

Research Article Microstructure Models with Short-Term Inertia and Stochastic Volatility

Michael A. Kouritzin

Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, AB, Canada T6G 2G1

Correspondence should be addressed to Michael A. Kouritzin; michaelk@ualberta.ca

Received 2 March 2015; Revised 9 July 2015; Accepted 15 July 2015

Academic Editor: Quanxin Zhu

Copyright © 2015 Michael A. Kouritzin. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Partially observed microstructure models, containing stochastic volatility, dynamic trading noise, and short-term inertia, are introduced to address the following questions: (1) Do the observed prices exhibit statistically significant inertia? (2) Is stochastic volatility (SV) still evident in the presence of dynamical trading noise? (3) If stochastic volatility and trading noise are present, which SV model matches the observed price data best? Bayes factor methods are used to answer these questions with real data and this allows us to consider volatility models with very different structures. Nonlinear filtering techniques are utilized to compute the Bayes factor on tick-by-tick data and to estimate the unknown parameters. It is shown that our price data sets all exhibit strong evidence of both inertia and Heston-type stochastic volatility.

1. Introduction

Financial analysts list speculation, finiteness of assets, interest rates, tick size, price inertia, price clustering, belief heterogeneity, asymmetric information, greed and fear, and so forth as causes for price fluctuations over time. Yet, popular models like geometric Brownian motion (GBM) (e.g., Black and Scholes [1], Merton [2]) or the Cox-Ross-Rubinstein model [3] try to handle all these factors in an overly simple framework, resulting in unnatural phenomena like the *volatility smile*. Consequently, stochastic volatility, which has been observed in real prices, is often added to the price value evolution (e.g., Heston [4], Jachwerth and Rubinstein [5], Hull and White [6], and Nelson [7]) to avoid the volatility smile. However, which stochastic volatility model fits the market data best?

Nowadays, many authors talk about the misspecification of stochastic price-volatility models (including the Heston model which we show favorably herein) so much. It leads us to wonder whether there are missing ingredients to these very simple models. Even combined stochastic value-volatility models do not address tick size, price inertia, price clustering, hidden liquidity, and fear-greed cycles that traders, especially high frequency traders, must deal with. To handle these issues, one is drawn to tick-by-tick microstructure models and left with the perplex question: How should one model price inertia in continuous time? We are using the term price inertia instead of the related term price momentum because we are not weighting transaction prices by volume. Fractional Brownian motion (FBM), best known for its long memory properties, exhibits inertia and has been used to model markets (Mandelbrot [8], Shiryaev [9]) even though these models allow arbitrage strategies. We speculate that FBM's success in modeling observed data is more attributable to inertia than long memory. However, we introduce an alternative inertia process and show that this new process better satisfies the desired properties of inertia than FBM. We then show strong statistical evidence of price inertia that lasts for hours or days using Bayes estimates and Bayes factor on real price data. We do not consider the possibility of arbitrage nor determine derivative prices for our models but rather leave these interesting mathematical finance questions to the experts. (See Capinski and Zastawniak [10] for an excellent introduction to these types of questions and to mathematical finance in general.) Also, we leave the difficult task of obtaining theoretical error bounds for our particle filter methods to other works. (See, e.g., Kouritzin and Zeng [11] and Del Moral et al. [12] for related work on approximate

filters.) Our focus is solely on modeling observed stock price data and the methodology of determining which of a class of models best fits the observed data.

High frequency data contains complete marketparticipant trading activities (Engle [13]) and is modeled using microstructure (Black [14], Chan and Lakonishok [15], Hasbrouck [16, 17], Engle and Russell [18], Engle [13], and Bandi and Russell [19]). Unlike the macrostructure market, the trading noise in the microstructure market is not negligible; thus, the intrinsic asset value is not readily discernable. In this paper, we introduce a class of dynamic microstructure models, where the transaction price is formulated as a distorted and color-noise corrupted variant of the intrinsic asset value with the intrinsic asset value being a traditional stochastic value-volatility process. Indeed, we view the transaction price data as random counting-measure observations of intrinsic value corrupted by microstructure trading noise with such things as inertia and fear-greed cycles built in. However, trading noise sources themselves introduce volatility to transaction prices. This raises the question, "Do we need to model stochastic volatility explicitly in the presence of dynamic microstructure trading noise?" We will give strong evidence of the presence of stochastic volatility through Bayes factor methods and stochastic filtering theory. Moreover, we also utilize model selection to provide strong evidence of Heston-type volatility over competing stochastic volatility models based on the observed transaction data in a microstructure market. This suggests that the common viewpoint of the Heston model being highly misspecified might be better stated as overly simplistic macrostructureonly models are underspecified. Bayes factor (see, e.g., Kass and Raftery [20]) is our preferred model selection method since it provides statistical comparisons in real time as to which model best fits the market data while allowing the stochastic value-volatility (signal) models to be singular to one another. Indeed, to use the Bayes factor method, we need only to be able to transform all microstructure asset-price observation models of interest into the same canonical process via Girsanov-type measure change.

Previously, Zeng [21] studied a filtering equation for inferring the intrinsic value process in a microstructure model while Xiong and Zeng [22] proposed a branching particle approximation to this equation. Kouritzin and Zeng [23] derived a Bayes factor equation and discussed the Bayesian model selection problem to determine whether financial data, such as stock prices, display jump-type stochastic volatility. However, all these works are based on a restricted microstructure model and thus cannot be applied to our general setting. Moreover, our problems of showing statistical evidence of inertia and determining which of the classical stochastic volatility models best represents real data in the presence of microstructure noise were not considered. We also propose a new inertia process, explain its role in modeling prices, and show its statistical significance with real tick-by-tick data.

Section 2 is devoted to explaining our model. First, our five standard value-volatility models (GBM, Hull-White, Log Ornstein-Uhlenbeck, continuous GARCH, and Simplified Heston) are given followed by our microstructure inertia process and its properties and then the other components of our dynamic microstructure model. Together the valuevolatility and microstructure components form our price evolution model, which, at the end of Section 2, is interpreted as a filtering model. In Section 3, we discuss model calibration and fair price/value estimation through Bayesian filter estimation. A filtering equation and a branching particle filter approximation algorithm are first given and explained. Then, their use to identify parameters and come up with initial state estimates is discussed. Finally, numeric parameter and initial state estimates for each model are given. As a byproduct, it is demonstrated that proper modeling and estimation of fair price (as is done herein) can provide information about overbought conditions and help avoid financial loss (see Figure 4). Section 4 is dedicated to Bayesian model selection. We first motivate the use of Bayes factor for model selection and explain how to estimate Bayes factor from unnormalized particle filters. Then, we establish strong statistical evidence of inertia and Heston-type volatility in all our price data through model selection using the Bayes factor method to test which fair price-volatility model and what amount of inertia best fit the observed price data.

2. The Partially Observed Market Model

In this section, we build our stochastic model that has macrostructure and microstructure components and interpret this model in terms of a *signal* that needs to be estimated in real time and *observations* which are used to form the signal estimates. The macrostructure model consists of fair price, volatility, and related parameters and will be denoted by (X, θ) in the sequel, with X = (S, V) being price and volatility and θ being the parameters for this model. Unlike macrostructure models, we do not assume access to (X, θ) , but rather we take it to be part of the signal to be estimated. Indeed, a model would be judged to be better if the macrostructure price *S* (which represents a "fair" price) is quite different than the observed price and we can use filtering to determine overbought and oversold situations.

The microstructure price construction converts the macrostructure model into the observed price. Such things as inertia (or momentum), fear-greed cycles, and wholeprice clustering (or rounding), which are not part of the fair price, are incorporated into the microstructure model. A distinguishing feature in our microstructure is dynamic state: To allow the microstructure to influence price over a period of time so that the observed microstructure price can differ from fair price significantly, one needs to add and then estimate microstructure state Z. In particular, the *inertia* process, characterized by a parameter h, is introduced to capture price inertia that might be caused by hidden liquidity; various reaction and access times to information as well as momentum traders themselves. This inertia process is not Markov, so we will have to consider the historical version \widehat{Z}^h of this state. Further, \widehat{Z}^h is also unobservable and hence must be added to the signal along with microstructure parameters ϑ and all must be estimated as nuisance parameters.

The nondynamic part of the microstructure noise consists of rounding and clustering noise. It is widely observed in markets that more trades occur at more even prices like whole nickel or whole dollar levels. Therefore, to match observed prices well, we should have a mechanism to convert evenly distributed raw prices into whole-price-biased observed prices. This is done by binning raw prices into sets D_1 , D_2 , D_3 , D_4 , and D_5 depending on how even they are and then randomly moving raw prices in the less even bins to close prices in the more even bins in order to match the observed prices.

The observations then become the marked counting process of the number of trades that occur at the various prices. We will later use these observations to select and calibrate models and to estimate the augmented signal:

$$(X, \theta, \vartheta, \widehat{Z}^h).$$
 (1)

The whole point of the microstructure is to allow the macrostructure price to distinguish itself from the observations and rather to represent fair value. We then use filtering on asset prices to estimate implied value (hereafter called fair price) and thereby judge whether an asset is overbought or oversold.

2.1. General Notation. Let [0, T] be a fixed time period and let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{0 \le t \le T}, \mathbb{P})$ be a complete filtered probability space. For any stochastic process ρ , its natural filtration, defined as $\mathcal{F}_t^{\rho} \doteq \sigma \{\rho_u : 0 \le u \le t\}$, represents the information in ρ up to time t. \mathbb{N}_0 denotes the set of nonnegative integers and, for any Polish space E, B(E) is the set of all bounded measurable \mathbb{R} -valued functions on E.

2.2. Common Macrostructure State Models. We use a macrostructure model $M = (X, \theta)$ for the unobservable fair price together with its volatility and parameters. Here, $X \in \mathbb{R}^{n_x}$ is the macrostructure financial state (fair price plus volatility) with macrostructure parameter $\theta \in \mathbb{R}^{n_{\theta}}$ for some $n_x, n_{\theta} \in \mathbb{N}_0$. We let μ be a probability distribution on $\mathbb{R}^{n_x+n_{\theta}}$, take **A** to be a generator with domain $\mathcal{D}(\mathbf{A}) \subset B(\mathbb{R}^{n_x+n_{\theta}})$, and assume (X, θ) satisfies the martingale problem.

Definition 1. (X, θ) is the unique solution of the $\mathbb{R}^{n_x+n_\theta}$ -valued martingale problem for **A** with initial distribution μ . That is,

(i)
$$\mu = \mathbb{P} \circ (X_0, \theta)^{-1}$$
,
(ii) $M_t^f = f(X_t, \theta) - f(X_0, \theta) - \int_0^t \mathbf{A} f(X_s, \theta) ds$
(2)

is $\{\mathscr{F}_t^{X,\theta}\}$ -martingale for each $f \in \mathscr{D}(\mathbf{A})$. Moreover, if $(\widetilde{X}, \widetilde{\theta})$ also satisfies (i) and (ii), then (X, θ) and $(\widetilde{X}, \widetilde{\theta})$ have the same finite dimensional distributions.

Remark 2. While θ does not vary in time, we include it in our macrostructure model to be estimated because it is still unknown. Nevertheless, the operator **A** does not act on the variable θ since $d\theta_t/dt = 0$ for our fixed parameters.

The martingale problem formulation (2) (see Stroock and Varadhan [24], Ethier and Kurtz [25] for more details) is

general enough to cover most interesting financial models. In this paper, the macrostructure state *X* consists of two components: the fair price *S* and the stochastic volatility *V* (if any). The most common example of (S, V, θ) in finance is the "geometric Brownian motion" (GBM) utilized in the classical Black-Scholes option pricing formula. Throughout this section, *W* and *B* are two independent standard Brownian motions and $(s, v, \theta) \in \mathbb{R}^{n_x+n_\theta}$.

Example 3 (GBM model; see Black and Scholes [1], Merton [2]). We have that

$$\frac{dS_t}{S_t} = \mu dt + \sigma dW_t,\tag{3}$$

with parameters $\theta = (\mu, \sigma)$, corresponds to our martingale problem with the generator

$$\mathbf{A}^{(1)}f = \frac{1}{2}\sigma^2 s^2 \frac{d^2 f}{ds^2} + \mu s \frac{df}{ds}.$$
 (4)

In GBM model, the volatility σ is a constant. To account for the "volatility smile" commonly observed in market option prices (see Jackwerth and Rubinstein [5] for a detailed survey), the GBM model is generalized to stochastic volatility (SV) models, where σ itself is replaced by a stochastic process $\{V_t^{1/2}, t \ge 0\}$. Some of the popular SV models include the following.

Example 4 (Hull-White model; see Hull and White [6]). Consider

$$\frac{dS_t}{S_t} = \mu dt + V_t^{1/2} dW_t,$$

$$\frac{dV_t}{V_t} = \nu dt + \kappa dB_t,$$
(5)

with parameters $\theta = (\mu, \nu, \kappa)$ and generator

$$\mathbf{A}^{(2)}f = \frac{1}{2}\nu s^2 \frac{\partial^2 f}{\partial s^2} + \mu s \frac{\partial f}{\partial s} + \frac{1}{2}\kappa^2 \nu^2 \frac{\partial^2 f}{\partial \nu^2} + \nu \nu \frac{\partial f}{\partial \nu}.$$
 (6)

Example 5 (Logarithmic Ornstein-Uhlenbeck model; see Scott [26]). We have that

$$\frac{dS_t}{S_t} = \mu dt + V_t^{1/2} dW_t,$$

$$\frac{dV_t^{1/2}}{V_t^{1/2}} = \left(\frac{1}{2}\nu^2 - \rho\left(\ln V_t^{1/2} - \bar{\omega}\right)\right) dt + \kappa dB_t,$$
(7)

with parameters $\theta = (\mu, \nu, \varrho, \omega, \kappa)$ and generator

$$\mathbf{A}^{(3)}f = \frac{1}{2}\nu^2 s^2 \frac{\partial^2 f}{\partial s^2} + \mu s \frac{\partial f}{\partial s} + \frac{1}{2}\kappa^2 \nu^2 \frac{\partial^2 f}{\partial \nu^2} + \nu \left(\frac{1}{2}\nu^2 - \rho \left(\ln \nu - \omega\right)\right) \frac{\partial f}{\partial \nu}.$$
(8)

Example 6 (continuous GARCH model; see Nelson [7]). We have that

$$\frac{dS_t}{S_t} = \mu dt + V_t^{1/2} dW_t,$$

$$dV_t = (\nu - \varrho V_t) dt + \kappa V_t dB_t,$$
(9)

with parameters $\theta = (\mu, \nu, \varrho, \kappa)$ and generator

$$\mathbf{A}^{(4)}f = \frac{1}{2}\nu s^2 \frac{\partial^2 f}{\partial s^2} + \mu s \frac{\partial f}{\partial s} + \frac{1}{2}\kappa^2 \nu^2 \frac{\partial^2 f}{\partial \nu^2} + \left(\nu - \rho \nu\right) \frac{\partial f}{\partial \nu}.$$
(10)

Example 7 (simplified Heston model; see Heston [4]). We have that

$$\begin{aligned} \frac{dS_t}{S_t} &= \mu dt + V_t^{1/2} dW_t, \\ dV_t &= \left(\nu - \varrho V_t\right) dt + \kappa V_t^{1/2} dB_t, \end{aligned} \tag{11}$$

with parameters $\theta = (\mu, \nu, \varrho, \kappa)$ and generator

$$\mathbf{A}^{(5)}f = \frac{1}{2}\nu s^2 \frac{\partial^2 f}{\partial s^2} + \mu s \frac{\partial f}{\partial s} + \frac{1}{2}\kappa^2 \nu \frac{\partial^2 f}{\partial \nu^2} + (\nu - \varrho \nu) \frac{\partial f}{\partial \nu}.$$
 (12)

We label this example as simplified because we do not allow B and W to be correlated as Heston did. There is no mathematical issue by including this correlation, but it would add a parameter to the model, which increases computation time. The Heston model already performed the best without this parameter. GBM (with microstructure) plays a special role in our study as it is our *no stochastic volatility* model. We will compare our other models against it on real data to determine whether stochastic volatility is present. In summary, refer to Table 1.

Remark 8. The continuous GARCH model is the continuoustime limit of many classical GARCH-type discrete-time processes (Nelson [7], Drost and Werker [27]). We did not consider jumping stochastic volatility models (e.g., Elliott et al. [28], Kouritzin and Zeng [23], Duffie et al. [29], Eraker et al. [30], and Eraker [31]) or models where *W*, *B* are correlated, due to our need to dedicate our limited computer resources to handling our complicated (non-Markov) microstructure with inertia. Still, we want to emphasize that the computational complexity we experienced is fundamental to the fact that we are using non-Markov (inertia) models and has little to do with our particular methods. Indeed, our Bayes factor filtering methods are what makes the computations possible on an inexpensive contemporary desktop computer.

2.3. Construction of Microstructure Price. The fair pricevolatility models account for the random variances of the intrinsic asset value; thus, the selection of proper SV model is crucial for investing, derivative pricing, and hedging. On the other hand, microstructure noise (see Black [14], Hansen and Lunde [32], Duan and Fulop [33], etc.) causes random

TABLE 1

Name	Model	Macrostate	Macroparameter	Generator
GBM	$M^{(1)}$	S	(μ, σ)	$\mathbf{A}^{(1)}$
Hull-White	$M^{(2)}$	(S, V)	(μ, ν, κ)	$\mathbf{A}^{(2)}$
Log O-U	$M^{(3)}$	(S, V)	$(\mu, \nu, \varrho, \varpi, \kappa)$	$A^{(3)}$
GARCH	$M^{(4)}$	(S, V)	(μ, ν, ϱ, κ)	$\mathbf{A}^{(4)}$
Heston	$M^{(5)}$	(S, V)	(μ, ν, ϱ, κ)	$\mathbf{A}^{(5)}$

perturbations of transaction price from its intrinsic value and the disregard of such trading noise introduces severe bias into stochastic volatility estimation (see Duan and Fulop [33]). We incorporate microstructure trading noise into traditional fair price-volatility models and use statistical filtering to reveal such things as short-term inertia in the trading noise and stochastic volatility in the intrinsic value.

In microstructure markets, the price changes occur only at irregularly spaced transaction times t_1, t_2, \ldots with total trading intensity a(t) (see Engle [13]). Here, we assume a(t)is just a time-varying measurable function as the empirical analysis illustrates that there is no need to consider more general structures. At each transaction time t_i , the transaction price Y_{t_i} is formulated as

$$Y_{t_i} = F\left(X_{t_i}, t_i\right),\tag{13}$$

where F is some nonlinear random field modeling the trading noise. Formulation (13) is similar to that of Hasbrouck [16], where X is the intrinsic and permanent component while F introduces the transitory component.

The empirical evidence reported by Hansen and Lunde [32] suggests strongly that the trading noise is serially correlated. Similar results can be found in Aït-Sahalia et al. [34]. Indeed, there exist situations in which the trading noise variance estimate is zero if the trading noise is simply assumed to be independent (see Duan and Fulop [33]). This does not mean there is no trading noise but rather that the trading noise is autocorrelated. To characterize this correlation, Hansen and Lunde [32] assume the trading noise to be some Gaussian random sequence with stationary covariance and finite dependence. However, this model is most suitable for the low-frequency data and ignores many crucial microstructure effects. We build correlation into our microstructure information noise through inertia and mean-reversion while utilizing microstructure rounding and clustering noise to explain the discreteness and whole-price biasing.

2.3.1. Inertia. The idea of momentum or inertia has been used in many studies (see Jegadeesh and Titman [35], Moskowitz and Grinblatt [36], Grundy and Martin [37], Grundy et al. [38], etc.). Basically, there is the tendency for a stock to continue to move in one direction. To illustrate our approach, we introduce the following definition.

Definition 9. A process (Z_t) is said to have stochastic inertia at time *t* if

$$I_t^Z \doteq \lim_{u \searrow t} \frac{\partial}{\partial t} \frac{\partial}{\partial u} \mathbb{E} \left[Z_u Z_t \right] \in (0, \infty] \,. \tag{14}$$

 I_t^Z is called the inertia function.

The idea behind our definition is that for inertia we should expect $Z_{u+r} - Z_u$ and $Z_t - Z_{t-k}$ to have the same sign for u > t, but close to t and r, k > 0 small. We strengthen this condition to

$$\lim_{u \searrow t} \lim_{k \to 0} \lim_{r \to 0} \frac{E\left[\left(Z_{u+r} - Z_{u}\right)\left(Z_{t} - Z_{t-k}\right)\right]}{rk} > 0$$

$$\iff \lim_{u \searrow t} \lim_{k \to 0} \left[\lim_{r \to 0} \frac{E\left[\left(Z_{u+r} - Z_{u}\right)Z_{t}\right]}{rk} - \lim_{r \to 0} \frac{E\left[\left(Z_{u+r} - Z_{u}\right)Z_{t-k}\right]}{rk}\right] > 0$$

$$\iff \lim_{u \searrow t} \frac{\partial}{\partial t} \frac{\partial}{\partial u} E\left[Z_{u}Z_{t}\right] > 0.$$
(15)

Many processes have inertia. However, to model the stock price effect of the information reaching all market participants, we want the following five properties: (1) Z_t is Gaussian and driftless and $Var(Z_t)$ is proportional to t so Z resembles Brownian motion; (2) I_t^Z is finite, not infinite, indicating that the influence of past values on immediate future is not too strong; (3) Z_t makes sense from informational and hidden liquidity points of view; more precisely, it can explain well the price effects due to the reactions of all market participants to information and rumor being diffused and simulated over a period of time as well as due to the purchases or sales of an agent spreading out a large change in his/her position over time; (4) Z is easy to simulate using, for example, the Gaussian property; (5) Z is easy to analyze.

Neither a Brownian motion *B* nor more generally a square integrable martingale has inertia. Brownian motion with drift $Z_t = Z_0 + \int_0^t m(Z_s) ds + B_t$ has inertia $I_t^Z = E[m^2(Z_t)]$ but we do not want drift. For fractional Brownian motion (FBM) B^h ,

$$\mathbb{E}\left[B_{t}^{h}B_{u}^{h}\right] = \frac{1}{2}\left(\left|t\right|^{2h} + \left|u\right|^{2h} - \left|u - t\right|^{2h}\right),\tag{16}$$

where $h \in (0, 1)$ is the Hurst parameter. Therefore,

$$\lim_{u \searrow t} \frac{\partial}{\partial t} \frac{\partial}{\partial u} \mathbb{E} \left[B_t^h B_u^h \right] = \lim_{u \searrow t} \left((2h-1) h \left(u - t \right)^{2h-2} \right) = \infty$$
(17)
if $h > \frac{1}{2}$.

Thus, the inertia function of B^h is infinity for all t if h > 1/2 (and is $-\infty$ if h < 1/2). Neither case satisfies our five properties. Still, standard representations of FBM motivate the creation of driftless inertia by convolving a Brownian motion with the desired impulse response for information dissemination. With this in mind, we consider the following inertia process.

Definition 10. Our stochastic inertia process is

$$\xi_t^h = \sqrt{h} \int_0^t \tanh\left(\frac{(t-s)}{\Delta}\right) dB_s^{\xi} + \sqrt{1-h} W_t^{\xi}, \quad (18)$$

where (B^{ξ}, W^{ξ}) is a 2-dimensional standard Brownian motion, $\Delta > 0$, and $0 \le h \le 1$.

Remark 11. ξ_t^h is a weighted average of the historical information (the first term) and fundamental information (the second term). In fact, $tanh(t/\Delta)$ can be viewed as the impulse response on price created by market participants receiving and simulating the "information" dB_t^{ξ} and Δ determines the diffusion speed in the market. This formulation captures the idea that news or rumor and its ramifications require time to be fully disseminated and understood. When h = 1, it represents the case of only historical information resulting in the strongest inertia in prices. Alternatively, we can use inertia to explain "hidden liquidity." If everybody knew that an agent was going to make a big change in a position, then the price would immediately jump. However, if the agent breaks up the desired change into small transactions, then it takes time for this extra buying or selling pressure to be recognized in the market. In this case, h = 1 represents the case, where all changes in position are done over a period of time and Δ represents the time to effect 58% of the positional change.

Note that ξ_t^h is a centered Gaussian process such that the autocovariance

$$\mathbb{E}\left[\xi_t^h \xi_u^h\right] = h \int_0^t \tanh\left(\frac{(t-s)}{\Delta}\right) \tanh\left(\frac{(u-s)}{\Delta}\right) ds + (1-h)t$$
(19)

is positive for any $u \ge t$. In particular,

$$\frac{\operatorname{Var}\left(\xi_{t}^{h}\right)}{t} = \frac{h}{t} \int_{0}^{t} \tanh^{2}\left(\frac{(t-s)}{\Delta}\right) ds + (1-h)$$

$$= 1 - h\Delta \frac{\tanh\left(t/\Delta\right)}{t}.$$
(20)

Thus, $\operatorname{Var}(\xi_t^h)/t$ converges to 1 as $t \to \infty$ with speed determined by Δ . (Hence, informational noise increases at the same asymptotic rate as Brownian motion.) Moreover,

$$\frac{\partial}{\partial t} \frac{\partial}{\partial u} \mathbb{E} \left[\xi_t^h \xi_u^h \right]$$

$$= \frac{h}{\Delta^2} \int_0^t \operatorname{sech}^2 \left(\frac{(t-s)}{\Delta} \right) \operatorname{sech}^2 \left(\frac{(u-s)}{\Delta} \right) ds$$
(21)

and, using standard antiderivatives,

$$\lim_{u \to t} \frac{\partial}{\partial t} \frac{\partial}{\partial u} \mathbb{E} \left[\xi_t^h \xi_u^h \right] = \frac{h}{\Delta^2} \int_0^t \operatorname{sech}^4 \left(\frac{s}{\Delta} \right) ds$$

$$= \frac{h}{\Delta} \left[\tanh\left(\frac{t}{\Delta}\right) - \frac{\tanh^3\left(t/\Delta\right)}{3} \right].$$
(22)

Thus, the inertia function of our inertia process is $I_t^{\xi^h} = (h/\Delta)[\tanh(t/\Delta) - \tanh^3(t/\Delta)/3]$, the steady-state inertia is

$$\lim_{t \to \infty} \frac{h}{\Delta} \left[\tanh\left(\frac{t}{\Delta}\right) - \frac{\tanh^3\left(t/\Delta\right)}{3} \right] = \frac{2h}{3\Delta} , \qquad (23)$$

and this happens quickly for small Δ . We can thus verify that ξ^h , defined in (18), satisfies our five desired properties. One can also look upon Δ as the time for new information to be disseminated to fifty-eight percent of the market. Below, we consider three different dissemination times: $\Delta = 40$ minutes, $\Delta = 2$ hours, and $\Delta = 1/2$ day on real stock data. Finally, the fact that ξ^h is Gaussian eases its simulation greatly.

2.3.2. Information Noise and Augmented State. Hitherto, we have focused on constructing inertia processes. Now, we include all informational noise into asset prices. Information noise is introduced to represent trading noises due to things like inertia, fear-greed cycles, belief heterogeneity, and asymmetric information. For the *i*th-transaction occurring at t_i , the raw price \mathcal{Y}_{t_i} is defined by

$$\ln \mathcal{Y}_{t_i} = \begin{cases} \ln S_{t_i} + Z_{t_i}^{h,\Delta} + \epsilon \zeta_i, & \text{dynamical microstructure,} \end{cases}$$
(24)
$$\ln S_{t_i} + \epsilon \zeta_i, & \text{nondynamical,} \end{cases}$$

$$dZ_{t}^{h,\Delta} = -\phi Z_{t}^{h} dt + d\xi_{t}^{h}, \quad Z_{0}^{h,\Delta} = z_{0},$$
(25)

where $Z^{h,\Delta}$ is the dynamical part of the microstructure through which inertia is introduced (with our inertia process ξ^h) and X = (S, V). The case $Z^{h,\Delta} \equiv 0$ is of particular importance in the sequel as it represents the *nondynamical* microstructure case and is used as a calibration model.

The information noise consists of two parts: $\zeta = \{\zeta_i\}_{i=1}^{\infty}$ is a sequence of independent standard Gaussian random variables, $\epsilon > 0$; Z^h is Ornstein-Uhlenbeck- (O-U-) like inertia velocity process with mean-reverting parameter ϕ > 0. Here, ξ^h , ζ , and X are independent and z_0 is a constant. Z^h provides an intuitive continuous-time model that accommodates the joint presence of the inertia and mean-reversion. Our information noise is more reasonable than that of Zeng [21] in that (1) we preclude the possibility of negative prices by using multiplicative noise; (2) the stochastic inertia process ξ^h captures the empirical feature of the inertia observed in transaction prices (e.g., Jegadeesh and Titman [35]); (3) the mean-reverting structure of Z^h when combined with the inertia captures the cyclic property of prices (e.g., Black [14]). Z^h is not a Markov process, so we introduce its historical process as

$$\widehat{Z}_t^h(\tau) \doteq Z_{t\wedge\tau}^h,\tag{26}$$

which is Markovian. Moreover, $\widehat{Z}_t^h \in C[0,T]$, the space of all continuous functions on [0,T], since the paths of Z^h are continuous. Consequently, we augment the state vector to be

$$(X, \theta, \vartheta, \widehat{Z}^h),$$
 (27)



FIGURE 1: Clustering for 3 stocks in April 2010.

where $\vartheta = (\epsilon, \phi)$ is the microstructure noise parameter set. The advantage of this formulation is that we can estimate \widehat{Z}^h and thus Z^h jointly with other components using particle filtering methods. The generalized state incorporates fair price, volatility, parameters, and the historical trading noise \widehat{Z}^h while keeping the tractability of a Markovian framework.

Remark 12. We include neither h nor Δ into the model parameters but rather consider different models corresponding to different values of h and Δ as well as different SV models 1–5. Indeed, we will provide evidence of inertia in the sequel by using Bayesian methods to select a model with a large value of h based upon tick-by-tick stock data.

2.3.3. Rounding and Clustering Noise. Our final modeling goal is to convert uniform raw price into observed wholeprice-biased price. While raw price \mathcal{Y}_{t_i} can take any value, the trading price Y_{t_i} is restricted to multiples of the *tick*, $\{y_0 = 0, y_1 = 1/M, \dots, y_j = j/M, \dots\}$, for some positive integer *M*. The tick size in New York Stock Exchange (NYSE) was switched to \$1/16 from \$1/8 in June 24, 1997, and then further to \$0.01 from January 29, 2001. The empirical studies suggest that the tick size 1/M plays an important role in microstructure market analysis (e.g., Huang and Stoll [39]). Since we are concerned with price clustering for decimal pricing in stock markets, we let M = 100.

It is well documented that there is price clustering to more whole prices. To quantify this price clustering, we examine the price behavior for three NYSE-listed stocks over April 2010 (Figure 1 and Table 2). (In a larger study, we considered eight NYSE stocks in different sectors. However, we only report on three here to conserve space. The results for the other five were similar in nature.)

The transaction data of these stocks shows there is modest clustering at multiples of 5 cents as shown in Figure 1, plotted in terms of pennies. Supposing the raw price \mathcal{Y}_{t_i} falls in the interval $[y_j - 1/2M, y_j + 1/2M)$, then if there was no clustering noise, the trading price Y_{t_i} would just be y_j . Thus,

TABLE 2

NYSE stock	Ticker symbol
Goldman Sachs	GS
International Business Machines Corporation	IBM
PepsiCo Inc.	PEP

the probability of trading at y_j with no clustering noise given $X_{t_i} = x, Z_{t_i} = z$ would be

$$R\left(y_{j} \mid x, z, \theta\right) \doteq P\left(\mathcal{Y}_{t_{i}} = y_{j} \mid X_{t_{i}} = x, Z_{t_{i}} = z, \theta\right)$$

$$= \begin{cases} \int_{\ln\left((y_{j}+1/2M)/(xe^{z})\right)}^{\ln\left((y_{j}+1/2M)/(xe^{z})\right)} \frac{1}{\sqrt{2\pi\epsilon}} e^{-u^{2}/2\epsilon^{2}} du & \text{dynamic microstructure (28)} \\ \int_{\ln\left((y_{j}+1/2M)/x\right)}^{\ln\left((y_{j}+1/2M)/x\right)} \frac{1}{\sqrt{2\pi\epsilon}} e^{-u^{2}/2\epsilon^{2}} du & \text{nondynamical.} \end{cases}$$

Equivalently, we can write R in terms of the historical process as

$$R\left(y_{j} \mid X_{t_{i}}, \Pi_{t_{i}}\widehat{Z}_{t_{i}}^{h}, \vartheta\right)$$

$$= \int_{\ln((y_{j}+1/2M)/X_{t_{i}}e^{\Pi_{t_{i}}\widehat{Z}_{t_{i}}^{h}})}^{\ln((y_{j}+1/2M)/X_{t_{i}}e^{\Pi_{t_{i}}\widehat{Z}_{t_{i}}^{h}})} \frac{1}{\sqrt{2\pi\epsilon}}e^{-u^{2}/2\epsilon^{2}}du,$$
(29)

where Π_{t_i} is the projection onto time t_i ; that is,

$$\Pi_{t_i} \widehat{Z}_{t_i}^h = \widehat{Z}_{t_i}^h (t_i) = Z_{t_i \wedge t_i}^h = Z_{t_i}^h.$$
(30)

Clearly, $R(y_j | x, z, \vartheta)$ is a smooth function of (x, z, ϑ) for each fixed y_i .

To build the observed whole-price bias into our model, we introduce the following sets:

 $D_1 = \{$ The integers in (0, 100)

that are not multiples of 5},

 $D_2 = \{$ The integers in (0, 100)

that are multiples of 5 but not of 25, (31)

$$D_3 = \{25, 75\},$$

$$D_4 = \{50\},$$

$$D_5 = \{100\}.$$

While the raw price will be uniformly distributed over $D_1 \cup D_2 \cup D_3 \cup D_4 \cup D_5$ (or rather the continuous interval (0, 100]), the observed price model must bias D_2 over D_1 , D_3 over either D_2 or D_1 , and so forth. We distribute the observed price randomly over $D_1 \cup D_2 \cup D_3 \cup D_4 \cup D_5$ based upon the raw price in a biased manner favoring the more whole-price ticks in $D_2 \cup D_3 \cup D_4 \cup D_5$. In particular, if the fractional part of the raw price *y* rounded to the nearest cent is in D_1 , then the observed value will stay at the same price with probability $1 - \alpha$ or move to the closest multiple of 5 cents, that is, the closest tick level in $D_2 \cup D_3 \cup D_4 \cup D_5$ with probability α . Then, if the fractional part of the price *y* is in D_2 , it will stay in the same level with probability $1 - \beta$ or move to the closest tick level in $D_3 \cup D_4 \cup D_5$ with probability β . Finally, if the fractional part of the price *y* is in D_3 , then it will stay in the same level with probability $1 - \gamma_1 - \gamma_2$ or move to the closest tick level in D_4 with probability γ_1 and the closest tick level in D_5 with probability γ_2 . In summary, the transition probability function is obtained iteratively by the following.

Case 1. If the fractional part of y_i belongs to D_1 ,

$$p(y_j \mid x, z, \vartheta) = R(y_j \mid x, z, \vartheta)(1 - \alpha).$$
(32)

Case 2. If the fractional part of y_i belongs to D_2 ,

$$p(y_j \mid x, z, \vartheta) = R^*(y_j \mid x, z, \vartheta)(1 - \beta), \qquad (33)$$

where

$$R^{*}(y_{j} | x, z, \vartheta)$$

$$\doteq R(y_{j} | x, z, \vartheta)$$

$$+ \alpha \left(R(y_{j-1} | x, z, \vartheta) + R(y_{j-2} | x, z, \vartheta) \right)$$

$$+ \alpha \left(R(y_{j+1} | x, z, \vartheta) + R(y_{j+2} | x, z, \vartheta) \right).$$
(34)

Case 3. If the fractional part of y_i belongs to D_3 ,

$$p(y_j \mid x, z, \vartheta) = R^{**}(y_j \mid x, z, \vartheta)(1 - \gamma_1 - \gamma_2), \quad (35)$$

where

$$R^{**}(y_{j} | x, z, \vartheta)$$

$$\stackrel{=}{=} R^{*}(y_{j} | x, z, \vartheta)$$

$$+ \beta \left(R^{*}(y_{j-5} | x, z, \vartheta) + R^{*}(y_{j-10} | x, z, \vartheta) \right)$$

$$+ \beta \left(R^{*}(y_{j+5} | x, z, \vartheta) + R^{*}(y_{j+10} | x, z, \vartheta) \right).$$
(36)

Case 4. If the fractional part of y_i belongs to D_4 ,

$$p(y_{j} | x, z, \vartheta) = R^{**}(y_{j} | x, z, \vartheta)$$

$$+ \gamma_{1}(R^{**}(y_{j-25} | x, z, \vartheta) + R^{**}(y_{j+25} | x, z, \vartheta)).$$

$$(37)$$

Case 5. If the fractional part of y_i belongs to D_5 ,

$$p(y_{j} \mid x, z, \vartheta) = R^{**}(y_{j} \mid x, z, \vartheta)$$

$$+ \gamma_{2} \left(R^{**}(y_{j-25} \mid x, z, \vartheta) + R^{**}(y_{j+25} \mid x, z, \vartheta) \right).$$

$$(38)$$

Moreover, we have to handle the case j = 0 separately to avoid negative prices.

Case 6. For
$$j = 0$$
,

$$p(y_0 \mid x, z, \vartheta)$$

$$= R(y_0 \mid x, z, \vartheta)$$

$$+ \alpha (R(y_1 \mid x, z, \vartheta) + R(y_2 \mid x, z, \vartheta))$$

$$+ \beta (R^*(y_5 \mid x, z, \vartheta) + R^*(y_{10} \mid x, z, \vartheta))$$

$$+ y_2 R^{**}(y_{25} \mid x, z, \vartheta).$$
(39)

Clustering p

	TABLE 3	
arameters		

α	0.060475
β	0.046883
γ_1	0.03883
γ_2	0.16525

Estimate

Remark 13. Our clustering setup is designed to work well for intrinsic prices over \$1. For real penny stocks, our setup would introduce positive bias and should be modified slightly.

Using relative frequency analysis on the aggregate of our three stocks, we found the values presented in Table 3.

The large degree of clustering exhibited, especially to the whole dollar, might be considered surprising. However, earlier studies of Huang and Stoll [39], Chung et al. [40], and Chung et al. [41] also showed significant clustering. Moreover, the degree of price clustering in NYSE is weaker than that of NASDAQ. For example, Barclay [42] examined 472 stocks from NASDAQ before and after their listing in NYSE or American Stock Exchange (AMEX): before the listing, the average fraction of even-eighths (0, 1/4, 1/2, 3/4) is 78% while thereafter it drops to about 56%.

2.4. Nonlinear Filtering Model. Our price process can be formulated as a marked point process \vec{Y} : a sequence of random vectors $\vec{Y} = (t_i, Y_{t_i}, i \ge 1)$, where $t_i \in [0, T]$ denotes the time of *i*th-trade and Y_{t_i} the corresponding trading price. Accordingly, the mark space of \vec{Y} is (E, \mathcal{C}) , where $E = \mathbb{N}_0$ and \mathcal{C} is all its subsets. Here, $j \in E$ corresponds to the *j*th-tick level j/M. For each $A \in \mathcal{C}$, we associate the counting process

$$Y_t(A) \doteq \sum_{i \ge 1} \mathbb{1}_{\{Y_{t_i} \in A\}} \mathbb{1}_{\{t_i \le t\}}$$
(40)

to count the trades in tick level set *A* up to time *t*. In particular, for $j \in E$,

$$Y_{j}(t) \doteq Y_{t}(\{j\}) = \sum_{i\geq 1} \mathbb{1}_{\{Y_{t_{i}}=j\}} \mathbb{1}_{\{t_{i}\leq t\}}$$
(41)

denotes the total trades at *j*th-tick level j/M until time *t*. Equivalently, we can introduce the random counting measure $Y(dz \times dt)$ on $\mathcal{C} \otimes \mathcal{B}[0, T]$ by

$$Y(\omega, A \times (s, t]) \doteq Y_t(\omega, A) - Y_s(\omega, A),$$

$$\forall \omega \in \Omega, \ s \le t \in [0, T], \ A \in \mathscr{C}.$$
(42)

The natural filtration, that is, information content, of *Y* is

$$\mathscr{F}_{t}^{Y} \doteq \sigma\left(Y_{s}\left(A\right), \ 0 \le s \le t, \ A \in \mathscr{C}\right).$$

$$(43)$$

Now, we assume the following.

(C1) The total trade process $Y_t = Y_t(E)$ admits an intensity a(t) for some positive measurable function a.

Therefore, using the conditional probabilities defined in the previous subsection, we find that $Y_i(t)$ has intensity

$$\lambda_j \left(X_t, Z_t^h, \vartheta, t \right) = a \left(t \right) \cdot p \left(y_j \mid X_t, Z_t^h, \vartheta \right).$$
(44)

To simplify the notation, we rewrite (44) as $\lambda_i = a \cdot p_i$.



FIGURE 2: Intertrade duration in seconds.

For our present work, we estimated total intensity function a(t) from intertrade data allowing for intraday variation. Figure 2 is the intertrade duration histogram of our 3 NYSElisted stocks averaged over all times of the day. We divided the intertrade data into half-hour periods over the course of the day and took *a* to be constant over these half-hour periods:

$$a(t) = \frac{\text{Average number of trades in period}}{1800 \text{ seconds}}$$
(45)

for *t* in that daily period.

(C2) There exist some positive constants δ , *C* such that $\delta \le a(t) \le C$ for all *t*.

Based on representation (40), (44), $(X, \theta, \vartheta, \widehat{Z}^h; Y)$ is framed by a partial-observation model, where $(X, \theta, \vartheta, \widehat{Z}^h)$ is the state (signal), which is partially observed through the infinite dimensional counting process Y. One difficulty in calibrating these models is that their transition probability functions are usually unknown in closed form, so maximum likelihood estimation (MLE) methods are difficult to use (see Aït-Sahalia and Kimmel [43] for further details). Instead, we use Bayesian filtering because (1) Bayes estimates do not require the availability or regularity of the full likelihood functions; (2) Bayes estimates can be computed recursively for our tick-by-tick data; (3) Bayesian hypothesis tests can be conducted through Bayes factor, which is the ratio of marginal likelihoods and is easily computed even when the signals are of different dimension or, more generally, singular to each other.

3. Model Calibration

Our foremost goal is to contribute to the process of model building for financial markets both by suggesting elements to be included in the models and proposing methods to select models based on real observation data. To be able to do this effectively, we need to be able to tune each possible model effectively to get good prior (probability distribution) estimates for the complete signal $(X, \theta, \theta, \widehat{Z}^h)$ before the test period. We do this through nonlinear filtering and in particular through particle filtering. In this section, we first introduce the filtering equations for our problem. Then, we introduce a branching particle filter algorithm that is an approximation to the unnormalized filter and can be implemented on a computer. Next, we explain how we did the calibration (i.e., came up with this prior distribution) and finally we give the results for the models of interest herein.

3.1. Nonlinear Filtering Equations. The available information about $(X_t, \theta, \theta, \widehat{Z}_t^h)$ is the observation filtration $\mathscr{F}_t^Y \subset \mathscr{F}_t$, defined in (43), and the primary goal of nonlinear filtering is to characterize the conditional distribution

$$\pi_t\left(\cdot\right) = \mathbb{P}\left[\left(X_t, \theta, \vartheta, \widehat{Z}_t^h\right) \in \cdot \mid \mathscr{F}_t^Y\right] \tag{46}$$

or, equivalently,

$$\pi_t(f) = \mathbb{E}\left[f\left(X_t, \theta, \vartheta, \widehat{Z}_t^h\right) \mid \mathscr{F}_t^Y\right]$$
(47)

for $f \in B(\mathbb{R}^{n_x+n_\theta+2} \times C[0,T])$. Here, $\vartheta = (\epsilon, \phi)$, \widehat{Z}^h is the long memory portion of our information noise and (X, θ) is the state and parameter of our fair price-volatility martingale problem.

Remark 14. Actually, we often only want to estimate $\mathbb{P}[(X_t, \theta) \in \cdot | \mathscr{F}_t^Y]$, but there is no simple recursive formula for this marginal. The filter is naturally model dependent, so we can produce different filtering processes for each model, that is, for each SV choice (1–5), each value of Δ , and each value of *h* in our inertia process.

Suppose κ_z is a positive constant for each $z \in \mathbb{N}_0$ such that $\kappa \doteq \sum_{z=0}^{\infty} \kappa_z < \infty$, and consider the continuous-time likelihood function

$$L_{t} = L_{t} \left(X, \vartheta, Z^{h} \right)$$

= $\exp\left(\int_{0}^{t} \int_{E} \ln \left| \frac{\lambda_{z} \left(X_{s}, \vartheta, Z_{s}^{h}, s \right)}{\kappa_{z}} \right| Y (dz, ds) - \int_{0}^{t} \left(a \left(s \right) - \kappa \right) ds \right).$ (48)

 L_t is a martingale under Condition (C2) and \mathbb{Q} , defined by

$$\frac{d\mathbb{Q}}{d\mathbb{P}}\Big|_{\mathscr{F}_{T}} \qquad (49)$$

$$= L_{T}^{-1}\left(\text{i.e. }\mathbb{Q}\left(A\right) = \int_{A} L_{T}^{-1} d\mathbb{P} \quad \text{for } A \in \mathscr{F}_{T}\right),$$

is called the reference measure. Under \mathbb{Q} , the observations are just a Poisson measure, independent of the state vector $(X, \theta, \vartheta, \widehat{Z}^h)$, with mean measure $\eta(A \times (0, t]) = \sum_{z \in A} \kappa_z \times$ (0, t]. To make the likelihoods more manageable in the particle filters to follow, we choose κ to be a long time average value $(1/T) \int_0^T a(s) ds$ of a(s) and $z \to \kappa_z$ to be highest where the trades will be more concentrated. Bayes Theorem (see Bremaud [44], p. 165) then links the desired (real-world) conditional distribution π_t with the unnormalized filter σ_t by

$$\pi_t(f) = \frac{\sigma_t(f)}{\sigma_t(1)},\tag{50}$$

where the unnormalized filter σ_t is defined by

$$\sigma_t(f) \doteq \mathbb{E}^{\mathbb{Q}}\left[f\left(X_t, \theta, \vartheta, \widehat{Z}_t^h\right)L_t \mid \mathscr{F}_t^Y\right]$$
(51)

for all $f \in B(\mathbb{R}^{n_x+n_{\theta}+2} \otimes C[0,T])$. Now, we can give the evolution equation for σ_t .

Theorem 15. Under (C1) and (C2), the unnormalized filter σ_t is the unique measure-valued solution of the stochastic filtering equation

$$\sigma_{t}(f) = \sigma_{0}(f) + \int_{0}^{t} \sigma_{s}\left(\left(\overline{\mathbf{A}} - a(s) + \kappa\right)f\right) ds + \int_{0}^{t} \int_{E} \sigma_{s-}\left(\left(\frac{\lambda_{z}(s-)}{\kappa_{z}} - 1\right)f\right)Y(dz, ds),$$
(52)

for t > 0 *and* $f \in \mathcal{D}(\overline{\mathbf{A}})$ *.*

This theorem is a modest generalization of prior results and can be obtained in much the same manner as results in Kouritzin and Zeng [23] and Xiong and Zeng [22]. Here, $\overline{\mathbf{A}}$ is the generator of the joint martingale problem to $(X, \theta, \vartheta, \widehat{Z}^h)$ obtained from \mathbf{A} , the generator of state (X, θ) and \mathbf{A}^Z , the generator of the historical process \widehat{Z}^h . We do not need an explicit formula for $\overline{\mathbf{A}}$. Instead, we can use particle filters to approximate σ_t .

Henceforth, it is convenient to think of the reference measure \mathbb{Q} as the standard measure from which we can construct the measure $\mathbb{P}^{k,\theta,\vartheta,h,\Delta}$ corresponding to model $k \in \{1,\ldots,5\}$ with parameters θ and microstructure with parameters ϑ , h, and Δ .

3.2. Particle Filter. The weighted filter is the simplest of particle filters. The idea behind the weighted filter is that, by the independence of signal $(X, \theta, \vartheta, \widehat{Z}^h)$ from the observations Y under \mathbb{Q} , we can create an infinite collection of particles $\{P^k\}_{k=1}^N = \{(X^k, \theta^k, \vartheta^k, \widehat{Z}^{h,k})\}_{k=1}^N$, each having the same law as $(X, \theta, \vartheta, \widehat{Z}^h)$ that are also independent of the observations. Then, it follows from the law of large numbers that for \mathbb{Q} -almost all Y we have the weak convergence of finite measures

$$\sigma_t^{N,W} \doteq \frac{1}{N} \sum_{k=1}^N L_t \left(X^k, \vartheta^k, Z^{h,k} \right) \delta_{(X_t^k, \theta^k, \vartheta^k, \widehat{Z}_t^{h,k})} \Longrightarrow \sigma_t.$$
(53)

Unfortunately, it is well known that the weighted particle filter may not work well for a fixed number of particles *N*. Roughly speaking, most of the particles diffuse away, do not track the signal well, are assigned low likelihoods, and do not really affect the average $\sigma_t^{N,W}$. Meanwhile, very few particles do match the observations better and have likelihoods that are orders of magnitude higher than of the majority of particles. $\sigma_t^{N,W}$ essentially becomes an average over too few particles to reflect σ_t well.

To fix the weighted filter particle spread problem, we add particle resampling, resulting in following novel particle filter. (See Gordon et al. [45], Del Moral et al. [46], Del Moral et al. [12], and Ballantyne et al. [47] for earlier algorithms.) For some large $N \in \mathbb{N}_0$, the particle system $\{P^k\}_{k=1}^N = \{(X^k, \theta^k, \vartheta^k, \widehat{Z}^{h,k})\}_{k=1}^N$ is constructed as follows.

3.2.1. Initialization. At the initial time $t_0 = 0$, we generate independent particles $\{P_0^k\}_{k=1}^N$ from the joint prior distribution $\pi_0(\cdot)$ of $(X_0, \theta, \theta, \widehat{Z}_0^h) \in \mathbb{R}^{n_x+n_\theta+2} \times C[0, T]$. The empirical measure at t_0 is

$$\sigma_0^N = \frac{1}{N} \sum_{k=1}^N \delta_{P_0^k} \left(\cdot \right),$$
 (54)

where $\delta_x(\cdot)$ is the Dirac measure at *x*. By the strong law of large numbers,

$$\lim_{N \to \infty} \left(\sigma_0^N, f \right) = \sigma_0 \left(f \right)$$

$$\forall f \in B \left(\mathbb{R}^{n_x + n_\theta + 2} \otimes C \left[0, T \right] \right),$$
(55)

so $\sigma_0^N \Rightarrow \sigma_0$ for almost all *Y*. Here, $(\mu, f) \doteq \int f(y)\mu(dy)$ for measures μ so

$$\left(\sigma_{0}^{N},f\right) = \frac{1}{N}\sum_{k=1}^{N}f\left(P_{0}^{k}\right).$$
 (56)

Remark 16. When $L_0 = 1$, $\pi_0(f) = \sigma_0(f)$. Note that \widehat{Z}_0^h is a constant function defined on [0, T]. Whereas most particle filters approximate the filter π_t , we will approximate the unnormalized filter σ_t to facilitate Bayesian model selection without the storage of prior filter estimates.

We also initialize the number of particles to $\mathbb{N}_0 = N$ and particle likelihoods all to $\mathbb{A}_0 = 1$.

3.2.2. Evolution. Between observations at t_{i-1} and t_i , the particles, $\{(X^k, \theta^k, \vartheta^k, \widehat{Z}^{h,k})\}_{k=1}^{\mathbb{N}_{i-1}}$, move independently as samples from the transition probability of $(X, \theta, \vartheta, \widehat{Z}^h)$. In particular, we use the Euler scheme (see, e.g., Kloeden and Platen [48]) to evolve the dynamics, Examples 3–7 and (25). We let \widehat{P}_{t_i} denote the evolved version of $P_{t_{i-1}}$.

3.2.3. Particle Weights and Average Weight. We simulate using the reference measure Q and we incorporate the observations

based upon (48). At the *i*th observation (t_i, Y_{t_i}) , the *k*th particle's weight is multiplied by

$$\omega_{i}^{k} \doteq \exp\left(\int_{t_{i-1}}^{t_{i}} \int_{E} \ln \frac{\lambda_{z}\left(X_{s}^{k}, \vartheta^{k}, Z_{s}^{h,k}, s\right)}{\kappa_{z}} Y\left(dz, ds\right)$$
$$-\int_{t_{i-1}}^{t_{i}} \left(a\left(s\right) - \kappa\right) ds\right)$$
$$= \exp\left(\ln \frac{\lambda_{z_{i}}\left(X_{t_{i}}^{k}, \vartheta^{k}, Z_{t_{i}}^{h,k}, t_{i}\right)}{\kappa_{z_{i}}}\right)$$
$$-\int_{t_{i-1}}^{t_{i}} \left(a\left(s\right) - \kappa\right) ds\right) \triangleq \alpha_{i}\left(\widehat{P}_{t_{i}}^{k}, t_{i}, t_{i-1}\right),$$
(57)

where $z_i = Y_{t_i}$. Hence, the *k*th particle's weight becomes

$$\widehat{\mathbb{L}}_{i}^{k} = \omega_{i}^{k} \mathbb{A}_{i-1} \tag{58}$$

and the average weight is

$$\mathbb{A}_i = \frac{1}{N} \sum_{k=1}^{\mathbb{N}_{i-1}} \widehat{\mathbb{L}}_i^k.$$
(59)

Note that $(X_{t_i-}^k, Z_{t_i-}^{h,k}, t_i-) = (X_{t_i}^k, Z_{t_i}^{h,k}, t_i)$ in (57) by continuous paths. Here, ω_i^k depends on the observation Y and the increment of likelihood ratio of measure \mathbb{P} over measure \mathbb{Q} defined by (48) given the simulated particle path realized on the interval $[t_{i-1}, t_i)$. These weights do not depend upon the parameters θ directly. This is common and is why the observations are often called partial observations. We still can estimate θ and include these parameters as part of the particles' states since they do affect stock price S, which is observed in the presence of noise and distortion. The weights are stored along with the states of particles before resampling.

3.2.4. Resampling. After weighting, we resample the particles pruning the unlikely ones and duplicating the better ones in an unbiased manner. In particular, we let ρ_i^k be $(\widehat{\mathbb{L}}_i^k/\mathbb{A}_i - \lfloor \widehat{\mathbb{L}}_i^k/\mathbb{A}_i \rfloor)$ -Bernoulli random variable independent of everything and produce $\lfloor \widehat{\mathbb{L}}_i^k/\mathbb{A}_i \rfloor + \rho_i^k$ particles at location $\widehat{P}_{t_i}^k$. We then give all the particles weight A_i and let

$$\mathbb{N}_{i} \doteq \sum_{k=1}^{\mathbb{N}_{i-1}} \left\{ \left\lfloor \frac{\widehat{\mathbb{L}}_{i}^{k}}{\mathbb{A}_{i}} \right\rfloor + \rho_{i}^{k} \right\}.$$
(60)

3.2.5. Unnormalized Filter. Now, we can estimate the unnormalized filter at the *i*th observation time, σ_{t_i} , by

$$\sigma_{t_i}^N = A_i \sum_{k=1}^{N_i} \delta_{P_{t_i}^k}.$$
(61)

The actual algorithm that was implemented is as follows.

Initialize. $\{P_0^k\}_{k=1}^N$ are independent samples of π_0 , $\mathbb{N}_0 = N$, $\mathbb{N}_n = 0$, for all $n \in \mathbb{N}$, and $\mathbb{L}_0^k = 1$ for k = 1, ..., N.

Repeat. For n = 0, 1, 2, ..., do

- (1) evolve $P_{t_{i}}^{k}$ to $\hat{P}_{t_{i+1}}^{k}$ independently of other particles;
- (2) weight by observation: $\widehat{\mathbb{L}}_{n+1}^k = \alpha_{n+1}(\widehat{P}_{t_{n+1}}^k, t_{n+1}, t_n) \mathbb{A}_n$ for $k = 1, 2, ..., \mathbb{N}_n$;
- (3) estimate $\sigma_{t_{n+1}}$ by $\sigma_{t_{n+1}}^N = \frac{1}{N} \sum_{k=1}^{\mathbb{N}_n} \widehat{\mathbb{L}}_{n+1}^k \delta_{\widehat{P}_{t_{n+1}}^k}$;
- (4) average weight: $\mathbb{A}_{n+1} = \sigma_{t_{n+1}}^N(1);$
- (5) repeat: for $k = 1, 2, \ldots, \mathbb{N}_n$ do
 - (a) offspring number: $\mathbb{N}_{n+1}^k = \lfloor \widehat{\mathbb{L}}_{n+1}^k / \mathbb{A}_{n+1} \rfloor + \rho_n^k$, with ρ_n^k being $((\widehat{\mathbb{L}}_{n+1}^k / \mathbb{A}_{n+1}) - \lfloor \widehat{\mathbb{L}}_{n+1}^k / \mathbb{A}_{n+1} \rfloor)$ -Bernoulli independent of everything;
 - (b) resample: $P_{t_{n+1}}^{\mathbb{N}_{n+1}+j} = \widehat{P}_{t_{n+1}}^k$ for $j = 1, ..., \mathbb{N}_{n+1}^k$;
 - (c) add offspring number: $\mathbb{N}_{n+1} = \mathbb{N}_{n+1} + \mathbb{N}_{n+1}^k$

Remark 17. (i) We extract our estimate before resampling to avoid excess noise. (ii) The key step is (5) that determines the new number of particles \mathbb{N}_{n+1} and weights \mathbb{L}_{n+1}^k in an unbiased manner. The result is zero or more particles all having the average weight at the same location as the parent. (iii) The particle evolution would typically be done via Newton's or Milstein's method.

Since the above algorithm produces unbiased resampling of the weighted particle filter, it is quite reasonable to believe the following result.

Theorem 18. Under (C1) and (C2), $\sigma_{t_i}^N \Rightarrow \sigma_{t_i}$ for any *i* and almost all observation paths.

The technicality of this result's proof would detract from our applications so it is omitted.

3.2.6. Bayesian Estimation. By Bayes rule (50), the particle approximation of the normalized filter $\pi(\cdot)$ is

$$\pi_{t_i}^N(f) = \frac{\sigma_{t_i}^N(f)}{\sigma_{t_i}^N(1)}$$
(62)

for all $f \in B(\mathbb{R}^{n_x+n_{\theta}+2} \times C[0,T])$. To get our parameter estimates, we can just set $f(X,\theta,\vartheta,\widehat{Z}^h)$ to one component of these parameters, that is, θ^i or ϑ^j .

3.3. Calibration and Historical Training. To keep the problem size manageable, we just used the clustering parameter estimates of α , β , γ_1 , and γ_2 given above as the actual values throughout our simulations.

One is often faced with the problem of estimating initial distributions for fair price, volatility, and the parameters prior to filtering over the time interval of interest (April 2010 here). Our approach was to make arbitrary assignments very far in the past (January 3, 2000, to be precise) and then do an excessive amount of prior particle filtering, relying on the ability of the filter to forget its starting point and to produce



FIGURE 3: Long-term value estimation of PEP.

reasonable distributions at a much later point, April 1, 2010. (See, e.g., Ocone and Pardoux [49], Delyon and Zeitouni [50], and Atar [51] for mathematical results regarding this phenomenon.) This had to be done for every model, namely, every combination of our three stocks, five SV models, and multiple microstructure models, characterized by inertia parameters. Our main purpose in this historical training was to get a starting joint distribution for $(X, \theta, \vartheta, \widehat{Z}^h)$ as of April 1, 2010, under each model combination. Due to the large number of cases this produced, we first display and discuss two models: the nondynamical microstructure Heston case and the *median inertia* dynamical case where h = 1/2 and $\Delta = 7200 \text{ s}$ (i.e., 2 hrs) in the inertia microstructure model. Also, to ensure that θ and ϑ did not converge to a single value, we made them vary slightly in a random manner; that is, we replaced the equation $d\theta = 0$ with $d\theta_t = dv_t$ for a very low variance Brownian motion v.

In Figure 3, we illustrate our prior filtering of PepsiCo. The choppiest curve is the actual stock price while the smoothest curve is the filter's fair price estimate $E[S_t | \mathscr{F}_t^Y]$ using the Heston SV model with (median) microstructure inertia. The middle curve is the filter's fair price estimate $E[S_t | \mathscr{F}_t^Y]$ using the Heston SV model without dynamics in the microstructure; that is, $Z^h = 0$. These curves go beyond April 1, 2010. However, the required initial distributions were taken from the filter at that point.

Notice from Figure 3 that the implied fair price process estimate is far less volatile in the presence of dynamical microstructure than without. This lower volatility for fair price is highly desirable. It does not make sense that the fair price of a stock should fluctuate dramatically from day to day or within a day in the absence of an event, but rather these short-term fluctuations are better explained by trading noise. Moreover, fair price is a mathematically more optimal version of moving averages, which are used to judge value and momentum from, and so fair price estimates should inherit the smooth nature of such moving averages.

TABLE 4

_					
PE	P GBM	HW	LOU	Nelson	Heston
μ	1.51E - 06	1.47E - 06	1.52E - 06	1.44E-06	1.49E - 06
ν	$\sigma=2.86E-06$	1.17E - 09	9.55E-06	1.06E-10	1.07E - 11
κ		1.59E - 03	1.80E-03	1.94E-03	2.58E - 07
ę	_	_	4.75E - 03	6.51E - 03	6.02E - 03
Ŵ	_	_	4.84E - 06	_	_

TABLE	5
-------	---

$\begin{array}{c ccccccccccccccccccccccccccccccccccc$						
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	PEF	P GBM	HW	LOU	Nelson	Heston
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	μ	1.05E - 06	1.02E - 06	9.92E - 07	1.03E - 06	1.01E - 06
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	ν	$\sigma=2.21E-06$	5.50E-10	5.13E-06	6.32E-11	5.94E - 12
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	κ	_	2.18E-03	1.87E-03	2.12E-03	2.26E - 07
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	ę	_	_	2.25E-03	2.90E-03	3.23E - 03
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	Ŵ	_	—	2.60E-06	—	—
ϕ 2.13 <i>E</i> - 09 2.31 <i>E</i> - 09 2.23 <i>E</i> - 09 2.33 <i>E</i> - 09 2.31 <i>E</i> - 0	е	2.43E - 09	2.05E - 09	2.33E-09	2.31E - 09	2.46E - 0.9
	ϕ	2.13E - 09	2.31E-09	2.23E-09	2.33E-09	2.31E - 0.9

From a modeling perspective, this fair price smoothness indicates that dynamical microstructure (with inertia) can replace much of what stochastic volatility tries to do and leads to one of our central questions addressed below. Is stochastic volatility necessary in the presence of dynamical microstructure?

3.4. Numerical Results. The data is one month (April 2010) of transaction prices of our three NYSE-listed stocks. Our filter produces Bayes estimates to the macro- and microparameter vectors θ and ϑ , respectively. These estimates in the nondynamical microstructure case (i.e., using the simpler form in (24)) for PepsiCo are as shown in Table 4.

All parameters are estimated using time in seconds. Our PepsiCo Bayes estimates in the *median* inertia case are as shown in Table 5.

While it is difficult to read much from these numbers, we can see that the main volatility parameters ν , κ , and ρ are mostly smaller when dynamics is included in the microstructure. This further justifies our conjecture that at least some stochastic volatility is better replaced by microstructure with dynamics.

Figures 4 and 5 show the conditional expectation fair price estimation for Goldman Sachs and PepsiCo, respectively, in the cases of no dynamics and median inertia dynamics for each of our SV models. There are a total of eleven curves in both figures. The most volatile curve is the stock price itself over this month. The smoothest curves somewhat separated from the stock price are the fair price estimates using the five SV models with (median inertia) dynamical microstructure. The remaining five curves (that hug the stock price in Figures 4 and 5) are our fair price estimates for our five SV models with nondynamical microstructure. In this last case, the microstructure does not have the power to separate the fair price and actual stock price to any large degree.



FIGURE 4: Value estimation of GS, April 2010.



FIGURE 5: Value estimation of PEP, April 2010.

It is important to realize that these pictures are really just a one-month snapshot of a much bigger multiyear filtering process. This explains why many of the fair price processes are significantly different than the actual stock price on April 1, 2010: The filter is estimating that the difference is due to the microstructure. It is apparent that adding dynamics to the microstructure allows the estimated fair price process to differ significantly from the stock price. Indeed, there is a significant correction of all three stock prices (especially Goldman Sachs) towards estimated fair price of the models with (median inertia) dynamical microstructure. This produces a compelling reason to use models with microstructure dynamics. You would be estimating that the stocks were significantly overvalued before the correction if you used the model with microstructure dynamics and this could be

TABLE 6: Volatility estimation, April 2010.

	Without dynamics	With dynamics
PEP (2 hrs, $h = 0.6$)	1.58416E - 09	1.01312E - 11
GS (1/2 day, $h = 0.4$)	4.14645E - 08	4.3005E - 10
IBM (1/2 day, $h = 1$)	8.21731E - 10	4.03211E - 11

used as a warning to lessen ones exposure. You have no such warning when the microstructure does not contain (inertia) dynamics as the estimated fair price is very close to the observed price. It is interesting to ponder what this possible discrepancy would mean to option prices.

The filters provide conditional distributions and estimates for more than just fair price and parameters. Table 6 shows the average volatility estimates without microstructure dynamics (see (24)) and with (the best performing) microstructure inertia using the simplified Heston SV model. We only highlighted Heston here because (1) we will show evidence below that Heston performs the best and (2) the volatility estimates of the other SV models behave similarly. The amount of stochastic volatility estimated when there is (median inertia) dynamics in the microstructure shrank to a couple of percent of what it was without. This really suggested that by far the primary use of stochastic volatility is as a proxy for microstructure with dynamics and further raises the question about the need for stochastic volatility in the presence of microstructure dynamics.

The final and most difficult quantity the filter estimates (in the dynamical microstructure case) is the historical noise. For practical purposes, we can not let the historical path go back all the way to year 2000, but we found that there is not much loss if we just update discrete samples over the previous three years, which is still a tremendous amount of data. Also, we can not plot these historical paths so we just plot the projection onto the current time; that is, we just plot Z_t^h even though we must propagate the Markov process \widehat{Z}_t^h in the filter. Figure 6 shows the noise estimate for PepsiCo. In this graph, we look at the effect of inertia. The curves where h = 0 represent the no-inertia case, so Z_t^0 is just an Ornstein-Uhlenbeck process. Conversely, the case h = 1 represents the one hundred percent inertia case and Z_t^1 is not Markov. We see from these graphs that the amount of estimated noise is very similar indicating that the amount of inertia modeled might not be that significant. However, the noise processes where h = 1 are far smoother due to the inertia. Below, we will produce strong evidence that inertia is important and find that the best h is in the range [0.4, 1], depending upon the stock. We compare the behavior of our models in terms of the SV models and the inertia parameters *h* and Δ within the Bayesian model selection framework in the following section.

4. Evidence for Inertia and Stochastic Volatility

The main objective of this section is to use Bayes factor to investigate the model selection in microstructure markets. To use the Bayes factor method, we need only to be able to



FIGURE 6: Noise process estimation of PEP, April 2010.

transform all observation models of interest into the same canonical process via Girsanov measure change. The signal models can be singular to one another. Kouritzin and Zeng [23] discuss the Bayesian model selection problem. However, their equations do not apply to our models.

4.1. *Model Selection and Bayes Factor.* Consider our five SV macrostructure fair price-volatility models

$$M^{(k)} \doteq \left(X^{(k)}, \theta^{(k)}\right) \in \mathbb{R}^{n_x^{(k)} + n_\theta^{(k)}}, \tag{63}$$

where the generators of the martingale problem to $M^{(k)}$ are, respectively, $\mathbf{A}^{(k)}$ for k = 1, 2, 3, 4, 5. Normally, we would have to consider a multitude of parameters θ resulting in a plethora of models. However, by our calibration process we have reduced the setting to one parameter set per martingale problem so we have a base of five models. However, we still have to consider the various choices for our inertia. For simplicity, we restrict ourselves to three distinct values for Δ , eleven choices for h, and we use the calibration process to estimate the other microstructure parameters ϑ . Therefore, we have a total of $5 \times 3 \times 11 = 165$ models to test.

The likelihood of *Y* being produced by model (k, h, Δ) up until time *t* is

$$L_{t}^{(k,h,\Delta)} = 1 + \int_{0}^{t} \int_{E} \left(\frac{\lambda_{z} \left(X_{s}^{(k)}, Z_{s}^{h,\Delta}, \vartheta, s \right)}{\kappa_{z}} - 1 \right)$$

$$\cdot L_{s-}^{(k,h,\Delta)} \left(Y \left(dz, ds \right) - \kappa_{z} m \left(dz \right) ds \right).$$
(64)

Here, m(dz) is the counting measure on $E = \mathbb{N}_0$ and the same observations and observation rate information are used for all models. One can think of $L_t^{(k,h,\Delta)}$ as the likelihood ratio of the model $M^{(k,h,\Delta)}$ with distribution $\mathbb{P}^{(k,h,\Delta)}$ characterized by (k, h, Δ) to the simple (or null) model M^0 with distribution \mathbb{Q} where the observation prices just arrive according to a Poisson measure with intensity measure $\mu(A) = \int_A k_z m(dz)$, that is, with rate independent of any macrostructure model and independent of any microstructure state. In other words, $(L_t^{(k,h,\Delta)})^{-1} = (d\mathbb{Q}/d\mathbb{P}^{(k,h,\Delta)})|_{\mathscr{F}_t}$ then transforms the observations into the same Poisson measure with intensity measure $\mu(A) = \int_A k_z m(dz)$ regardless of (k, h, Δ) . Unfortunately, $L_t^{(k,h,\Delta)}$ depends upon $X_s^{(k)}, Z_s^{h,\Delta}$, which are unknown so we can not select models via the likelihood.

4.2. *Bayes Factor.* The available information in microstructure market is the observation process *Y*, which represents the cumulative transaction records throughout all tick price levels. The normalized filter $\pi_t^{(k,h,\Delta)}$, $k = 1, 2, 3, 4, 5, h \in [0, 1]$, $\Delta > 0$, satisfies

$$\pi_t^{(k,h,\Delta)}\left(f_k\right) = \frac{\sigma_t^{(k,h,\Delta)}\left(f_k\right)}{\sigma_t^{(k,h,\Delta)}\left(1\right)},\tag{65}$$

where $f_k \in B(\mathbb{R}^{n_x^{(k)}+n_\theta^{(k)}+2} \otimes C[0,T])$ for k = 1, 2, 3, 4, 5, the unnormalized filter $\sigma_t^{(k,h,\Delta)}$ is

$$\begin{aligned} &\sigma_t^{(k,h,\Delta)}\left(f_k\right) \\ &\doteq \mathbb{E}^{\mathbb{Q}}\left[f_k\left(X_t^{(k)}, \theta^{(k)}, \vartheta, \widehat{Z}_t^{h,\Delta}\right) L_t^{(k,h,\Delta)} \mid \mathscr{F}_t^Y\right], \end{aligned}$$
(66)

and $\sigma_t^{(k,h,\Delta)}(1)$ is the integrated (or marginal) likelihood of *Y* for model (k, h, Δ) .

Now, we use Bayes factor to compare models. The Bayes factor determines which model best fits this observed data by doing pairwise comparisons. We define Bayes factor of model $M^{(k,h,\Delta)}$ to the null model by the conditional likelihood:

$$\mathbb{E}^{\mathbb{Q}}\left[\frac{d\mathbb{P}^{(k,h,\Delta)}}{d\mathbb{Q}} \mid \mathscr{F}_{t}^{Y}\right] = \sigma_{t}^{(k,h,\Delta)}\left(1\right),\tag{67}$$

which is consistent with more basic definitions of Bayes factor. It then follows that the Bayes factors for two models, characterized by (k_1, h_1, Δ_1) and (k_2, h_2, Δ_2) , are the ratios

$$B_{12}(t) = \frac{\sigma_t^1(1)}{\sigma_t^2(1)},$$

$$B_{21}(t) = \frac{\sigma_t^2(1)}{\sigma_t^1(1)},$$
(68)

with the integrated likelihoods $\sigma_t^1(1) = \sigma_t^{(k_1,h_1,\Delta_1)}(1), \sigma_t^2(1) = \sigma_t^{(k_2,h_2,\Delta_2)}(1)$ that can be approximated using the algorithm of Section 3.2.5. Kass and Raftery [20] demonstrate how to interpret Bayes factor shown in Table 7.

4.3. Numerical Results on Stochastic Volatility. First, we consider the problem of selecting the best of our five fair price-volatility models,

$$M^{(k)} \doteq \left(X^{(k)}, \theta^{(k)}\right),\tag{69}$$

and the resulting partially observed market models,

$$\left(X^{(k)}, \widehat{Z}^{h,\Delta}, \theta^{(k,h,\Delta)}, \vartheta; Y\right).$$
(70)

TABLE 7

B ₁₂	Evidence against Model 2
1–3	Barely mentionable
3–12	Positive
12–150	Strong
>150	Decisive



FIGURE 7: Bayes factor for PEP SV model determination, April 2010.

TABLE 8: Bayes factor for model determination, April 2010.

	Heston	GARCH	Log O-U	HW	GBM
PEP (2 hrs, $h = 0.6$)	27.30	25.23	21.31	17.04	6.08
GS (1/2 day, $h = 0.4$)	19.14	18.91	18.77	18.59	10.18
IBM $(1/2 \text{ day}, h = 1)$	49.94	45.43	40.75	37.07	16.16
PEP* (2 hrs, $h = 0.6$)	1.08	1.07	1.06	1.04	1.00
GS^* (1/2 day, $h = 0.4$)	1.06	1.04	1.06	1.06	1.00
IBM^* (1/2 day, $h = 1$)	1.06	1.04	1.06	1.03	1.00
*					

*Without dynamics

We compare these five models to determine which can best represent the market data. More precisely, we run all unnormalized filters as explained in Section 3.2 with the optimal parameters discovered and reported earlier. Then, we choose Model *i* if $\sigma_T^{(i,h,\Delta)}$ is the largest. Naturally, this corresponds to the model whose Bayes factor ends up greater than one when compared to any other model. While we have five basic models, we also consider different market ingestion times Δ and inertia magnitude parameters *h* for each model.

Using GBM with nondynamic microstructure (i.e., $Z^h = 0$) as the benchmark, we determine which combination of SV model and inertia parameters outperforms GBM most. We first focus on the candidate models (Examples 3–7). In each case, we pick the inertia parameters from the sets $\Delta \in \{30 \text{ mins}, 2 \text{ hrs}, 1/2 \text{ day}\}$ and $h \in \{0, 0.1, 0.2, \dots, 0.9, 1\}$ that would yield the highest Bayes factor against the calibration model. The data is the transaction price of PepsiCo, IBM, and Goldman Sachs, April 2010. Figure 7 and Table 8 summarize

PEP: Bayes factor for Heston model with h = 0.6 (* indicates models without dynamics)



FIGURE 8: Bayes factor for PEP ingestion time determination, April 2010.

TABLE 9: Bayes factor for ingestion time determination, April 2010.

	Heston*	40 mins	2 hrs	1/2 day
GS (<i>h</i> = 0.4)	1.000	15.083	17.578	18.100
PEP ($h = 0.6$)	1.000	19.066	25.259	24.187
IBM $(h = 1)$	1.00	31.152	42.744	46.988
*				

* Without dynamics.

the Bayes factor performance. The Bayes factors computed in this table give strong evidence (based upon the Kass and Raftery criterion mentioned before) for the Heston model based on a full month of real tick-by-tick stock price data. Indeed, as we will see below, there would still be strong evidence supporting Heston if we used different values of h and Δ . It is also interesting that the order of the models did not change over our three stock selections, with Heston always being preferred and GBM always performing the worst. Recall that all models are tuned to have their best parameters θ and ϑ .

4.4. Numerical Results on Inertia. Next, we look at the ingestion time Δ using nondynamic microstructure Heston as the calibration model. Figure 8 and Table 9 show the effect of varying Δ over {30 mins, 2 hrs, 1/2 day} for $h \in \{0, 0.1, 0.2, ..., 0.9, 1\}$ fixed to give the highest Bayes factor. There is a drop in the Bayes factor values from the model determination experiment which is entirely due to the change of calibration model from GBM with nondynamic microstructure to Heston with nondynamic microstructure. Our results show that the best ingestion times for Goldman Sachs, PepsiCo, and International Business Machines stocks are, respectively, 1/2 day, 2 hours, and 1/2 day. The fact that the data supports long-time ingestion might add merit to the case of the momentum trader.

15



FIGURE 9: Bayes factor for PEP inertia determination, April 2010.

TABLE 10: Bayes factor for inertia determination, April 2010.

h	*	0	0.1	0.2	0.3	0.4
PEP (2 hrs)	1.00	3.745	5.875	6.950	11.693	16.733
GS (1/2 day)	1.00	11.578	13.507	16.194	17.746	18.100
IBM (1/2 day)	1.00	3.822	7.100	8.816	10.927	13.522
h	0.5	0.6	0.7	0.8	0.9	1.0
PEP (2 hrs)	23.524	25.259	24.386	22.322	19.347	17.548
GS (1/2 day)	17.878	17.184	16.515	16.225	16.008	15.612
IBM (1/2 day)	16.707	20.611	25.388	31.225	38.345	46.988

*Without dynamics.

Finally, we investigate the optimal amount of inertia. Figure 9 and Table 10 show the effect of varying the amount of inertia *h* over $\{0, 0.1, 0.2, ..., 0.9, 1\}$ for $\Delta \in \{30 \text{ mins}, 2 \text{ hrs}, 1/2 \text{ day}\}$ fixed to give the highest Bayes factor. The table shows that inertia is important. In fact, the best *h* was always at least h = 0.4 and was even h = 1 in the case of IBM so all microstructure dynamics should be driven by the inertia process.

5. Conclusions

Herein, we considered five popular SV models to represent intrinsic or fair price and stochastic volatility of this price. These SV models are free of inertia or momentum. We then added microstructure noise with possible dynamics and inertia to these SV models to accommodate trading noise, trend following, information dispersion, and the slow unwinding of big positions. We used Bayesian model selection techniques to determine which of these combined models fits real NYSE data best. In the process of selecting the best model we also investigated characteristics like microstructure dynamics, inertia, and stochastic volatility. For the stock data considered, we can conclude the following:

- (1) Bayesian model selection through particle filtering provides a computationally effective means to identify the best finance models on real tick-by-tick data.
- (2) The SV and inertia components of the financial models compared can be singular to each other as long as the microstructure can be changed into the same canonical Poisson measure process for all models considered.
- (3) There is strong evidence of *dynamical* microstructure noise.
- (4) Adding dynamics to the microstructure allowed much greater deviations of price from intrinsic value, which can be detected by filtering and used as a warning sign to investors and traders.
- (5) The simplified Heston stochastic volatility model with microstructure dynamics and significant inertia performed the best in all cases.
- (6) There is strong statistical evidence that such simplified Heston stochastic volatility models with microstructure dynamics and inertia match the data better than the classical geometrical Brownian motion.
- (7) The amount of inertia h and the time it lasted Δ varied a little from stock to stock but in all cases there was significant inertia that lasted for hours.

More complicated SV models can be investigated in our future work. One could also postulate more complicated microstructure dynamics and consider additional real data analysis.

Conflict of Interests

The author declares that there is no conflict of interests regarding the publication of this paper.

References

- F. Black and M. Scholes, "The pricing of options and corporate liabilities," *Journal of Political Economy*, vol. 81, no. 3, pp. 637– 654, 1973.
- [2] R. C. Merton, "Theory of rational option pricing," *The Bell Journal of Economics and Management Science*, vol. 4, no. 1, pp. 141–183, 1973.
- [3] J. C. Cox, S. A. Ross, and M. Rubinstein, "Option pricing: a simplified approach," *Journal of Financial Economics*, vol. 7, no. 3, pp. 229–263, 1979.
- [4] S. L. Heston, "A closed-form solution for options with stochastic volatility with applications to bond and currency options," *Review of Financial Studies*, vol. 6, no. 2, pp. 327–343, 1993.
- [5] J. C. Jackwerth and M. Rubinstein, "Recovering probability distributions from option prices," *The Journal of Finance*, vol. 51, no. 5, pp. 1611–1631, 1996.
- [6] J. Hull and A. White, "The pricing of options on assets with stochastic volatilities," *The Journal of Finance*, vol. 42, no. 2, pp. 281–300, 1987.

- [7] D. B. Nelson, "ARCH models as diffusion approximations," *Journal of Econometrics*, vol. 45, no. 1-2, pp. 7–38, 1990.
- [8] B. B. Mandelbrot, Fractals and Scaling in Finance: Discontinuity, Concentration, Risk, Springer, 1997.
- [9] A. Shiryaev, *Essentials of Stochastic Finance*, World Scientific, 1999.
- [10] M. Capinski and T. Zastawniak, *Mathematics for Finance. An Introduction to Financial Engineering*, Springer, London, UK, 2nd edition, 2011.
- [11] M. A. Kouritzin and Y. Zeng, "Weak convergence for a type of conditional expectation: application to the inference for a class of asset price models," *Nonlinear Analysis: Theory, Methods & Applications*, vol. 60, no. 2, pp. 231–239, 2005.
- [12] P. Del Moral, M. A. Kouritzin, and L. Miclo, "On a class of discrete generation interacting particle systems," *Electronic Journal of Probability*, vol. 6, pp. 1–26, 2001.
- [13] R. F. Engle, "The econometrics of ultra-high-frequency data," *Econometrica*, vol. 68, no. 1, pp. 1–22, 2000.
- [14] F. Black, "Noise," Journal of Finance, vol. 41, pp. 529-543, 1986.
- [15] L. K. C. Chan and J. Lakonishok, "Institutional trades and intraday stock price behavior," *Journal of Financial Economics*, vol. 33, no. 2, pp. 173–199, 1993.
- [16] J. Hasbrouck, Handbook of Statistics, vol. 14 of Modelling Market Microstructure Time Series, North-Holland, Amsterdam, The Netherlands, 1996.
- [17] J. Hasbrouck, "Security bid/ask dynamics with discreteness and clustering: simple strategies for modeling and estimation," *Journal of Financial Markets*, vol. 2, no. 1, pp. 1–28, 1999.
- [18] R. F. Engle and J. R. Russell, "Autoregressive conditional duration: a new model for irregularly spaced transaction data," *Econometrica*, vol. 66, no. 5, pp. 1127–1162, 1998.
- [19] F. M. Bandi and J. R. Russell, "Separating microstructure noise from volatility," *Journal of Financial Economics*, vol. 79, no. 3, pp. 655–692, 2006.
- [20] R. E. Kass and A. E. Raftery, "Bayes factor and model uncertainty," *Journal of the American Statistical Association*, vol. 90, pp. 773–795, 1995.
- [21] Y. Zeng, "A partially observed model for micromovement of asset prices with bayes estimation via filtering," *Mathematical Finance*, vol. 13, no. 3, pp. 411–444, 2003.
- [22] J. Xiong and Y. Zeng, "A branching particle approximation to a filtering micromovement model of asset price," *Statistical Inference for Stochastic Processes*, vol. 14, no. 2, pp. 111–140, 2011.
- [23] M. A. Kouritzin and Y. Zeng, "Bayesian model selection via filtering for a class of micro-movement models of asset price," *International Journal of Theoretical and Applied Finance*, vol. 8, no. 1, pp. 97–121, 2005.
- [24] D. W. Stroock and S. R. S. Varadhan, *Multidimensional Diffusion Processes*, Springer, Berlin, Germany, 1979.
- [25] S. N. Ethier and T. G. Kurtz, Markov Processes: Characterization and Convergence, John Wiley & Sons, New York, NY, USA, 1986.
- [26] L. O. Scott, "Option pricing when the variance changes randomly: theory, estimation, and an application," *The Journal of Financial and Quantitative Analysis*, vol. 22, no. 4, pp. 419–438, 1987.
- [27] F. C. Drost and B. J. M. Werker, "Closing the GARCH gap: continuous time GARCH modeling," *Journal of Econometrics*, vol. 74, no. 1, pp. 31–57, 1996.
- [28] R. J. Elliott, W. P. Malcolm, and A. H. Tsoi, "Robust parameter estimation for asset price models with Markov modulated

volatilities," *Journal of Economic Dynamics and Control*, vol. 27, no. 8, pp. 1391–1409, 2003.

- [29] D. Duffie, J. Pan, and K. Singleton, "Transform analysis and asset pricing for affine jump-diffusions," *Econometrica*, vol. 68, no. 6, pp. 1343–1376, 2000.
- [30] B. Eraker, M. Johannes, and N. Polson, "The impact of jumps in volatility and returns," *The Journal of Finance*, vol. 58, no. 3, pp. 1269–1300, 2003.
- [31] B. Eraker, "Do stock prices and volatility jump? Reconciling evidence from spot and option prices," *The Journal of Finance*, vol. 59, no. 3, pp. 1367–1403, 2004.
- [32] P. R. Hansen and A. Lunde, "Realized variance and market microstructure noise," *Journal of Business and Economic Statistics*, vol. 24, no. 2, pp. 127–161, 2006.
- [33] J.-C. Duan and A. Fulop, "Estimating the structural credit risk model when equity prices are contaminated by trading noises," *Journal of Econometrics*, vol. 150, no. 2, pp. 288–296, 2009.
- [34] Y. Aït-Sahalia, P. Mykland, and L. Zhang, "How often to sample a continuous-time process in the presence of market microstructure noise," *Review of Financial Studies*, vol. 18, pp. 351–416, 2005.
- [35] N. Jegadeesh and S. Titman, "Returns to buying winners and selling losers: implications for stock market efficiency," *The Journal of Finance*, vol. 48, no. 1, pp. 65–91, 1993.
- [36] T. J. Moskowitz and M. Grinblatt, "Do industries explain momentum?" *The Journal of Finance*, vol. 54, no. 4, pp. 1249– 1290, 1999.
- [37] B. D. Grundy and J. S. Martin, "Understanding the nature of the risks and the source of the rewards to momentum investing," *Review of Financial Studies*, vol. 14, no. 1, pp. 29–78, 2001.
- [38] J. M. Griffin, X. Ji, and J. S. Martin, "Momentum investing and business cycle risk: evidence from Pole to Pole," *Journal of Finance*, vol. 58, no. 6, pp. 2515–2547, 2003.
- [39] R. D. Huang and H. R. Stoll, "Tick size, bid-ask spreads, and market structure," *Journal of Financial and Quantitative Analysis*, vol. 36, no. 4, pp. 503–522, 2001.
- [40] K. H. Chung, B. F. Van Ness, and R. A. Van Ness, "Trading costs and quote clustering on the NYSE and NASDAQ after decimalization," *Journal of Financial Research*, vol. 27, no. 3, pp. 309–328, 2004.
- [41] K. H. Chung, K. A. Kim, and P. Kitsabunnarat, "Liquidity and quote clustering in a market with multiple tick sizes," *Journal of Financial Research*, vol. 28, no. 2, pp. 177–195, 2005.
- [42] M. J. Barclay, "Bid-ask spreads and the avoidance of odd-eighth quotes on nasdaq: an examination of exchange listings," *Journal* of Financial Economics, vol. 45, no. 1, pp. 35–60, 1997.
- [43] Y. Ait-Sahalia and R. Kimmel, "Maximum likelihood estimation of stochastic volatility models," *Journal of Financial Economics*, vol. 83, no. 2, pp. 413–452, 2007.
- [44] P. Bremaud, Point Process and Queues: Martingale Dynamics, Springer, New York, NY, USA, 1981.
- [45] N. J. Gordon, D. J. Salmond, and A. F. M. Smith, "Novel approach to nonlinear/non-gaussian Bayesian state estimation," *IEE Proceedings, Part F: Radar and Signal Processing*, vol. 140, no. 2, pp. 107–113, 1993.
- [46] P. Del Moral, J.-C. Noyer, and G. Salut, "Maslov optimisation theory: stochastic interpretation, particle resolution," in 11th International Conference on Analysis and Optimization of Systems Discrete Event Systems, vol. 199 of Lecture Notes in Control and Information Sciences, pp. 312–318, Springer, Berlin, Germany, 1994.

- [47] D. J. Ballantyne, H. Y. Chan, and M. A. Kouritzin, "Novel branching particle method for tracking," in *Signal and Data Processing of Small Targets*, O. E. Drummond, Ed., vol. 4048 of *Proceedings of SPIE*, pp. 277–287, July 2000.
- [48] P. E. Kloeden and E. Platen, *Numerical Solution of Stochastic Differential Equations*, Springer, 1992.
- [49] D. Ocone and E. Pardoux, "Asymptotic stability of the optimal filter with respect to its initial condition," *SIAM Journal on Control and Optimization*, vol. 34, no. 1, pp. 226–243, 1996.
- [50] B. Delyon and O. Zeitouni, "Lyapunov exponents for filtering problems," in *Applied Stochastic Analysis*, M. H. A. Davis and R. J. Elliott, Eds., pp. 511–521, Gordon and Breach Science Publishers, London, UK, 1991.
- [51] R. Atar, "Exponential stability for nonlinear filtering of diffusion processes in a noncompact domain," *Annals of Probability*, vol. 26, no. 4, pp. 1552–1574, 1998.



Advances in **Operations Research**



The Scientific World Journal







Hindawi

Submit your manuscripts at http://www.hindawi.com



Algebra



Journal of Probability and Statistics



International Journal of Differential Equations





Complex Analysis





Mathematical Problems in Engineering



Abstract and Applied Analysis



Discrete Dynamics in Nature and Society



International Journal of Mathematics and Mathematical Sciences





Journal of **Function Spaces**



International Journal of Stochastic Analysis

