Statistical Analysis of the Filtering Models with Marked Point Process (MPP) Observations: Applications to Ultra-High Frequency (UHF) Data

Yong Zeng

zeng@mendota.umkc.edu

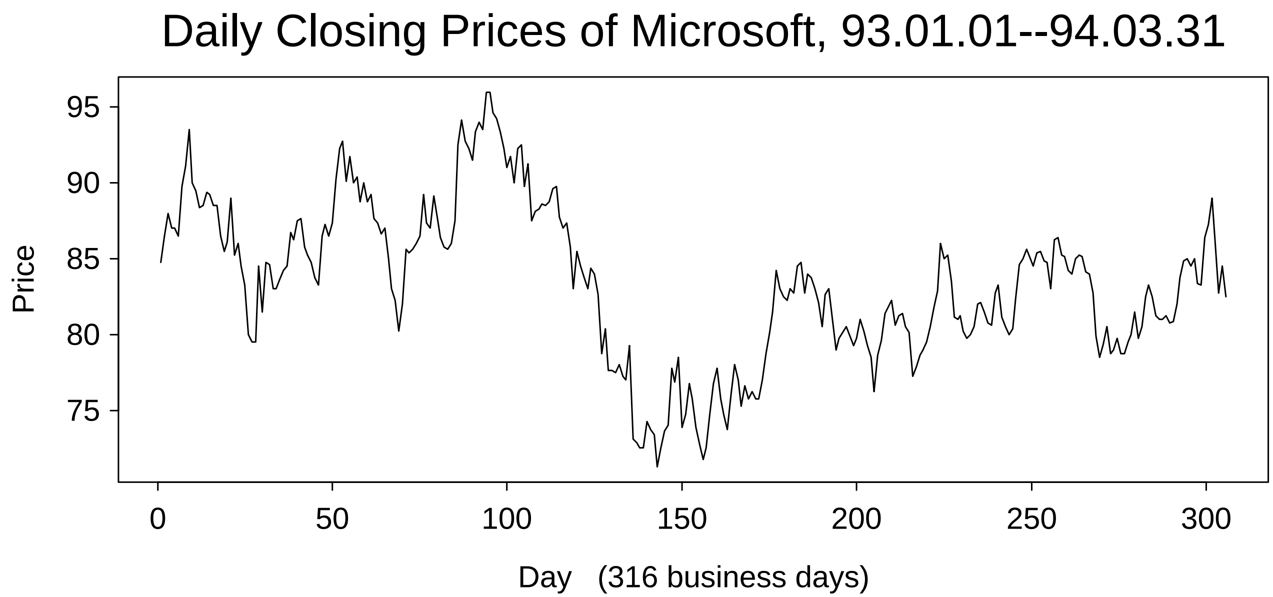
Department of Mathematics and Statistics

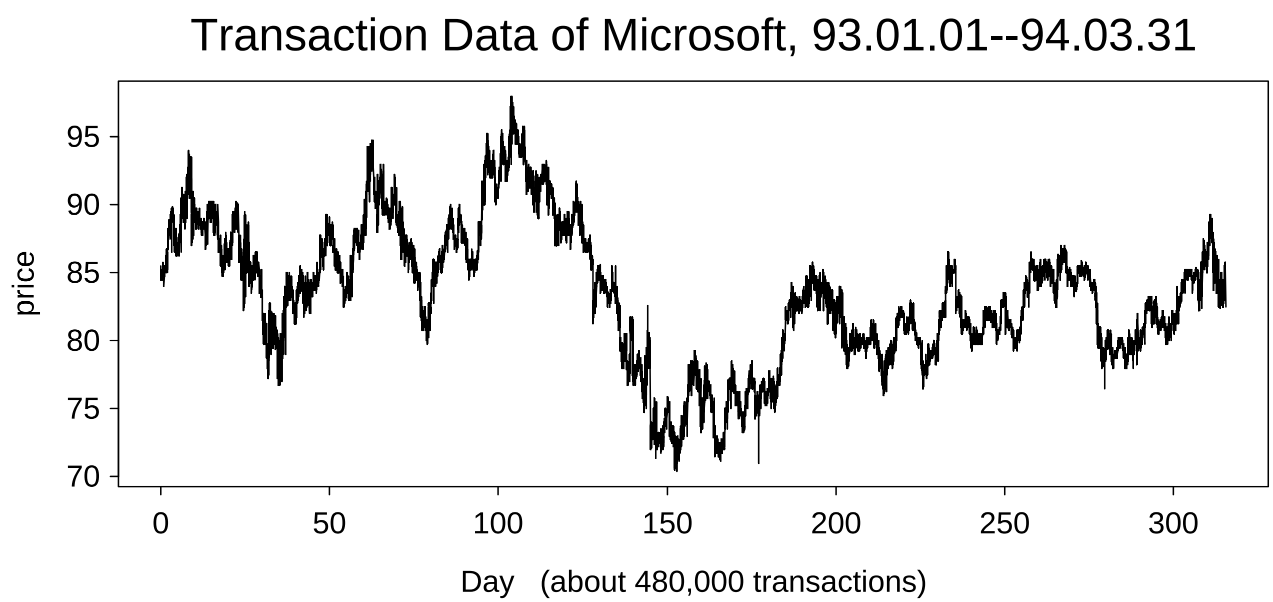
University of Missouri at Kansas City

Outline

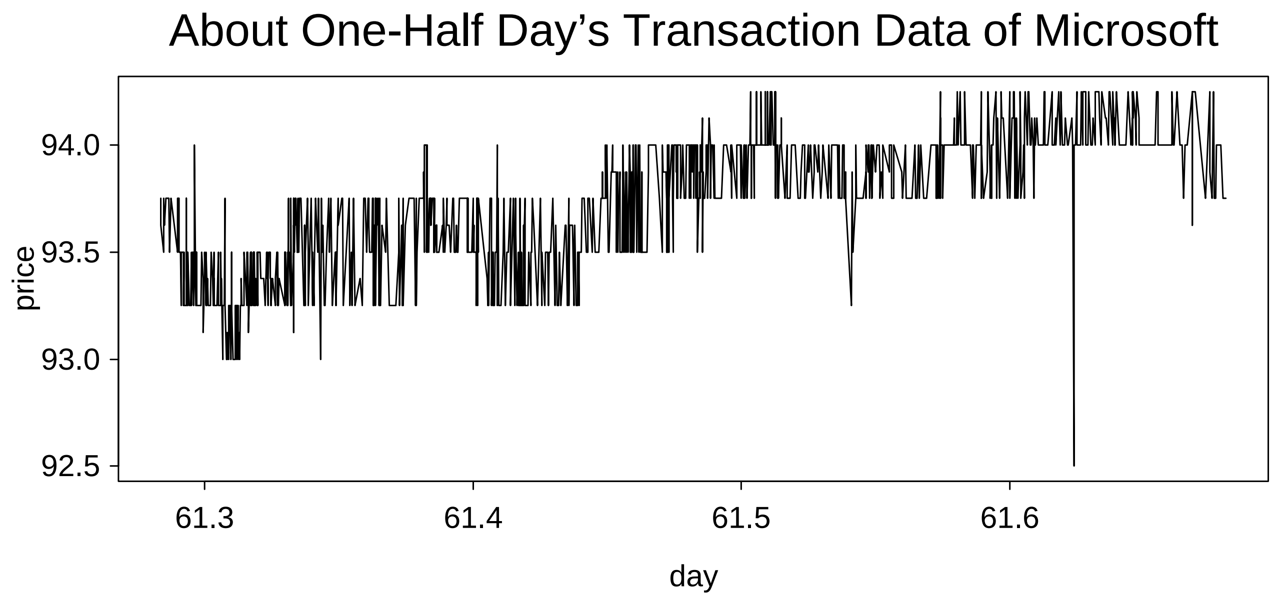
- UHF data and Marked Point Process
- Two Different Views of UHF data
  - An Irregularly-Spaced Time Series
  - A Realized Sample Path of MPP
- A Model with two Equivalent Representations
  - Filtering with MPP (counting process) Observations
  - Random-arrival-time State Space Model
- Bayesian Inference via Filtering
  - Continuous-time Likelihoods, Posterior, and Bayes Factors
  - Filtering Equations and Evolution Equations for Bayes Factors
  - Two Computational Approaches and their Consistency
    - The Markov Chain Approx. Method and nearly likelihood etc.
    - Particle Filtering (or Sequential Monte Carlo)
- Future Works

Macro- and Micro-movements





Microscopic Picture



Two Characteristics of UHF Data

- Observations occur at varying random time intervals
- Trading (or market microstructure) noises are in price data

Marked Point Process (MPP)

- Point process: R.V. $\{T_n\}$ satisfying $T_n \leq T_{n+1}$.
- Mark, X_n: are random elements associated with these times.
- Marked Point: (T_n, X_n); MPP: $\{(T_n, X_n)\}$.

An Irregularly-Spaced Time Series

- Engle (2000) calls the transaction (tick-by-tick) financial time series *Ultra-High Frequency* Data.

- Engle (2000) – Date: $\{(\Delta t_i, y_i), i = 1, ..., N\}$ where $\Delta t_i = t_i - t_{i-1}$.
- **Log likelihood:**

$$(\Delta t_i, y_i)|\mathcal{F}_{i-1} \sim f(\Delta t_i, y_i|\breve{\Delta} t_{i-1}, \breve{y}_{i-1}; \theta)$$

where $\breve{z}_i = \{z_i, z_{i-1}, ..., z_1\}$.

$$f(\Delta t_i, y_i|\breve{\Delta} t_{i-1}, \breve{y}_{i-1}; \theta) = g(\Delta t_i|\breve{\Delta} t_{i-1}, \breve{y}_{i-1}; \theta) q(y_i|\Delta t_i, \breve{\Delta} t_{i-1}, \breve{y}_{i-1}; \theta) \quad (1)$$

$$L(\Delta, Y; \theta) = \sum_{i=1}^{N} \log g(\Delta t_i|\breve{\Delta} t_{i-1}, \breve{y}_{i-1}; \theta) + \sum_{i=1}^{N} \log q(y_i|\Delta t_i, \breve{\Delta} t_{i-1}, \breve{y}_{i-1}; \theta)$$

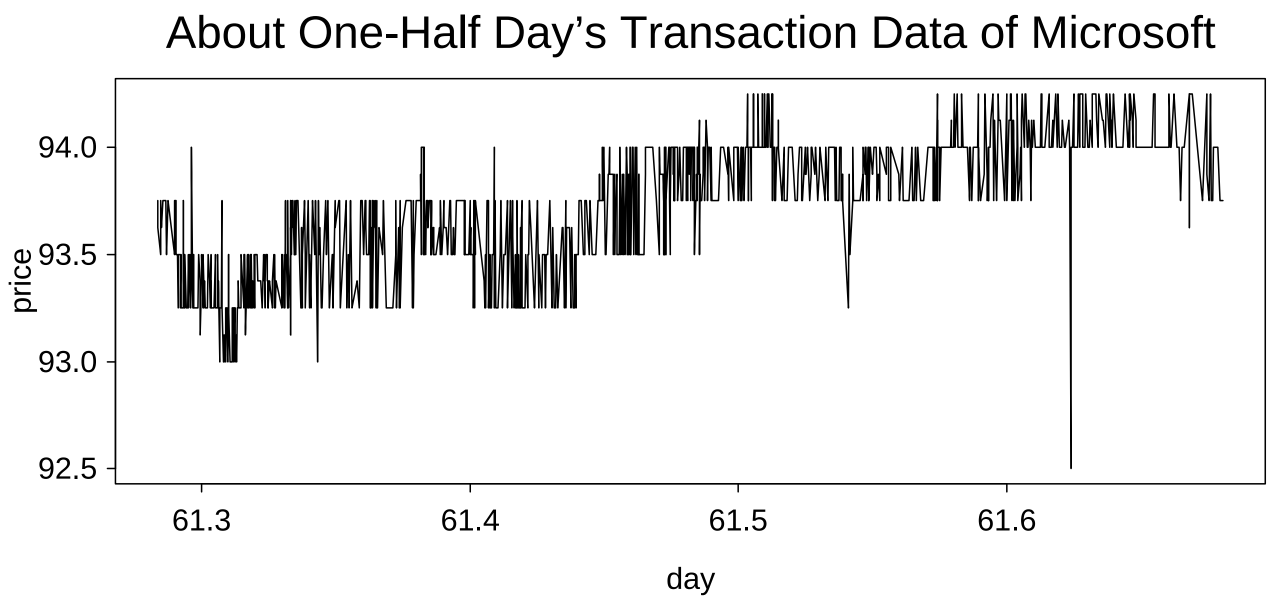
Developments within Engle (2000)

- Autoregressive Conditional Duration (ACD) model by Engle and Russell (1998) and logarithmic ACD by Bauwens and Giot (2000),threshold ACD by Zhang, Russell and Tsay (2001), Asymmetric ACD by Bauwens and Giot (2003)
- Bivariate Point process model by Engle and Lunde (2003), and Autoregressive Conditional Intensity model by Russell (1999)
- Ordered Probit model by Hausman, Lo and Mackinlay (1992) Autoregressive Conditional Multinomial by Russell (1998), Activity-Direction-Size model by Rydberg and Shepard (2003), Price Change Duration model by McCulloch and Tsay (2001)
- UHF-GARCH by Engle (2000), ACD-GARCH by Ghysels and Jasiak (1998)

- **The Second View:**

A Realized Sample Path of MPP: Zeng (2003)

A Collection of Counting Processes



$$\vec{Y}(t) = \begin{pmatrix} N_1(\int_0^t \lambda_1(\theta(s), X(s), s)ds) \\ N_2(\int_0^t \lambda_2(\theta(s), X(s), s)ds) \\ \vdots \\ N_n(\int_0^t \lambda_n(\theta(s), X(s), s)ds) \end{pmatrix}, \quad (2)$$

where $Y_j(t) = N_j(\int_0^t \lambda_j(\theta(s), X(s), s)ds)$ records the cumulative # of trades that have occurred at the jth price level up to time t.

Assumptions: Model I

Filtering with counting process observations

- **Assumption 1.1:** Markov process, (θ, X), is the solution of a martingale problem for a generator $\mathbf{A}$ such that

$$M_f(t) = f(\theta(t), X(t)) - \int_0^t \mathbf{A} f(\theta(s), X(s)) ds$$

is a $\mathcal{F}_t^{\theta, X}$-martingale. $X(t)$ is the intrinsic value process of an asset.
- **Assumption 1.2:** $(N_1, \ldots, N_n)$ are unit Poisson processes under measure P.
- **Assumption 1.3:** $(\theta, X), N_1, \ldots, N_n$ are independent under P.
- **Assumption 1.4:** $0 \le a(\theta(t), X(t), t) \le C$ for some $C > 0$ and all $\theta(t), X(t), t > 0$.
- **Assumption 1.5:** Intensities: $\lambda_j(\theta, x, t) = a(\theta, x, t) p(y_j | x)$, where $a(x, \theta, t)$ is the total trading intensity, and $p_j = p(y_j | x)$ is the transition probability from x to y_j.

Filtering with MPP Observations

- **Setup:** Mark space: U; measure space: $(U, \mathcal{U}, \mu)$, μ: finite measure; ξ is a Poisson Random Measure (PRM) on $\mathcal{U} \times \mathcal{B}[0,\infty) \times \mathcal{B}[0,\infty)$ with mean measure $\mu \times m \times m$. For $A \in \mathcal{U}$,

$$Y(A,t) = \int_{A\times[0,t]\times[0,+\infty)} \mathbf{I}_{[0,\lambda(\theta(s),X(s),V(s),Z(s);u,s)]}(v)\xi(du \times ds \times dv),$$

where $Y(A,t)$ is a counting process recording the cumulative number of events that have occurred in the set A up to time t.

$$\tilde{Y}(A,t) = Y(A,t) - \int_{A\times[0,t]} \lambda(\theta(s), X(s), V(s), Z(s); u, s)\mu(du)ds$$

is a martingale.

Signal: (θ, X) **Observation:** (Y, V) or (Z, V).

Assumptions: Model II

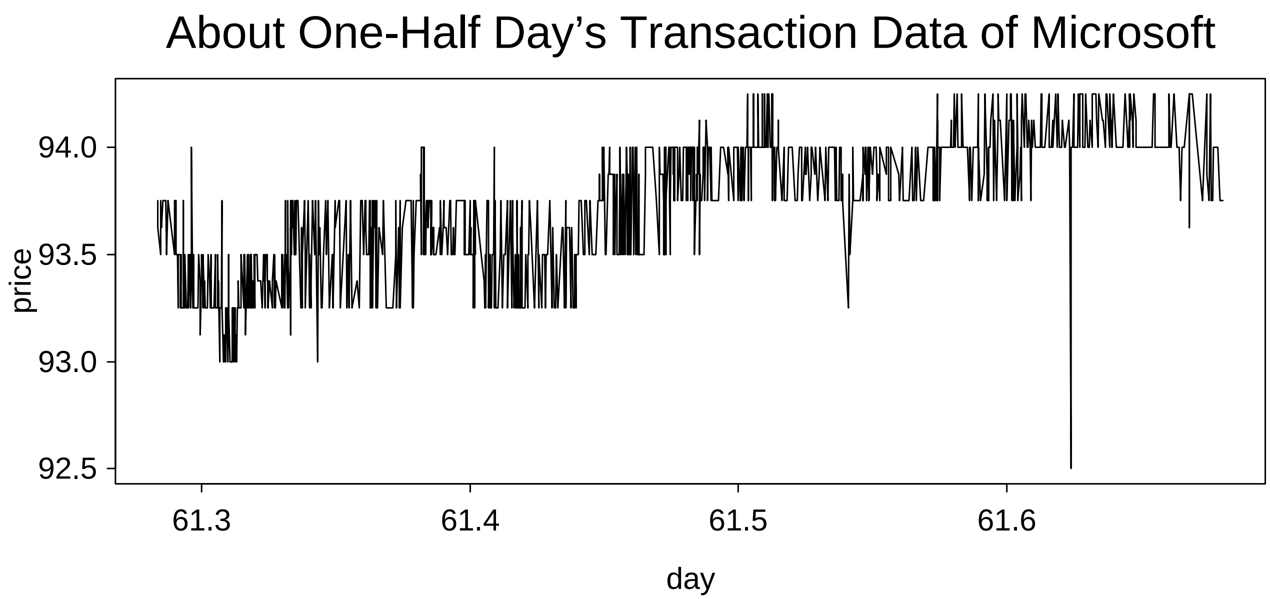
- **Assumption 2.1:** is the same as Assumption 1.1, but both θ and X can be vector processes. Or, (θ, X) is a semimartingale vector process.
- **Assumption 2.2:** ξ is a PRM with the mean measure $\mu \times m \times m$ under P, where μ is finite measure.
- **Assumption 2.3:** (θ, X) and ξ are independent under measure P.
- **Assumption 2.4:** $0 \leq a(\theta(t), X(t), V(t), Z(t), t) \leq C$ for some $C > 0$ and all possible $\theta(t), X(t), V(t), Z(t), t$.
- **Assumption 2.5:** Stochastic intensity kernel:

$$\lambda(\theta, x, v, z; u, t-) = a(\theta, x, v, z, t-)p(u|x; \theta, v, z) \tag{3}$$

where $p(u|x; \theta, v, z) = p(u|X(t); \theta(t-), V(t-), Z(t-))$ is the transition probability from $X(t)$ to u.

Remark: Note the similarity between Eq.(3) and Eq.(1).

Random-arrival-time State-Space Model



Three-Step Construction:

- State process: (θ, X) as in Assumption 2.1.
- Event times, $t_1, t_2, \ldots, t_i, \ldots$ follows a conditional Poisson process with $a(\theta(t), X(t), V(t), Z(t), t)$ in Assumption 2.4.
- Observation at t_i: $Z(t_i) = F(X(t_i); \theta(t_i), V(t_i), Z(t_{i-1}))$, where $F(\cdot; \cdots)$ is a random transformation with the transition probability $p(Z(t_i)|X(t_i); \theta(t_i), V(t_i), Z(t_{i-1}))$ as in Assumption 2.5.

Examples

- Zeng (2003) and its extension to multi-stocks.
- Many models under the framework of Engle (2000) such as Exponential ACD model, UHF-GARCH and more.
- Estimating Volatility via filtering: Frey and Runggaldier (2001) and Cvitanic, Liptser and Rozovskii (2003).
- Estimating Markov process sampled at conditional Poisson time: Duffie and Glenn (2004).
- Classical examples of MPP filtering problems in books: Bremaud (1981), Liptser and Shiryayev (2002, 2nd Ed.), and Last and Brandt (1995).

An Integral Form of Price

- Let $Z(t)$ be the price of the most recent transaction at or before time t.

$$Z(t) = Z(0) + \int_{[0,t]\times U} (u - Z(s-))Y(du \times ds).$$

$$dZ(t) = \int_U (u - Z(t-))Y(du \times dt).$$

Remarks :

- This is the telescoping sum: $Z(t) = Z(0) + \sum_{t_i \le t}(Z(t_i) - Z(t_{i-1}))$.
- If there is a price change from $Z(t-)$ to u occurs at time t, then $Z(t) - Z(t-) = (u - Z(t-))$ implying $Z(t) = u$.

- This form is essential for the risk minimization hedging (Lee and Zeng 2006, Model I), and the mean-variance portfolio selection problem of the model (Xiong and Zeng 2006, Model I).

Another Example

- Intrinsic value process:

$$\frac{dX_t}{X_t} = \mu(\theta, X_t, V_t, Z_t)dt + \sigma(\theta, X_t, V_t, Z_t)dB_t$$

- Price:

$$dZ(t) = \int_U (u - Z(t-))Y(du \times dt).$$

where the stochastic intensity kernel for $Y(\cdot, \cdot)$ at (u, t) is:

$$\lambda_Y(u, t-; \theta, X_{t-}, V_{t-}, Z_{t-}) =$$

$$a(\theta, X_{t-}, V_{t-}, Z_{t-}, t)p(u|\theta, X_{t-}, V_{t-}, Z_{t-})$$

and θ: parameters; and V: other observable factors.

Joint Likelihood Function

- Continuous-time joint likelihood function of (θ, X, Y):
- For Model I,

$$L(t) = \frac{dP}{dQ}(t) = \prod_{k=1}^{n} \exp\Big\{ \int_0^t \log \lambda_k(\theta(s-), X(s-), s-) dY_k(s) - \int_0^t \Big[\lambda_k(s) - 1\Big] ds \Big\}.$$

- For Model II,

$$L(t) = \exp\Big\{ \int_0^t \int_U \log \lambda(\theta(s-), X(s-), V(s-), Z(s-); u, s-) Y(du \times ds) - \int_0^t \int_U \Big[\lambda(u,s) - 1\Big] \mu(du) ds \Big\}$$

Why $L(t)$ has this form?

Under P, $\tilde{Y}(t)$: a conditional Poisson; under Q, a unit Poisson process.

$$\tilde{L}(t) = \frac{d\tilde{P}}{d\tilde{Q}}(t) = \exp\left\{\int_0^t \log \tilde{\lambda}(\tilde{X}(s-), s-)d\tilde{Y}(s) - \int_0^t \left[\tilde{\lambda}(\tilde{X}(s), s) - 1\right]ds\right\}.$$

$$\frac{d\tilde{P}}{d\tilde{Q}}(t) = \frac{d\tilde{P}/dR}{d\tilde{Q}/dR} \qquad \Big(P(X=0) = e^{-\lambda}, \;\; P(X=1) = e^{-\lambda}\lambda\Big)$$

$$\approx \frac{e^{-(\int_0^{t_1-h} + \sum_{i=1}^{m-1}\int_{t_i+}^{t_{i+1}-h} + \int_{t_m+}^{t})\tilde{\lambda}ds} \cdot \prod_{i=1}^m e^{-\int_{t_i-h}^{t_i}\tilde{\lambda}ds}\int_{t_i-h}^{t_i}\tilde{\lambda}ds}{e^{-[(t_1-h)+\sum_{i=1}^{m-1}(t_{i+1}-h-t_i)+(t-t_m)]} \cdot \prod_{i=1}^m e^{-h}h}$$

$$\approx e^{-(\int_0^{t_1-h} + \sum_{i=1}^{m-1}\int_{t_i+}^{t_{i+1}-h} + \int_{t_m+}^{t})(\tilde{\lambda}-1)ds}\prod_{i=1}^m e^{-h[\tilde{\lambda}-1]}\tilde{\lambda}(\tilde{X}(t_i-), t_i-)$$

$$\to \exp\left\{-\int_0^t [\tilde{\lambda}-1]ds + \int_0^t \log\tilde{\lambda}(\tilde{X}(s-), s-)d\tilde{Y}(s)\right\}$$

Likelihoods and Posterior

Define: $\phi(f,t) = E^Q[f(\theta(t), X(t))L(t)|\mathcal{F}_t^{Y,V}]$. Then, $\phi(1,t) = E^Q[L(t)|\mathcal{F}_t^{Y,V}]$ is the likelihood of Y *or* the *integrated (marginal) likelihood* of Y after assigning a prior to $(\theta(0), X(0))$.

- To see this, suppose $p(\theta, x, y)$ is the joint density for real R.V. (θ, X, Y) w.r.t. Lebesgue measure Q'. Then the marginal density of Y is:

$$p_Y(y) = \int\int p(\theta, x, y)d\theta dx = E^{Q'}[p(\theta, X, Y)|Y = y].$$

Define: π_t is the conditional distribution of $(\theta(t), X(t))$ given $\mathcal{F}_t^{Y,V}$.
π_t becomes the *posterior* after a prior is assigned.
Define: $\pi(f,t) = E^P[f(\theta(t), X(t))|\mathcal{F}_t^{Y,V}] = \int f(\theta, x)\pi_t(d\theta, dx)$.

- Kallianpur-Striebel (Bayes) Formula gives: $\pi(f,t) = \frac{\phi(f,t)}{\phi(1,t)}$.

Bayes Factor and Likelihood Ratio

Suppose there are two models: Model 1 and Model 2.

Define:

$$q_1(f_1, t) = \frac{\phi_1(f_1, t)}{\phi_2(1, t)} \quad \text{and} \quad q_2(f_2, t) = \frac{\phi_2(f_2, t)}{\phi_1(1, t)}$$

The Bayes Factors:(BF: the ratio of two integrated likelihoods)

$$BF_{12} = \frac{\phi_1(1, t)}{\phi_2(1, t)} = q_1(1, t) \quad \text{and} \quad BF_{21} = \frac{\phi_2(1, t)}{\phi_1(1, t)} = q_2(1, t)$$

- *Strongly Reject Model 1* if BF_{21} is larger than 12.
- *Decisively Reject Model 1* if BF_{21} is larger than 150.

Advantages: (1) BF do not require the two models to be nested, nor their distributions to be absolutely continuous w.r.t. each other.
(2) Under some conditions, *BF* $\approx$ *BIC*, which penalizes according to both the number of parameters and the number of data.

Filtering Equations

- **Theorem 1:** *Under Assumptions 2.1–2.5,*

$$\phi(f,t) = \phi(f,0) + \int_0^t \phi(\mathbf{A}f,s)ds - \int_0^t \int_U \phi(f(\lambda(u)-1),s)\mu(du)ds$$

$$+ \int_0^t \int_U \phi(f(\lambda(u)-1),s-)Y(du \times ds).$$

and

$$\pi(f,t) = \pi(f,0) + \int_0^t \pi(\mathbf{A}f,s)ds + \int_0^t \pi(f,s) \int_U \pi(\lambda(u),s)\mu(du)ds$$

$$- \int_0^t \int_U \pi(f\lambda(u),s)\mu(du)ds + \int_0^t \int_U \Big[\frac{\pi(f\lambda(u),s-)}{\pi(\lambda(u),s-)} - \pi(f,s-)\Big] dY(du \times ds)$$

Evolution Equations for BF

- **Theorem 2:** *Assume Model 1 has* $(\mathbf{A_1}, \lambda_1, \mu_1)$ *and Model 2 has* $(\mathbf{A_2}, \lambda_2, \mu_2)$. *Both models satisfy Assumptions 2.1–2.5*

$$q_1(f_1, t) = q_1(f_1, 0) + \int_0^t q_1(\mathbf{A_1} f_1, s) ds$$

$$+ \int_0^t \frac{q_1(f_1, s)}{q_2(1, s)} \int_U q_2(\lambda_2(u), s)\mu_2(du) ds - \int_0^t \int_U q_1(f_1\lambda_1(u), s)\mu_1(du) ds$$

$$+ \int_0^t \int_U \Big[\frac{q_1(f_1\lambda_1(u), s-)}{q_2(\lambda_2(u), s-)} q_2(1, s-) - q_1(f_1, s-) \Big] dY(du \times ds)$$

and

$$q_2(f_2, t) = \ldots$$

A Consistency Theorem

- **Theorem 3:** *Suppose that Assumptions 2.1 to 2.5 hold for* (θ, X, Y) *and* $(\theta_\epsilon, X_\epsilon, Y_\epsilon)$*. If* $(\theta_\epsilon, X_\epsilon) \Rightarrow (\theta, X)$ *as* $\epsilon \to 0$*, then for bounded continuous functions,* f*,*
(i) $Y_\epsilon \Rightarrow Y$*,* *(ii)* $\phi_\epsilon(f,t) \Rightarrow \phi(f,t)$*,* *(iii)* $\pi_\epsilon(f,t) \Rightarrow \pi(f,t)$*.*
In the two-model case for model selection, then
(iv) $q_{k,\epsilon}(f_k, t) \Rightarrow q_k(f_k, t)$ *for* $k = 1, 2$ *simultaneously.*

Sketch of Proof:

- First, use Kurtz and Protter (1991)'s theorem on convergence of stochastic integral and the Continuous Mapping theorem to prove $L_\epsilon \Rightarrow L$. Then, $((\theta_\epsilon, X_\epsilon), Y_\epsilon, L_\epsilon) \Rightarrow ((\theta, X), Y, L)$.
- Second, use Goggin (1994)'s or Kouritzin and Zeng (2005)'s theorems to convergence of conditional expectations and the Continuous Mapping theorem to prove (ii), (iii) and (iv).

Markov Chain Approximation Method

Three-Step Construction of Recursive Algorithms – *for computing nearly posterior, integrated likelihood and Bayes factors*

For Example, to compute the *nearly* posterior:

- Construct a continuous-time Markov chain $(\theta_\epsilon, X_\epsilon)$ to approximate (θ, X).
- Derive the filtering (or evolution) equations for $(\theta_\epsilon, X_\epsilon, Y_\epsilon)$.
- Convert the equation for $(\theta_\epsilon, X_\epsilon, Y_\epsilon)$ to recursive algorithms by
  - (a) representing $\pi_\epsilon(\cdot, t)$, for example, as a finite array with components being $\pi_\epsilon(f, t)$ for lattice-point indicator f;
  - (b) approximating the time integral with an Euler scheme.

Two Micromovement Models

- Value Processes of the two Micromovement Models

1. GBM: (Zeng 2003)

$$\frac{dX_t}{X_t} = \mu dt + \sigma dW_t,$$

$$\mathbf{A} f(x) = \frac{1}{2}\sigma^2 x^2 \frac{\partial^2}{\partial x^2} f(x) + \mu x \frac{\partial}{\partial x} f(x).$$

2. JSV-GBM: (Zeng 2004)

$$\frac{dX_t}{X_t} = \mu dt + \sigma(t) dW_t,$$

$$d\sigma(t) = (U_{N(t)} - \sigma(t-))dN(t)$$

where $N(t)$ is a Poisson process with intensity λ_σ and the jump size, $\{U_i\}$, are i.i.d random variables with uniform distribution on $[\alpha_\sigma, \beta_\sigma]$.

Noise for the Two Models

$$Y(t_i) = F(X(t_i)) = b_i(R[X(t_i), \frac{1}{8}] + V_i)$$

- **Discrete noise:** $R[x, \frac{1}{8}]$, rounding function.
- **Non-clustering noise:** $\{V_i\}$, has a doubly-geometric distribution:

$$P\{V = v\} = \begin{cases} (1-\rho) & \text{if } v = 0 \\ \frac{1}{2}(1-\rho)\rho^{8|v|} & \text{if } v = \pm\frac{i}{8} \text{ for } i = 1, 2, 3\ldots. \end{cases}.$$

- **Clustering noise:** $b_i(\cdot)$, a random biasing function

biasing rule: Set $y' = R[X(t_i), \frac{1}{8}] + V_i$ and $y = Y(t_i) = b(y')$.

· If the fractional part of y' is an even eighth, then y stays on y' w. p. 1.

· If the fractional part of y' is an odd eighth, then

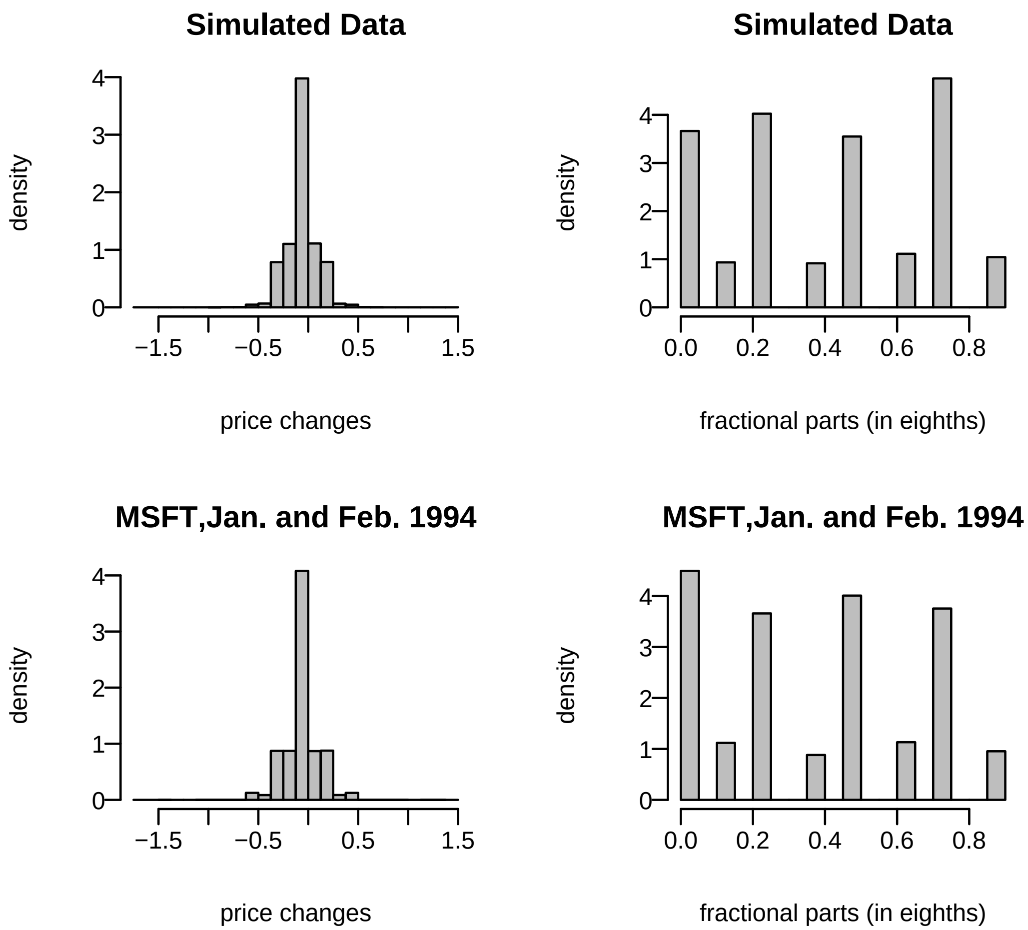
y' moves to the closest odd quarter w.p. α,

or y' moves to the closest half or integer w.p. β,

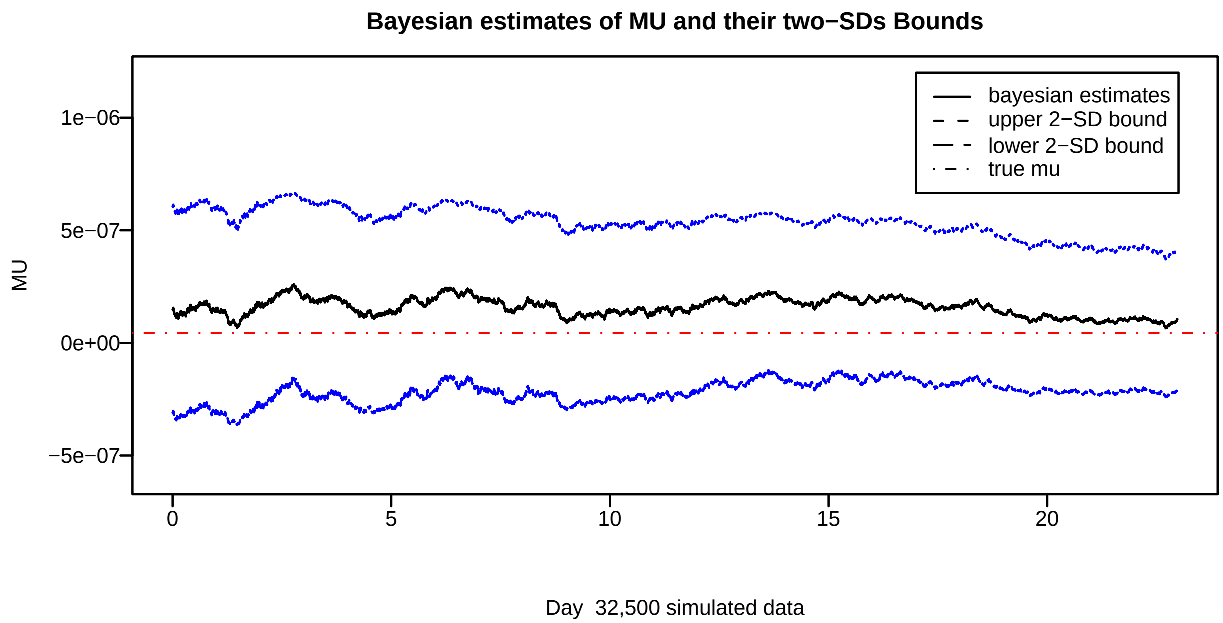
or y stays on y' w.p. $1 - \alpha - \beta$.

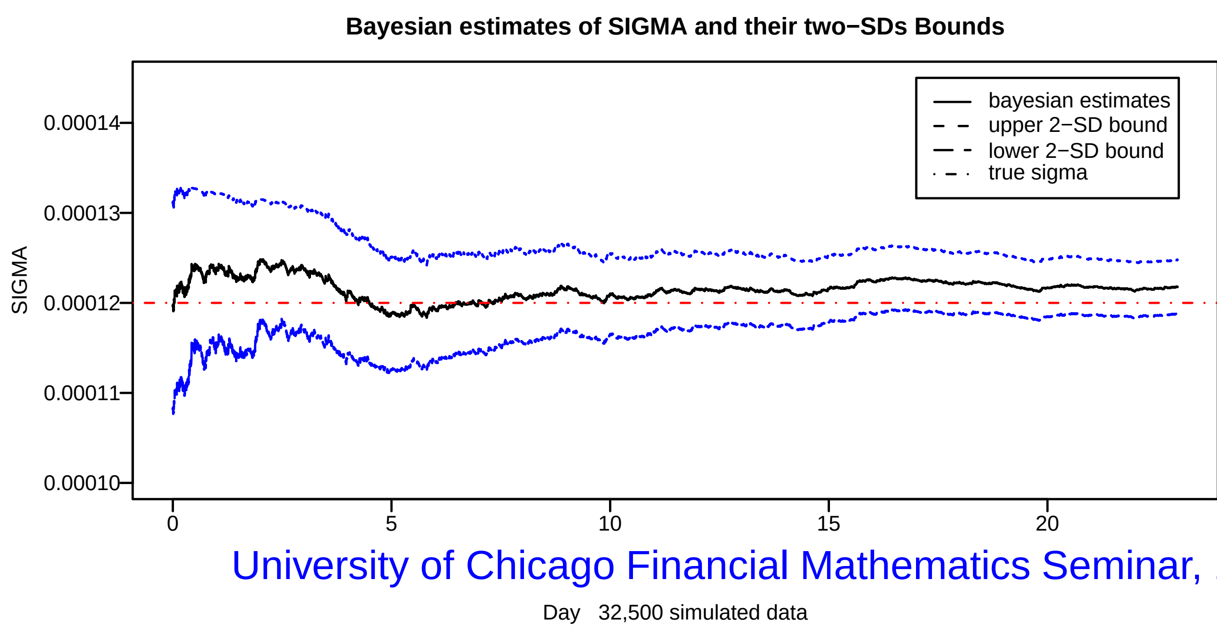
- Model 1: $(\mu, \sigma, \rho, \alpha, \beta)$. Model 2: $(\mu, \sigma(t), \lambda_\sigma, \rho, \alpha, \beta)$.

Noise Fitting

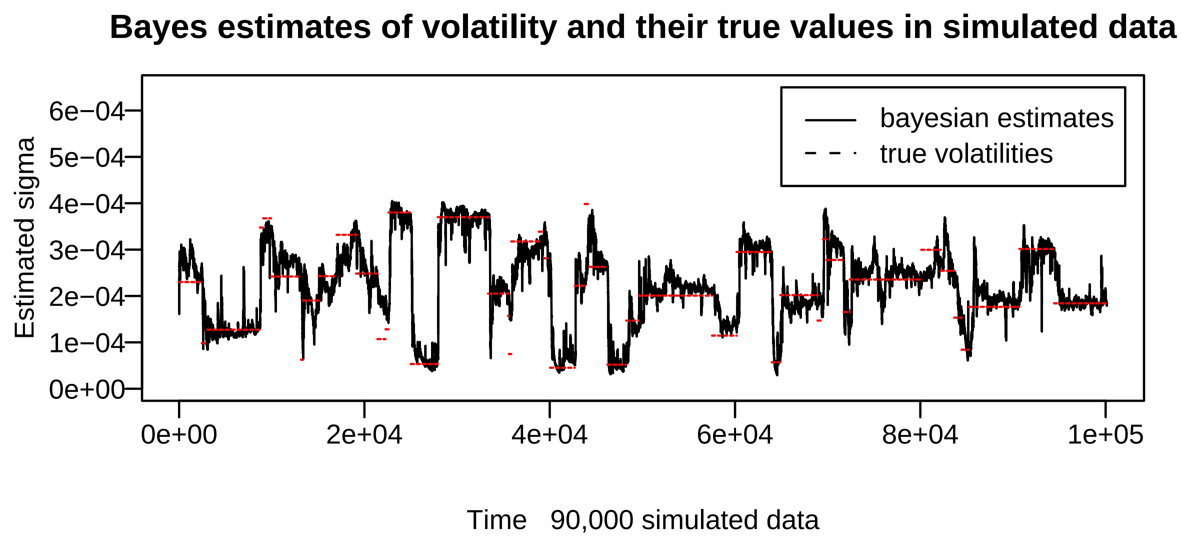


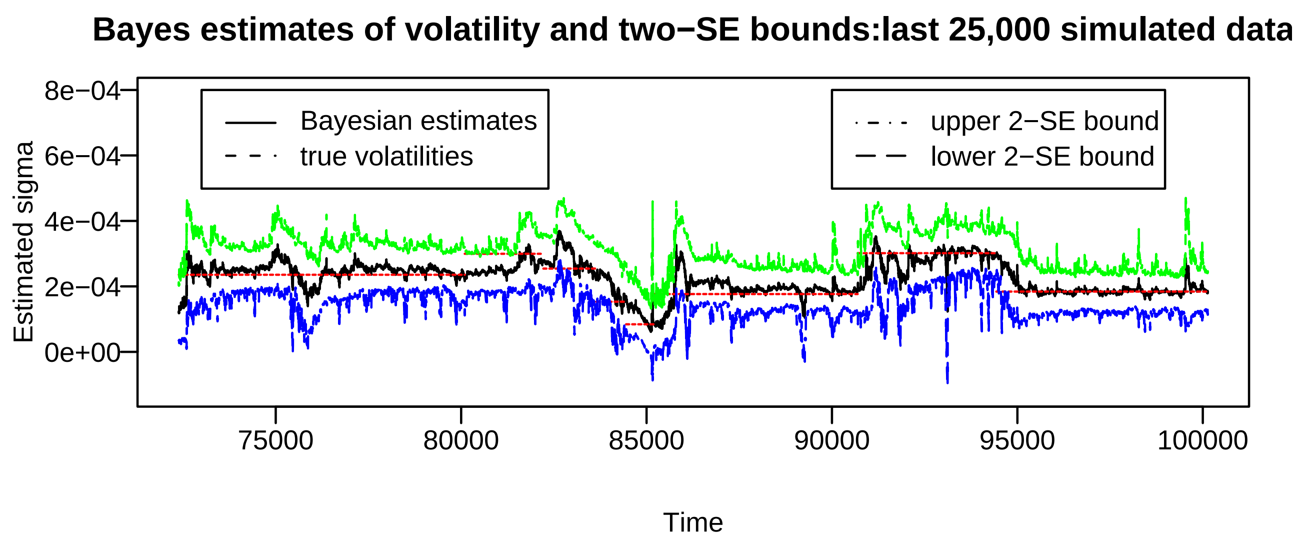
Bayes Estimates I: Simulated Data



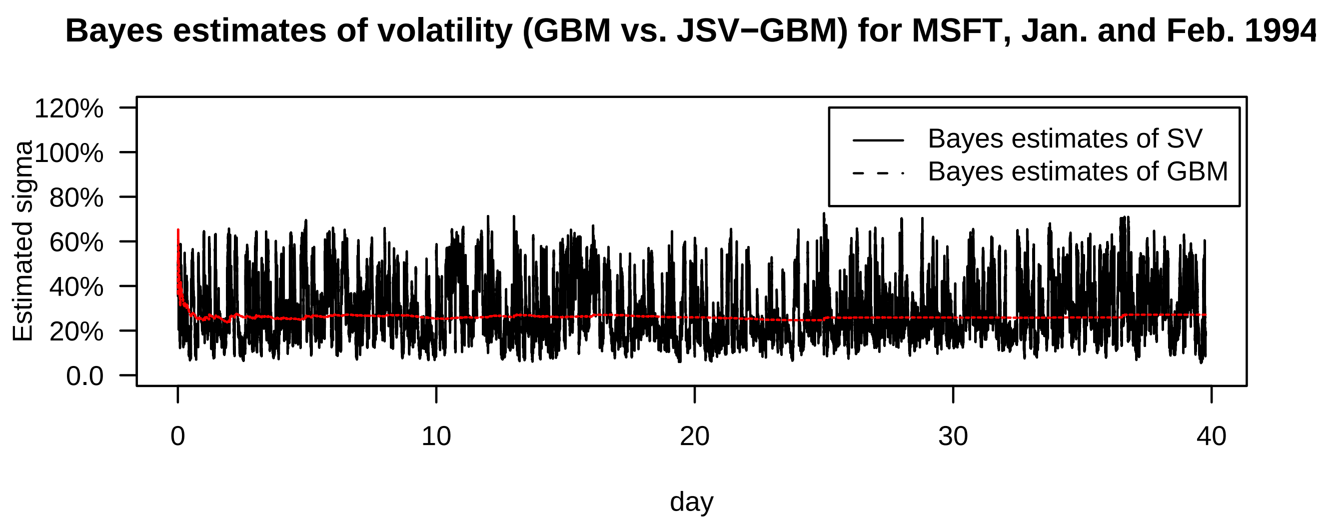


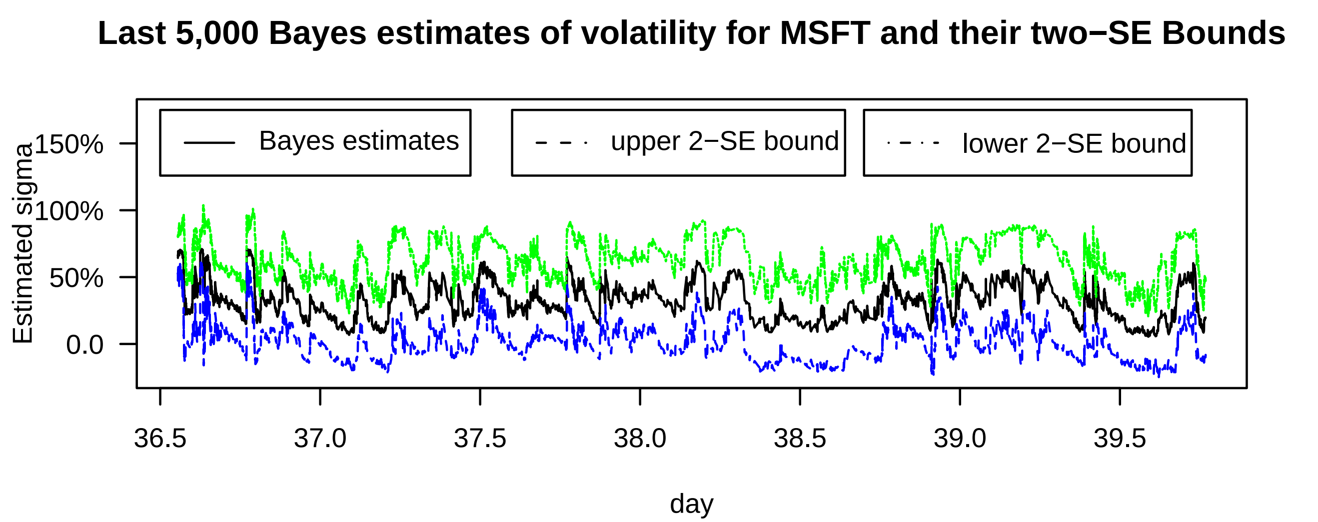
Bayes Estimates II: Simulated Data



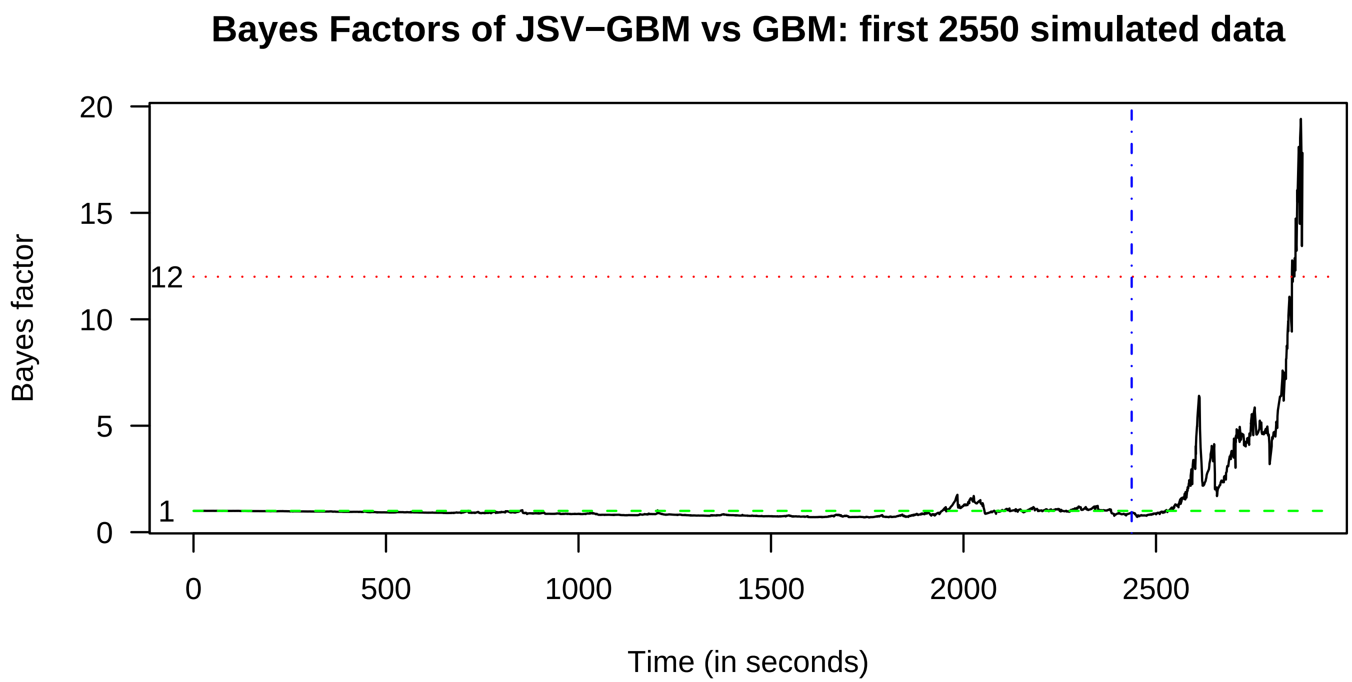


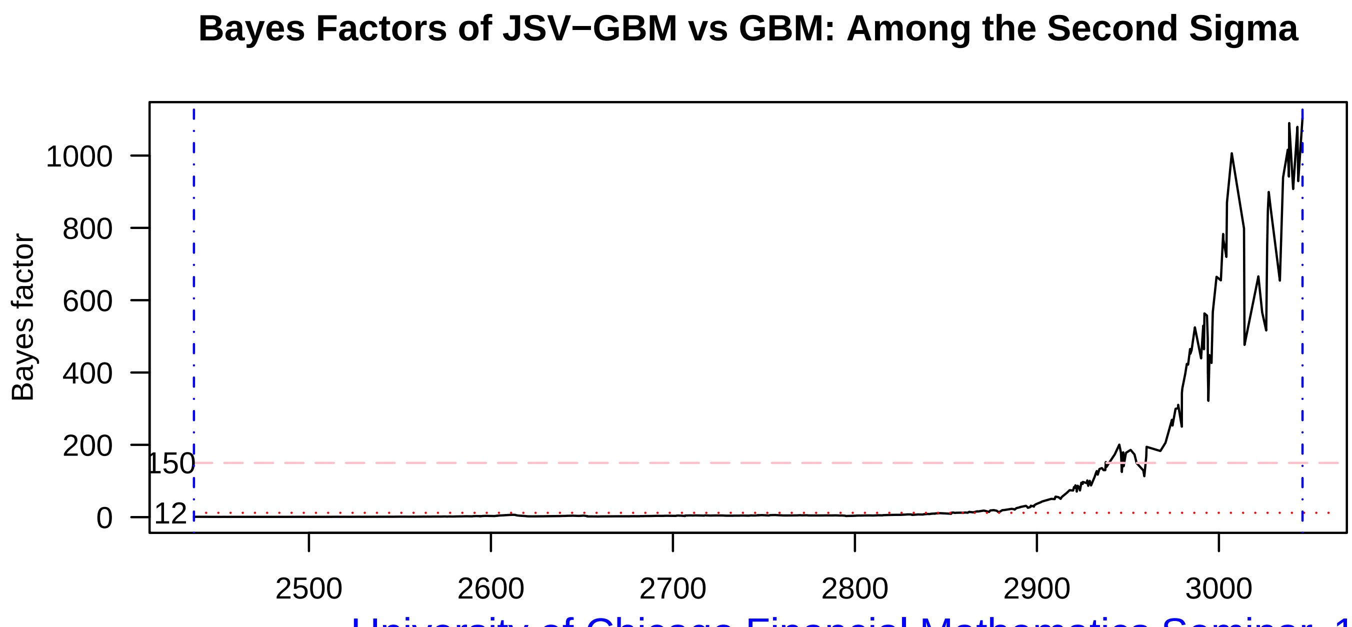
Bayes Est. III: MSFT, Jan/Feb 1994





Bayes Factor I: Simulated Data





Bayes Factor II: Simulated Data

Table 1: Bayes Factors for a Simulated Data

Position before σ changes	2166	2676	7790	8113	90000
Bayes Factor: B_{21}	0.9358	1103.70	1.134e+10	1.255e+10	1.089e+194

Bayes Factor III: MSFT, Jan/Feb. 1994

Table 2: Summary Statistics for BF_{21} of the First Day in MSFT Data

Position	NO. of Data	Min.	Median	Mean	Max.
1st Quarter	375	0.9133	48.19	104.20	931.30
2nd Quarter	164	28.64	69.40	660.00	11280.00
3rd Quarter	130	2178	7472	67060	584400
4th Quarter	287	24250	41360	75680	297800

Particle Filtering (*or SMC*)

Suppose parameters are known. To estimate the value process, X,

- **Simulate** 100 independent sample paths of X following GBM: $V_j(t)$, $j = 1, 2, \cdots, 100$, at each trading times.
- At some time, t_i, calculate the **importance weight** for each V_j:

$$w_i^j(V_j(t_i)) = L^j(t_i) = \prod_{k=1}^{n} \exp\left\{\int_{t_{i-1}}^{t_i} \log \lambda_k(V_j(s-), s-) dY_k(s)\right\}$$

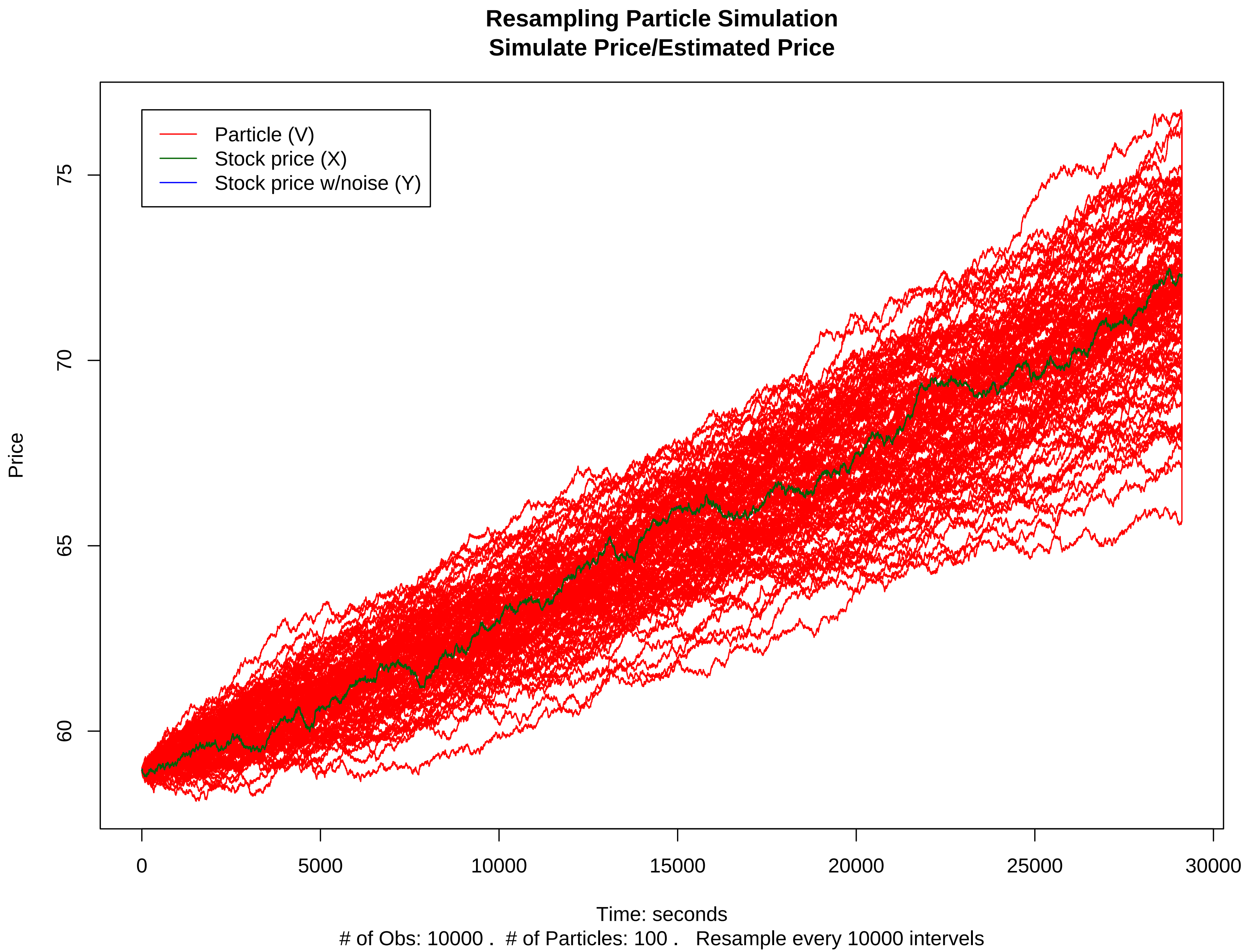
$$\times \prod_{k=1}^{n} \exp\left\{-\int_{t_{i-1}}^{t_i} \Big[\lambda_k - 1\Big] ds\right\}.$$

- **Resample** the sample paths according to the distribution proportional to the importance weights at times.

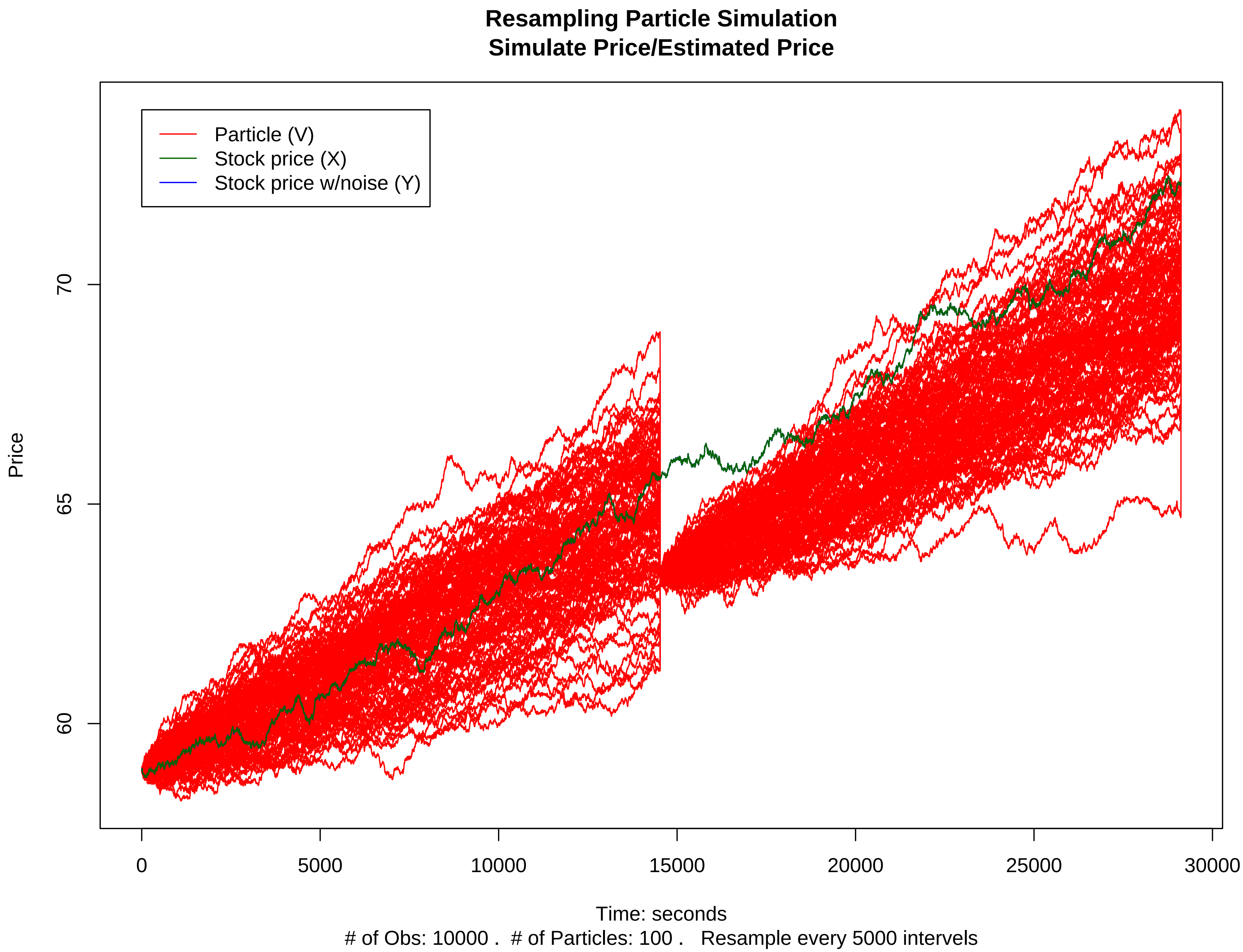
Or

- **Branch** each particle to a random number of particles proportional to the importance weights at times. (Xiong and Zeng 2006)

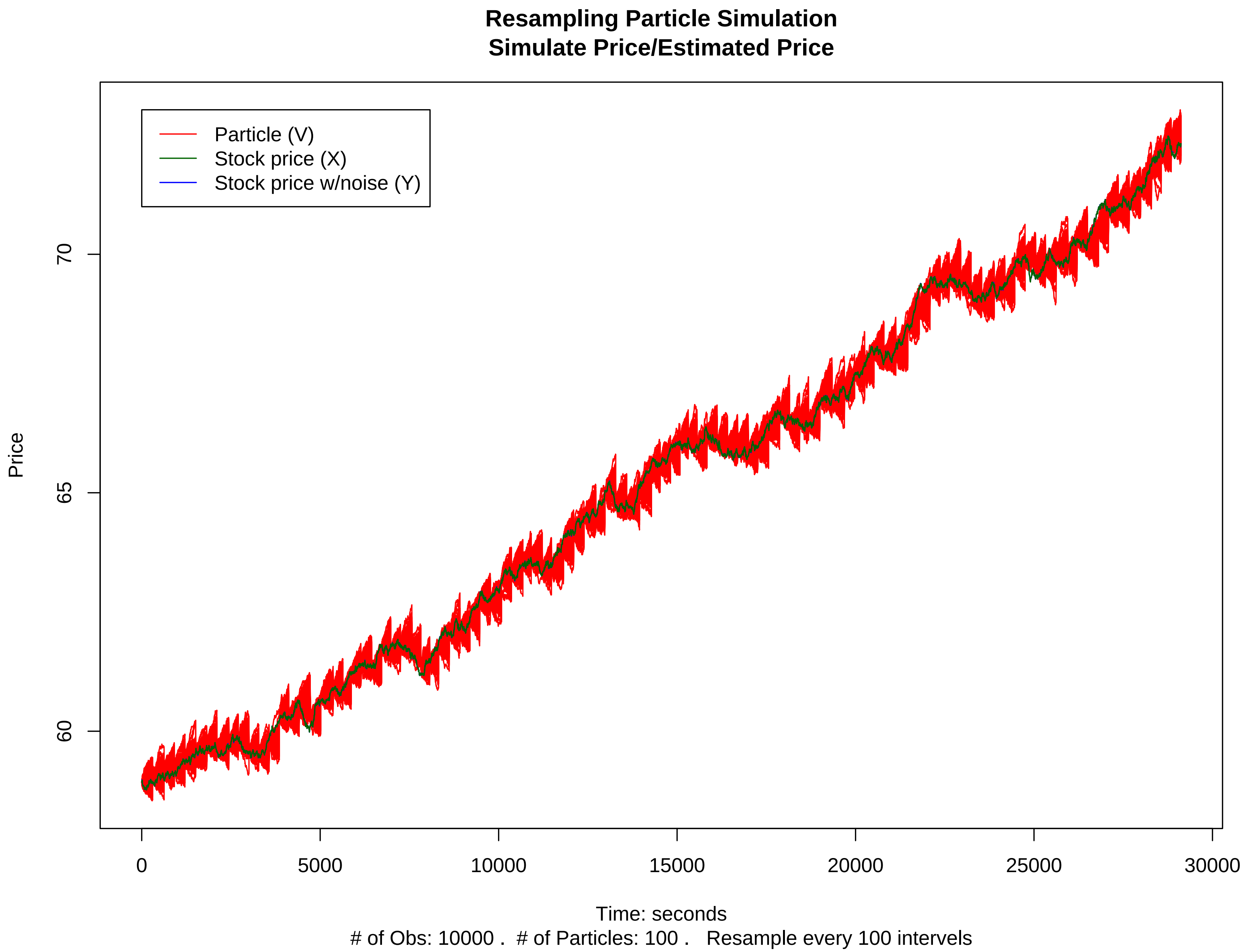
Simulation: No Resampling



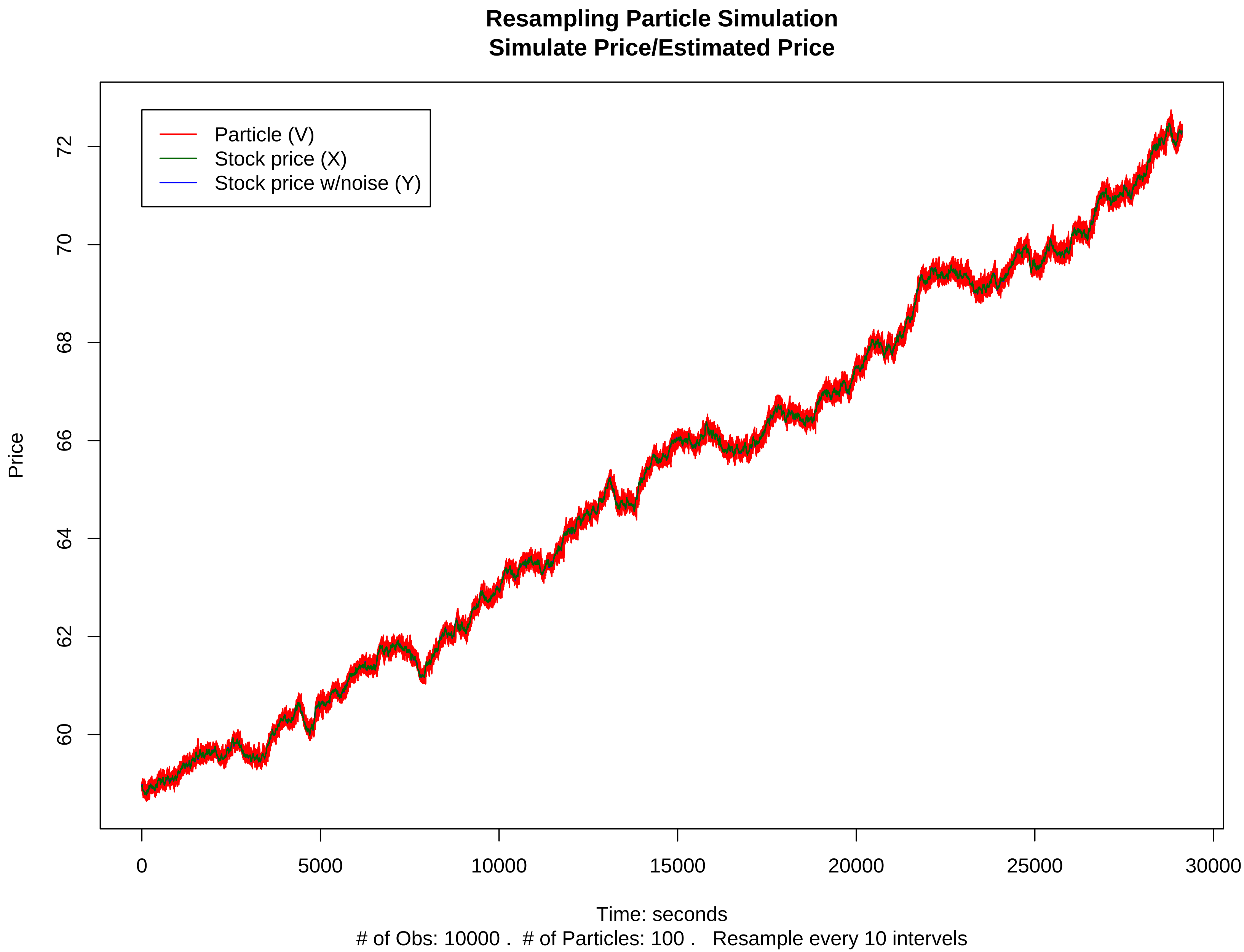
Simulation: One Resampling



Simulation: Resampling Every 100 Trades



Simulation: Resampling Every 10 Trades



Conclusions and Future Works

- Mathematical finance:
  - Option pricing and hedging, portfolio optimization, and utility maximization.
- Financial applications on market microstructure theory
- Particle filtering or sequential Monte Carlo
  - Convergence, convergence rate and large deviation
- Statistics:
  - Consistency, CLT for the estimators of parameters.
- To allow long-range dependence by using fractional signal.

Related papers, real data, Fortran codes are available at
`http://mendota.umkc.edu/paper-tick.html`