MICROSTRUCTURE MODELS WITH SHORT-TERM INERTIA AND STOCHASTIC VOLATILITY

RUNNING TITLE = MICROSTRUCTURE MODELS

MICHAEL A. KOURITZIN

Abstract. 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 so, which SV model matches the observed price data best? Bayes factor methods are chosen for determining best-fit to allow volatility models with very different structures to be considered. Nonlinear filtering techniques are utilized to compute the Bayes factor on tick-by-tick data and to estimate the unknown parameters. 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, etc.; as causes for price fluctuations over time. Yet, simple mathematical models like geometric Brownian motion (GBM) (e.g. Black and Scholes [8], Merton [43]) or the Cox-Ross-Rubinstein model (Cox, Ross and Rubinstein [13]) lump these factors together resulting in unnatural phenominom like the *volatility smile*. Stochastic volatility has been observed in real prices and is often added into the price value evolution (e.g. Heston [34], Jachwerth and Rubinstein [37], Hull and White [36], Nelson [45]) to handle the volatility smile. However, which stochastic volatility model fits the market data best? Moreover, even combined stochastic value-volatility models do not address tick size, price inertia, price clustering and fear-greed cycles. To handle these issues, one is drawn to tick-by-tick

Key words and phrases. Asset Price, Stochastic Volatility, Trading Noise, Price Inertia, Bayes Factor, Counting Process, Microstructure, Model Selection.

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 Bownian motion (FBM), best known for its long memory properties, exhibits inertia and has been used successfully to model markets (Mandelbrot [42], Shiryaev [49]). We speculate that FBM's success is more attributable to inertia than long memory, 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.

High frequency data contains complete market-participant trading activities (Engle [24]) and is modeled using microstructure (Black [7], Chan and Lakonishok [10], Hasbrouck [32], [33], Engle and Russell [23], Engle [24], Bandi and Russell [4] etc.). 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 microstructure models where the transaction price is formulated as a distorted and 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 price, raising 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 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. Bayes factor (see e.g. Kass and Raftery [39]) 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 be able to transform all microstructure asset-price observation models of interest into the same canonical process via Girsanov-type measure change.

Previously, Zeng [52] studied a filtering equation for inferring the intrinsic value process in a microstructure model while Xiong and Zeng [51] proposed a branching particle approximation to this equation. Kouritzin and Zeng [41] derived a Bayes factor equation and discussed the Bayesian model selection problem to determine if financial data, such as stock price, displays 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 proposing a new inertia process, 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.

Section 2 contains our inertia process and properties, the five standard value-volatility models (GBM, Hull-White, Log Ornstein-Uhlenbeck, GARCH, Heston) that we use and our novel microstructure model. Together the value-volatility and microstructure components form our price evolution model. In Section 3, we introduce a model with dynamic microstructure noise and estimate the parameters in our stochastic volatility (SV) models using Bayes estimation. Furthermore, we present an evolution equation that characterize the Bayes filter and a novel, efficient particle filtering algorithm to implement these equations. In Section 4, 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 value-volatility model and what amount of inertia best fits the observed price data.

2. The Partially-Observed Market Model

Let $[0,T]$ be a fixed time period and $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{0\le t\le T}, \mathbb{P})$ be a complete filtered probability space. For any stochastic process S, its natural filtration, defined as $\mathcal{F}_t^S \triangleq \sigma\{S_u : 0 \le u \le t\}$, represents the information in S 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.

2.1. **Construction of Macrostructure State.** We introduce the macrostructure model $M = (X, \theta)$ for unobservable intrinsic (or fair) asset value together with its volatility and parameters. $X \in \mathbb{R}^{n_x}$ is the macrostructure financial state (value 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 $\mathbf{A}$ 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 2.1. *$(X,\ \theta)$ is the unique solution of the $\mathbb{R}^{n_x+n_\theta}-$valued martingale problem for $\mathbf{A}$ with initial distribution μ. That is,*

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

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

Remark 2.1. *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 $\mathbf{A}$ does not act on the variable θ since $\frac{d\theta_t}{dt} = 0$ for our fixed parameters.*

The martingale problem formulation (1) (see Stroock and Varadhan [50], Ethier and Kurtz [27] for more details) is general enough to cover most interesting financial models. In our paper, the macrostructure state X consists of two components: the intrinsic value 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 1. (GBM model) *(see Black and Scholes* [8]*, Merton* [43]*)*

$$\frac{dS_t}{S_t} = \mu dt + \sigma dW_t, \tag{2}$$

with parameters (μ, σ), corresponds to our martingale problem with the generator

$$\mathbf{A}^{(1)} f = \frac{1}{2}\sigma^2 s^2 \frac{d^2 f}{ds^2} + \mu s \frac{df}{ds}. \tag{3}$$

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 [37] for a detailed survey), the GBM model is generalized to stochastic volatility (SV) models, where σ itself is replaced by a stochastic process $\{V_t^{\frac{1}{2}}, t \geq 0\}$. Some of the popular SV models include:

Example 2. (Hull-White model) *(see Hull and White* [36]*)*

$$\frac{dS_t}{S_t} = \mu dt + V_t^{\frac{1}{2}} dW_t, \quad \frac{dV_t}{V_t} = \nu dt + \kappa dB_t, \tag{4}$$

with parameters (μ, ν, κ) and generator

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

Example 3. (Logrithmetic Ornstein-Uhlenbeck model) *(see Scott* [48]*)*

$$\frac{dS_t}{S_t} = \mu dt + V_t^{\frac{1}{2}} dW_t, \quad \frac{dV_t^{\frac{1}{2}}}{V_t^{\frac{1}{2}}} = (\frac{1}{2}\nu^2 - \varrho(\ln V_t^{\frac{1}{2}} - \varpi))dt + \kappa dB_t, \tag{6}$$

with parameters $(\mu, \nu, \varrho, \varpi, \kappa)$ *and generator*

$$\mathbf{A}^{(3)} f = \frac{1}{2} v^2 s^2 \frac{\partial^2 f}{\partial s^2} + \mu s \frac{\partial f}{\partial s} + \frac{1}{2} \kappa^2 v^2 \frac{\partial^2 f}{\partial v^2} + v \left(\frac{1}{2} \nu^2 - \varrho (\ln v - \varpi) \right) \frac{\partial f}{\partial v}. \tag{7}$$

Example 4. (GARCH model) *(see Nelson* [45]*)*

$$\frac{dS_t}{S_t} = \mu dt + V_t^{\frac{1}{2}} dW_t, \quad dV_t = (\nu - \varrho V_t) dt + \kappa V_t dB_t, \tag{8}$$

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

$$\mathbf{A}^{(4)} f = \frac{1}{2} v s^2 \frac{\partial^2 f}{\partial s^2} + \mu s \frac{\partial f}{\partial s} + \frac{1}{2} \kappa^2 v^2 \frac{\partial^2 f}{\partial v^2} + (\nu - \varrho v) \frac{\partial f}{\partial v}. \tag{9}$$

Example 5. (Heston model)*(see Heston* [34]*)*

$$\frac{dS_t}{S_t} = \mu dt + V_t^{\frac{1}{2}} dW_t, \quad dV_t = (\nu - \varrho V_t) dt + \kappa V_t^{\frac{1}{2}} dB_t, \tag{10}$$

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

$$\mathbf{A}^{(5)} f = \frac{1}{2} v s^2 \frac{\partial^2 f}{\partial s^2} + \mu s \frac{\partial f}{\partial s} + \frac{1}{2} \kappa^2 v \frac{\partial^2 f}{\partial v^2} + (\nu - \varrho v) \frac{\partial f}{\partial v}. \tag{11}$$

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 if stochastic volatility is present. In summary, we have

Name	Model	Macro-State	Macro-Parameter	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)	$(\mu, \nu, \varrho, \varpi, \kappa)$	$\mathbf{A}^{(3)}$
GARCH	$M^{(4)}$	(S, V)	$(\mu, \nu, \varrho, \kappa)$	$\mathbf{A}^{(4)}$
Heston	$M^{(5)}$	(S, V)	$(\mu, \nu, \varrho, \kappa)$	$\mathbf{A}^{(5)}$

.

Remark 2.2. *The GARCH model is the continuous-time limit of many classical GARCH-type discrete-time processes (Nelson* [45]*, Drost and Werker* [18]*). We did not consider jumping stochastic volatility models (e.g. Elliot, Malcolm and Tsoi* [22]*, Kouritzin and Zeng* [41] *Duffie, Pan* *and Singleton* [20]*, Eraker, Johannes and Polson* [25]*, Eraker* [26]*) 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 fundmental 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 make the computations possible on an inexpensive contemporary desktop computer.*

2.2. **Construction of Microstructure Price.** The value-volatility models account for the random variances of the intrinsic asset value thus the selection of proper SV model is crucial for the derivative pricing and hedging. On the other hand, microstructure noise (Black [7], Hansen and Lunde [31], Duan and Fulop [19], Grothe and Müller [30], etc.), causes random perturbations of transaction price from its intrinsic value and the disregard of such trading noise introduces severe bias into stochastic volatility estimation (Duan and Fulop [19]). We incorporate microstructure trading noise into traditional value-volatility models and use statistical filtering to reveal such things as short-term inertia and stochastic volatility.

In microstructure markets, the price changes occur only at irregularly spaced transaction times $t_1, t_2, \cdots$ with total trading intensity $a(t)$ (see Engle [24]). 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

$$Y_{t_i} = F(X_{t_i}, t_i), \tag{12}$$

where F is some nonlinear random field depending on the trading noise to be specified. The formulation (12) is similar to that of Hasbrouck [32] in which X is the intrinsic and permanent component while F introduces the transitory component.

The empirical evidence reported by Hansen and Lunde [31] suggests strongly that the trading noise is serially correlated. Similar results can be found in Aït-Sahalia, Mykland and Zhang [1], Barndorff-Nielsen, Hansen, Lunde and Shephard [6]. Indeed, there exist situations in which the trading noise variance estimate is zero when the trading noise is simply assumed to be independent (Duan and Fulop [19]). This does not mean there is no trading noise but rather that the trading noise is autocorrelated. To characterize this correlation, Hansen and Lunde [31] assume the noise to be some Gaussian random sequence with stationary covariance and finite dependence. However, this model is most suitable for the low-frequency setup and ignores many crucial microstructure effects. We build correlation into our microstucture information noise through inertia and mean-reversion while utilizing microstructure rounding and clustering noise to explain the discreteness and whole price biasing.

2.2.1. *Inertia and Information Noise.* The idea of momentum or inertia has been used in many studies (see [38], [44], [29], [28], etc.). Basically, it is the tendency for a stock to continue to move in one direction. To make our presentation manifest, we formalized this idea:

Definition 2.2. *A process (Z_t) is said to have stochastic inertia if*

$$I_t^Z \triangleq \lim_{u \searrow t} \frac{d}{dt}\frac{d}{du}\mathbb{E}[Z_u Z_t] \in (0, \infty] \tag{13}$$

for all t. I_t^Z is called the inertia function.

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

$$\begin{aligned}
(14) \qquad & \lim_{u \searrow t} \lim_{k \to 0} \lim_{h \to 0} \frac{E[(Z_{u+h} - Z_u)(Z_t - Z_{t-k})]}{kh} > 0 \\
\leftrightarrow \quad & \lim_{u \searrow t} \lim_{k \to 0} \left[\lim_{h \to 0} \frac{E[(Z_{u+h} - Z_u) Z_t]}{hk} - \lim_{h \to 0} \frac{E[(Z_{u+h} - Z_u) Z_{t-k}]}{hk} \right] > 0 \\
\leftrightarrow \quad & \lim_{u \searrow t} \frac{\partial}{\partial t} \frac{\partial}{\partial u} E[Z_u Z_t] > 0.
\end{aligned}$$

Many processes have inertia; however, we prefer ones with the following five properties: (1) $Var(Z_t)$ should be proportional to t so it grows in a similar way as 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 make senses from informational and hidden liquidity points of view. More precisely, it can well explain the price effects of diffusion and assimulation of information and rumor as well as the purchases or sales of a large agent changing his/her position over time; (4) Z is easy to simulate; (5) Z is easy to analyze.

Brownian motion B does not have inertia since $I_t^B \equiv 0$. For fractional Brownian motion (FBM) B^h,

$$\mathbb{E}[B_t^h B_u^h] = \frac{1}{2}(t^{2h} + u^{2h} - |u - t|^{2h}),$$

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

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

Thus, the inertia function of B^h is infinity for all t if $h > \frac{1}{2}$ (and is $-\infty$ if $h < \frac{1}{2}$). Neither case satisfies our five properties. Our stochastic inertia process is

$$(15) \qquad \xi_t^h = \sqrt{h} \int_0^t \tanh((t - s)/\Delta) dB_s^\xi + \sqrt{1 - h} W_t^\xi$$

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

Remark 2.3. *The inertia in ξ_t^h is introduced through a weighted average (by the choice of h) of the historical information (the first term) and fundamental information (the second term). In fact, the* $\tanh(t/\Delta)$ *can be viewed as the impulse response on price created by market participants receiving and assimulating the "information" dB_t^ξ so Δ determines the diffusion speed in the market. This formulation captures the idea that news or rumor and its ramifications can require time to be fully disseminated and understood. $h = 1$ then represents the case of only historical information resulting in the strongest short-term inertia in prices. Further, we can use this explanation 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.*

ξ_t^h is a centered Gaussian process such that for any $u \geq t$

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

In particular,

$$\frac{Var(\xi_t^h)}{t} = \frac{h}{t} \int_0^t \tanh^2((t-s)/\Delta) ds + (1-h) = 1 - h\Delta \frac{\tanh(t/\Delta)}{t}$$

thus $\frac{Var(\xi_t^h)}{t}$ converges to 1 with speed determined by Δ. Moreover,

$$\frac{d}{dt}\frac{d}{du}\mathbb{E}[\xi_t^h \xi_u^h] = \frac{h}{\Delta^2} \int_0^t \operatorname{sech}^2((t-s)/\Delta) \operatorname{sech}^2((u-s)/\Delta) ds$$

and using standard antiderivatives

$$\lim_{u \searrow t} \frac{d}{dt}\frac{d}{du}\mathbb{E}[\xi_t^h \xi_u^h] = \frac{h}{\Delta^2} \int_0^t \operatorname{sech}^4(s/\Delta) ds = \frac{h}{3\Delta}[\cosh(2t/\Delta) + 2] \tanh(t/\Delta) \operatorname{sech}^2(t/\Delta).$$

Note that

$$\lim_{t\to\infty} \frac{h}{3\Delta}[\cosh(2t/\Delta)+2]\tanh(t/\Delta)\mathrm{sech}^2(t/\Delta) = \frac{2h}{3\Delta}$$

and this happens quickly for small Δ. We can thus verify ξ^h, defined in (15), 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 $\Delta = \frac{1}{2}$ day on real stock data.

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 intermediate price $\mathcal{Y}_{t_i}$ is defined by

$$\ln \mathcal{Y}_{t_i} = \begin{cases} \ln S_{t_i} + Z^{h,\Delta}_{t_i} + \epsilon\zeta_i, & \text{dynamical microstructure} \\ \ln S_{t_i} + \epsilon\zeta_i, & \text{non-dynamical} \end{cases} \tag{16}$$

$$dZ^{h,\Delta}_t = -\alpha_Z Z^h_t dt + d\xi^h_t, \quad Z^{h,\Delta}_0 = z_0, \tag{17}$$

where $X = (S, V)$ and $Z^{h,\Delta}$ is the dynamical part of the microstructure. The case $Z^{h,\Delta} \equiv 0$ is of particular importance in the sequel as it represents the *non-dynamical* 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 a Ornstein-Uhlenbeck (O-U) like inertia velocity process with mean-reverting parameter $\alpha_Z > 0$. 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 [52] 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 [38]); (3) The mean-reverting structure of Z^h when combined with the inertia we incorporated captures the cyclic property of prices (e.g. Black [7]). Z^h is not a Markov process so we introduce its historical process as

$$\widehat{Z}^h_t(\tau) \triangleq Z^h_{t\wedge\tau}, \tag{18}$$

which is Markov. Moreover, $\widehat{Z}^h_t \in C[0,T]$, the space of all continuous functions on $[0,T]$, since the paths of Z^h are continuous. Consequently, we generalize the state vector to be $(X, \theta, \vartheta, \widehat{Z}^h)$, where $\vartheta = (\epsilon, \alpha_Z)$ is the microstructure noise parameter set. The advantage of this formulation is that we can estimate $\widehat{Z}^h$ thus Z^h jointly with other components using particle filtering methods. The generalized state incorporates value, volatility, parameters and the historical trading noise $\widehat{Z}^h$ while keeping the tractability of a Markovian framework.

Remark 2.4. *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.*

2.2.2. *Rounding and Clustering Noise.* While $\mathcal{Y}_{t_i}$ can take any value, the trading price Y_{t_i} is restricted to multiples of the *tick*, $\{y_0 = 0, y_1 = \frac{1}{M}, \cdots y_j = \frac{j}{M}, \cdots\}$, for some positive integer M. The tick size in New York Stock Exchange (NYSE) was switched to $\$\frac{1}{16}$ from $\$\frac{1}{8}$ in June 24, 1997 and then further adjusted to \$0.01 beginning from January 29, 2001. The empirical studies suggest that the tick size $\frac{1}{M}$ plays an important role in microstructure market analysis (e.g. Huang and Stoll [35]). As we are concerned with price clustering for post-decimal pricing in stock markets, we let $M = 100$.

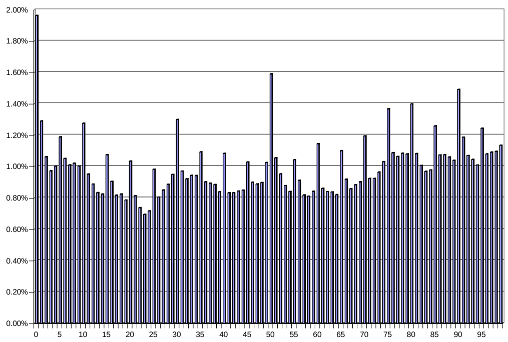

Figure 1: Clustering for 3 NYSE stocks in April 2010.

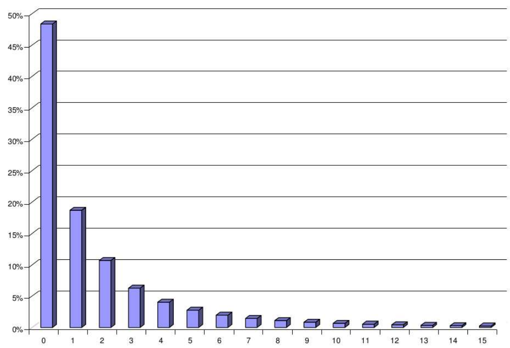

Figure 2: Inter-trade Time Frequencies in seconds.

It is well documented that there exists price clustering to more whole prices. To quantify this price clustering, we examine the tick price behavior for three NYSE listed stocks over April 2010:

NYSE Stock	Ticker Symbol
Morgan Stanley	MS
International Business Machines Corporation	IBM
PepsiCo Inc.	PEP

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. Suppose the intermediate price $\mathcal{Y}_{t_i}$ falls in the interval $[y_j - \frac{1}{2M}, y_j + \frac{1}{2M})$, then if there was no clustering noise, the trading price Y_{t_i} would just be y_j. Thus, the probability of trading at y_j with no clustering noise given $X_{t_i} = x$, $Z_{t_i} = z$ would be

$$(19) \qquad R(y_j|x,z,\vartheta) \triangleq P(\mathcal{Y}_{t_i} = y_j | X_{t_i} = x, Z_{t_i} = z, \vartheta) = \begin{cases} \int_{\ln\left(\frac{y_j - \frac{1}{2M}}{x \cdot e^z}\right)}^{\ln\left(\frac{y_j + \frac{1}{2M}}{x \cdot e^z}\right)} \frac{1}{\sqrt{2\pi\epsilon}} e^{-\frac{u^2}{2\epsilon^2}} du & \text{dynamical microstructure} \\ \int_{\ln\left(\frac{y_j - \frac{1}{2M}}{x}\right)}^{\ln\left(\frac{y_j + \frac{1}{2M}}{x}\right)} \frac{1}{\sqrt{2\pi\epsilon}} e^{-\frac{u^2}{2\epsilon^2}} du & \text{non-dynamical} \end{cases}.$$

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

$$R(y_j|X_{t_i},\Pi_{t_i}\widehat{Z}^h_{t_i},\vartheta)=\int_{\ln\left(\frac{y_j-\frac{1}{2M}}{X_{t_i}e^{\Pi_{t_i}\widehat{Z}^h_{t_i}}}\right)}^{\ln\left(\frac{y_j+\frac{1}{2M}}{X_{t_i}e^{\Pi_{t_i}\widehat{Z}^h_{t_i}}}\right)}\frac{1}{\sqrt{2\pi}\epsilon}e^{-\frac{u^2}{2\epsilon^2}}du, \tag{20}$$

where Π_{t_i} is the projection onto time t_i, i.e.,

$$\Pi_{t_i}\widehat{Z}^h_{t_i}=\widehat{Z}^h_{t_i}(t_i)=Z^h_{t_i\wedge t_i}=Z^h_{t_i}. \tag{21}$$

Clearly, $R(y_j|x,z,\vartheta)$ is a smooth function of (x,z,ϑ) for each fixed y_j. Since $M=100$, it is convenient to introduce the following notation:

$$D_1=\{\text{The integers in } (0,100] \text{ that are not multiples of } 5\},$$

$$D_2=\{\text{The integers in } (0,100] \text{ that are multiples of 5 but not of } 25\},$$

$$D_3=\{25,75\},\quad D_4=\{50\},\quad D_5=\{100\}.$$

If the fractional part of the price y is in D_1, then it will stay in the same level 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

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

$$p(y_j|x,z,\vartheta)=R(y_j|x,z,\vartheta)(1-\alpha). \tag{22}$$

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

$$p(y_j|x,z,\vartheta) = R^*(y_j|x,z,\vartheta)(1-\beta), \tag{23}$$

where

$$R^*(y_j|x,z,\vartheta) \triangleq R(y_j|x,z,\vartheta) + \alpha\left(R(y_{j-1}|x,z,\vartheta) + R(y_{j-2}|x,z,\vartheta)\right) + \alpha\left(R(y_{j+1}|x,z,\vartheta) + R(y_{j+2}|x,z,\vartheta)\right) \tag{24}$$

Case 3. *If the fractional part of y_j belong to D_3,*

$$p(y_j|x,z,\vartheta) = R^{**}(y_j|x,z,\vartheta)(1-\gamma_1-\gamma_2), \tag{25}$$

where

$$R^{**}(y_j|x,z,\vartheta) \triangleq R^*(y_j|x,z,\vartheta) + \beta(R^*(y_{j-5}|x,z,\vartheta) + R^*(y_{j-10}|x,z,\vartheta)) + \beta(R^*(y_{j+5}|x,z,\vartheta) + R^*(y_{j+10}|x,z,\vartheta)). \tag{26}$$

Case 4. *If the fractional part of y_j belong to D_4,*

$$p(y_j|x,z,\vartheta) = R^{**}(y_j|x,z,\vartheta) + \gamma_1(R^{**}(y_{j-25}|x,z,\vartheta) + R^{**}(y_{j-25}|x,z,\vartheta)). \tag{27}$$

Case 5. *If the fractional part of y_j belong to D_5,*

$$p(y_j|x,z,\vartheta) = R^{**}(y_j|x,z,\vartheta) + \gamma_2(R^{**}(y_{j-25}|x,z,\vartheta) + R^{**}(y_{j+25}|x,z,\vartheta)). \tag{28}$$

Moreover, we have to handle the case $j=0$ separately since there are no negative prices

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

$$p(y_0|x,z,\vartheta) = R(y_0|x,z,\vartheta) + \alpha(R(y_1|x,z,\vartheta) + R(y_2|x,z,\vartheta)) + \beta(R^*(y_5|x,z,\vartheta) + R^*(y_{10}|x,z,\vartheta)) + \gamma_2 R^{**}(y_{25}|x,z,\vartheta). \tag{29}$$

Remark 2.5. *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

Clustering Parameters	Estimate
α	0.060475
β	0.046883
γ_1	0.03883
γ_2	0.16525

.

The large degree of clustering exhibited, especially to the whole dollar, might be considered surprising. However, earlier studies of Huang, Stoll [35], Chung, Van Ness [12], Chung, Kim and Kitsabunnarat [11] also showed significant clustering. Moreover, the degree of price clustering in NYSE is weaker than that of NASDAQ. For example, Barclay [5] 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-eigthes is 78% while after, it drops to about 56%.

2.3. **Nonlinear Filtering Model.** Our price process can be formulated as a marked point process $\overrightarrow{Y}$: a sequence of random vectors $\overrightarrow{Y} = (t_i, Y_{t_i},\ i \geq 1)$, 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 $\overrightarrow{Y}$ is $(E, \mathcal{E})$ where $E = \mathbb{N}_0$ and $\mathcal{E}$ is all its subsets. Here, $j \in E$ corresponds to the $j^{th}-$tick level $\frac{j}{M}$. For each $A \in \mathcal{E}$, we associate the counting process $Y_t(A)$

$$Y_t(A) \triangleq \sum_{i\geq 1} 1_{\{Y_{t_i}\in A\}} 1_{\{t_i \leq t\}} \tag{30}$$

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

$$Y_j(t) \triangleq Y_t(\{j\}) = \sum_{i\geq 1} 1_{\{Y_{t_i}=j\}} 1_{\{t_i \leq t\}}$$

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

$$Y(\omega, A \times (s,t]) \triangleq Y_t(\omega, A) - Y_s(\omega, A), \qquad \forall \omega \in \Omega, \quad s \le t \in [0,T], \quad A \in \mathcal{E}. \tag{31}$$

The natural filtration, i.e. information content, of Y is

$$\mathcal{F}_t^Y \triangleq \sigma(Y_s(A), \quad 0 \le s \le t, \quad A \in \mathcal{E}). \tag{32}$$

Now we assume:

A 1. *The total trade process $Y_t = Y_t(E)$ admits an intensity $a(t)$ for some measurable function a.*

Therefore, $Y_j(t)$ has intensity

$$\lambda_j(X_t, Z_t^h, \vartheta, t) = a(t) \cdot p(y_j | X_t, Z_t^h, \vartheta). \tag{33}$$

To simplify the notation, we rewrite (33) as $\lambda_j = a \cdot p_j$.

A 2. *There exists some positive constants δ, C such that $\delta \le a(t) \le C$ for all t.*

Based on representation (30), (33), $(X, \widehat{Z}^h, \theta, \vartheta; Y)$ is framed by a partial-observation model, where $(X, \widehat{Z}^h, \theta, \vartheta)$ 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 as explained in Aït-Sahalia and Kimmel [2]. 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.

3. Calibration through Nonlinear Filtering

3.1. **Nonlinear Filtering and Particle Filter.** The available information about $(X_t, \theta, \vartheta, \widehat{Z}_t^h)$ is the observation filtration $\mathcal{F}_t^Y \subset \mathcal{F}_t$, defined in (32), and the primary goal of nonlinear filtering is to characterize the conditional distribution

$$\pi_t(\cdot) = \mathbb{P}[(X_t, \theta, \vartheta, \widehat{Z}_t^h) \in \cdot\, |\mathcal{F}_t^Y] \tag{1}$$

or equivalently,

$$\pi_t(f) = \mathbb{E}[f(X_t, \theta, \vartheta, \widehat{Z}_t^h)|\mathcal{F}_t^Y] \tag{2}$$

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

Remark 3.1. *Actually, we only want to estimate $\mathbb{P}[(X_t, \theta) \in \cdot\, |\mathcal{F}_t^Y]$ but there is no simple recursive formula for this marginal. The filter is naturally model dependent so we produce different filtering processes for each model, that is, for each SV choice* 1-5*, each value of Δ and each value of h.*

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

$$L_t = \exp\left(\int_0^t \int_E \ln \frac{\lambda_z(X_s, Z_s^h, \vartheta, s)}{\kappa_z} Y(dz, ds) - \int_0^t (a(s) - \kappa) ds\right). \tag{3}$$

L_t is a martingale under Condition **A** 2 and $\mathbb{Q}$, defined by

$$\frac{d\mathbb{Q}}{d\mathbb{P}}\Big|_{\mathcal{F}_T} = L_T^{-1} \quad \text{(i.e. } Q(A) = \int_A L_T^{-1} dP \quad \text{for } A \in \mathcal{F}_T), \tag{4}$$

is called the reference measure. Bayes Theorem (see Bremaud [9, p. 165]) links the desired (real-world) conditional distribution π_t with the unnormalized filter σ_t by

$$\pi_t(f) = \frac{\sigma_t(f)}{\sigma_t(1)}, \tag{5}$$

where the unnormalized filter σ_t is defined by

$$\sigma_t(f) \triangleq \mathbb{E}^{\mathbb{Q}}[f(X_t, \theta, \vartheta, \widehat{Z}_t^h) L_t | \mathcal{F}_t^Y] \tag{6}$$

for all $f \in B(\mathbb{R}^{n_x+n_\theta+2} \otimes C[0,T])$. Under $\mathbb{Q}$, the state vector $(X, \theta, \vartheta, \widehat{Z}^h)$ is independent of the observation Y and we have:

Theorem 3.1. *Under* A 1 *and* A 2*, the unnormalized filter σ_t is the unique measure-valued solution of the stochastic filtering equation*

$$\sigma_t(f) = \sigma_0(f) + \int_0^t \sigma_s\left((\bar{\mathbf{A}} - a(s) + \kappa)f\right) ds + \int_0^t \int_E \sigma_{s-}\left(\left(\frac{\lambda_z(s-)}{\kappa_z} - 1\right) f\right) Y(dz, ds), \tag{7}$$

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

This theorem is a modest generalization of prior results and can be obtained in much the same manner as results in Kouritzin and Zeng [41] and Xiong and Zeng [51]. Here, $\bar{\mathbf{A}}$ 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 $\bar{\mathbf{A}}$. Instead, to implement the evolution equation (7), we apply the following novel particle filter that can be thought of as a generalization of Del Moral, Noyer and Salut [16]. For some large $N \in \mathbb{N}_0$ the particle system $\{P_t^k\}_{k=1}^N$ is constructed as follows:

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

$$\varphi_N(0) = \frac{1}{N} \sum_{k=1}^N \delta_{P_0^k}(\cdot),$$

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

$$\lim_{N \longrightarrow \infty} (\varphi_N(0), f) = \pi_0(f) \quad \forall f \in B(\mathbb{R}^{n_x+n_\theta+2} \otimes C[0,T]).$$

Here, $(\mu, f) \triangleq \int f(y)\mu(dy)$ for measures μ so

$$(\varphi_N(0), f) = \frac{1}{N}\sum_{k=1}^{N} f(P_0^k).$$

Remark 3.2. *Note that* $L_0 = 1$ *so* $\pi_0(f) = \sigma_0(f)$. *When there is no special information, it is convenient to assign uniform distributions to* $(X_0, \theta, \vartheta, \widehat{Z}_0^h)$. *Note that* $\widehat{Z}_0^h$ *is a constant function defined on* $[0, T]$.

3.1.2. *Evolution.* Between observations all particles move independently as samples from the transition probability of $(X, \theta, \vartheta, \widehat{Z}^h)$. In particular, we use the Euler scheme (Kloeden and Platen [40]) to evolve the dynamics Examples 2.2-2.6 and (17).

3.1.3. *Particle Weights.* At the i^{th} observation (t_i, Y_{t_i}), each particle is given a weight observation

$$\omega_i^k = \omega_i^k(t_i) \quad \triangleq \quad \exp\left(\int_{t_{i-1}}^{t_i}\int_E \ln\frac{\lambda_z(P_s^k, s)}{\kappa_z} Y(dz, ds) - \int_{t_{i-1}}^{t_i}(a(s) - \kappa)ds\right).$$

ω_i^k depends on the observation Y and the likelihood ratio of measure $\mathbb{P}$ over measure $\mathbb{Q}$ defined by (3) given the simulated particle path realized on the interval $[t_{i-1}, t_i)$. These weights are stored along with the states of particles before re-sampling.

3.1.4. *Re-sampling.* The average particle weight at t_i is

$$\omega_i \triangleq \frac{1}{N}\sum_{k=1}^{N}\omega_i^k. \tag{8}$$

If a particle has a weight $\omega_i^k = r^k\omega_i + z^k$, where $r^k \in \{0, 1, 2, \cdots\}$ and $z^k \in [0, \omega_i)$ before the re-sampling, then there will be r^k or r^k+1 particles at this state after the re-sampling with a probability selected in order to leave the system unbiased, meaning there will $r^k + 1$ particles with probability $\frac{z^k}{\omega_i}$. In particular, the extra particles can be placed according to a $\left(N - \sum_{k=1}^{N} r^k; \frac{z^1}{\omega_i}, \frac{z^2}{\omega_i}, ..., \frac{z^N}{\omega_i}\right)$-multinomial distribution. It is this average-weight usage in our resampling that differentiates this procedure from the earlier, popular one of Del Moral, Noyer and Salut [16]. This simple change leads to dramatic outperformance (see Del Moral, Kouritzin and Miclo [15] for an illustration of how resampling can change performance).

3.1.5. *Bayesian Estimation.* By the strong law of large numbers and (5), the particle approximation of the normalized filter $\pi(\cdot)$ is

$$\pi_{N,\,t}(f) = \frac{\frac{1}{N}\sum_{k=1}^{N} f(P_t^k)\exp(\int_E \ln(\frac{\lambda_z}{\kappa_z})(P_t^k, t))Y(dz,t))}{\frac{1}{N}\sum_{k=1}^{N} \exp(\int_E \ln(\frac{\lambda_z}{\kappa_z})(P_t^k, t))Y(dz,t))}$$

for all $f \in B(\mathbb{R}^{n_x+n_\theta+2} \times C[0,T])$.

3.2. **Calibration and Historical Training.** To keep the problem size manageable, we just used the clustering parameter estimates given above as the actual values throughout our simulations.

We also took the total intensity function $a(t)$ to be the hazard function

$$a(t) = -\frac{d}{dt}\ln \mathbb{P}(\mathcal{T} > t).$$

Here, $\mathcal{T}$ represents the inter-trade duration of the tick data. Figure 2 is the inter-trade duration histogram of our 3 NYSE-listed stocks from which the hazard rate was estimated.

One is always faced with the problem of estimating initial distributions for value, volatility and the parameters prior to filtering over the time interval of interest (April 2010 here). Our approach was to make arbitrary uniform 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 reasonable distributions at a much later point, April 1, 2010. (See e.g. Ocone and Pardoux [47], Delyon and Zeitouni [17], Atar [3] 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

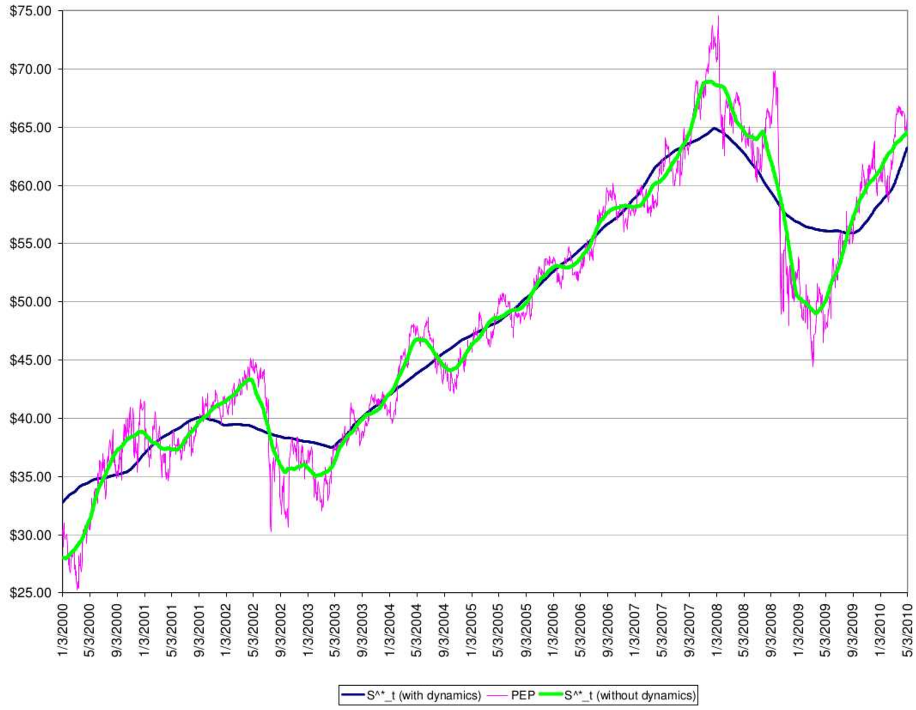

Figure 3: Longterm Value Estimation of PEP.

parameters. Our main purpose in this historical training was to get a starting joint distribution for $(X, \theta, \vartheta, \hat{Z}^h)$ 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 non-dynamical microstructure Heston case and the *median inertia* dynamical case where $h = \frac{1}{2}$ and $\Delta = 7200$ 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, e.g. 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 Pepsi. The choppiest curve is the actual stock price while the smoothest curve is the filter's value estimate $E[S_t|\mathcal{F}_t^Y]$ using the Heston SV model with (median) microstructure inertia. The middle curve is the filter's value estimate $E[S_t|\mathcal{F}_t^Y]$ using the

Heston SV model without dynamics in the microstructure, i.e. $Z^h = 0$. 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 value process estimate is far less volatile in the pressence of dynamical microstructure than without. This indicates 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?

3.3. **Numerical Results.** The data is one month (April, 2010) of transaction prices of our three NYSE-listed stocks. Our filter produces Bayes estimates (in seconds) to the macro- and mirco-parameter vectors θ and ϑ respectively. These estimates in the non-dynamical microstructure case for PepsiCo are as follows:

PEP	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	$-$	$1.59E-03$	$1.80E-03$	$1.94E-03$	$2.58E-07$
ϱ	$-$	$-$	$4.75E-03$	$6.51E-03$	$6.02E-03$
ϖ	$-$	$-$	$4.84E-06$	$-$	$-$

All parameters are estimated using time in seconds. Our Pepsico Bayes estimates in the *median* inertia case are as follows:

PEP	GBM	HW	LOU	Nelson	Heston
μ	$1.05E-06$	$1.02E-06$	$9.92E-07$	$1.03E-06$	$1.01E-06$
ν	$\sigma = 2.21E-06$	$5.50E-10$	$5.13E-06$	$6.32E-11$	$5.94E-12$
κ	$-$	$2.18E-03$	$1.87E-03$	$2.12E-03$	$2.26E-07$
ϱ	$-$	$-$	$2.25E-03$	$2.90E-03$	$3.23E-03$
ϖ	$-$	$-$	$2.60E-06$	$-$	$-$
ϵ	$2.43E-09$	$2.05E-09$	$2.33E-09$	$2.31E-09$	$2.46E-09$
α_Z	$2.13E-09$	$2.31E-09$	$2.23E-09$	$2.33E-09$	$2.31E-09$

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

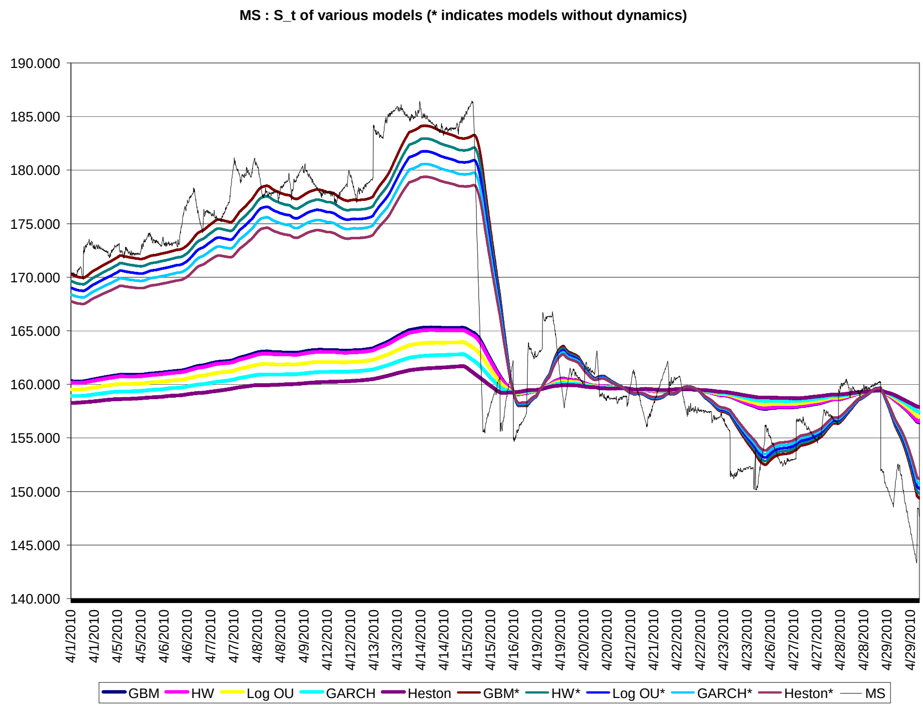

Figure 4: Value Estimation of MS, April 2010.

Figures 4 − 5 show the conditional expectation value estimation for Morgan Stanley 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 on each of three graphs. The most volatile curve is the stock price itself over this month. The smoothest curves somewhat separated from the stock price are the value estimates using the five SV models with (median inertia) dynamical microstructure. The remaining five curves (that hug the stock price) are our value estimates for our five SV models with non-dynamical microstructure. In this last case, the microstructure does not have the power to separate the value and actual stock price to any large degree.

It is important to realize that these pictures are really just a one month snapshot of a much bigger multi-year filtering process. This explains why many of the value processes are significantly

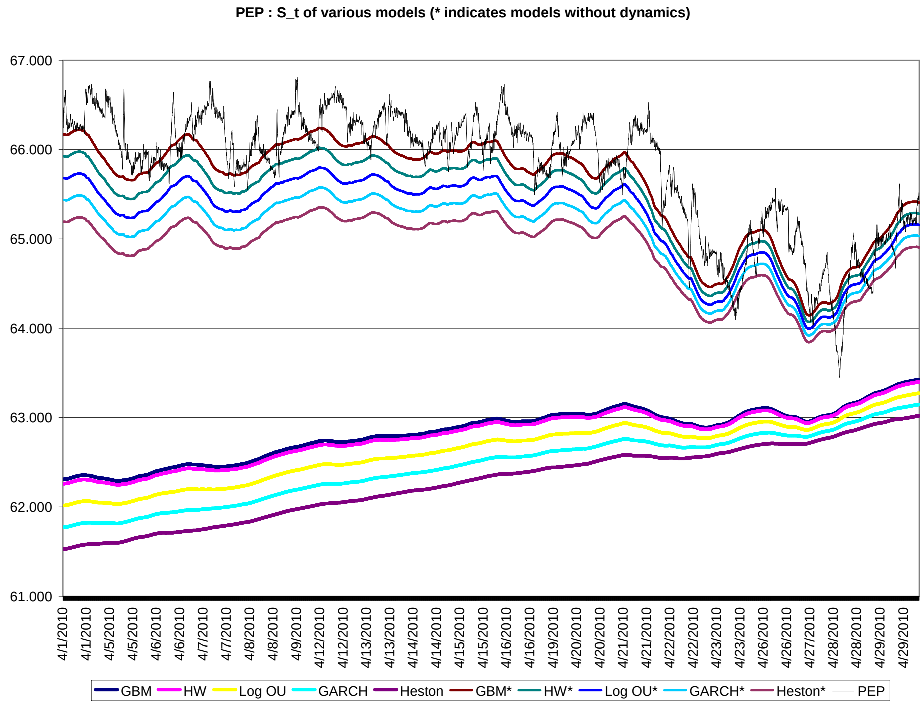

Figure 5: Value Estimation of PEP, April 2010.

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 into the microstructure allows the estimated value process really to differ from the stock price. Indeed, there is a significant correction of all three stock prices (especially Morgan Stanley) towards estimated value 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 but otherwise would have had no such warning. It is an interesting to ponder what this possible discrepancy would mean to option prices.

The filters provide conditional distributions and estimates for more than just value and parameters. Table 1 shows the volatility estimates with (median) microstructure inertia and without

	without dynamics	with dynamics
PEP (2 Hrs., $h = 0.6$)	1.58416E-09	1.01312E-11
MS (1/2 Day, $h = 0.4$)	4.14645E-08	4.3005E-10
IBM (1/2 Day, $h = 1$)	8.21731E-10	4.03211E-11

Table 1: Volatility Estimation, April 2010

microstructure dynamics using the 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 suggest that by far the primary use of stochastic volatiliy 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 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 path so we just plot the projection onto the current time, i.e. we just plot Z_t^h even though we must propagate the Markov process $\hat{Z}_t^h$ in the filter. Figures 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

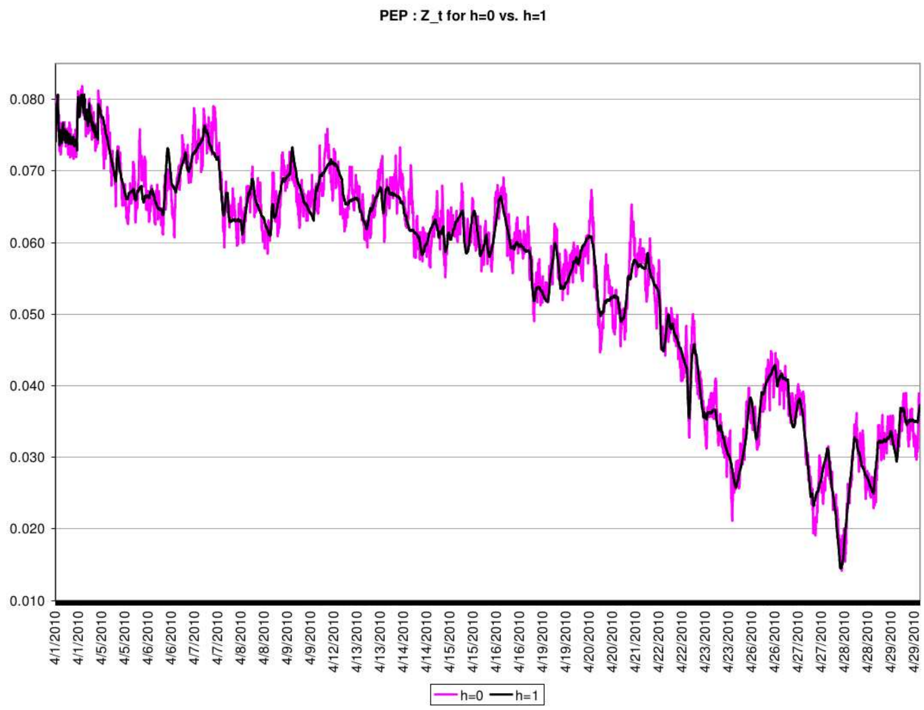

Figure 6: Noise Process Estimation of PEP, April 2010.

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. Selecting the Best Volatility Model by Bayes Factor

4.1. **Model Selection and Bayes Factor.** 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 be able to transform all observation models of interests into the same canonical process via Girsanov measure change. The signal models can be singular to one another. Kouritzin and Zeng [41] discuss the Bayesian model selection problem. However, their equations do not apply to our models. The available information in microstructure market is the observation process Y, which represents the cumulative transaction records throughout all tick price levels. The Bayes factor determines which model best fits this observed data by doing pairwise comparisons. Consider our 5 SV macrostructure value-volatility models

$$M^{(k)} \triangleq (X^{(k)}, \theta^{(k)}) \in \mathbb{R}^{n_x^{(k)}+n_\theta^{(k)}},$$

where the generators of the martingale problem to $M^{(k)}$ are respectively $\mathbf{A}^{(k)}$ for $k = 1, 2, 3, 4, 5$. The likelihood of Y at time t is

$$L_t^{(k,h,\Delta)} = 1 + \int_0^t \int_E \left(\frac{\lambda_z(X_s^{(k)}, Z_s^{h,\Delta}, \vartheta, s)}{\kappa_z} - 1 \right) L_{s-}^{(k,h,\Delta)} \left(Y(dz, ds) - \kappa_z m(dz) ds \right).$$

Here, $m(dz)$ is the counting measure on E and the same observations and observation rate information is used for all models. $(L_t^{(k,h,\Delta)})^{-1}$ then transforms the observations into the same Poisson measure with intensity measure $\mu(A) = \int_A k_z m(dz)$. The normalized filter $\pi_t^{(k,h,\Delta)}$, $k = 1, 2, 3, 4, 5$; $h \in [0, 1]$; $\Delta > 0$ satisfies

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

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

$$\sigma_t^{(k,h,\Delta)}(f_k) \triangleq \mathbb{E}^{\mathbb{Q}}[f_k(X_t^{(k)}, \theta^{(k)}, \vartheta, \widehat{Z}_t^{h,\Delta}) L_t^{(k,h,\Delta)} | \mathcal{F}_t^Y]$$

and $\sigma_t^{(k,h,\Delta)}(1)$ is the integrated (or marginal) likelihood of Y. Now, we use Bayes factor to compare models. To calculate the Bayes factor, we select two complete models characterized by (k_1, h_1, Δ_1) and (k_2, h_2, Δ_2), calculate the integrated likelihoods $\sigma_t^1(1) = \sigma_t^{(k_1,h_1,\Delta_1)}(1)$, $\sigma_t^2(1) = \sigma_t^{(k_2,h_2,\Delta_2)}(1)$

and then take the Bayes' factor ratios:

$$B_{12}(t) = \frac{\sigma_t^1(1)}{\sigma_t^2(1)}, \quad B_{21}(t) = \frac{\sigma_t^2(1)}{\sigma_t^1(1)}. \tag{1}$$

σ_t^1 and σ_t^2 are computed by the unnormalized filtering equation. Kass and Raftery [39] demonstrate how to interpret Bayes factor:

B_{12}	Evidence against Model 2
$1-3$	Barely mentionable
$3-12$	Positive
$12-150$	Strong
>150	Decisive

Now, we consider the problem of selecting the best of our five value-volatility models:

$$M^{(k)} \triangleq (X^{(k)}, \theta^{(k)})$$

and the resulting partially-observed market models:

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

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 3.1 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.

4.2. **Numerical Results.** Using GBM with non-dynamic 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 $2.2-2.6$). In each case, we pick the inertia parameters from the sets $\Delta \in \{30 \text{ mins}, 2 \text{ hrs}, 1/2 \text{ day}\}$ and $h \in \{0, 0.1, 0.2, ..., 0.9, 1\}$ that would yield the highest Bayes factor against the calibration model. The data is the transaction price

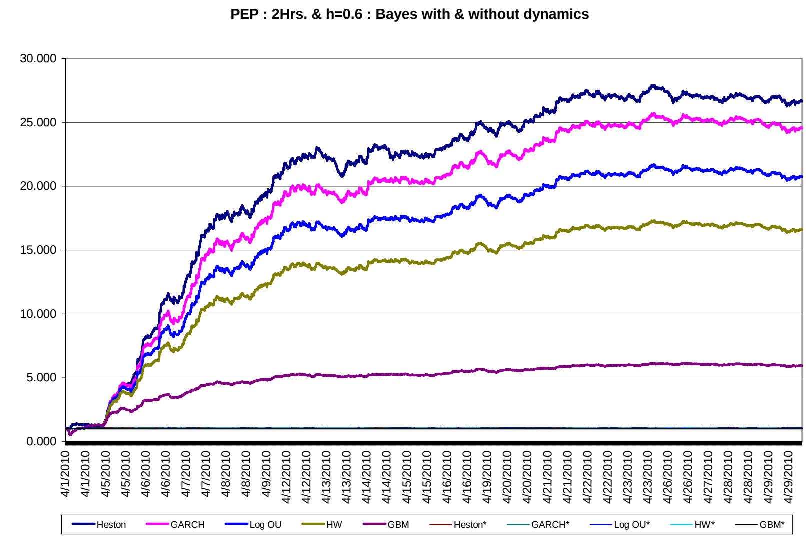

Figure 7: Bayes Factor for PEP SV Model Determination, April 2010

of Pepsi, IBM and Morgan Stanley, April, 2010. Table 2 summarizes the Bayes factor performance. The Bayes factors computed in this table gives strong evidence for the Heston model based on only

	Heston	GARCH	LogOU	HW	GBM
PEP (2 Hrs., $h = 0.6$)	27.30	25.23	21.31	17.04	6.08
MS (1/2 Day, $h = 0.4$)	19.14	18.91	18.77	18.59	10.18
IBM (1/2 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
MS* (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

Table 2: Bayes Factor for Model Determination, April 2010
* indicates without dynamics

one month of real 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

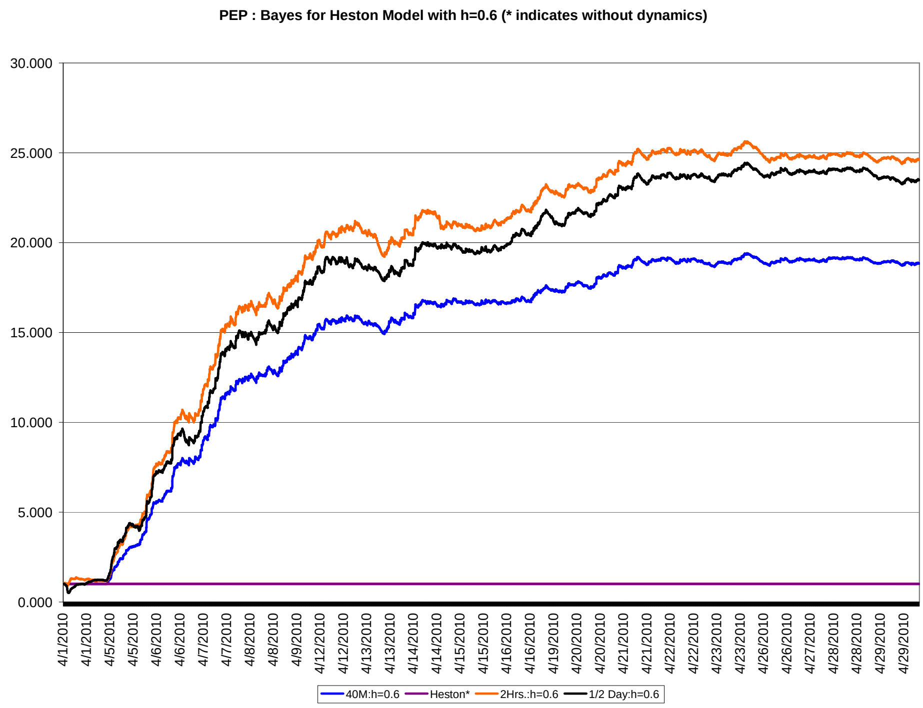

Figure 8: Bayes Factor for PEP Ingestion Time Determination April 2010.

the models did not change over our three stock selections, with Heston always being preferred and GBM always performing the worst.

Next, we look at the ingestion time Δ using non-dynamic microstructure Heston as the calibration model. Figure 8 and Table 3 show the effect of varying Δ over $\{30 \text{ mins}, 2 \text{ hrs}, 1/2 \text{ day}\}$ for the

	Heston*	40 Mins.	2 Hrs.	1/2 Day
MS ($h = 0.4$)	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

Table 3: Bayes Factor for Ingestion Time Determination, April 2010
* indicates without dynamics

$h \in \{0, 0.1, 0.2, ..., 0.9, 1\}$ fixed to give the highest Bayes factor. There is a drop in the Bayes factor

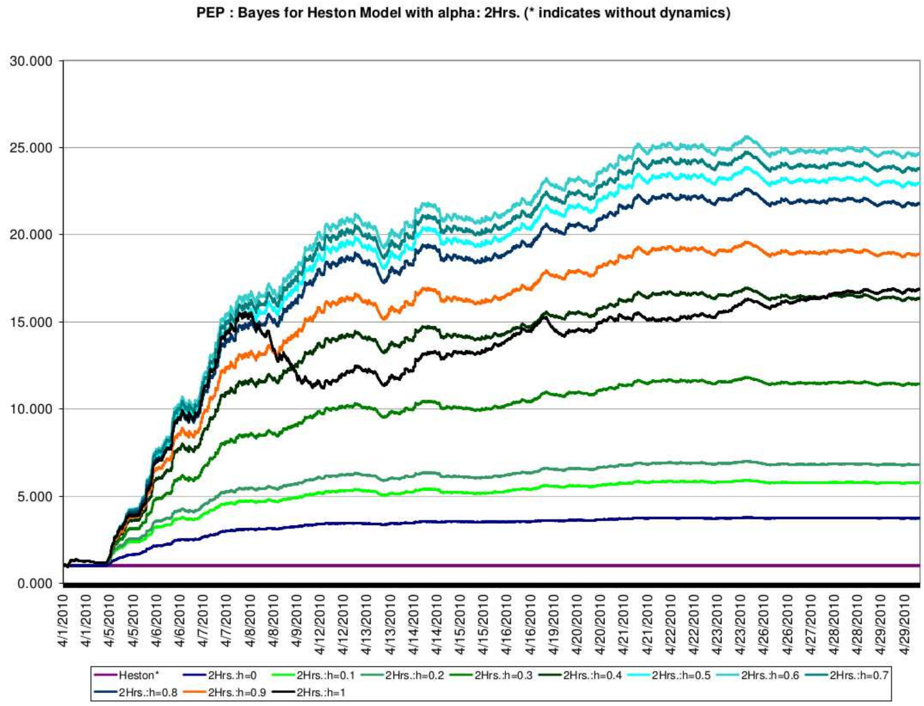

Figure 9: Bayes Factor for PEP Inertia Determination April 2010.

values from the model determination experiment which is entirely due to the change of calibration model from GBM with non-dynamic microstructure to Heston with non-dynamic microstructure. Our results show that the best ingestion times for Morgan Stanley, 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.

Finally, we investigate the optimal amount of inertia. Figure 9 and Table 4 show the effect of varying the amount of inertia h over $\{0, 0.1, 0.2, ..., 0.9, 1\}$ for the $\Delta \in \{30 \text{ mins}, 2 \text{ hrs}, 1/2 \text{ day}\}$ fixed to give the highest Bayes factor. The table shows 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.

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
MS (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
MS (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

Table 4: Bayes Factor for Inertia Determination, April 2010
* indicates without dynamics

Here, we considered five popular SV models for illustration purpose. More complicated SV models can be investigated in our future work. One could also postulate more complicated microstructure dynamics.

References

[1] Aït-Sahalia, Y., Mykland, P., Zhang, L. (2005). How Often to Sample a Continuous-Time Process in the Presence of Market Microstructure Noise. *Review of Financial Studies* **18**, 351-416.

[2] Aït-Sahalia, Y., Kimmel, R. (2007). Maximum Likelihood Estimation of Stochastic Volatility Models. *Journal of Financial Economics* **83**, 413-452.

[3] Atar, R. (1998). Exponential stability for non-linear filtering of diffusion processes in a noncompact domain. *Annals of Probability* **26** No. 4, 1552-1574.

[4] Bandi, F. M., Russell, J. R. (2006). Separating Microstructure Noise from Volatility. *Journal of Financial Economics* **79**, 655-692.

[5] Barclay, M. (1997). Bid-ask Spreads and the Avoidance of Odd-eighth Quotes on Nasdaq: An Examination of Exchange Listings. *Journal of Financial Economics* **45**, 35-60.

[6] Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., Shephard, N. (2008). Subsampling Realised Kernels. Working paper.

[7] Black, F. (1986). Noise. *Journal of Finance* **41**, 529-543.

[8] Black, F., Scholes, M. (1973). The Pricing of Options and Corporate Liabilities. *Journal of Political Economy* **81**, 637-654.

[9] Bremaud, P. (1981). *Point Process and Queues: Martingale Dynamics.* Springer Verlag, New York.

[10] Chan, L. K. C., Lakonishok, J. (1993). Institutes Trades and Intraday Stock Prices Behavior. *Journal of Financial Econometrics* **33**, 173-199.

[11] Chung, K. H., Kim, K. A., Kitsabunnarat (2005). Liquidity and Quote Clustering in a Market with Multiple Tick Sizes. *Journal of Financial Research* **28**, 177-195.

[12] Chung, K. H., B. Van Ness, and R. Van Ness (2004). Trading Costs and Quotes Clustering on the NYSE and NASDAQ after Decimalization. *Journal of Financial Research* **27**, 309-328.

[13] Cox, J. C., Ross, S. A., and Rubinstein, M. (1979). Option Pricing: a Simplified Approach. *Journal of Financial Economics* **7**, 229-263. Cox-Ross Rubinstein (1979)

[14] Del Moral, P., Jacod, J., Protter, Ph. (2001). The Monte-Carlo Method for Filtering with Discrete-time Observations. *Probability Theory and Realted Fields* **120**, 346-368.

[15] Del Moral P., Kouritzin M.A. and Miclo L. (2001). On a class of discrete generation interacting particle systems. *Electron. J. Probab.* **6** 1-26.

[16] Del Moral, P., Noyer, J. C., Salut, G. (1995). Maslov Optimization Theory: Stochastic Interpretation, Particle Resolution. *Lecture Note in Control and Information Sciences* **199**, Springer-Verlag, Berlin.

[17] Delyon, B. Zeitouni, O. (1991) Lyapunov exponents for filtering problems. in *Applied Stochastic Analysis* edited by M. H. A. Davis and R. J. Elliot, Gordon and Breach Science Publishers, London, 511-521.

[18] Drost, F. C., Werker, J. M. (1996). Closing the GARCH Gap: Continuous Time GARCH Modeling. *Journal of Econometrics* **74**, 31-57.

[19] Duan, J. C., Fulop, A. (2008). Estimating the Structural Credit Risk Model When Equity Prices are Contaminated by Trading Noises. *Journal of Econometrics* Forthcoming.

[20] Duffie, D., Pan, J., Singleton, K. (2000). Transform Analysis and Asset Pricing for Affine Jump-Diffusions. *Econometrica* **68**, 1343V1376.

[21] Duncan, T. E., Hu, Y. Z., Pasik-Duncan, B. (2000). Stochastic Calculus for Fractional Brownian Motion. *SIAM J. Control Optim.* **38**, 582-612.

[22] Elliot, R. J., Malcolm, W. P., Tsoi, A. H. (2003). Robust Parameter Estimation for Asset Price Models with Markov Modulated Volatilities. *Journal of Econ. Dyna. Control* **27**, 1391-1409.

[23] Engle, R., Russell, J. (1998). Autoregressive Conditional Duration: A New Model for Irregularly Spaced Transaction Data. *Econometrica* **66**, 1127-1162.

[24] Engle, R. (2000). The Econometrics of Ultra-high-frequency Data. *Econometrica* **68**, 1-22.

[25] Eraker, B., Johannes, M., Polson, N. (2003). The Impact of Jumps in Volatility and Returns. *Journal of Finance* **58**, 1269-1300.

[26] Eraker, B. (2004). Do Equity Prices and Volatility Jumps? Reconciling Evidence from Spot and Option Prices. *Journal of Finance* **59**, 1367V1403.

[27] Ethier, S. N., Kurtz, T. G. (1986). *Markov Processes: Characterization and Convergence.* John Wiley & Sons, New York.

[28] Griffin, J. M., Ji, S., Martin, J. S. (2003). Momentum Investing and Business Cycle Risk: Evidence from Pole to Pole. *Journal of Finance* **58**, 2515-2547.

[29] Grundy, B. D., Martin, J. S. (2001). Understanding the Nature of the Risks and the Source of the Rewards to Momentum Investing. *Review of Financial Studies* **14**, 29-78.

[30] Grother, O., Müller, C. (2008). Liquidity-Based Estimation of Stochastic Volatility under Microstructure Noise. Working paper.

[31] Hansen, P. R., Lunde, A. (2006). Realized Variance and Market Microstructure Noise. *Journal of Business and Economic Statistics* **24**, 1-58.

[32] Hasbrouck, J. (1996). Modelling Market Microstructure Time Series. *Handbook of Statistics. North Holland, Amsterdam.* **14**, 647-692.

[33] Hasbrouck, J. (1999). Security Bid/Ask Dynamics with Discreteness and Clustering: Simple Strategies for Modeling and Estimation. *Journal of Financial Markets* **2**, 1-28.

[34] Heston, S. (1993). A Closed-form Solutions for Options with Stochastic Volatility with Applications to Bond and Currency Options. *Rev. Finan. Stud.* **6**, 327-343.

[35] Huang, R., Stoll, A. (2001). Tick Size, Bid-ask Spreads, and Market Structure. *Journal of Financial and Quantitative Analysis* **36**, 503-522.

[36] Hull, J., White, A. (1987). The Pricing of Options on Assets with Stochastic Volatilities. *Journal of Finance* **42**, 281-300.

[37] Jackwerth, J., Rubinstein, M. (1996). Recovering Probability Distributions from Contemporaneous Security Prices. *Journal of Finance* **51**, 1611-1631.

[38] Jegadeesh, N., Titman, S. (1993). Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency. *Journal of Finance* **48**, 65-91.

[39] Kass, R. E., Raftery, A. E. (1995). Bayes Factor and Model Uncertainty. *Journal of the American Statistical Association* **90**, 773-795.

[40] Kloeden, P. E., and Platen, E. (1992). Numerical Solution of Stochastic Differential Equations, Springer.

[41] Kouritzin, M. A., Zeng, Y. (2005). Bayesian Model Selection via Filtering for a Class of Micro-movement Models of Asset Price. *International Journal of Theoretical and Applied Finance* **8**, 97-121.

[42] Mandelbrot, B. B. (1997). *Fractals and Scaling in Finance: Discontinuity, Concentration, Risk.* Springer-Verlag.

[43] Merton, R. C. (1973). The Theory of Rational Option Pricing. *The Bell Journal of Economics and Management Science* **4**, 141-183.

[44] Moskowitz, T. J., Grinblatt, M. (1999). Do Industries Explain Momentum? *Journal of Finance* **54**, 1249-1290.

[45] Nelson, D. B. (1990). ARCH Models as Diffusion Approximations. *Journal of Econometrics* **45**, 7-39.

[46] Nielsen, J., Vestergaard, M. (2000). Estimation in Continuous-time Stochastic Volatility Models Using Nonlinear Filters. *International Journal of Theoretical and Applied Finance* **3**, 279-308.

[47] Ocone, D. Pardoux, E. (1996). Asymptotic stability of the optimal lter with respect to its initial condition. *SIAM J. Control Optim.* **34**, no. 1, 226-243.

[48] Scott, L. O. (1987). Option Pricing when the Variance Changes Randomly: Theory, Estimation and Application. *Journal of Financial Quantitative Analysis* **22**, 419-438.

[49] Shiryaev, A. (1999). *Essentials of Stochastic Finance*. World Scientific.

[50] Stroock, D. W., Varadhan, S. R. S. (1979). *Multidimensional Diffusion Processes*. Springer-Verlag, Berlin.

[51] Xiong, J., Zeng, Y. (2011). A Branching Particle Approximation to the Filtering Problem with Counting Process Observations. Statistical Inference for Stochastic Processes **14**, 111-140..

[52] Zeng, Y. (2003). A Partially Observed Model for Micromovement of Asset Prices with Bayes Estimation via Filtering. *Mathematical Finance* **13**, 411-444.

Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, Alberta T6G 2G1 Canada. 780-492-2704

E-mail address: `michaelk@ualberta.ca`