Accurate asset price modeling and related statistical problems under microstructure noise

José E. Figueroa-López[1]

[1]Department of Statistics
Purdue University

Exploring Statistical Science Research Seminar

December 8, 2009

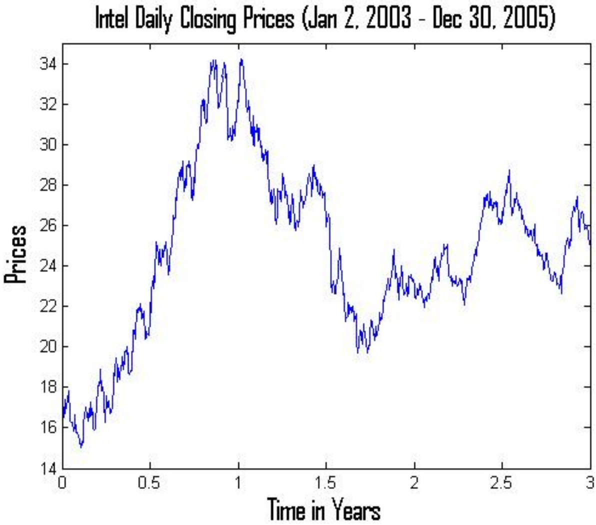
Outline

1. Motivation
   - Asset price modeling
   - Stochastic models for asset prices
2. High-frequency based statistics
   - General idea
   - Examples
3. Market Microstructure
4. Open problems

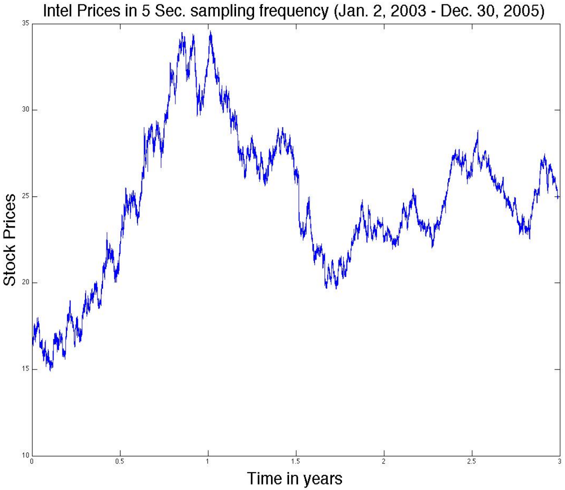
Outline

How does the price of a stock behaves in time?

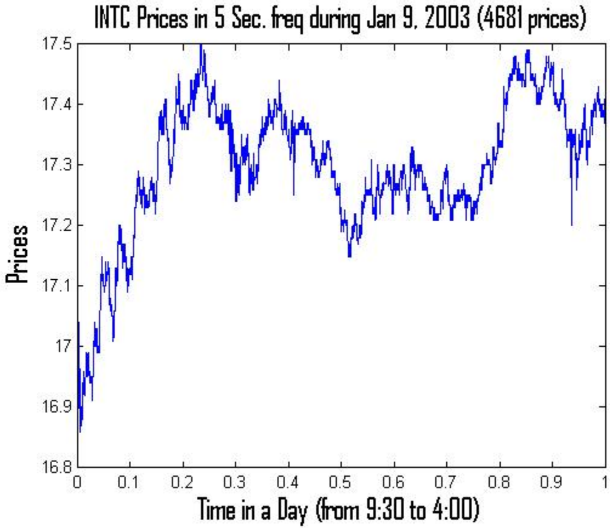
How does the price of a stock behaves in time?



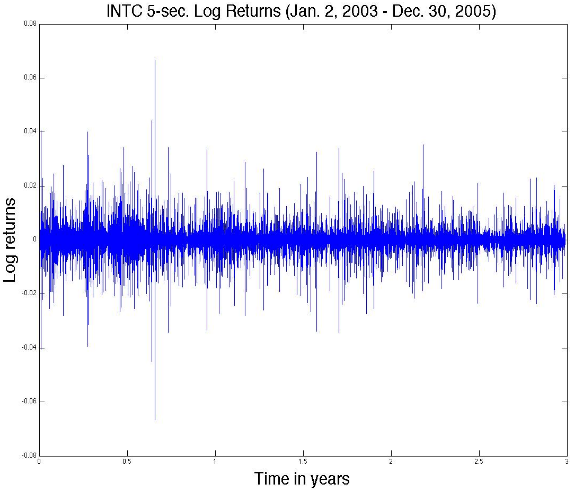
How does the price of a stock behaves in time?



How does the price of a stock behaves in time?



How does the price of a stock behaves in time?



Stylized empirical features of stock prices

1. “Sudden big" changes in the price levels (Jumps)
2. Volatility clustering (intermittency)
3. Log returns with heavy-tails and high-kurtosis distributions
4. Leverage phenomenon

Outline

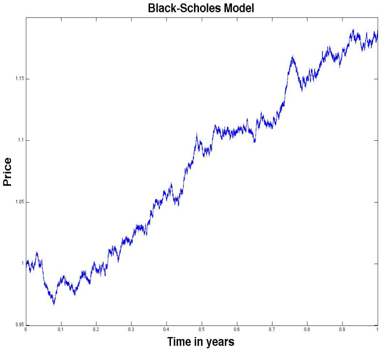
What are good models for the price process $\{S_t\}_{t\geq 0}$?

What are good models for the price process $\{S_t\}_{t\geq 0}$?

Black-Scholes model (1973); Samuelson (1965):

$$R_i^{(\delta)} = \log \frac{S_{(i+1)\delta}}{S_{i\delta}} = b\delta + \sqrt{\delta v}\, Z_i, \quad \text{where} \quad Z_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1).$$

Jiggling motion, no jump-like changes, no clustering or leverage

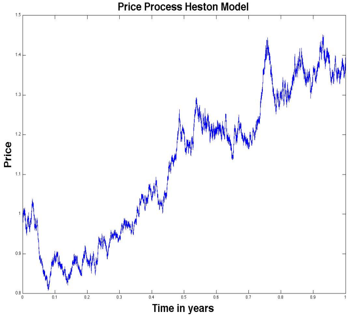


What are good models for the price process $\{S_t\}_{t\geq 0}$?

Stochastic volatility Heston model (1993):

$$R_i^{(\delta)} = b\delta + \sqrt{\delta\, v_{i-1}^{\delta}}\, Z_i,$$

$$v_i^{\delta} = v_{i-1}^{\delta} + \alpha(m - v_{i-1}^{\delta})\delta + \gamma\sqrt{\delta v_{i-1}^{\delta}}\, Z_i', \quad \rho = \mathrm{Corr}(Z, Z').$$



What are good models for the price process $\{S_t\}_{t\geq 0}$?

Stochastic volatility with jumps:

$$R_i^{(\delta)} = b\delta + \sqrt{\delta\, v_{i-1}^{\delta}}\, Z_i + \theta\delta^{1/\beta} J_i^{(\beta)},$$
$$v_i^{\delta} = v_{i-1}^{\delta} + \alpha(m - v_{i-1}^{\delta})\delta + \gamma\sqrt{\delta v_{i-1}^{\delta}}\, Z_i'.$$

where $0 < \beta < 2$ and $J_i^{(\beta)}$ are i.i.d. symmetric with heavy tails:

$$P(J_i^{(\beta)} \geq x) \sim cx^{-\beta}, \quad \text{as} \quad x \to \infty.$$

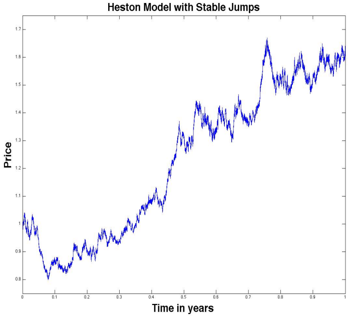
Remarks:

- $P(Z_i \geq x) \ll e^{-x}/x$; hence, J_i feels like jumps compared to Z_i;
- The larger β, the lighter the tails, and the smaller $J_i^{(\beta)}$ will tend to be;
- β is called the index of jump activity.

What are good models for the price process $\{S_t\}_{t\geq 0}$?

Stochastic volatility with jumps:

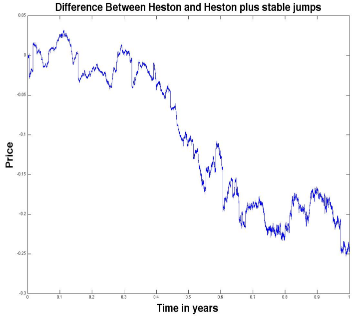
$$R_i^{(\delta)} = b\delta + \sqrt{\delta\, v_{i-1}^{\delta}}\, Z_i + \theta\delta^{1/\beta} J_i^{(\beta)},$$
$$v_i^{\delta} = v_{i-1}^{\delta} + \alpha(m - v_{i-1}^{\delta})\delta + \gamma\sqrt{\delta v_{i-1}^{\delta}}\, Z_i'.$$



What are good models for the price process $\{S_t\}_{t\geq 0}$?

Stochastic volatility with jumps:

$$R_i^{(\delta)} = b\delta + \sqrt{\delta\, v_{i-1}^{\delta}}\, Z_i + \theta\delta^{1/\beta} J_i^{(\beta)},$$
$$v_i^{\delta} = v_{i-1}^{\delta} + \alpha(m - v_{i-1}^{\delta})\delta + \gamma\sqrt{\delta v_{i-1}^{\delta}}\, Z_i'.$$

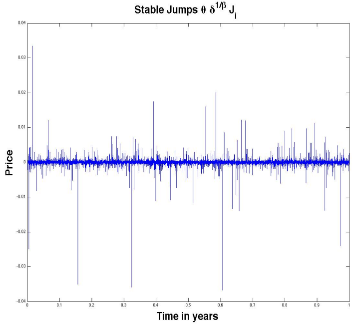


What are good models for the price process $\{S_t\}_{t\geq 0}$?

Stochastic volatility with jumps:

$$R_i^{(\delta)} = b\delta + \sqrt{\delta\, v_{i-1}^{\delta}}\, Z_i + \theta\delta^{1/\beta} J_i^{(\beta)},$$

$$v_i^{\delta} = v_{i-1}^{\delta} + \alpha(m - v_{i-1}^{\delta})\delta + \gamma\sqrt{\delta v_{i-1}^{\delta}}\, Z_i'.$$



Outline

Estimation methods based on high-frequency data

- Set-up:
  - We collect n log returns at intervals of size δ:

    $$R_i^{(\delta)} := \log \frac{S_{i\delta}}{S_{(i-1)\delta}},$$

    for $i = 1, \ldots, n$.
  - Overnight returns are filtered out;
  - The sampling times are then $t_i = i\delta$ and the time horizon is $T = n\delta$.
- What is it?

  Any estimator $\hat{\theta}^{\delta,n}$ which is consistent for a “parameter" θ when $\delta \to 0$:

  $$\hat{\theta}^{\delta,n} \xrightarrow{P} \theta,$$

  as $\delta \to 0$, keeping T fixed!

Estimation methods based on high-frequency data

- Set-up:
  - We collect n log returns at intervals of size δ:

$$R_i^{(\delta)} := \log \frac{S_{i\delta}}{S_{(i-1)\delta}},$$

    for $i = 1, \ldots, n$.
  - Overnight returns are filtered out;
  - The sampling times are then $t_i = i\delta$ and the time horizon is $T = n\delta$.
- What is it?

  Any estimator $\hat{\theta}^{\delta,n}$ which is consistent for a "parameter" θ when $\delta \to 0$:

$$\hat{\theta}^{\delta,n} \xrightarrow{P} \theta,$$

  as $\delta \to 0$, keeping T fixed!

Estimation methods based on high-frequency data

- Set-up:
  - We collect n log returns at intervals of size δ:

$$R_i^{(\delta)} := \log \frac{S_{i\delta}}{S_{(i-1)\delta}},$$

    for $i = 1, \ldots, n$.
  - Overnight returns are filtered out;
  - The sampling times are then $t_i = i\delta$ and the time horizon is $T = n\delta$.
- What is it?

  Any estimator $\hat{\theta}^{\delta,n}$ which is consistent for a “parameter" θ when $\delta \to 0$:

$$\hat{\theta}^{\delta,n} \xrightarrow{P} \theta,$$

  as $\delta \to 0$, keeping T fixed!

Outline

Examples

1. Quadratic variation of the continuous component:

$$v_t := \lim_{\delta \to 0} \sum_{i:\delta i \le t} v^{\delta}_{i-1}\delta, \quad 0 \le t \le T.$$

Then, for any $\alpha > 0$ and $0 > \omega < 1/2$,

$$v_t^{\delta,n,\alpha,\omega} := v_t^{\delta} := \sum_{i:\delta i \le t} |R_i^{(\delta)}|^2 \mathbf{1}_{\left\{|R_i^{(\delta)}| \le \alpha\delta^{\omega}\right\}} \xrightarrow{P} v_t,$$

2. Index of jump activity β:

$$\beta^{\delta,n,\alpha,\alpha',\omega} := \beta^{\delta} := \frac{\log\left(\frac{\sum_{i:\delta i \le t} |R_i^{(\delta)}| \mathbf{1}_{\left\{|R_i^{(\delta)}| \ge \alpha\delta^{\omega}\right\}}}{\sum_{i:\delta i \le t} |R_i^{(\delta)}| \mathbf{1}_{\left\{|R_i^{(\delta)}| \ge \alpha'\alpha\delta^{\omega}\right\}}}\right)}{\log(\alpha')} \xrightarrow{P} \beta.$$

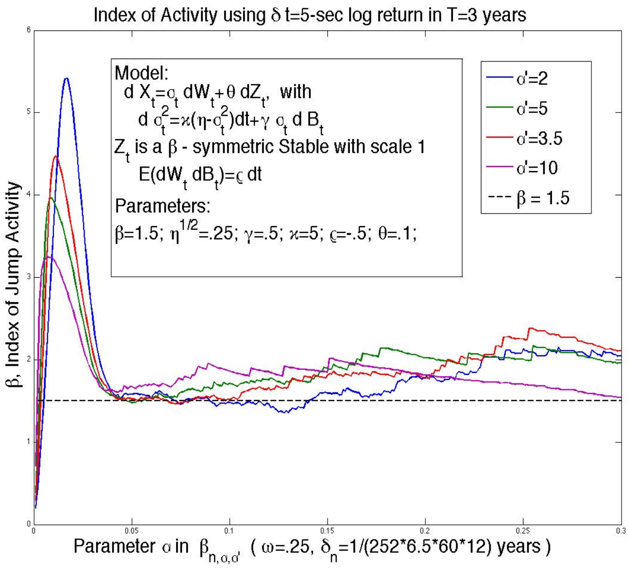
Index of jump activity for Heston with stable jumps

Stochastic volatility with jumps:

$$R_i^{(\delta)} = b\delta + \sqrt{\delta\, v_{i-1}^{\delta}}\, Z_i + \theta\delta^{1/\beta} J_i^{(\beta)},$$

$$v_i^{\delta} = v_{i-1}^{\delta} + \alpha(m - v_{i-1}^{\delta})\delta + \gamma\sqrt{\delta v_{i-1}^{\delta}}\, Z_i'.$$

Index of jump activity for Heston with stable jumps

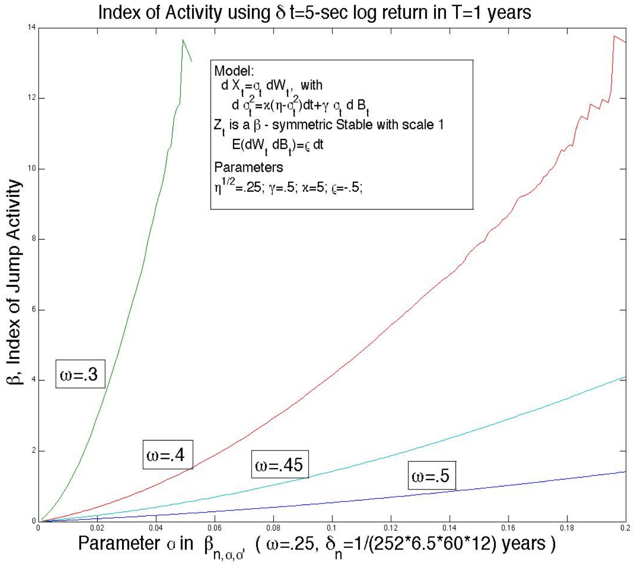


Index of jump activity for Heston model

Stochastic volatility Heston model:

$$R_i^{(\delta)} = b\delta + \sqrt{\delta\, v_{i-1}^{\delta}}\, Z_i,$$
$$v_i^{\delta} = v_{i-1}^{\delta} + \alpha(m - v_{i-1}^{\delta})\delta + \gamma\sqrt{\delta\, v_{i-1}^{\delta}}\, Z_i'.$$

Index of jump activity for Heston model



Jump index for Heston with time-changed NIG jumps

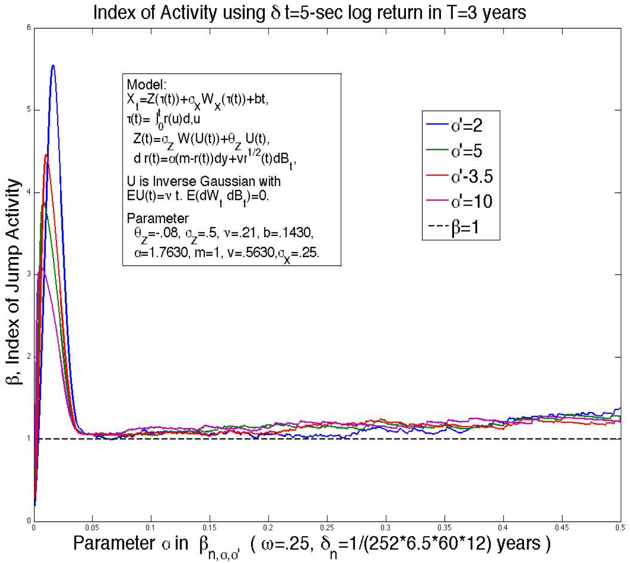
Model:

$$R_i^{(\delta)} = b\delta + \sqrt{\delta\, v_{i-1}^{\delta}}\, Z_i + J_i^{(\delta v_{i-1}^{\delta})},$$

$$v_i^{\delta} = v_{i-1}^{\delta} + \alpha(m - v_{i-1}^{\delta})\delta + \gamma\sqrt{\delta v_{i-1}^{\delta}}\, Z_i',$$

$$\mathrm{Var}(J_i^{(t)}) = t, \quad J_i^{(t)} \overset{\text{i.i.d.}}{\sim} \text{"Heavy tail distribution"}, \quad \beta = 1.$$

Jump index for Heston with time-changed NIG jumps



Jump index for pure time-changed NIG jumps

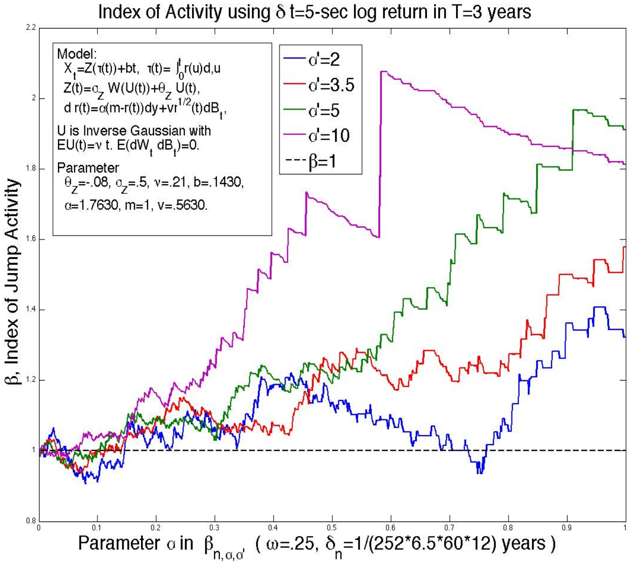
Model:

$$R_i^{(\delta)} = b\delta + J_i^{(\delta v_{i-1}^{\delta})},$$

$$v_i^{\delta} = v_{i-1}^{\delta} + \alpha(m - v_{i-1}^{\delta})\delta + \gamma\sqrt{\delta v_{i-1}^{\delta}}\, Z_i',$$

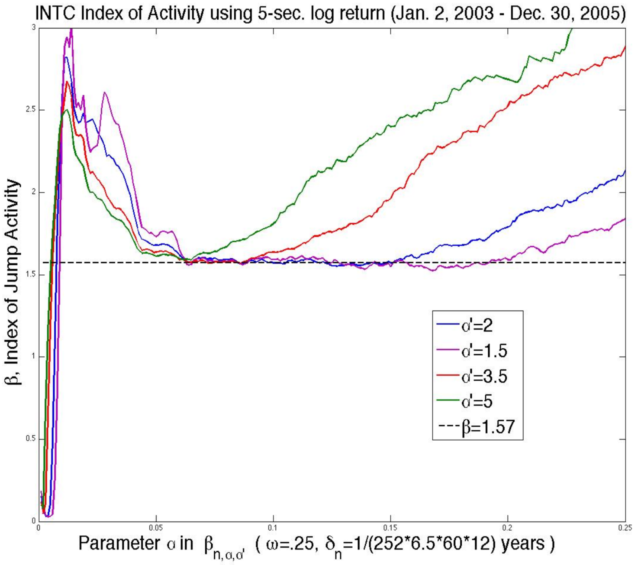
$$\mathrm{Var}(J_i^{(t)}) = t, \quad J_i^{(t)} \overset{\text{i.i.d.}}{\sim} \text{"Heavy tail distribution"}, \quad \beta = 1.$$

Jump index for pure time-changed NIG jumps



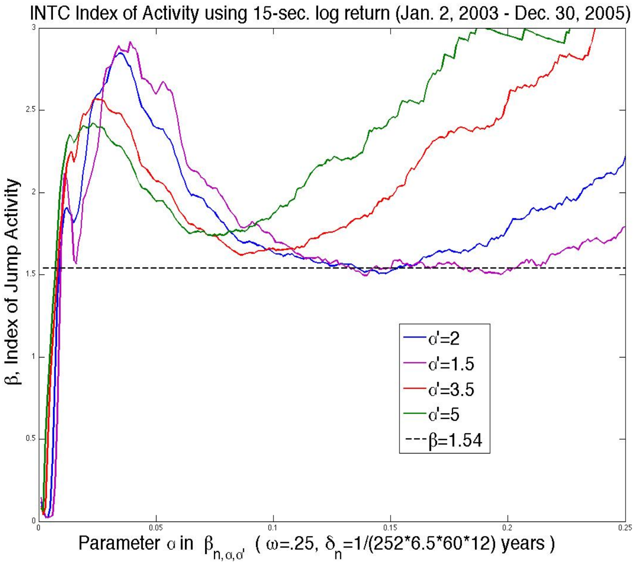
Index of jump activity for INTC with 5-sec returns

Index of jump activity for INTC with 5-sec returns



Index of jump activity for INTC with 15-sec returns

Index of jump activity for INTC with 15-sec returns



Market microstructure features

1. High-frequency estimation depend heavily on an accurate description of the stock price evolution at a very small-time scale.
2. However, real stock prices exhibit several features inherited from the way trading takes place in the market:
   - (i) Nontrading effects
   - (ii) Clustering noise

     e.g. Prices tend to fall more often on whole-dollar multiples than on half-dollar multiples, or than 1/4-dollar multiples, etc.
   - (iii) Bid/ask bounce effect

     Recorded stock prices can be at the bid or at the ask prices. Bid/ask price bouncing creates spurious correlation in returns.
3. One modeling approach: Microstructure noise

Market microstructure features

1. High-frequency estimation depend heavily on an accurate description of the stock price evolution at a very small-time scale.
2. However, real stock prices exhibit several features inherited from the way trading takes place in the market:
   (i) Nontrading effects
   (ii) Clustering noise
   e.g. Prices tend to fall more often on whole-dollar multiples than on half-dollar multiples, or than 1/4-dollar multiples, etc.
   (iii) Bid/ask bounce effect
   Recorded stock prices can be at the bid or at the ask prices. Bid/ask price bouncing creates spurious correlation in returns.
3. One modeling approach: Microstructure noise

Market microstructure features

1. High-frequency estimation depend heavily on an accurate description of the stock price evolution at a very small-time scale.
2. However, real stock prices exhibit several features inherited from the way trading takes place in the market:
   (i) Nontrading effects
   (ii) Clustering noise
   e.g. Prices tend to fall more often on whole-dollar multiples than on half-dollar multiples, or than $1/4$-dollar multiples, etc.
   (iii) Bid/ask bounce effect
   Recorded stock prices can be at the bid or at the ask prices. Bid/ask price bouncing creates spurious correlation in returns.
3. One modeling approach: Microstructure noise

Market microstructure features

1. High-frequency estimation depend heavily on an accurate description of the stock price evolution at a very small-time scale.
2. However, real stock prices exhibit several features inherited from the way trading takes place in the market:
   (i) Nontrading effects
   (ii) Clustering noise
   e.g. Prices tend to fall more often on whole-dollar multiples than on half-dollar multiples, or than $1/4$-dollar multiples, etc.
   (iii) Bid/ask bounce effect
   Recorded stock prices can be at the bid or at the ask prices. Bid/ask price bouncing creates spurious correlation in returns.
3. One modeling approach: Microstructure noise

Market microstructure features

1. High-frequency estimation depend heavily on an accurate description of the stock price evolution at a very small-time scale.
2. However, real stock prices exhibit several features inherited from the way trading takes place in the market:
   (i) Nontrading effects
   (ii) Clustering noise
   e.g. Prices tend to fall more often on whole-dollar multiples than on half-dollar multiples, or than $1/4$-dollar multiples, etc.
   (iii) Bid/ask bounce effect
   Recorded stock prices can be at the bid or at the ask prices. Bid/ask price bouncing creates spurious correlation in returns.
3. One modeling approach: Microstructure noise

Market microstructure features

1. High-frequency estimation depend heavily on an accurate description of the stock price evolution at a very small-time scale.
2. However, real stock prices exhibit several features inherited from the way trading takes place in the market:
   - (i) Nontrading effects
   - (ii) Clustering noise

     e.g. Prices tend to fall more often on whole-dollar multiples than on half-dollar multiples, or than $1/4$-dollar multiples, etc.
   - (iii) Bid/ask bounce effect

     Recorded stock prices can be at the bid or at the ask prices. Bid/ask price bouncing creates spurious correlation in returns.
3. One modeling approach: Microstructure noise

Market microstructure features

1. High-frequency estimation depend heavily on an accurate description of the stock price evolution at a very small-time scale.
2. However, real stock prices exhibit several features inherited from the way trading takes place in the market:
   (i) Nontrading effects
   (ii) Clustering noise
   e.g. Prices tend to fall more often on whole-dollar multiples than on half-dollar multiples, or than $1/4$-dollar multiples, etc.
   (iii) Bid/ask bounce effect
   Recorded stock prices can be at the bid or at the ask prices. Bid/ask price bouncing creates spurious correlation in returns.
3. One modeling approach: Microstructure noise

$$R_i^{\text{Obs}} = R_i^{(\delta)} + \varepsilon_i, \quad \text{where} \quad \varepsilon_i \quad \text{is a stationary sequence.}$$

Market microstructure features

1. High-frequency estimation depend heavily on an accurate description of the stock price evolution at a very small-time scale.
2. However, real stock prices exhibit several features inherited from the way trading takes place in the market:
   (i) Nontrading effects
   (ii) Clustering noise
   e.g. Prices tend to fall more often on whole-dollar multiples than on half-dollar multiples, or than $1/4$-dollar multiples, etc.
   (iii) Bid/ask bounce effect
   Recorded stock prices can be at the bid or at the ask prices. Bid/ask price bouncing creates spurious correlation in returns.
3. One modeling approach: Microstructure noise

$$R_i^{\text{Obs}} = \text{Quantization}(R_i^{(\delta)} + \varepsilon_i), \quad \text{where} \quad \varepsilon_i \quad \text{is a stationary sequence.}$$

Estimation under microstructure “noise"

1. Ultra high-frequency sampling will eventually recover the tick-by-tick data.
2. How frequently to sample?

   The higher sampling frequency, the smaller the theoretical standard error of the estimation methods (under absence of noise), but the higher the microstructure noise.
3. Need to analyze the performance of the methods towards “microstructure noise” (or other kind of noise).
4. Need to devise methods that are robust against marker microstructure noise.
5. There are some models for tick-by-tick data, but the “bridge” between these models and semimartingale models is not well-understood yet.

Estimation under microstructure “noise"

1. Ultra high-frequency sampling will eventually recover the tick-by-tick data.
2. How frequently to sample?
   The higher sampling frequency, the smaller the theoretical standard error of the estimation methods (under absence of noise), but the higher the microstructure noise.
3. Need to analyze the performance of the methods towards “microstructure noise” (or other kind of noise).
4. Need to devise methods that are robust against marker microstructure noise.
5. There are some models for tick-by-tick data, but the “bridge” between these models and semimartingale models is not well-understood yet.

Estimation under microstructure “noise"

1. Ultra high-frequency sampling will eventually recover the tick-by-tick data.
2. How frequently to sample?

   The higher sampling frequency, the smaller the theoretical standard error of the estimation methods (under absence of noise), but the higher the microstructure noise.
3. Need to analyze the performance of the methods towards “microstructure noise” (or other kind of noise).
4. Need to devise methods that are robust against marker microstructure noise.
5. There are some models for tick-by-tick data, but the “bridge” between these models and semimartingale models is not well-understood yet.

Estimation under microstructure “noise"

1. Ultra high-frequency sampling will eventually recover the tick-by-tick data.
2. How frequently to sample?

   The higher sampling frequency, the smaller the theoretical standard error of the estimation methods (under absence of noise), but the higher the microstructure noise.
3. Need to analyze the performance of the methods towards “microstructure noise” (or other kind of noise).
4. Need to devise methods that are robust against marker microstructure noise.
5. There are some models for tick-by-tick data, but the “bridge” between these models and semimartingale models is not well-understood yet.

Estimation under microstructure “noise"

1. Ultra high-frequency sampling will eventually recover the tick-by-tick data.
2. How frequently to sample?

   The higher sampling frequency, the smaller the theoretical standard error of the estimation methods (under absence of noise), but the higher the microstructure noise.
3. Need to analyze the performance of the methods towards “microstructure noise” (or other kind of noise).
4. Need to devise methods that are robust against marker microstructure noise.
5. There are some models for tick-by-tick data, but the “bridge” between these models and semimartingale models is not well-understood yet.

Estimation under microstructure “noise"

1. Ultra high-frequency sampling will eventually recover the tick-by-tick data.
2. How frequently to sample?

   The higher sampling frequency, the smaller the theoretical standard error of the estimation methods (under absence of noise), but the higher the microstructure noise.
3. Need to analyze the performance of the methods towards “microstructure noise” (or other kind of noise).
4. Need to devise methods that are robust against marker microstructure noise.
5. There are some models for tick-by-tick data, but the “bridge” between these models and semimartingale models is not well-understood yet.

Empirical distribution of returns

$$\text{Return during a given time period} = \log \frac{\text{Final price}}{\text{Initial price}}.$$

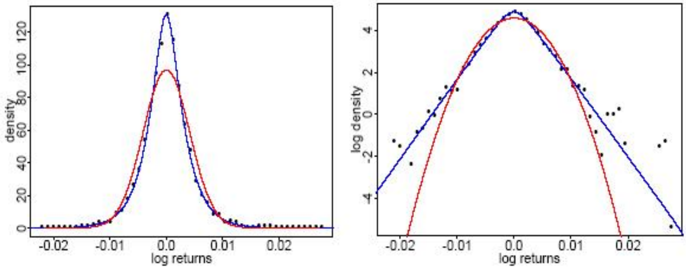


Figure 1: Empirical density of one-hour returns (Bayer) vs. density of fitted hyperbolic (blue) and fitted normal distribution (red).

Back

Dynamics of the price process

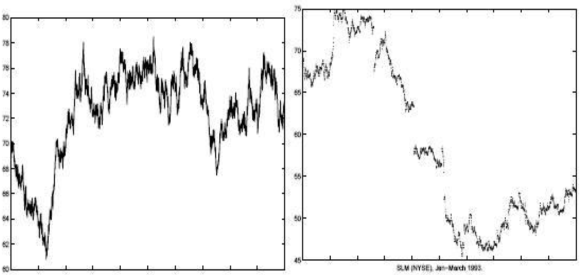
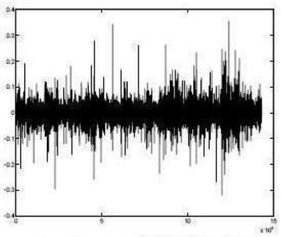


FIGURE 1.2: Evolution of SLM (NYSE), January-March 1993, compared with a scenario simulated from a Black-Scholes model with same annualized return and volatility.

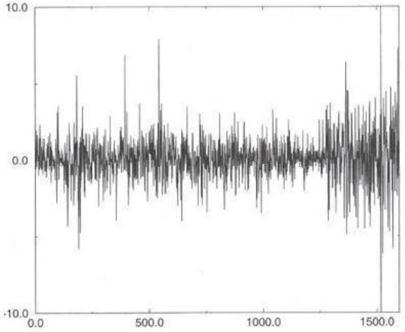
Back

Times series of returns

Figures taken from Cont (2001)



Five-minute log-return for Yen/Deutschemark exchange rate, 1992-1995



BMW daily log-returns

Back