante beliefs and ex post outcomes, and could have important asset pricing implications.

Statistical learning and its implications for asset pricing have attracted an enormous amount of attention. A recent survey has been provided by Pastor and Veronesi (2009). One of the key implications is that Bayesian learning generates persistent and long-term changes to the agent's beliefs, which have important influence on stock valuation, risk measures, and time series predictability. Among others, Timmermann (1993, 1996) and Lewellen and Shanken (2002) show that learning may generate excess volatility and predictability in stock returns. Veronesi (2004) studies implications of learning about a peso state in a Markov-switching model. Pastor and Veronesi (2003, 2006) investigate stock valuation and learning about profitability. Cogley and Sargent (2008) provide an alternative explanation of the observed equity risk premium from

the learning perspective. Benzoni, Collin-Dufresne, and Goldstein (2011) show that updating of beliefs on jump parameters may cause permanent shifts in option prices.

However, most of the existing studies focus only on learning about either state variables or a single parameter. In contrast, in this paper, we concurrently learn about parameters and state variables. Simultaneous learning in an asset pricing model with a complex structure remains difficult, as the large number of unknowns complicates inference and slows down the learning process. We implement Bayesian learning on our self-exciting model by following the marginalized resample-move approach proposed by Fulop and Li (2013), and then study the implications of learning for risk measures, volatility forecasting, and option pricing. To highlight the effects of parameter learning and uncertainty, we compare most results in three cases: (i) parameter learning and uncertainty are present; (ii) only parameter uncertainty is present—the full-sample posterior distributions of parameters are used in analysis; and (iii) both parameter learning and uncertainty are ignored—the full-sample posterior means of parameters are used in analysis. Recently, similar to our approach, several papers have investigated the implications for asset pricing when the agent jointly learns about parameters and states. Johannes, Korteweg, and Polson (2014) investigate sequential learning and return predictability. Johannes, Lochstoer, and Mou (Forthcoming) focus on learning about consumption dynamics. Collin-Dufresne, Johannes, and Lochstoer (2013) study parameter learning in a general equilibrium setup and its implications for asset pricing.

We use the S&P 500 index for inference. The data range from January 2, 1980, to December 31, 2012, and have 8,325 daily observations in total. This dataset includes the 1987 market crash, the 1997 Asian financial crisis, the 2002 dot-com bubble burst, the 2008 global financial crisis, and the recent European debt crisis of 2010 to 2012. A number of important results emerge from our empirical analysis.

First, we find that the evidence of co-jumps between diffusion volatility and asset returns is robust since the market crash of 1987. However, while the data call for co-jumps in returns and jump intensities, the parameters driving the jump intensity are hard to identify. The self-exciting jump intensity has become more important since the 2008 global financial crisis. We find that the speed of learning for the diffusion parameters is remarkably faster than that for the jump parameters. The slow speed of learning and large uncertainty on the jump parameters can be explained by the low arrival rate of extreme events.

Second, learning generates excess volatility, and does so through the jump component. For example, in the full self-exciting model, the average annualized total return volatility is about 18%, of which the jump volatility is about 9.4%. However, if we ignore learning and simply use the full-sample posterior means of the parameters, the average annualized total volatility decreases to 16.6%. This decrease is only from the jump volatility, as the diffusion volatility is

approximately the same as before, whereas the jump volatility reduces to about 7.0%. Having investigated the higher conditional moments of the predictive return distribution, we find that learning makes the predictive distribution more left skewed and leptokurtic. Furthermore, we observe a strong asymmetry in the amount of learning over the tails: the left tail of the return distribution can be well pinned down after the 1987 market crash, whereas there is a great deal of uncertainty on the right-tail behavior throughout the sample.

Third, from a volatility forecasting perspective, the self-exciting jump intensity is found to be important. In comparison with the more restricted specifications, the full self-exciting model always generates smaller root-mean-square-errors (RMSEs) and larger Mincer-Zarnowitz R^2s, whether learning is present or not. Learning also has an important implication for volatility forecasting. For example, when learning is taken into account, the RMSE from the full model is about 4.9%, whereas it reduces to 4.3% when the full-sample posterior means of parameters are used. This comparison allows us to quantify the cost of not knowing parameters in volatility forecasting. However, similar to P. Hansen and Lunde (2005), we find that the GARCH(1,1) model cannot be beaten by the more sophisticated self-exciting models in volatility forecasting.

Fourth, learning and self-exciting jumps have important implications for option pricing. In general, we find that the existence of the self-exciting jump intensity makes the model more flexible in capturing high levels of volatility during periods of financial crisis. This feature is particularly important in pricing short maturity out-of-money put and/or in-the-money call options. Furthermore, learning has a first-order effect on pricing in-the-money and out-of-money short maturity call options, and its effect on the deep out-of-money call options is even stronger than on the deep in-the-money ones. This is closely related to the fact that learning alters the tail behaviors and introduces even larger uncertainty on the right tail of return distributions. Our results are consistent with Benzoni, Collin-Dufresne, and Goldstein (2011), who argue that updating of beliefs on jump parameters can cause permanent shifts in option prices.

The last set of results relates model-implied option prices to observed option prices between January 1996 and December 2012. First, we find that the model with self-exciting jumps fits option prices better than a specification excluding this channel, especially for short-term out-of-the-money puts and during periods of financial crisis. Second, we document that updates in model-implied option prices due to parameter learning are significantly related to observed option prices, even after controlling for model-implied option prices computed with the fixed parameter estimates. These results extend those in Johannes, Lochstoer, and Mou (Forthcoming) to option prices, and suggest that parameter learning is a significant issue for option market participants.

Our work makes two contributions to the literature. First, we conduct a real-time sequential analysis to examine the importance of self-exciting jump intensity. Interestingly, even though the data call for simultaneous jumps

between asset returns and jump intensities from the 1987 market crash onward, the self-exciting jump intensity becomes more important since the onset of the 2008 global financial crisis. Second, we provide novel results on the implications of learning for risk measures, volatility forecasting, and option pricing. Such results are quite relevant in practice, as the agent needs to update her beliefs sequentially over time when new market information arrives.

1. The Self-Exciting Asset Pricing Model

Under a probability space $(\Omega, \mathcal{F}, P)$ and the complete filtration $\{\mathcal{F}_t\}_{t\geq 0}$, the dynamics of asset price, S_t, are governed by the following time-changed stochastic process,

$$\ln S_t/S_0 = \int_0^t \mu_s ds + \left(W_{T_{1,t}} - k_W(1)T_{1,t}\right) + \left(J_{T_{2,t}} - k_J(1)T_{2,t}\right), \tag{1}$$

where μ_t captures the instantaneous mean rate, $W_{T_{1,t}}$ is a time-changed Brownian motion, $J_{T_{2,t}}$ is a time-changed jump component that is time-inhomogeneous, $T_{i,t}$ represents business time and will be discussed below, and $k_W(1)$ and $k_J(1)$ are convexity adjustments for the Brownian motion and the time-homogeneous jump process, respectively, and can be computed from their respective cumulant exponents: $k(u) \equiv \frac{1}{t}\ln\left(E[e^{uL_t}]\right)$, where L_t is either W_t or J_t.

The dynamics in Equation (1) indicate two distinct types of shocks to asset returns: small continuous shocks, captured by a Brownian motion, and large discontinuous shocks, generated by a jump component. In this paper, the time-homogeneous jump component is modeled by the variance gamma process of Madan, Carr, and Chang (1998), which is a stochastic process in the class of infinite-activity Lévy processes. The jump component is important for capturing extreme events and generating return non-normality and implied volatility smile/skew. The empirical study by Li, Wells, and Yu (2008) shows that the infinite-activity Lévy models outperform the affine Poisson jump models. Furthermore, the recent nonparametric works by Aït-Sahalia and Jacod (2009, 2011) and Lee and Hannig (2010) provide strong evidence on infinite-activity jumps in asset returns.

The variance gamma process can be constructed through subordinating a Brownian motion with drift using an independent subordinator

$$J_t = \omega \mathcal{S}_t + \eta \tilde{W}(\mathcal{S}_t), \tag{2}$$

where $\tilde{W}_t$ is a standard Brownian motion, and $\mathcal{S}_t$ is a gamma subordinator $\mathcal{S}_t = \Gamma(t; 1, v)$ with a unit mean rate and variance rate of v. Alternatively, it can be decomposed into the upside component, J_t^+, and the downside component, J_t^-, such that

$$\begin{aligned} J_t &= J_t^+ + J_t^- \\ &= \Gamma_u(t; \mu_u, v_u) - \Gamma_d(t; \mu_d, v_d), \end{aligned} \tag{3}$$

where Γ_u is a gamma process with mean rate μ_u and variance rate v_u, Γ_d a gamma process with mean rate μ_d and variance rate v_d, and

$$\mu_u = \frac{1}{2}\left(\sqrt{\omega^2 + 2\eta^2/v} + \omega\right), \quad v_u = \mu_u^2 v, \tag{4}$$

$$\mu_d = \frac{1}{2}\left(\sqrt{\omega^2 + 2\eta^2/v} - \omega\right), \quad v_d = \mu_d^2 v. \tag{5}$$

The decay rates and the fatness of the right and the left tails are governed by $\lambda_+ = \mu_u / v_u$ and $\lambda_- = \mu_d / v_d$, respectively.

The stochastic business time, $T_{i,t} \equiv \int_0^t V_{i,s-} ds$, captures the randomness of the diffusion variance ($i = 1$) or of the jump intensity ($i = 2$) over a time interval $[0, t]$ (Clark 1973; Carr, Geman, et al. 2003; Carr and Wu 2004). $V_{i,t}$, which should be nonnegative, is the instantaneous variance rate ($i = 1$) or the jump arrival rate ($i = 2$), both of them reflecting the intensity of economic activity and information flow. Stochastic volatility or stochastic jump intensity is generated by replacing calendar time t with business time $T_{i,t}$. The time-changed jump component has the decomposition of $J_{T_{2,t}} = J^+_{T_{2,t}} + J^-_{T_{2,t}}$ and its convexity adjustment term is $k_J(1) T_{2,t} = \left(k_J^+(1) + k_J^-(1)\right) T_{2,t}$.

Recent empirical studies find that a big negative jump in asset prices tends to be associated with an abrupt move in asset variance—that is, co-jumps of prices and volatility (Jacod and Todorov 2010; Todorov and Tauchen 2011). Furthermore, market turmoils seem to indicate that an extreme movement in markets tends to be followed by another extreme movement, resulting in self-exciting jump clustering (Carr and Wu 2011; Aït-Sahalia, Cacho-Diaz, and Laeven Forthcoming). Thus, we propose to allow negative return jumps entering into both diffusion variance and the jump intensity and model the instantaneous variance rate, $V_{1,t}$, and the jump arrival rate, $V_{2,t}$, as follows:

$$dV_{1,t} = \kappa_1(\theta_1 - V_{1,t})dt + \sigma_{11}\sqrt{V_{1,t}}\, dZ_t - \sigma_{12} dJ^-_{T_{2,t}}, \tag{6}$$

$$dV_{2,t} = \kappa_2(\theta_2 - V_{2,t})dt - \sigma_2 dJ^-_{T_{2,t}}. \tag{7}$$

Equation (6) captures stochastic variance of the continuous shocks, where Z_t is a standard Brownian motion and is allowed to be correlated to W_t with a correlation parameter ρ in order to accommodate the diffusion leverage effect. Diffusion variance also depends on the negative return jumps, indicating that there will be an abrupt increase in $V_{1,t}$ once there is a negative jump in asset return. Equation (7) models the stochastic intensity of jumps, which is a mean-reverting pure jump process. The specification implies that the jump intensity relies only on the negative jumps in asset returns. Dependence of diffusion variance and the jump intensity on negative return jumps is consistent with the well-documented empirical regularity in financial markets that investors react more strongly to bad macroeconomic surprises than to good surprises (Andersen, Bollerslev, and Diebold 2007).

The conditional expectation of the jump intensity in Equation (7) can be found as follows:[1]

$$E[V_{2,t}|\mathcal{F}_0]=\frac{\kappa_2\theta_2}{\kappa_2-\sigma_2\mu_d}\left(1-e^{-(\kappa_2-\sigma_2\mu_d)t}\right)+e^{-(\kappa_2-\sigma_2\mu_d)t}V_{2,0}, \tag{8}$$

from which its long-run mean can be obtained by letting $t\to+\infty$,

$$\bar{V}_2=\frac{\kappa_2\theta_2}{\kappa_2-\sigma_2\mu_d}. \tag{9}$$

Equations (8) and (9) indicate that the conditional expectation of the jump intensity is a weighted average between the current intensity, $V_{2,0}$, and its long-run mean, $\bar{V}_2$, and the speed of mean reversion of the jump intensity is controlled by $\kappa_2-\sigma_2\mu_d$. Using Equations (8) and (9), the conditional expectation of diffusion variance in Equation (6) can also be found as follows:

$$\begin{aligned}E[V_{1,t}|\mathcal{F}_0]=e^{-\kappa_1 t}V_{1,0}+\theta_1\left(1-e^{-\kappa_1 t}\right)+\sigma_{12}\mu_d\Bigg[\frac{1-e^{-\kappa_1 t}}{\kappa_1}\bar{V}_2\\ +\frac{e^{-(\kappa_2-\sigma_2\mu_d)t}-e^{-\kappa_1 t}}{\kappa_2-\sigma_2\mu_d-\kappa_1}\left(\bar{V}_2-V_{2,0}\right)\Bigg],\end{aligned} \tag{10}$$

and its long-run mean is given by

$$\bar{V}_1=\theta_1+\frac{\sigma_{12}}{\kappa_1}\mu_d\bar{V}_2. \tag{11}$$

The conditional expectation of diffusion variance consists of two parts, one arising from the square-root diffusion part (the first two terms on the right-hand side in Equation (10)) and the other from negative return jumps (the last term on the right-hand side in Equation (10)). If the jump intensity is constant, the contribution of jumps to the conditional diffusion variance becomes constant over time.

The above model (hereafter SE-M1) indicates that time-varying aggregate return volatility can be traced back to two sources: one arising from time-varying diffusion volatility and the other from the time-varying jump intensity. In this model, the self-exciting behavior is captured through two channels: (i) a negative jump in asset return pushes up the jump intensity, which in turn triggers more jumps in future asset returns; (ii) a negative jump in asset return makes diffusion volatility jump, and this high diffusion volatility tends to facilitate big movements in future asset returns. In contrast, the existing literature allows only one of these channels at a time and is unable to compare

[1] Define $f(t)=e^{\kappa_2 t}E[V_{2,t}|V_{2,0}]$. $f(t)$ can be analytically found by solving the ordinary differential equation

$$f'(t)=\sigma_2\mu_d f(t)+\kappa_2\theta_2 e^{\kappa_2 t},$$

from which we obtain the conditional expectation in Equation (8).

their relative importance. In particular, Eraker, Johannes, and Polson (2003) and Eraker (2004) allow for comovement of return jumps and diffusion volatility through a synchronized Poisson process, while Aït-Sahalia, Cacho-Diaz, and Laeven (Forthcoming) and Carr and Wu (2011) link only the jump intensity to jumps in asset returns.

One of the central questions we are concerned with in the current paper is the dynamic structure of extreme movements in asset returns. In order to explore the issue, we also investigate the following restricted models:

- SE-M2: the self-exciting model where diffusion volatility does not jump—that is, $\sigma_{12}=0$, and the total volatility jump and the self-exciting effect are only from the time-varying jump intensity;
- SE-M3: the self-exciting model where the jump intensity is constant—that is, $V_{2,0}=1, \kappa_2=0$, and $\sigma_2=0$, and the total volatility jump is only from the diffusion volatility process;
- SE-M4: the model that has no volatility jumps and no self-exciting effect—that is, $\sigma_{12}=0$, $V_{2,0}=1, \kappa_2=0$, and $\sigma_2=0$.

2. Bayesian Learning and Belief Updating

Following the suggestion by L. Hansen (2007), we assume that the agent in the market is Bayesian and faces the same belief updating problem as the econometrician. She simultaneously learns about parameters, hidden states, and models sequentially over time when new market observations arrive.

For a given self-exciting model $\mathcal{M}_i$, there is a set of unknown static parameters, Θ, and a vector of the hidden states, $x_t=\{V_{1,t}, V_{2,t}, J_{u,t}, J_{d,t}\}$, where $V_{1,t}$ denotes diffusion variance, $V_{2,t}$ the jump intensity, $J_{u,t}$ the upside jump, and $J_{d,t}$ the downside jump. The market observations include a time series of (log) stock prices, $y_{1:t}=\{\ln S_s\}_{s=1}^{t}$. For each time t, Bayesian learning consists of forming the joint posterior distribution of the hidden states and the static parameters based on information available up to time t,

$$p(x_t,\Theta|y_{1:t},\mathcal{M}_i)=p(x_t|\Theta,y_{1:t},\mathcal{M}_i)p(\Theta|y_{1:t},\mathcal{M}_i), \tag{12}$$

where $p(x_t|y_{1:t},\Theta,\mathcal{M}_i)$ solves the state filtering problem, and $p(\Theta|y_{1:t},\mathcal{M}_i)$ addresses the parameter inference issue. The update of the agent's beliefs therefore corresponds to updating this posterior distribution.

Our self-exciting models are nonlinear and non-Gaussian. Therefore, we design a hybrid particle filter, which is capable of efficiently handling outliers (see Appendix, section A.1, for the detailed algorithm). The decomposition (Equation (12)) suggests a hierarchical framework for model inference and learning. At each time, for a given set of model parameters proposed from some proposal, we can run a particle filter, which delivers the empirical distribution of the hidden states, $p(x_t|\Theta,y_{1:t},\mathcal{M}_i)$, and the estimate of the likelihood, $p(y_{1:t}|\Theta,\mathcal{M}_i)$, that can be used for parameter learning,

$p(\Theta|y_{1:t},\mathcal{M}_i)\propto p(y_{1:t}|\Theta,\mathcal{M}_i)p(\Theta,\mathcal{M}_i)$. To achieve this aim, we rely on the marginalized resample-move approach developed by Fulop and Li (2013). The key point used in this approach is that the likelihood estimate from the particle filter is unbiased (Del Moral 2004). Furthermore, in contrast to traditional Bayesian methods, this approach can easily be made parallel, making it computationally fast and convenient to use.

This particle-based learning approach provides as a natural output an estimate of the marginal likelihood of the new observation

$$p(y_t|y_{1:t-1},\mathcal{M}_i)=\int p(y_t|x_t,\Theta,y_{1:t-1},\mathcal{M}_i)p(x_t|\Theta,y_{1:t-1},\mathcal{M}_i)$$
$$p(\Theta|y_{1:t-1},\mathcal{M}_i)dx_t d\Theta, \tag{13}$$

which summarizes model fit over time (model learning) and can be used to construct a sequential Bayes factor for sequential model comparison. For any models $\mathcal{M}_1$ and $\mathcal{M}_2$, the Bayes factor at time t has the following recursive formula

$$\mathcal{BF}_t\equiv\frac{p(y_{1:t}|\mathcal{M}_1)}{p(y_{1:t}|\mathcal{M}_2)}=\frac{p(y_t|y_{1:t-1},\mathcal{M}_1)}{p(y_t|y_{1:t-1},\mathcal{M}_2)}\mathcal{BF}_{t-1}, \tag{14}$$

which is completely out-of-sample, and can be used for sequential comparison of both nested and non-nested models.

Bayesian learning and belief updating generate persistent and long-term shocks to the agent beliefs. To see this, define $\theta_t=E[\theta|y_{1:t}]$ as the posterior mean of a parameter θ obtained using information up to time t. The iterated expectation indicates

$$E[\theta_{t+1}|y_{1:t}]=E[E[\theta|y_{1:t+1}]|y_{1:t}]=E[\theta|y_{1:t}]=\theta_t. \tag{15}$$

Therefore, θ_t is a martingale, indicating that shocks to the agent beliefs on this parameter are not only persistent but also permanent. Thus, in Bayesian learning, the agent gradually updates her beliefs that the value of a parameter is higher or lower than that previously thought and/or that a model fits the data better than the other.

The Bayesian learning process is initialized by an agent's initial beliefs or the prior distributions. We move the fixed parameters in one block using a Gaussian mixture proposal. Given that in our marginalized approach the likelihood estimate is a complicated nonlinear function of the fixed parameters, conjugate priors are not available. In general, we assume normal distributions for the priors. However, if a parameter under consideration has a finite support, we take a truncated normal as its prior. The hyper-parameters of the prior distributions are calibrated using a training sample—that is, an initial dataset is used to provide information on the location and scale of the parameters. This procedure is initialized by priors with very large variances. The training-sample approach is a common way to generate the objective prior distributions

(O'Hagan 1994). We find that most model parameters, except those controlling the self-exciting jump intensity, κ_2 and σ_2, are not so sensitive to the selection of the priors. Therefore, based on information from the training sample, we give relatively informative priors to κ_2 and σ_2, but give quite flat priors to other parameters. See the Appendix, section A.2, for details of the selection of functional forms and hyper-parameters for the priors, and section A.3 for Monte Carlo and sensitivity studies.

3. Information Flow and Learning

Our Bayesian agent learns about and updates her beliefs on fixed parameters, hidden states, and models as information arrives sequentially over time. We initialize the Bayesian learning process using the priors described in the Appendix, section A.2. Section 3.1 presents the data used for inference. Section 3.2 implements model learning and sequential model comparison, and Section 3.3 presents results on parameter and state learning. More statistical results can be found in the Appendix, section A.4.

3.1 The data

The data used are the S&P 500 stock index ranging from January 2, 1980, to December 31, 2012, with 8,325 daily observations in total. This dataset contains the recent European debt crisis of 2010 to 2012, the global financial crisis in late 2008, the market crash on October 19, 1987 (-22.9%), and other market turmoils. The upper panel of Figure 1 plots the S&P 500 index returns. A striking feature of the data is the high non-normality of the return distribution, with a skewness of -1.2 and a kurtosis of 29.7. The Jarque-Bera test easily rejects the null hypothesis of normality of returns with a very small p-value (less than 0.001).

The lower panel presents realized volatility (RV_t), computed from the previous 21-day (one-month) returns at each time, $RV_t = \sqrt{\frac{252}{21}\sum_{j=0}^{20} R_{t-j}^2}$. The simultaneity of abrupt moves in realized volatility and extreme events in returns is very clear, and turbulent periods tend to be realized through many consecutive large up-and-down return moves. What is hard to gauge is the extent to which these are due to high diffusion volatility or persistent fat tails. The model inference that follows will shed more light on this issue from a Bayesian learning perspective.

3.2 Model learning and sequential comparison

In the Bayesian framework, model comparison can be made by the Bayes factor, defined as the ratio of the marginal likelihoods of two models.[2] This Bayesian

[2] In Bayesian statistics, Jeffreys (1961) gave a scale for interpretation of Bayes factors. For two given models, $\mathcal{M}_1$ and $\mathcal{M}_2$, if the value of the log Bayes factor is between 0 and 1.1, $\mathcal{M}_1$ is barely worth mentioning; if it is

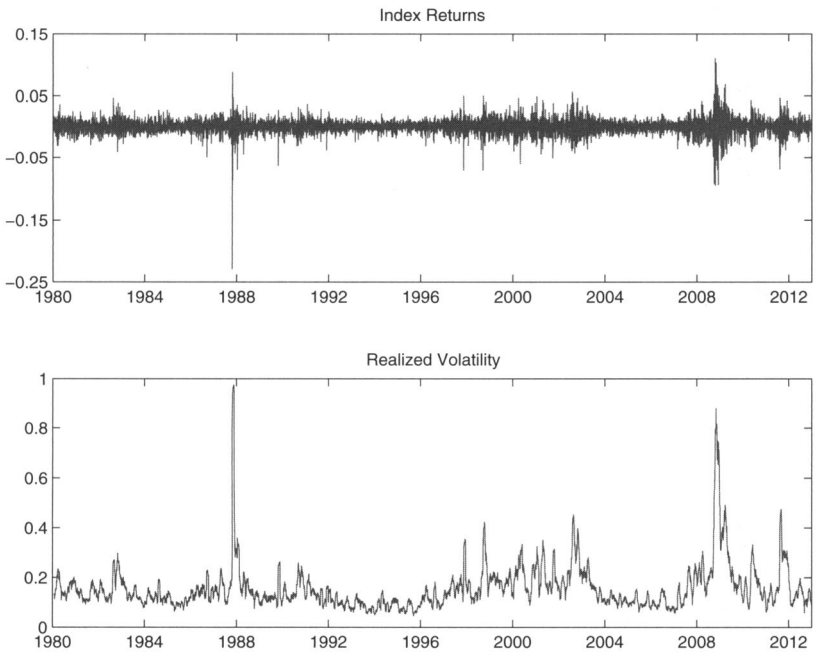

Figure 1
S&P 500 index returns and realized volatility
The figure plots S&P 500 index returns (upper panel) and realized volatility (lower panel). The data range from January 2, 1980, to December 31, 2012. In total, there are 8,325 daily observations. Realized volatility at each time is computed from the previous 21-day (one-month) returns at each point in time, $RV_t = \sqrt{\frac{252}{21}\sum_{j=0}^{20} R_{t-j}^2}$.

Table 1
The full sample log Bayes factors

	SE-M1	SE-M2	SE-M3	SE-M4
SE-M1	0.00	—	—	—
SE-M2	19.9	0.00	—	—
SE-M3	11.4	−8.42	0.00	—
SE-M4	25.5	5.64	14.1	0.00

The table presents the log Bayes factor of the column model to the row model using all available S&P 500 index data from January 2, 1980, to December 31, 2012. The interpretation of values in the table is given in note 2.

approach penalizes unnecessarily complicated models and is completely out-of-sample. Table 1 presents the full information Bayes factors (in log) for the four models investigated using all available data. We find that the SE-M1 model and the SE-M3 model, both of which allow negative return jumps to affect

between 1.1 and 2.3, $\mathcal{M}_1$ is substantially better than $\mathcal{M}_2$; if it is between 2.3 and 3.4, $\mathcal{M}_1$ is strongly better than $\mathcal{M}_2$; if it is between 3.4 and 4.6, $\mathcal{M}_1$ is very strongly better than $\mathcal{M}_2$; and if it is larger than 4.6, $\mathcal{M}_1$ is decisively better than $\mathcal{M}_2$.

diffusion volatility, outperform the SE-M2 model and the SE-M4 model that exclude this channel. For example, the log Bayes factors between the SE-M1 model and the SE-M2/SE-M4 models are about 19.9 and 25.5, respectively, and the log Bayes factors between the SE-M3 model and the SE-M2/SE-M4 models are about 8.4 and 14.1, respectively. Thus, there is decisive evidence in the data for negative return jumps affecting diffusion volatility and co-jumps of returns and volatility. Furthermore, there exists very strong evidence for negative return jumps affecting the jump intensity. Comparing the SE-M1 model, where both self-exciting channels are allowed, to the SE-M3 model where only diffusion volatility is influenced by return jumps, the former is very decisively preferred with a log Bayes factor of 11.4.

The above batch comparison does not tell us how market information accumulates and how different models perform over time. Does one model outperform another in a certain state of economy, but underperform it in other states? Our Bayesian learning approach has a recursive nature and produces the sequential marginal likelihood at each time for each model. We can then construct the sequential Bayes factors and use them for real-time model analysis and comparison.

Figure 2 presents the sequential log Bayes factors that give us a richer picture of model performance over time. A number of important features emerge. First, when market information is scarce in the beginning of the sample, the SE-M1 model performs nearly the same as the other three models despite the fact that it is the best model according to Bayes factors in Table 1.

Second, as the market information accumulates over time, in particular, after the 1987 market crash, the data strongly favor the SE-M1 model that allows negative return jumps to affect both diffusion volatility and jump intensity.

Third, the relative importance of diffusion volatility jumps and self-exciting jump intensities changes over time. This can be seen by comparing the SE-M2 model with the SE-M3 model at the top-right panel of Figure 2. The self-exciting jump intensity is more important over the period from 1992 up to 2001, whereas diffusion volatility jumps begin to dominate after 2001. Furthermore, the lower-left panel presents the sequential comparison between the SE-M1 model and the SE-M3 model. We clearly see the importance of introducing the self-exciting jump intensity after the 1987 market crash, and it becomes even more important after Lehman Brothers' bankruptcy in September 2008, as the log Bayes factor very quickly moves up to about 9.0 from about 3.0.

Fourth, most of the up-moves in Bayes factors happen during market turmoils. This phenomenon is particularly obvious during the 1987 market crash and the 2008 global financial crisis in and indicates that the market participants mainly update their beliefs on model specifications during market turmoils.

3.3 Parameter and state learning

Different from batch estimation, our Bayesian learning approach provides us with the whole picture of how parameter posteriors evolve over time

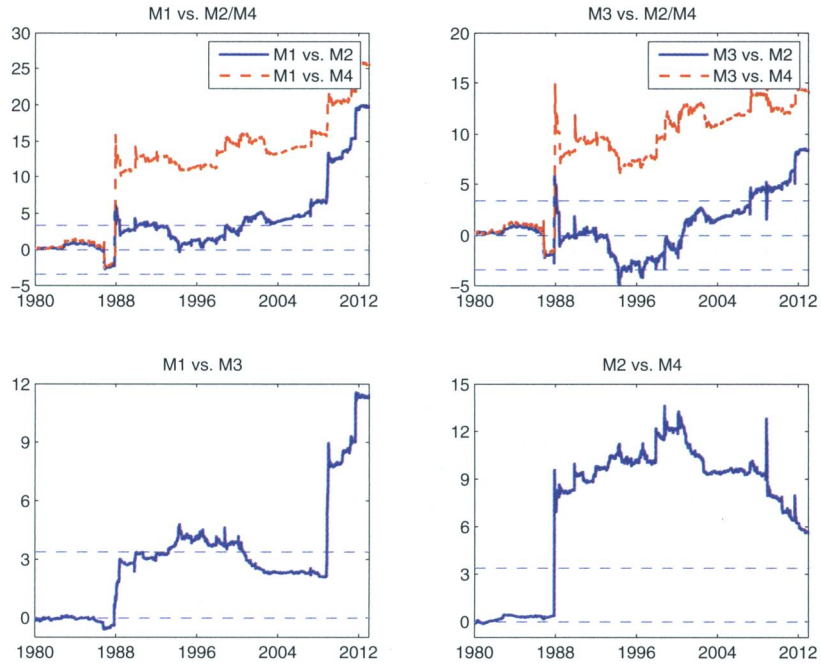

Figure 2
Sequential model momparison
The figure plots the sequential log Bayes factors for sequential model comparison. The straight dashed lines in each panel represent −3.4, 0, and 3.4, respectively, which determine how strongly one model outperforms the other. The statistical interpretation of these values is given in note 2.

with respect to accumulation of information. Figure 3 presents the sequential learning of the fixed parameters in the SE-M1 model, which is the best-performing one. For each parameter, the posterior mean (solid line) and the (5, 95)% credible interval (dashed lines) are reported. We group the model parameters into the diffusion parameter set, $\Theta_D = (\kappa_1, \theta_1, \sigma_{11}, \rho, \sigma_{12})$, and the jump parameter set, $\Theta_J = (\omega, \eta, \upsilon, \kappa_2, \sigma_2)$.

There are a number of notable features. First, the agent's beliefs are quite uncertain in the beginning before the 1987 market crash with large credible intervals for all parameters. Then, as information accumulates, the credible intervals of most parameters become narrower and narrower over time and parameter uncertainty diminishes.[3]

[3] The plots in Figure 2 and Figure 3 seem to suggest that there are "spikes" in model probabilities and parameter posteriors, especially during market turmoils. However, after we take a closer look at these estimates, we find that they are not really spikes per se since they rapidly move up but slowly go down. Take σ_{12} in Figure 3 as an example. If we look at the zoom-in of the 1987 market crash period, we can clearly see that after moving up, it takes more than 30 days for σ_{12} to go down. This is because during the 1987 market crash, as we can see in Figure 1, there is a big negative jump in the index return, and the return volatility suddenly moves up to a very high level. According to our model specification, we need a large value of σ_{12} to achieve the high level of

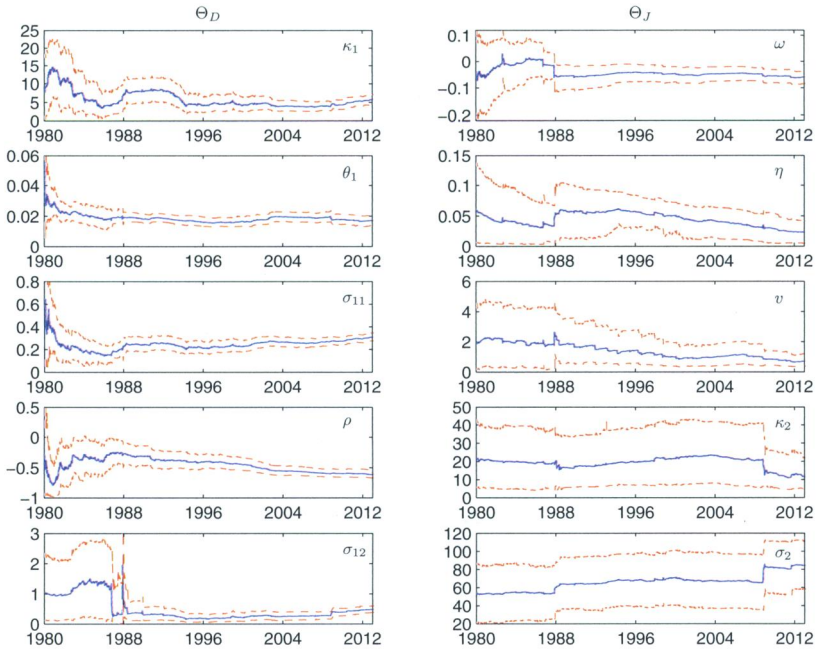

Figure 3
Parameter learning
The figure presents sequential parameter estimates over time in the SE-M1 model using the S&P 500 index starting from January 2, 1980, up to December 31, 2012. In each panel, the posterior means (the solid line) and (5, 95)% quantiles (the dashed lines) are reported.

Second, the speed of learning is quite different across parameters. Learning is remarkably faster for the diffusion parameters than for the jump parameters. We can see that after the 1987 market crash, most of the diffusion parameters are quickly pinned down and have narrow credible intervals. However, for the jump parameters, their credible intervals shrink very slowly. This is particularly obvious for parameters controlling the self-exciting jump intensity, κ_2 and σ_2. The credible intervals of these two parameters barely narrow down over time. We observe a sudden tightening of the credible interval of κ_2 only from the 2008 global financial crisis onward and a little shrinkage of the σ_2's credible interval. The slow learning and large parameter uncertainty of the jump parameters can be explained by the low arrival rate of extreme events and could be important for risk management.

Third, in the SE-M1 model, the total return volatility consists of two components, the diffusion volatility and the jump volatility, which behave quite differently. The diffusion volatility is more persistent and less volatile than the

volatility. Hence, the posterior distribution of σ_{12} moves up. Similar features are also found in the Monte Carlo study reported in the Appendix, section A.3.

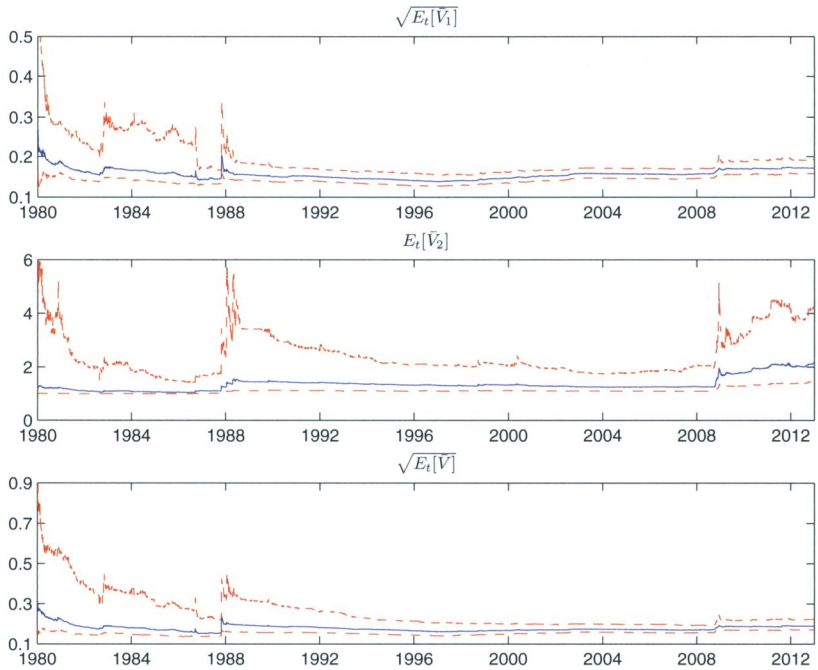

Figure 4
Learning about long-run volatility
The figure presents sequential estimates of the long-run diffusion volatility, the long-run jump intensity, and the long-run return volatility over time in the SE-M1 model, using the S&P 500 index from January 2, 1980, to December 31, 2012. In each panel, the (5, 50, 95)% quantiles are reported.

jump volatility because the learned value of κ_1 is almost always smaller than that of κ_2 and the value σ_2 is particularly large. The evidence of co-jumps between the volatility and the return through the diffusion volatility is robust ever since the 1987 market crash. However, even though the full dataset calls for self-exciting jump intensity as seen in Figure 2, it only becomes really important at the onset of the 2008 financial crisis. The parameters driving the intensity dynamics, in particular σ_2, remain hard to identify throughout.

Finally, the agent's beliefs on the long-run components of uncertainty vary over time. If the agent knows the fixed parameters, her beliefs on the long-run diffusion volatility, the long-run jump intensity, and the long-run return volatility should be constant over time as suggested by Equations (11) and (9). However, our agent is Bayesian, and therefore, parameter learning and uncertainty are directly transferred to her beliefs on the long-run risks. To investigate this point, we present the long-run return volatility and its components in Figure 4. We see from the upper panel that before the 1987 market crash, the long-run diffusion volatility varies dramatically, and the agent is quite uncertain about its value. Since then, it has much less variation and a

narrow 90% credible interval over time, though we observe a significant upward adjustment since 2008. In contrast, the long-run jump intensity in the middle panel changes over time and its 90% credible interval remains large, indicating that the agent is quite uncertain about future jumps. This pattern is even more striking during financial crises. For example, during the 2008 financial crisis, the long-run jump intensity suddenly increases and then slowly wanders up, and its 90% credible interval becomes larger than before. The agent's uncertainty on variance components is directly reflected in her beliefs on the long-run return volatility, which is presented in the lower panel.[4]

Embedded in our learning algorithm is an efficient hybrid particle filter. One advantage of this particle filter is that it can separate positive jumps and negative jumps. This separation is important from both a statistical and a practical perspective. Statistically, it makes our self-exciting models feasible to estimate since both the diffusion volatility and the jump intensity depend only on negative jumps. Practically, investors are mostly concerned about negative jumps. The ability to disentangle the negative jumps provides us with an important tool for risk management.

The left panels of Figure 5 present the filtered diffusion volatility, the jump intensity, and the return volatility using the parameters learned at each time. We can see that whenever there is a big negative return jump, the diffusion volatility and the jump intensity abruptly move up to a high level. However, there are some important differences between the two state variables. The diffusion volatility is well identified with a tight 90% credible interval. In contrast, our ability to pin down the jump intensity is limited as we can see that its credible intervals are wide during the crisis periods. Furthermore, there seems to be an abrupt change in the behavior of the jump intensity since the 2008 crisis. Prior to this episode, during a turbulent period, the credible interval of the jump intensity first widens and then quickly reverts to its long-run mean. Ever since the 2008 crisis, however, it has remained consistently high and wide. This suggests that, as far as the tails are concerned, the recent crisis is special, with a sustained probability of large extreme events going forward. The lower panel presents the return volatility, which is computed as $\sqrt{V_{1,t}+Var_t(J_1)V_{2,t}}$. It has a large credible interval in the beginning, and because of the information accumulation, its credible interval slowly becomes tighter and tighter.

The right panels of Figure 5 present the filtered positive, negative, and return jumps. The filtered negative jumps in the middle panel can effectively capture all market turmoils, such as the 1987 market crash, the 1997 Asian financial crisis, the 2008 financial crisis and the 2010 to 2012 European debt crisis.

---

[4] In Figure 4, we present the (5, 50, 95)% quantiles instead of the posterior means and (5, 95)% quantiles. This is because the long-run volatility components are nonlinear functions of model parameters. Any extreme values of model parameters could result in very unreasonable long-run components, which may dominate the computation of the posterior means such that they could be larger than the 95% quantiles. This issue is particularly striking in the early stage of learning when the market information is minimal and the posterior distributions have large dispersion.

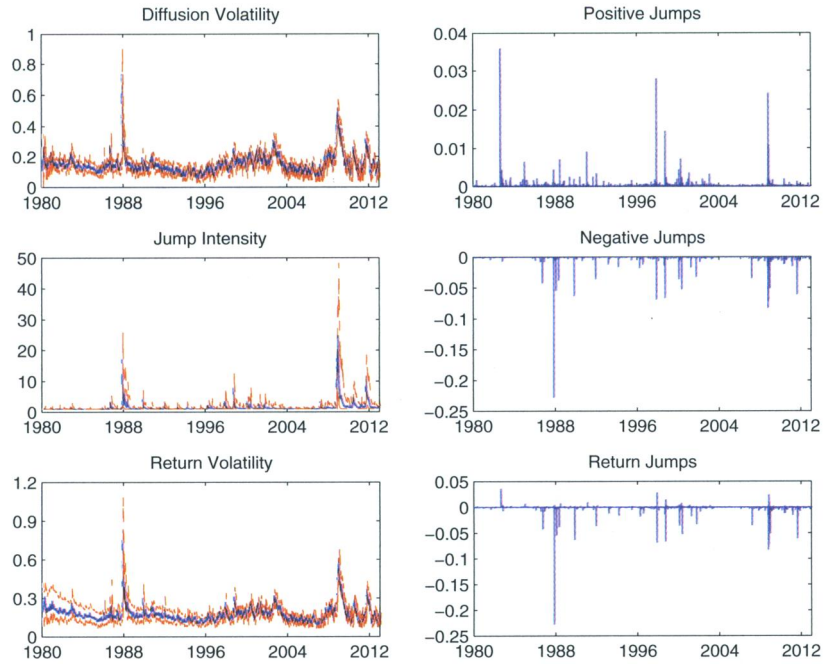

Figure 5
Volatility and jumps learning
The left panels present the sequential posterior means and the (5, 95)% quantiles of the diffusion volatility ($\sqrt{V_{1,t}}$), the jump intensity ($V_{2,t}$), and the return volatility ($\sqrt{V_{1,t}+Var_t(J_1)V_{2,t}}$). The right panels present the sequential posterior means of the positive jumps ($J_{u,t}$), the negative jumps ($J_{d,t}$) and the return jumps ($J_{u,t}+J_{d,t}$) in the SE-M1 model. The data used are the S&P 500 index from January 2, 1980, to December 31, 2012.

However, as shown in the upper and the lower panels, the positive jumps are quite small, most of them less than 1%. This is a new and potentially important empirical result, suggesting that whenever jumps in the diffusion volatility are taken into account, the positive jump component in the index return is not so important and the positive movements in the return can be captured by the diffusion component. This finding reinforces our choice of giving the negative jumps more prominence.

4. Economic and Empirical Implications

4.1 Excess volatility and tail behaviors

In Bayesian learning, the model parameters have quite large 90% credible intervals in the early stage of learning, and they slowly narrow when market information accumulates over time. It is therefore interesting to examine how learning affects return volatility. For this purpose, we consider the following three cases when estimating return volatility. Case I: we take into account both parameter learning and uncertainty. Case II: we only allow for parameter

Table 2
Excess volatility and jump contributions

		Total Vol.	Diff. Vol.	Jump Vol.	Jump Ctr.(%)
SE-M1	Case I	18.0	15.1	9.36	30.2
	Case II	16.7	15.0	7.16	19.9
	Case III	16.6	15.0	7.00	19.6
SE-M2	Case I	17.5	14.8	8.93	29.0
	Case II	16.7	15.0	6.98	20.6
	Case III	16.5	15.0	6.52	18.8
SE-M3	Case I	18.0	15.2	9.09	30.3
	Case II	16.8	15.2	6.66	19.9
	Case III	16.8	15.2	6.63	19.8
SE-M4	Case I	17.7	15.1	8.57	28.3
	Case II	16.4	15.3	5.58	15.0
	Case III	16.4	15.3	5.44	14.4

The table presents the average annualized total return volatility ($\sqrt{V_{1,t}+Var_t(J_1)V_{2,t}}$), its diffusion and jump components ($\sqrt{V_{1,t}}$ and $\sqrt{Var_t(J_1)V_{2,t}}$), and the jump contribution to total return variance in percentage ($Var_t(J_1)V_{2,t}/(V_{1,t}+Var_t(J_1)V_{2,t})$). Three cases are considered. Case I: both parameter learning and uncertainty is taken into account. Case II: only parameter uncertainty is allowed—that is, the full-sample posterior distributions of parameters are used. Case III: both parameter learning and uncertainty are ignored and simply the full-sample posterior means of parameters are used.

uncertainty but not for learning—that is, we use the full-sample posterior distributions of parameters. Case III: we ignore both parameter learning and uncertainty, and instead use the full-sample posterior means of parameters.

Table 2 presents the average annualized total volatility and its components. There are a number of notable findings. First, for all models, the total return volatility is the largest in Case I, when parameter learning and uncertainty are taken into account, and it is the smallest in Case III, when both parameter learning and uncertainty are ignored. For example, in the SE-M1 model, the total return volatility is 18.0% in Case I, but it is only 16.6% in Case III. Second, parameter learning and uncertainty do not have any impact on the diffusion volatility estimate. For example, the average annualized diffusion volatility is about 15.0% in the SE-M1 model, whether parameter learning and/or uncertainty are taken into account or ignored. A similar result can be found for other models as well.

Third, parameter learning and uncertainty have an important impact on the jump volatility estimate. We can clearly see that learning increases the importance of the jump component. For example, the average annualized jump volatility is about 9.4%, contributing 30.2% to the total volatility in the SE-M1 model in Case I. However, if we ignore learning and only take into account parameter uncertainty, the estimated jump volatility becomes 7.2%, which contributes 19.9% to the total volatility in Case II. When we move to Case III where there is no parameter learning and uncertainty, the jump volatility is only about 7.0%, accounting for about 19.6% of total volatility.[5] Comparing

[5] Our estimates of jump contributions to total volatility are larger than those obtained non-parametrically. For example, Huang and Tauchen (2005) find that the jump component takes about 7% of total volatility, and

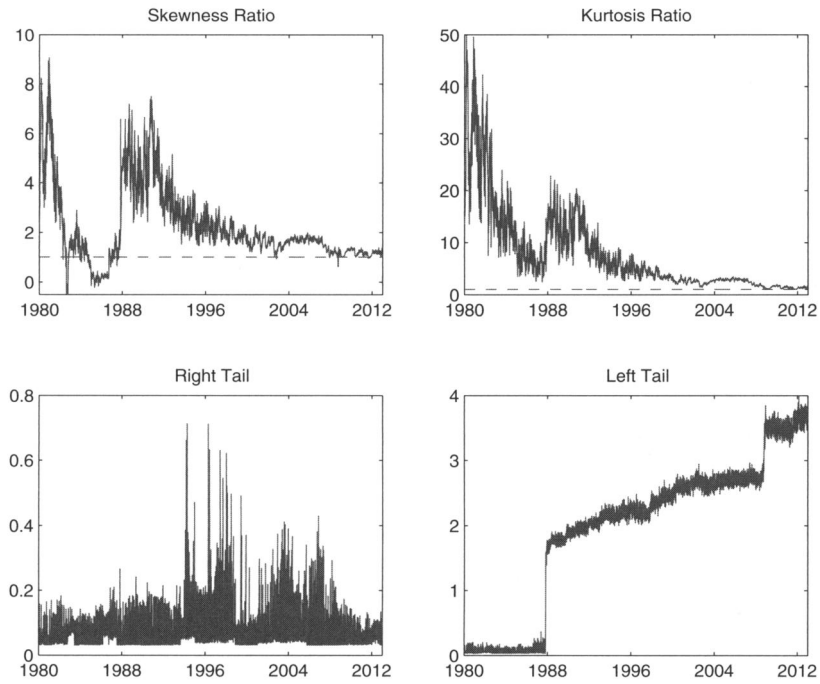

Figure 6
Skewness, kurtosis, and tail indices
The upper panels present the sequential mean ratios of the conditional skewness (left) and the sequential mean ratios of the conditional kurtosis (right), between with and without learning. The dashed lines indicate the level of 1. The lower panels present the sequential ratio of t-statistics of the right-tail index estimates (λ_+) and the left-tail index estimates (λ_-) in the SE-M1 model.

Cases I and II with Case III, it can be seen that learning has a first-order impact compared to uncertainty.

In particular, if we take a look at higher conditional moments, we find that learning generates even more left-skewed and leptokurtic predictive distributions. The upper panels of Figure 6 present the sequential mean ratios of the conditional skewness and the sequential mean ratios of the conditional kurtosis, between with and without learning (Case I and Case III). We can see that nearly all the ratios are larger than one, and this feature is particularly obvious in the early stage of learning when the market information is minimal and after the 1987 market crash. As the market information accumulates, the ratio slowly converges to one. Furthermore, in our model, the right and the left tails are determined by the positive and the negative jump components, which follow gamma processes. The fatness and the decay rates of the two

Andersen, Bollerslev, and Diebold (2007) find that it accounts for about 15% of total volatility. The main reason for this difference is that the data span in both studies is quite short and does not include the 1987 market crash, the 2008 global financial crisis, and the recent European debt crisis.

tails are controlled by λ_+ and λ_-, respectively, for the right and the left tails (Madan, Carr, and Chang 1998). The lower panels of Figure 6 present the sequential t-statistics, defined as the ratios of the posterior means to the posterior standard deviations, for the two tail indices. The t-ratio for the right tail index is always much less than 2. When we take a closer look at the posterior standard deviations of the same parameter, we find them very large. Hence, we can conclude that the information accumulation is not helpful in reducing the uncertainty on the right tail. However, the left tail can be pinned down very quickly after the 1987 market crash, since the t-ratio after October 1987 is larger than 2 and the posterior standard deviation (not reported) gets smaller as the market information accumulates. Clearly, the agent is more confident about the behavior of the left tail.

4.2 Learning and volatility forecasting

In this subsection, we evaluate the relative performance of the four jump models for predicting daily total volatility, and quantify the cost of not knowing the parameters. For each model, expected one-day ahead return variance at time t is given by $E_t[\int_t^{t+\tau} V_{1,s}ds]+Var_t(J_1)E_t[\int_t^{t+\tau} V_{2,s}ds]$, which is known analytically. As a comparison, we also take a look at the performance of the GARCH(1,1) model. True volatility is approximated by the realized one, computed from the previous 21-day (one-month) returns at each time, as in Subsection 3.1.

We again consider the three cases that were defined in the previous subsection. Table 3 reports the forecasting results including the RMSEs, the R^2s from the Mincer-Zarnowitz (MZ hereafter) regressions, and the Diebold-Mariano (DM hereafter) statistics. We have the following findings. First, the SE-M1 model, which takes into account both diffusion volatility jumps and self-exciting jump intensity, always outperforms the other three jump models, whether parameter learning and/or uncertainty are present or not. For example, in Case I, the RMSE and MZ R^2 from the SE-M1 model are 4.94% and 83.8%, respectively, whereas the other three jump models generate larger RMSEs and smaller MZ R^2s.

Second, the existence of parameter learning and/or uncertainty makes volatility more difficult to forecast. This feature holds for all models. The RMSEs (MZ R^2s) are the highest (smallest) in Case I and the smallest (highest) in Case III. For example, for the SE-M1 model, the RMSE is about 4.9% in Case I, whereas it decreases to 4.3% in Case III, and the MZ R^2 is about 83.8% in Case I, while it increases to 85.5% in Case III.

Third, we measure the forecast accuracy using the DM statistic, where the squared error loss function and the heteroscedasticity and autocorrelation consistent (HAC) variance (with 21 lags) are adopted and the benchmark model is chosen to be the SE-M1 model in each case. The forecasting errors in the DM statistic are measured using the residuals from the corresponding MZ regression. The SE-M1 model outperforms the other three jump models in all

Table 3
Volatility forecasting

		SE-M1	SE-M2	SE-M3	SE-M4	GARCH
Case I	RMSE	4.94	5.36	5.49	6.04	2.61
	MZ R^2	83.8	78.4	80.5	69.4	94.0
	DM	–	−1.94	−3.67	−1.55	5.94
Case II	RMSE	4.40	5.24	5.08	5.40	2.34
	MZ R^2	85.9	78.6	82.6	74.8	94.9
	DM	–	−2.00	−2.20	−1.88	3.82
Case III	RMSE	4.36	5.16	5.03	5.23	2.33
	MZ R^2	86.3	78.9	82.8	76.6	95.0
	DM	–	−1.92	−2.14	−1.82	3.91

The table presents the volatility forecasting results. True volatility is approximated by the realized one, computed from the previous 21-day (one-month) returns at each time—that is, $RV_t = \sqrt{\frac{252}{21}\sum_{j=0}^{20} R_{t-j}^2}$. RMSE is the root mean squared error between forecasted and realized volatility. MZ R^2 represents the R^2 from the Mincer-Zarnowitz regression, and DM stands for the Diebold-Mariano statistic, where the squared error loss function and the HAC-type variance with 21 lags are adopted and the benchmark model is chosen to be the SE-M1 model in each case. Three cases are considered. Case I: both parameter learning and uncertainty are taken into account. Case II: only parameter uncertainty is allowed—that is, the full-sample posterior distributions of parameters are used. Case III: both parameter learning and uncertainty are ignored and simply the full-sample posterior means of parameters are used.

three cases, but the significance level is different. For example, the SE-M1 model significantly outperforms the SE-M2 model and the SE-M3 model at the 10% level and the 5% level, respectively, for all three cases. However, the SE-M1 model significantly outperforms the SE-M4 model only at the 10% level in Cases II and III.

Fourth, when comparing the SE-M1 model to the GARCH(1,1) model, we find that the GARCH(1,1) model cannot be beaten by the more sophisticated model, even though the sequential Bayes factors indicate that the SE-M1 model performs much better than the GARCH(1,1) model in modeling the S&P 500 index returns.[6] This result is similar to what P. Hansen and Lunde (2005) find.

4.3 Option pricing implications

4.3.1 Simulation-based results. Now we begin to investigate how self-excitation and learning affect the implied volatility surface. As we only use the underlying return data to estimate the models, the problem of unavailability of the risk-premium parameters remains. For simplicity, we assume that the jump and the volatility parameters remain the same under the change of measure. In the next subsection, we will calibrate the risk-premium parameters using the observed options data. The risk-free interest rate is fixed at 4.00%. At each time starting from January 1981, whenever we obtain the parameter and volatility particles in learning, we use the Monte Carlo method to price call options with

[6] The likelihood function for the GARCH(1,1) model is available in closed form. This makes our Bayesian learning algorithm efficient and fast as we do not need any particle filtering methods. The full-sample log Bayes factor between the SE-M1 model and the GARCH(1,1) model is as large as 275. Furthermore, the outperformance of the GARCH(1,1) model is qualitatively unchanged for the one-week-ahead volatility forecasting.

Table 4
Effects of learning on option pricing

	SE-M1				SE-M3			
K/S	7 Days	30 Days	90 Days	250 Days	7 Days	30 Days	90 Days	250 Days
0.85	1.13	1.05	1.01	1.00	1.02	1.03	1.01	1.00
0.90	1.09	1.02	1.00	1.00	1.01	1.01	1.00	1.00
0.95	1.04	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1.00	1.01	1.01	1.01	1.00	1.01	1.00	1.00	1.00
1.05	1.11	1.06	1.02	1.01	1.06	1.05	1.01	1.01
1.10	1.37	1.12	1.05	1.02	1.30	1.09	1.04	1.01
1.15	1.78	1.21	1.09	1.03	1.74	1.16	1.07	1.02

The table presents the mean ratios of the BS implied volatility between with and without learning (Case I and Case III) in the SE-M1 model and the SE-M3 model. We consider call options with maturity 7, 30, 90, and 250 days and with moneyness (defined as K/S) 0.85, 0.90, 0.95, 1.00, 1.05, 1.10, and 1.15. Option prices are computed using the Monte Carlo simulation method.

maturity 7, 30, 90, and 250 days and with moneyness (defined as K/S) 0.85, 0.90, 0.95, 1.00, 1.05, 1.10, and 1.15. Therefore, both parameter and volatility uncertainties are taken into account at each time in this practice. We price the same cross-section of options again using the full-sample posterior means of the parameters and filtered volatility obtained from these estimates. Thus, in the latter, the parameter learning and uncertainty are ignored.

Table 4 presents the time-series mean of the implied volatility ratios, with and without learning, for each option in the SE-M1 model and the SE-M3 model. For both models, all ratios are either larger than or equal to one, indicating that learning does (positively) affect the option pricing. In particular, we find that the learning effect on the in-the-money and the out-of-the-money options than on the near-the-money options, and it is more pronounced for deep out-of-the-money options than for deep in-the-money options. These results are closely related to the results that learning alters the tail behaviors and introduces even larger uncertainty on the right tail of the predictive return distributions. We also find that in general the learning effect decreases with respect to maturity. Moreover, we find that the main difference between the SE-M1 model and the SE-M3 model is in the short-maturity in-the-money options, indicating the self-exciting jump intensity mainly affects the left tail of return distributions.

To further investigate the learning effect, Figure 7 plots the time series of the implied volatility computed from the SE-M1 model (left panels) and the SE-M3 model (right panels). The solid and the dashed lines plot the implied volatility for contracts with maturity seven days when learning is taken into account and when it is ignored, respectively. We first focus on the SE-M1 model. As shown in the middle panel, learning does not seem to have a first-order effect on pricing the at-the-money options, except during the market crash in October 1987 when the implied volatility from learning is much higher. However, the picture is starkly different in the upper and the lower panels for the implied volatility of the deep in-the-money and the deep out-of-the-money call options. Learning does have a first-order effect here.

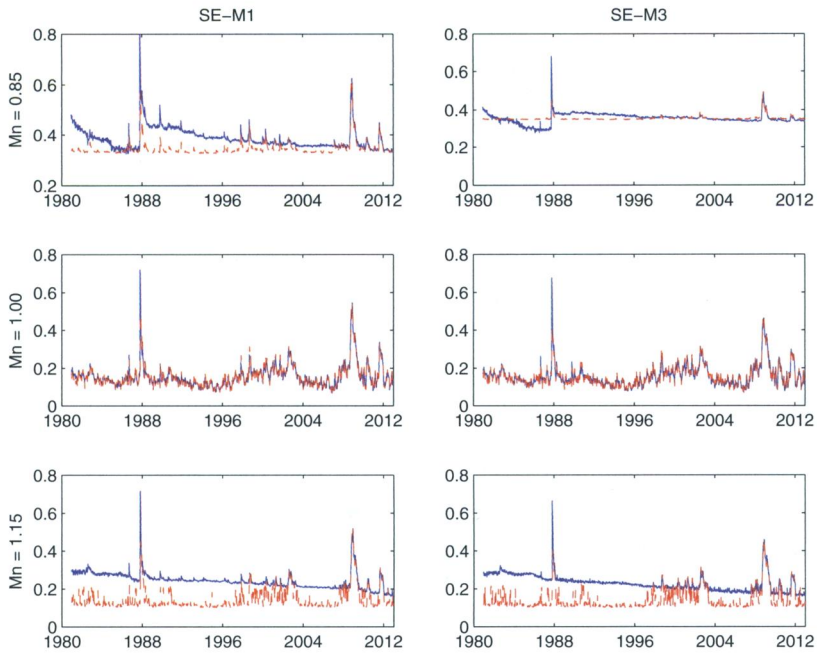

Figure 7
Learning and implied volatility
The figure plots the time series of implied volatility for call options with maturity seven days computed from the SE-M1 model (left panels) and the SE-M3 model (right panels). Options are priced using the Monte Carlo simulation method. The solid line plots the implied volatility when learning is taken into account, whereas the dashed (red) line plots the implied volatility when learning is ignored. We consider options with moneyness (K/S) equal to 0.85 in the upper panels, 1.00 in the middle panels, and 1.15 in the lower panels.

For deep in-the-money options, when learning is ignored, the implied volatility has little variation with abrupt bursts and drop-backs during the financial crisis periods, whereas when learning is taken into account, the implied volatility moves up to high levels during the crisis periods and stays there for a very long time. The deep in-the-money options are sensitive to extreme downside movements. These observations seem to indicate that learning leads to a long-lasting shift of beliefs on the left tail of the predictive return distribution. Benzoni, Collin-Dufresne, and Goldstein (2011) argue that updating of beliefs about jump parameters may cause a permanent shift in option prices. Our investigation reinforces this intuition. For the deep out-of-money options, the implied volatility is in general larger when learning is allowed than when learning is ignored, indicating learning can also generate the fat right tail. Overall, the above results suggest that parameter learning and uncertainty are likely to have important implications for pricing options that depend on the tails of the predictive return distribution. However, as the sample size grows, the effect of learning diminishes. But the rate of the diminishing effect differs for the in-the-money compared to the out-of-money options. For the in-the-money

options, the implied volatility is similar, whether learning is allowed or ignored. This indicates that the left-tail uncertainty vanishes at the end of the sample. However, this is not the case for the out-of-money options, indicating that the right-tail uncertainty remains. This result is consistent with what we have found in Figure 6.

We now further compare the SE-M1 model and the SE-M3 model and examine what roles the self-exciting jump intensity plays. Comparing the right panels to the left ones in Figure 7, we have the following findings. First, as seen from the middle panels, regardless of learning, the SE-M1 model prices the at-the-money options quite similarly to the SE-M3 model during the calm periods, whereas during the crisis periods, the SE-M1 model seems to be more flexible to capture high levels of volatility than the SE-M3 model. Second, when pricing the deep in-the-money options, the upper panels indicate that (i) the SE-M1 model is more flexible at tracking fluctuations of volatility and at capturing high levels of volatility during the financial crisis; and (ii) learning can generate an even more persistent and fatter left tail in the short term in the SE-M1 model compared to the SE-M3 model. Third, the SE-M1 model and the SE-M3 model produce similar deep out-of-money option prices, whether learning is considered or not.

However, for options with longer maturities, the learning effect becomes smaller, and both models perform quite similarly at different moneyness (not reported). The findings are consistent with what we have found in Table 4 and imply that the two models differ mainly in their ability to price the short maturity options.

4.3.2 Real data applications. In addition to the simulation-based results, we also use real option data to evaluate the model performance and learning effects. The S&P 500 index option data are obtained from the OptionMetrics volatility surface that provides daily call and put prices, BS implied volatility for the standardized maturities, and deltas between January 1996 and December 2012. In accordance with the option pricing literature, we only keep Wednesday put prices and exclude options with maturity larger than a half year.

Option prices not only reflect the agent's expectations over the evolution of the underlying but also contain risk-premium information. In derivative pricing theory, they are reflected in the stochastic discount factor (SDF) that drives a wedge between the real world and the pricing measures. We assume a simple reduced-form SDF with two free parameters: γ_J, the jump-risk-premium parameter distorting the Lévy density of the jump component, and γ_V, the risk-premium parameter connected to the diffusion volatility shocks (see Appendix, section A.5, for the change of measure).[7]

[7] Under this change of measure, only ω, η, κ_1, and θ_1 are different, and the other parameters remain unchanged. The jump-risk-premium parameter is bounded, $-\lambda_+ < \gamma_J < \lambda_-$, such that the risk-neutral tail indices, $\lambda_+^{\mathcal{Q}} = \lambda_+ + \gamma_J$ and $\lambda_-^{\mathcal{Q}} = \lambda_- - \gamma_J$, are positive.

To simplify the analysis, we assume that agents know the exact risk-premium parameters and only learn about the other parameters of the system, which are obtained from Section 3. Furthermore, we do not take into account the equilibrium implications of learning for risk premia, and hence, our treatment is in the spirit of Johannes, Korteweg, and Polson (2014) and Johannes, Lochstoer, and Mou (Forthcoming). The risk-premium parameters are calibrated using option prices. A consistent treatment of pricing should condition on the agent's information set. In our case, however, the complexity of learning over states and parameters results in an infinite dimensional state space similar to that in Johannes, Lochstoer, and Mou (Forthcoming), which makes the problem intractable. Hence, we resort to a simplified treatment of option pricing learning, where we integrate over state and parameter uncertainty, but risk aversion is taken into account through two fixed parameters, leading to a deterministic shift of the distribution under the risk-neutral measure.

To calibrate the risk-premium parameters, we compute model-implied option prices on a two-dimensional grid of risk premia using equidistant grid points in both directions. For each specification, we pick the risk-premium parameters minimizing the mean squared errors between the model-implied and observed option prices, weighted by the BS vega,

$$\hat{\gamma}_J, \hat{\gamma}_V = \arg\min_{\gamma_J, \gamma_V} \frac{1}{N} \sum_{i=1}^{N} \left(\frac{O_i^{obs} - O_i^M}{Vega_i} \right)^2 \tag{16}$$

where N is the number of options, O_i^{obs} is the observed option price, and O_i^M is the model-implied option price. Given that the SE-M1 model and the SE-M3 model dominate the other models, we focus on these two specifications in what follows.

Table 5 reports the calibrated risk premia and the implied risk-neutral parameters both for the case when learning is taken into account and for the case when parameter uncertainty is ignored and the full sample posterior means are plugged into the option pricing routine. For the SE-M1 model, the calibrated jump risk premium, γ_J, is positive under both cases. A positive γ_J indicates that the risk-neutral return distribution is more left-skewed and more leptokurtic than the physical counterpart. However, one can see in the table that the implied risk-neutral jump parameters, ω and η, stay well within the central 90% credible intervals of their physical counterparts, suggesting that there is a greater degree of statistical uncertainty about the exact location of the jump-risk-premium parameter. This result is unsurprising given that jumps are rare events and we only use underlying return data to estimate the physical jump parameters. By contrast, the diffusion volatility risk premium, γ_V, is negative, resulting in a slower mean reversion and a higher stationary mean risk-neutral diffusion volatility process than the physical one. Furthermore, the implied risk-neutral parameters are well outside the central 90% credible intervals of their physical counterparts, pointing towards the reliability of the calibrated risk premium.

Table 5
Risk premia and implied risk-neutral parameters

A. SE-M1

				Case I	Case III
γ_J				3.85	3.42
γ_V				−8.42	−9.47
	Full-Sample Posterior (Physical)				
	Mean	5th Prctile	95th Prctile	Implied Risk-Neutral Mean	
ω	−0.058	−0.084	−0.035	−0.072	−0.070
η	0.023	0.004	0.041	0.025	0.025
κ_1	5.793	4.555	7.062	3.174	2.847
θ_1	0.017	0.013	0.020	0.031	0.035

B. SE-M3

				Case I	Case III
γ_J				3.00	−1.14
γ_V				−10.52	−10.52
	Full-Sample Posterior (Physical)				
	Mean	5th Prctile	95th Prctile	Implied Risk-Neutral Mean	
ω	−0.062	−0.088	−0.039	−0.079	−0.058
η	0.023	0.004	0.043	0.026	0.022
κ_1	4.386	3.438	5.426	1.011	1.011
θ_1	0.021	0.017	0.025	0.092	0.092

This table presents the calibrated risk premia from a grid search on S&P 500 index option data obtained from the OptionMetrics volatility surface on Wednesdays between January 1996 and December 2012. Results are shown both for the SE-M1 model and the SE-M3 model and both when parameter learning is taken into account and when parameter learning and uncertainty are ignored. The risk-neutral parameters implied by the calibrated risk premia are also reported.

For the SE-M3 model, under the case of parameter learning, we obtain similar results to the SE-M1 model. However, when we ignore parameter learning and simply use the fixed full-sample posterior means, the calibration becomes more involved and we get a negative γ_J, further indicating unreliability of the jump-risk-premium calibration and misspecification of the SE-M3 model.

Table 6 reports the RMSEs between the model-implied and the observed implied volatility across maturities and strikes for the SE-M1 model and SE-M3 model, when parameter learning is taken into account. We find that the SE-M1 model always provides a better fit than the SE-M3 model, with a root mean squared error of 3.2%, compared to 3.4% for the SE-M3 model. Furthermore, in line with the simulation-based results, the advantage of the SE-M1 model is particularly important in pricing the short-term out-of-the-money puts for which the left tail of the predictive return distributions plays a critical role. For instance, for the 30-day options with the delta equal to −0.2, it has a RMSE of 4.7%, compared to 5.4% for the SE-M3 model.

Having taken a closer look at the behavior of the two models across different episodes, we see that the dominance of the SE-M1 model is concentrated in the period since the 2008 financial crisis. Allowing for the self-exciting jump intensity leads to substantially higher option prices and helps reduce the gap between the model-implied and the observed option prices. The

Table 6
Option pricing errors with learning

Δ	30 days	60 days	91 days	122 days	152 days
A. SE-M1					
−0.8	3.34	2.83	2.81	2.86	3.05
−0.7	2.88	2.67	2.72	2.82	3.04
−0.6	2.94	2.72	2.74	2.82	3.01
−0.5	3.16	2.88	2.85	2.89	3.02
−0.4	3.51	3.17	3.09	3.07	3.14
−0.3	4.02	3.66	3.49	3.41	3.42
−0.2	4.71	4.33	4.08	3.92	3.84
B. SE-M3					
−0.8	3.36	2.74	2.79	2.98	3.20
−0.7	2.94	2.63	2.73	2.95	3.20
−0.6	3.09	2.75	2.78	2.94	3.17
−0.5	3.42	2.99	2.93	3.00	3.15
−0.4	3.88	3.38	3.21	3.17	3.23
−0.3	4.52	3.97	3.68	3.53	3.48
−0.2	5.40	4.80	4.39	4.12	3.95

The table presents the root mean squared option pricing errors in percentage of the S&P 500 index put options for different strikes and maturities for the SE-M1 model and the SE-M3 model with parameter learning taken into account. The option data are from the OptionMetrics volatility surface on Wednesdays between January 1996 and December 2012. For each model, the risk-premium parameters are the optimal ones from the grid search.

increased importance of self-excitation since 2008 is not limited to the short-term OTM puts. The overall RMSE across maturities and strikes since the Lehman bankruptcy (September 15, 2008) is 3.3% for the SE-M1 model versus 3.7% for the SE-M3 model, mirroring a widening gap in overall performance between the two models. This finding reinforces our previous result indicated in Figure 2, which suggests that the self-exciting jump intensity becomes even more important after the 2008 financial crisis as measured by sequential Bayes factors.

When the full-sample posterior means of the fixed parameters are used in calibrating the risk-premium parameters and pricing options (not reported), we again find that the SE-M1 model dominates the SE-M3 model. However, the difference in RMSEs between learning and no-learning is smaller in the real data than in the simulated data. This is not surprising as our observed options data only start in January 1996, by which date parameter uncertainty has substantially decreased as seen from parameter learning in Figure 3.

Johannes, Lochstoer, and Mou (Forthcoming) propose a test procedure to detect the impact of parameter learning on asset prices. In particular, they regress equity returns on belief updates with parameter learning, while controlling for belief updates in the fixed-parameter case. A significant coefficient is interpreted as evidence for the importance of parameter learning for asset prices. Here we mimic their approach for option prices where the agent's information set consists of past stock returns. In particular, we are interested in whether updates in model-implied option prices, due to parameter learning, have any explanatory power for observed option prices over and above model-implied option prices in the fixed parameter case. To simplify exposition

Table 7
Belief updating and implied volatility

	30 days	60 days	91 days	122 days	152 days
Intercept	−0.00	−0.01	−0.02	−0.03*	−0.04*
	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)
$IV_{t,\tau}^{M,Fixed}$	1.09***	1.08***	1.10***	1.12***	1.15***
	(0.04)	(0.06)	(0.07)	(0.07)	(0.09)
$\hat{\varepsilon}_{t,\tau}$	0.89**	1.17***	1.26***	1.33***	1.13***
	(0.38)	(0.30)	(0.27)	(0.26)	(0.28)

The table presents the results from the regressions of the observed implied volatilities (IVs) on the innovations in model-implied IVs due to parameter learning. For each maturity, the dependent variable is the observed average IVs across strikes on each day. The control variable is the model-implied counterpart from the SE-M1 model when parameter uncertainty is ignored. The variable of interest is the model-implied counterpart with parameter learning. This latter is orthogonalized by using the residual from a regression of the model-implied IVs with learning on a constant and the model-implied IVs with the fixed parameters. The option data are from the OptionMetrics volatility surface on Wednesdays between January 1996 and December 2012. HAC standard errors (Newey-West, 30 lags) are reported in parentheses, * denotes the significance at the 10% level, ** denotes the significance at the 5% level, *** denotes the significance at the 1% level.

of the results, we collapse the cross-section of options by looking at the cross-sectional average option price for any given maturity date and focus on the SE-M1 model.[8]

Denote the cross-sectional average model-implied implied volatility at time t for a given maturity τ when learning is taken into account as $IV_{t,\tau}^{M,Learning}$. The same quantity, when the full sample posterior means are plugged into the option pricing routine, is denoted by $IV_{t,\tau}^{M,Fixed}$. And the observed counterpart is denoted by $IV_{t,\tau}^{Obs}$. In the first stage, we want to focus on the variability in model-implied IVs due solely to parameter learning. Hence, we run the following time-series regression:

$$IV_{t,\tau}^{M,Learning} = \alpha_{1,\tau} + \beta_{1,\tau} IV_{t,\tau}^{M,Fixed} + \varepsilon_{t,\tau}. \tag{17}$$

We estimate this regression separately for each maturity, τ, and take the residual from the ordinary least squares (OLS) regression, $\hat{\varepsilon}_{t,\tau}$, as the option price variability that is entirely due to parameter learning. Then in the second stage, we run an OLS regression of the observed IVs on a constant, the model-implied IVs with the fixed parameters, and the residuals obtained from the first-stage regression in Equation (17) as follows:

$$IV_{t,\tau}^{Obs} = \alpha_{2,\tau} + \beta_{2,\tau} IV_{t,\tau}^{M,Fixed} + \beta_{3,\tau} \hat{\varepsilon}_{t,\tau} + \xi_{t,\tau}. \tag{18}$$

The estimation results are reported in Table 7. We can clearly see that the coefficient of $\hat{\varepsilon}_{t,\tau}$, $\beta_{3,\tau}$, is highly statistically significant in each maturity. This indicates that updates in beliefs due to parameter learning have a highly significant effect on the observed IVs across all maturities, over and above the variability in model-implied IVs computed using the fixed parameters.

[8] The strike-specific results and results for the SE-M3 model are similar and are available upon request.

5. Concluding Remarks

We propose a self-exciting asset pricing model that takes into account co-jumps between prices and volatility and self-exciting jump clustering. A Bayesian learning approach is employed to implement a real-time sequential analysis. We find that the evidence of co-jumps between volatility and asset returns through diffusion volatility is robust ever since the 1987 market crash. Interestingly, while the data call for simultaneous jumps between asset returns and jump intensities from the 1987 market crash onward, the self-exciting jump intensity has become more important since the onset of the 2008 global financial crisis.

The new asset pricing model and the Bayesian learning approach allow us to investigate implications of learning for a variety of asset pricing applications. In this paper, we provide novel results on implications of learning for risk measures, volatility forecasting, and option pricing. Such results are quite relevant in practice as market participants need to update their beliefs sequentially over time when new market information arrives.

Our results suggest several interesting research directions. First, it would be interesting to examine what we can find if option prices are included in the learning procedure. This could help better identify the jump intensity and speed up the learning process. Second, the sequential nature of our joint parameter and state learning routine promises several practical applications, such as derivative pricing or portfolio allocation.

Appendix

A.1. A Hybrid Particle Filter

Our model can be cast into a state-space model framework. After discretizing the return process for a time interval τ using the Euler method, we have the following observation equation:

$$\ln S_t = \ln S_{t-\tau} + \left(\mu - \frac{1}{2} V_{1,t-\tau} - k(1) V_{2,t-\tau}\right)\tau + \sqrt{\tau V_{1,t-\tau}}\, w_t + J_{u,t} + J_{d,t}, \tag{A1}$$

where w_t is a standard normal noise, and $J_{u,t}$ and $J_{d,t}$ represent the upside and downside jump noises.

We take the diffusion variance $V_{1,t}$, the jump intensity $V_{2,t}$, and the upside/downside jumps $J_{u,t}/J_{d,t}$ as the hidden states. The diffusion variance and the jump intensity follow Equations (6) and (7), and the upside/downside jumps are gammas. After discretizing, we have the state equations as follows:

$$V_{1,t} = \kappa_1\theta_1\tau + (1-\kappa_1\tau)V_{1,t-\tau} + \sigma_{11}\sqrt{\tau V_{1,t-\tau}}\, z_t - \sigma_{12} J_{d,t}, \tag{A2}$$

$$V_{2,t} = \kappa_2\theta_2\tau + (1-\kappa_2\tau)V_{2,t-\tau} - \sigma_2 J_{d,t}, \tag{A3}$$

$$J_{u,t} = \Gamma(\tau V_{2,t-\tau}; \mu_u, v_u), \tag{A4}$$

$$J_{d,t} = -\Gamma(\tau V_{2,t-\tau}; \mu_d, v_d), \tag{A5}$$

where z_t is a standard normal noise, which is correlated to w_t in Equation (A1) with the correlation parameter ρ. In empirical analysis, we normalize θ_2 to 1 in order to alleviate the identification problem, as the time-homogeneous jump component has non–unit variance rate.

The above model is clearly nonlinear and non-Gaussian. Therefore, we use a particle filter to estimate the likelihood and the hidden states. The most commonly used particle filter is the bootstrap filter of Gordon, Salmond, and Smith (1993), which simply takes the state transition density as the proposal density. However, the bootstrap filter is known to perform poorly when the observation is informative on the hidden states. Our model has this feature because when we observe a large move in asset price, the jump can be largely pinned down by this observation. On the other hand, when the return is small, it is almost due to the diffusion component and contains little information on the jump. Hence, to provide an efficient sampler, we use an equally weighted two-component mixture as the proposal on the jump: the first component is a normal draw, equivalent to sampling from the transition density of the diffusion component, and the second component involves drawing from the transition law of the jump. We need this second component to stabilize the importance weights for small returns. Otherwise, we would compute the ratio of a normal and a gamma density in the importance weights which is unstable around zero. When the return is positive, we use this mixture as the proposal for the positive jump and the transition density for the negative jump, and vice versa.

The algorithm of the proposed hybrid particle filter consists of the following steps:

Step 1: Initialize at $t=0$: set initial particles to be $\left\{V_{1,0}^{(i)}=\theta_1; V_{2,0}^{(i)}=1; J_{u,0}^{(i)}=0; J_{d,0}^{(i)}=0\right\}_{i=1}^{M}$ and give each set of particles a weight $1/M$;

Step 2: For $t=1,2,\ldots$

- If $R_t=\ln S_t-\ln S_{t-\tau}>0$,
  - draw $J_{d,t}^{(i)}$ from its transition law (A5);
  - draw $J_{u,t}^{(i)}$ both from its transition law (A4) and its conditional posterior distribution $J_{u,t}=\ln S_t-\ln S_{t-\tau}-(\mu-\frac{1}{2}V_{1,t-\tau}-k(1)V_{2,t-\tau})\tau-J_{d,t}-\sqrt{\tau V_{1,t-\tau}}w_t$, which is normally distributed. Equal weights are attached to particles obtained from the transition law and the conditional posterior;
  - compute the particle weight by

$$w_t^{(i)}=\frac{p(\ln S_t|J_{u,t}^{(i)},J_{d,t}^{(i)},V_{1,t-\tau}^{(i)},V_{2,t-\tau}^{(i)})p(J_{u,t}^{(i)}|V_{2,t-\tau}^{(i)})}{0.5p(J_{u,t}^{(i)}|V_{2,t-\tau}^{(i)})+0.5\phi(\bar{\mu},\bar{\sigma})},$$

    where $\phi(\cdot,\cdot)$ represents the normal density with mean $\bar{\mu}=\ln S_t-\ln S_{t-\tau}-(\mu-\frac{1}{2}V_{1,t-\tau}^{(i)}-k(1)V_{2,t-\tau}^{(i)})\tau-J_{d,t}^{(i)}$ and standard deviation $\bar{\sigma}=\sqrt{\tau V_{1,t-\tau}^{(i)}}$;
- Otherwise, if $R_t=\ln S_t-\ln S_{t-\tau}<0$,
  - draw $J_{u,t}^{(i)}$ from its transition law (A4);
  - draw $J_{d,t}^{(i)}$ both from its transition law (A5) and its conditional posterior distribution $J_{d,t}=\ln S_t-\ln S_{t-\tau}-(\mu-\frac{1}{2}V_{1,t-\tau}-k(1)V_{2,t-\tau})\tau-J_{u,t}-\sqrt{V_{1,t-\tau}}w_t$, which is normally distributed. Equal weights are attached to particles obtained from the transition law and the conditional posterior;
  - compute the particle weight by

$$w_t^{(i)}=\frac{p(\ln S_t|J_{u,t}^{(i)},J_{d,t}^{(i)},V_{1,t-\tau}^{(i)},V_{2,t-\tau}^{(i)})p(J_{d,t}^{(i)}|V_{2,t-\tau}^{(i)})}{0.5p(J_{d,t}^{(i)}|V_{2,t-\tau}^{(i)})+0.5\phi(\bar{\mu},\bar{\sigma})},$$

    where $\phi(\cdot,\cdot)$ represents the normal density with mean $\bar{\mu}=\ln S_t-\ln S_{t-\tau}-(\mu-\frac{1}{2}V_{1,t-\tau}^{(i)}-k(1)V_{2,t-\tau}^{(i)})\tau-J_{u,t}^{(i)}$ and standard deviation $\bar{\sigma}=\sqrt{\tau V_{1,t-\tau}^{(i)}}$;

Table A1
The prior distributions

	F. Form	Support	(μ_0, σ_0)		F. Form	Support	(μ_0, σ_0)
μ	Normal	$(-\infty, \infty)$	(0.07, 0.15)				
κ_1	Tr. Normal	$(0, \infty)$	(5.00, 7.00)	ω	Normal	$(-\infty, \infty)$	(−0.05, 0.10)
θ_1	Tr. Normal	$(0, \infty)$	(0.03, 0.06)	η	Tr. Normal	$(0, \infty)$	(0.03, 0.06)
σ_{11}	Tr. Normal	$(0, \infty)$	(0.30, 0.60)	υ	Tr. Normal	$(0, \infty)$	(0.80, 2.00)
ρ	Tr. Normal	$[-1, 1]$	(-0.50, 0.60)	κ_2	Tr. Normal	$(0, \infty)$	(15.0, 10.0)
σ_{12}	Tr. Normal	$(0, \infty)$	(0.50, 1.00)	σ_2	Tr. Normal	$(0, \infty)$	(55.0, 20.0)

- Normalize the weight: $\tilde{w}_t^{(i)} = w_t^{(i)} / \sum_j^M w_t^{(j)}$;

Step 3: Resample (stratified resampling)

- Draw the new particle indexes by inverting the cumulative distribution function of the multinomial characterized by $\tilde{w}_t^{(i)}$ at the stratified uniforms $\frac{i+U^{(i)}}{M}$, where $U^{(i)}$ are i.i.d. uniforms;
- reset the weight to $1/M$;

Step 4: Update the diffusion variance and the jump intensity particles using Equations (A2) and (A3), where $z_t = \rho w_t + \sqrt{1-\rho^2}\tilde{z}_t$ with $\tilde{z}$ being an independent standard normal noise.

A.2. The Priors and Posteriors

The Bayesian learning procedure is initialized by the priors. In the full SE-M1 model, there are 11 parameters, among which κ_1, θ_1, σ_{11}, σ_{12}, η, υ, κ_2, and σ_2 need to be positive and ρ needs to be in $[-1, 1]$. We assume normal distributions for the priors. However, if a parameter under consideration has a finite support, we take a truncated normal as its prior.

The hyper-parameters are calibrated using a training sample from January 1975 to December 1979. As a result, we use quite flat priors for most of the parameters except those controlling dynamics of the jump intensity, κ_2 and σ_2. Table A1 presents the exact functional form and hyper-parameters for the prior distribution of each parameter.

Figure A1 presents the prior and full-sample posterior distributions for each parameter. We can see that for most parameters, even though the priors (dashed lines) are quite flat, the dispersions of the posterior distributions (solid lines) are very small. However, for the parameters controlling the jump intensity, κ_2 and σ_2, we use quite informative priors. Using the full sample, κ_2 seems to be well pinned down. This is consistent with what we got from learning that indicates that its credible interval shrinks dramatically after the 2008 financial crisis. However, σ_2 remains difficult to be identified because the posterior distribution still has quite large dispersion.

A.3. Monte Carlo and Sensitivity Studies

In implementation of our Bayesian learning approach, we need to choose the number of state particles, M, the number of parameter particles, N, and the thresholds for resample and move, N_1 and N_2, respectively. As discussed in Andrieu, Doucet, and Holenstein (2010) and Fulop and Li (2013), M is linearly related to the largest sample size (T) that one wants to tackle. Pitt et al. (2012) provide practical guidelines on how to choose the optimal number of state particles (M). However, there are not any guidelines on how to choose N, N_1, and N_2. In this Appendix, we implement Monte Carlo studies to see how well the algorithm works on the self-exciting model and how sensitive it is to N, N_1, and N_2.

We take the SE-M1 model as an example. There are 11 parameters in total, $\Theta = (\mu, \kappa_1, \theta_1, \sigma_{11}, \rho, \sigma_{12}, \omega, \eta, \upsilon, \kappa_2, \sigma_2)$. For each simulation in each Monte Carlo study, we generate a

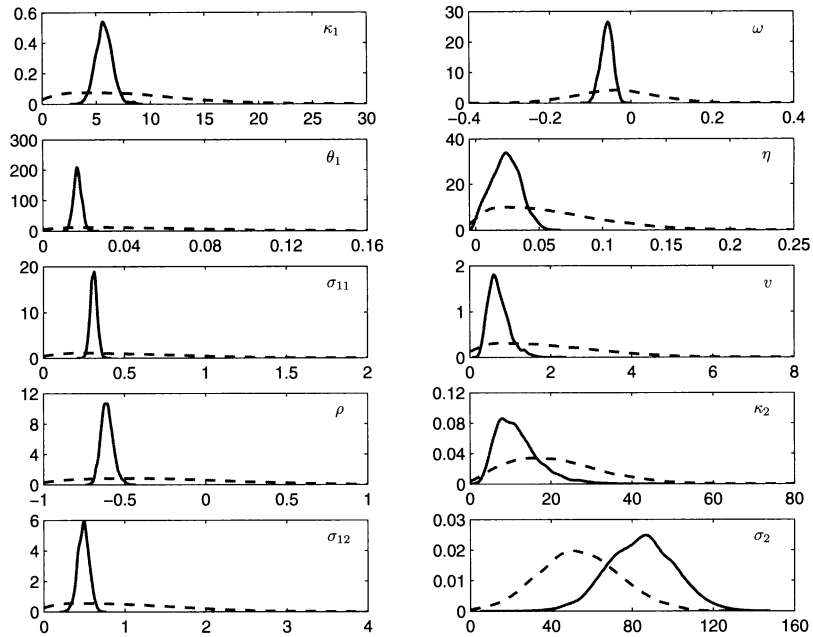

Figure A1
The prior and posterior distributions

sequence of daily observations with sample size $T=6,000$. The true values of the parameters are $\Theta^*=(0.10,\ 5.00,\ 0.02,\ 0.30,\ -0.60,\ 0.50,\ -0.05,\ 0.03,\ 0.80,\ 15.0,\ 50.0)$, which are close to the full-sample estimates in Section 3. The initial values of the stock price, S_0, diffusion variance, $V_{1,0}$, and the jump intensity, $V_{2,0}$, are given as 100, 0.03, and 1.00, respectively, and the priors are the same as in Table A1. Based on the guidelines of Pitt et al. (2012), we choose $M=5\times 1024$. Any increase of this number may result in higher acceptance rates, but it also increases the computational cost. As for the choices of N, N_1, and N_2, we consider the following three Monte Carlo studies:

- MC1: $N=2\times 1024$, and $N_1=N_2=N/2$;
- MC2: $N=2\times 1024$, and $N_1=N_2=N\times 2/3$;
- MC3: $N=4\times 1024$, and $N_1=N_2=N/2$.

The total number of simulated paths is 50 in each Monte Carlo study. We use the same simulated dataset across the three Monte Carlo studies. Graphical processor-based parallel architectures (GPUs) are used to speed up computations. Table A2 presents the Monte Carlo simulation results. The following findings are in order. First, our parameter learning algorithm is quite robust and not sensitive to the choices of N, N_1, and N_2 because the three Monte Carlo studies deliver quite similar results. Second, the diffusion parameters, Θ_D, and the jump parameters, ω and η, can be well identified by our learning algorithm because their means are quite close to the true values and their RMSEs are very small in all three Monte Carlo studies. Third, however, the jump parameters, υ, κ_2, and σ_2, which control the jump structure and intensity dynamics, are not easy to estimate. This is because the relatively large value of υ generates a small number of sizable jumps, making extreme events arrive at a very low frequency.

Based on the above results, in empirical applications, we choose $N=4\times 1024$, and $N_1=N_2=N/2$. However, the choice of the number of state particles (M) is more sensitive in real data applications than in simulations. In order to investigate this issue, we implement a sensitivity

Table A2
Monte Carlo studies

		MC1		MC2		MC3	
Θ	True Value	Mean	RMSE	Mean	RMSE	Mean	RMSE
μ	0.10	0.089	0.029	0.090	0.028	0.090	0.029
κ_1	5.00	5.204	0.709	5.180	0.721	5.212	0.710
θ_1	0.02	0.021	0.003	0.021	0.003	0.021	0.003
σ_{11}	0.30	0.301	0.022	0.301	0.023	0.301	0.022
ρ	−0.60	−0.595	0.046	−0.595	0.047	−0.594	0.046
σ_{12}	0.50	0.560	0.117	0.556	0.113	0.558	0.114
ω	−0.05	−0.055	0.014	−0.055	0.014	−0.055	0.014
η	0.03	0.027	0.006	0.027	0.006	0.027	0.006
v	0.80	1.314	0.668	1.300	0.656	1.331	0.687
κ_2	15.0	18.24	3.943	18.17	3.794	18.37	3.985
σ_2	50.0	48.32	5.264	48.65	5.058	48.40	5.023

Table A3
Sensitivity study

	$M=8*1024$		$M=16*1024$	
Θ	Mean	Std	Mean	Std
μ	0.045	0.004	0.043	0.004
κ_1	5.983	0.393	6.015	0.200
θ_1	0.017	0.000	0.017	0.000
σ_{11}	0.310	0.005	0.317	0.003
ρ	−0.606	0.006	−0.611	0.005
σ_{12}	0.498	0.024	0.510	0.019
ω	−0.058	0.003	−0.057	0.002
η	0.023	0.001	0.021	0.002
v	0.701	0.048	0.684	0.034
κ_2	10.76	1.154	9.789	0.512
σ_2	79.45	2.489	79.85	2.120
LMLH	2.719e4	1.094	2.720e4	0.723
Time	≈ 1.6 days		≈ 3.0 days	

study. Using the real data in section 3.1, we implement 15 independent runs of the algorithm for the SE-M1 model by setting $M_1 = 8 \times 1024$ and $M_2 = 16 \times 1024$, respectively. Table A3 presents means and standard deviations of the posterior means of the model parameters and the log marginal likelihoods across these runs. We clearly see that the log marginal likelihood (LMLH) under M_1 is nearly the same as that under M_2, even though its standard deviation under M_2 is a little bit reduced. Similar results can also be found for the parameter estimates. These results indicate that in the real data applications, model inference is not so sensitive to the choice of the number of state particles between M_1 and M_2, though we decide to choose $M = 16 \times 1024$ in our empirical analysis.

We program in MATLAB the main algorithm and offload the computational bottleneck of the algorithm, the particle filter, to the GPU, coded in CUDA. Relying on a Telsa K20 GPU, our Bayesian learning algorithm is quite fast. As we can see from Table A3, if we set the number of state particles equal to M_1, each run on the real data takes about 1.6 days. If we set it equal to M_2, each run takes about three days.

A.4. ESS and Acceptance Rates

Figure A2 presents the efficient sample sizes and the acceptance rates at the move steps for the four models. In general, we can see from the left panel that the algorithm takes more move steps in

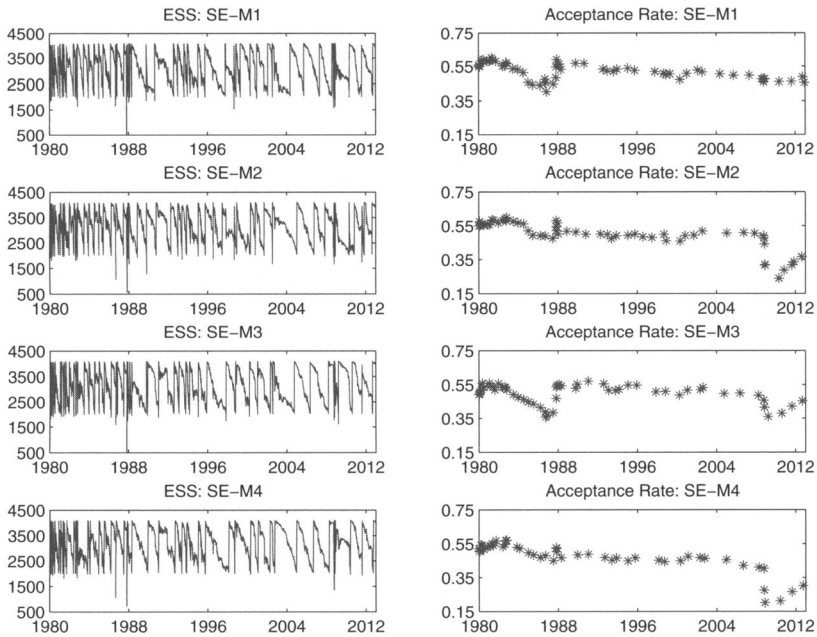

Figure A2
ESS and acceptance rates

the early stage of learning when the market information is minimal. We also find that the SE-M1 and the SE-M3 models take less move steps than the SE-M2 and the SE-M4 models, in particular, during the financial crises. From the right panel, we see that the acceptance rate remains high in the SE-M1 model over time, larger than 35%. However, in the SE-M3 model, it drops a little, and in the SE-M2 and the SE-M4 models it decreases to a low level during the 2008 financial crisis. Fewer move steps and higher acceptance rates in the SE-M1 model indicate that it can better adapt to the outliers.

A.5. Change of Measure

The no-arbitrage condition indicates that there exists at least one almost surely positive process, K_t, with $K_0 = 1$, such that the discounted gains process associated with any admissible trading strategy is a martingale (Harrison and Kreps 1979). K_t, which is assumed to be a semimartingale, is called the stochastic discount factor or the pricing kernel. We propose a class of models for the stochastic discount factor, K_t, such that the change of measure does not alter the model structure. Specifically,

$$K_t = \exp\Big(-\int_0^t r_s ds\Big)\mathcal{E}\Big(-\int_0^t \gamma_W(s) dW_s\Big)\mathcal{E}\Big(-\int_0^t \gamma_V(s) dZ_s\Big)$$

$$\times \mathcal{E}\Big(\int_0^t \int_{R-} (\gamma_J(s,x) - 1)\tilde{\pi}(dx,ds)\Big), \tag{A6}$$

where r_t is the risk-free rate of interest, $\mathcal{E}(\cdot)$ denotes the stochastic (*Doleans-Dade*) exponential operator, $\tilde{\pi}$ is a compensated random measure, and $\gamma_W(t)$, $\gamma_V(t)$, and $\gamma_J(t,x) - 1$ define market prices for the risk factors in the market.

The price for the diffusive volatility risk, Z_t, is assumed to have a form of $\gamma_V(t)=\gamma_V\sqrt{V_{1,t}}$, where γ_V is a constant. For the jump component, the above change of measure indicates that its risk-neutral Lévy density, $\nu^Q(dx)$, is related to its objective one, $\nu(dx)$, by $\nu^Q(dx)=\gamma_J(t,x)\nu(dx)$. We assume $\gamma_J(t,x)$ has an exponential form, $e^{-\gamma_J x}$, where γ_J is a constant, such that the risk-neutral Lévy density of the jump component is simply an exponential tilting of the objective one. In contrast, we leave $\gamma_W(t)$ unspecified.

The variance gamma process has a Lévy density under the objective measure, P, as follows:

$$\nu(dx)=\frac{1}{v}\left(\frac{\exp\left(-\frac{\mu_u}{v_u}x\right)}{x}1_{x>0}+\frac{\exp\left(-\frac{\mu_d}{v_d}|x|\right)}{|x|}1_{x<0}\right)dx. \tag{A7}$$

The above change of measure indicates that the risk-neutral Lévy density should have a form of

$$\nu^Q(dx)=\frac{1}{v}\left(\frac{\exp\left(-(\frac{\mu_u}{v_u}+\gamma_J)x\right)}{x}1_{x>0}+\frac{\exp\left(-(\frac{\mu_d}{v_d}-\gamma_J)|x|\right)}{|x|}1_{x<0}\right)dx. \tag{A8}$$

We therefore have the following risk-neutral model under the measure Q:

$$\ln S_t/S_0=\int_0^t r_s ds+\left(W^Q_{T_{1,t}}-k^Q_W(1)T_{1,t}\right)+\left(J^Q_{T_{2,t}}-k^Q_J(1)T_{2,t}\right), \tag{A9}$$

$$dV_{1,t}=\kappa^Q_1\left(\theta^Q_1-V_{1,t}\right)dt+\sigma_{11}\sqrt{V_{1,t}}dZ^Q_t-\sigma_{12}d(J^-_{T_{2,t}})^Q, \tag{A10}$$

$$dV_{2,t}=\kappa_2(\theta_2-V_{2,t})dt-\sigma_2 d(J^-_{T_{2,t}})^Q, \tag{A11}$$

where $\kappa^Q_1=\kappa_1+\sigma_{11}\gamma_V$, $\theta^Q_1=\kappa_1\theta_1/\kappa^Q_1$, and W^Q_t and Z^Q_t are two independent standard Brownian motions. The time-homogenous jump component is still the variance gamma process with the risk-neutral Lévy density, $\nu^Q(dx)$, given by Equation (A8). The risk-neutral jump parameters are now given as follows: $\omega^Q=(\omega-\gamma_J\eta^2)/A$, $\eta^Q=\eta/\sqrt{A}$, and $v^Q=v$, where $A=1+\gamma_J\omega v-\gamma_J^2\frac{\eta^2 v}{2}$.

References

Aït-Sahalia, Y., J. Cacho-Diaz, and R. Laeven. Forthcoming. Modeling financial contagion using mutually exciting jump processes. *Journal of Financial Economics*.

Aït-Sahalia, Y., and J. Jacod. 2009. Estimating the degree of activity of jumps in high frequency data. *Annals of Statistics* 37:2202–44.

———. 2011. Testing whether jumps have finite or infinite activity. *Annals of Statistics* 39:1689–1719.

Andersen, T., T. Bollerslev, and F. Diebold. 2007. Roughing it up: Including jump components in the measurement, modeling and forecasting of return volatility. *Review of Economics and Statistics* 89:701–20.

Andersen, T., T. Bollerslev, F. Diebold, and C. Vega. 2007. Real-time price discovery in stock, bond and foreign exchange markets. *Journal of International Economics* 73:251–77.

Andrieu, C., A. Doucet, and R. Holenstein. 2010. Particle Markov chain Monte Carlo (with Discussions). *Journal of the Royal Statistical Society: Series B* 72:269–342.

Benzoni, L., P. Collin-Dufresne, and R. Goldstein. 2011. Explaining asset pricing puzzles associated with the 1987 market crash. *Journal of Financial Economics* 101:552–73.

Carr, P., H. Geman, D. Madan, and M. Yor. 2003. Stochastic volatility for Lévy processes. *Mathematical Finance* 13:345–82.

Carr, P., and L. Wu. 2004. Time-changed Lévy processes and option pricing. *Journal of Financial Economics* 71:113–41.

———. 2011. Leverage effect, volatility feedback, and self-exciting market disruptions. Working Paper.

Clark, P. K. 1973. A subordinated stochastic process model with fixed variance for speculative prices. *Econometrica* 41:135–56.

Collin-Dufresne, P., M. Johannes, and L. Lochstoer. 2013. Parameter learning in general equilibrium: The asset pricing implications. Working Paper.

Cogley, T., and T. Sargent. 2008. The market price of risk and the equity premium: A legacy of the Great Depression. *Journal of Monetary Economics* 55:454–76.

Del Moral, P. 2004. *Feynman-Kac formulae: Genealogical and interacting particle systems with applications*. New York: Springer.

Eraker, B. 2004. Do stock prices and volatility jump? Reconciling evidence from spot and option prices. *Journal of Finance* 59:1367–403.

Eraker, B., M. Johannes, and N. Polson. 2003. The impact of jumps in equity index volatility and returns. *Journal of Finance* 58:1269–300.

Fulop, A., and J. Li. 2013. Efficient learning via simulation: A marginalized resample-move approach. *Journal of Econometrics* 176:146–61.

Gordon, N., D. Salmond, and A. Smith. 1993. Novel approach to nonlinear and non-Gaussian Bayesian state estimation. *IEEE Proceedings-F* 140:107–13.

Hansen, L. 2007. Beliefs, doubts and learning: Valuing macroeconomic risk. *American Economic Review* 97:1–30.

Hansen, P., and A. Lunde. 2005. A forecast comparison of volatility models: Does anything beat a GARCH(1,1)? *Journal of Applied Econometrics* 20:973–89.

Harrison, M., and D. Kreps. 1979. Martingales and arbitrage in multiperiod securities markets. *Journal of Economic Theory* 20:381-408.

Huang, X., and G. Tauchen. 2005. The relative contribution of jumps to total price variance. *Journal of Financial Econometrics* 3:456–99.

Jacod, J., and V. Todorov. 2010. Do price and volatility jump together? *Annals of Applied Probability* 20:1425–69.

Jeffreys, H. 1961. *The theory of probability* (3rd ed.). New York: Oxford University Press.

Johannes, M., A. Korteweg, and N. Polson. 2014. Sequential learning, predictability, and optimal portfolio returns. *Journal of Finance* 69:611–44.

Johannes, M., L. Lochstoer, and Y. Mou. Forthcoming. Learning about consumption dynamics. *Journal of Finance*.

Lee, S., and J. Hannig. 2010. Detecting jumps from Lévy jump diffusion processes. *Journal of Financial Economics* 96:271–90.

Lewellen, J., and J. Shanken. 2002. Learning, asset-pricing tests, and market efficiency. *Journal of Finance* 57:1113–45.

Li, H., M. Wells, and L. Yu. 2008. A Bayesian analysis of return dynamics with Lévy jumps. *Review of Financial Studies* 21:2345–78.

Madan, D., P. Carr, and E. Chang. 1998. The variance gamma process and option pricing. *European Finance Review* 2:79–105.

O'Hagan, A. 1994. *Kendall's advanced theory of statistics, vol. 2B: Bayesian inference*. New York: Halsted Press.

Pastor, L., and P. Veronesi. 2003. Stock valuation and learn about profitability. *Journal of Finance* 58:1749–89.

———. 2006. Was there a NASDAQ bubble in the late 1990s? *Journal of Financial Economics* 81:61–100.

———. 2009. Learning in financial markets. *Annual Review of Financial Economics* 1:361–81.

Pitt, M., R. Silva, P. Giordani, and R. Kohn. 2012. On some properties of Markov chain Monte Carlo simulation methods based on the particle filter. *Journal of Econometrics* 171:134–51.

Timmermann, A. 1993. How learning in financial markets generates excess volatility and predictability in stock prices. *Quarterly Journal of Economics* 108:1135–45.

———. 1996. Excess volatility and predictability of stock prices in autoregressive dividend models with learning. *Review of Economic Studies* 63:523–57.

Todorov, V., and G. Tauchen. 2011. Volatility jumps. *Journal of Business and Economic Statistics* 29:356–71.

Veronesi, P. 2004. The peso problem hypothesis and stock market returns. *Journal of Economic Dynamics and Control* 28:707–25.

Yu, J. 2004. Empirical characteristic function estimation and its applications. *Econometric Reviews* 23:93–123.