



High-Frequency Data and Volatility in Foreign-Exchange Rates

Author(s): Bin Zhou

Source: *Journal of Business & Economic Statistics*, Vol. 14, No. 1 (Jan., 1996), pp. 45-52

Published by: [American Statistical Association](#)

Stable URL: <http://www.jstor.org/stable/1392098>

Accessed: 14/04/2014 17:39

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at
<http://www.jstor.org/page/info/about/policies/terms.jsp>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



American Statistical Association is collaborating with JSTOR to digitize, preserve and extend access to *Journal of Business & Economic Statistics*.

<http://www.jstor.org>

High-Frequency Data and Volatility in Foreign-Exchange Rates

Bin ZHOU

Sloan School of Management, Massachusetts Institute of Technology, Cambridge, MA 02139

Exchange rates, like many other financial time series, display substantial heteroscedasticity. This poses obstacles in detecting trends and changes. Understanding volatility becomes extremely important in studying financial time series. Unfortunately, estimating volatility from low-frequency data, such as daily, weekly, or monthly observations, is very difficult. The recent availability of ultra-high-frequency observations, such as tick-by-tick data, to large financial institutions creates a new possibility for the analysis of volatile time series. This article uses tick-by-tick foreign-exchange rates to explore this new type of data. Unlike low-frequency data, high-frequency data have extremely high negative first-order autocorrelation in their return. In this article, I propose a model that can explain the negative autocorrelation and a volatility estimator for high-frequency data. The daily and hourly volatility estimates of exchange rate show some interesting patterns.

KEY WORDS: Financial time series; Heteroscedasticity; Tick-by-tick data.

There is considerable literature analyzing the behavior of exchange rates. Structural exchange-rate modeling has not been very successful, however. By studying monthly data, Meese and Rogoff (1983a,b) showed that a random-walk model fits at least as well as more complicated structural models.

Empirical studies such as those by Hsieh (1988) and Diebold (1988) have shown that the distribution of daily returns is approximately symmetric and leptokurtic (i.e., heavy tailed). There are weak autocorrelations in daily returns. One explanation for the heavy-tailed distribution is the hypothesis that data are independently distributed as a normal distribution whose mean and variance change over time (Diebold 1988; Friedman and Vandersteel 1982; Hsieh 1988). The argument for changing variance of returns is simple. The amount of "information flow" that causes changes in prices is not constant over time. Hence, there is no reason to believe that the variance of the price changes is constant over time. Clark (1973) and many others (Mandelbrot and Taylor 1969; Praetz 1972) have argued that observed returns come from a mixture of normal distributions. If the random variable X_t denotes the daily return of the price, the conditional distribution of X_t given information is

$$X_t|\omega_t \sim N(\mu, f(\omega_t)), \quad (1)$$

where ω_t is all the information available at time t . The quantity ω_t could be the number of transactions (Mandelbrot and Taylor 1969) or trading volume (Clark 1973).

The availability of high-frequency data has opened up new possibilities in estimating volatility. Tick-by-tick data provide us with a near-continuous observation of the process and give us the chance to study volatility in great detail. Understanding volatility is the key issue in the conditional heteroscedasticity model (1) and any other financial time series model. This article explores tick-by-tick data and uses the data to estimate volatilities and understand their behaviors.

1. HIGH-FREQUENCY DATA

Because of fast-growing computer power, gathering financial data is easier than ever. Data are no longer recorded daily or weekly. Many large institutions began to collect so-called *tick-by-tick* exchange rates in the early eighties.

In contrast to stock markets, foreign-exchange markets have no geographical location and no "business-hour" limitations. Traders negotiate deals and make exchanges over the telephone. The transaction prices and trading volume are not known to the public. The exchange rates used for most research are the quotes from large data suppliers such as Reuters, Telerate, or Knight Ridder. Any market-maker can submit new quotes to the data suppliers. The quotes are then conveyed to data subscribers' screens. The data suppliers cover the market information worldwide and 24 hours a day. The quotes are intended to be used by market participants as a general indication of where the exchange rate stands, but they do not necessarily represent the actual rate at which transactions are being conducted. It is possible for some participants to manipulate indicative prices occasionally and create a favorable market movement. Because a bank's reputation and credibility as a market-maker emerge from favorable relations with other market participants, however, it is generally felt that these indicative prices closely match the true prices experienced in the market. Readers who are unfamiliar with this type of data in foreign-exchange markets may want to read Goodhart and Figliuoli's (1991) article for details. They studied three days' minute-by-minute exchange rates (the closing tick of a minute) from Reuters. They found that the series exhibited (time-varying) leptokurtosis, unit roots, and first-order negative correlation.

In this study, three exchange rates of tick-by-tick data from October 1, 1992, to September 30, 1993, are used.

© 1996 American Statistical Association
Journal of Business & Economic Statistics
January 1996, Vol. 14, No. 1

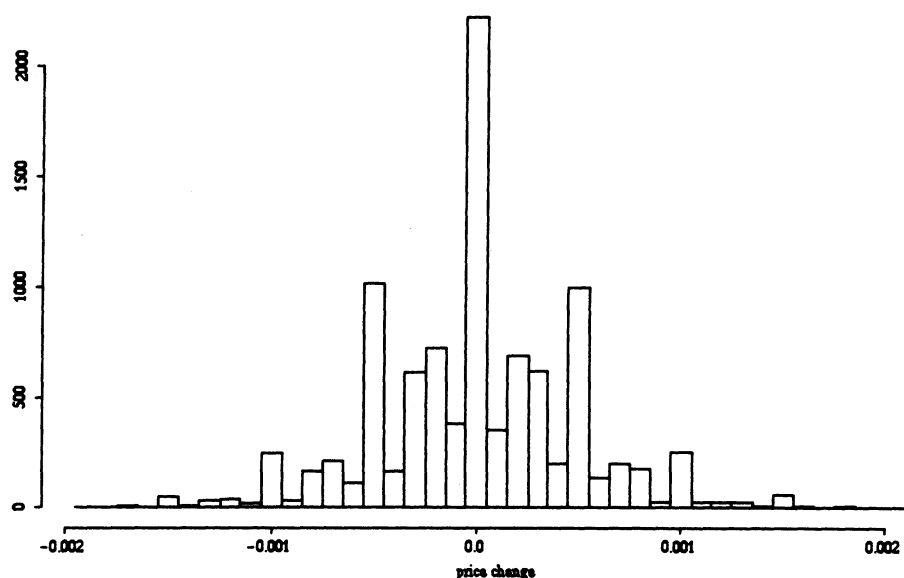


Figure 1. Histogram of DEM/USD Tick-by-Tick Price Change.

The data were provided by Olsen & Associates. They contain spot rate quotes of the Deutsche mark and U.S. dollar (DEM/USD), the Japanese yen and U.S. dollar (JPY/USD), and the Deutsche mark and Japanese yen (JPY/DEM) from Reuters. DEM/USD are the most active currencies traded in the market, followed by JPY/USD. Cross-rates, like JPY/DEM, are much less active. The data sets contain information about time, bid and ask prices, originator of the data, and location. In this article, only the bid prices are used because bid price is quoted in its entirety. Only the last two or three digits of the ask price are quoted. Interpreting ask price by computer is often troublesome. Because banks, in general, are not obligated to trade at the price they quoted on the Reuters screen, there are all kinds of errors in these quotes. Although Reuters uses a validation scheme to improve its data quality, its data still contain many spikes, in which one single quote deviates from general market level. Before the data can be used for statistical analysis, a validation procedure is used to remove these spikes. The validation procedure removes any sudden jumps with a reversal. The detailed program is listed in the Appendix.

The exchange rate is quoted in multiples of one point, with one point equal to .0001 in DEM/USD and .01 in JPY/USD and JPY/DEM. Price change from one tick to the next tick usually ranges from 0 to 20 points. A histogram of 10,000 DEM/USD tick-by-tick price changes is given in Figure 1. The consecutive quotes do not have to be different. A five-point change is the most frequent. Because the price change is often much larger than one point, the problem of discreteness is less severe than in other fi-

nanacial time series. This is the only place that I use the price itself without taking a logarithm transformation.

As an initial exploration, basic statistics of tick-by-tick returns are calculated. Disregarding any time stamp, price changes (logarithm) in consecutive price quotes are used in calculating mean, variance, skewness, kurtosis, first-order autocorrelation, and other statistics. The summary of these statistics is listed in Table 1. The number of ticks in each minute varies greatly. It ranges from 0 to several hundred. The average return of the tick-by-tick data is negligible in comparison to its standard deviation. Positive and negative moves are equally likely. The kurtosis is much higher than 3, the kurtosis of a normal distribution.

Although I expected the slightly negative first-lag autocorrelation that had been reported in other exchange-rate literature, over 40% negative autocorrelation in the tick-by-tick return was a surprise when I first saw the result, especially because I used only the bid price here. The bid-ask bounce cannot be blamed for negative autocorrelation (Roll 1984). To be cautious, I also calculated the autocorrelations of data in every three months. The autocorrelation coefficients were very similar. The negative autocorrelation was consistent. Obviously, high-frequency data does not follow a continuous Brownian motion as low-frequency data is assumed to. Is there any fundamental difference between the high- and low-frequency data?

After spending several days on a trading floor, I concluded that the difference between high- and low-frequency data is the level of noise. The noise is negligible compared to price shift in low-frequency data but becomes very sig-

Table 1. Summary Statistics of Tick-by-Tick Returns

Series	n	Min.	Max.	Mean	S.D.	Skewness	Kurtosis	ρ
DEM/USD	1,472,032	-.0066	.0054	9.92e-8	2.67e-4	.017(.002)	6.87(.004)	-.451
JPY/USD	570,689	-.0105	.0104	-2.18e-7	3.72e-4	-.029(.003)	32.17(.006)	-.425
JPY/DEM	158,958	-.0098	.0100	-1.70e-6	3.50e-4	-.385(.006)	23.26(.012)	-.107

NOTE: ρ is the first-order autocorrelation. Standard errors of skewness and kurtosis are given in parentheses.

nificant in high-frequency data. Validation schemes only remove bad data; they do not clean noises. There is substantial noise remaining in the quotes. For example, there is a fighting-screen effect. To keep their name on the Reuters screen, traders keep updating their quotes. The new update is often slightly different from the previous quotes even if the market level has not changed. Microactivities are another contribution to the noise. Small typographical errors or delayed quotes are all sources of noises. Summarizing these arguments, I propose the following process to describe the exchange rate:

$$S(t) = d(t) + B(\tau(t)) + \varepsilon_t, \quad (2)$$

where $S(t)$ is the logarithm of the exchange rate, $B(\cdot)$ is the standard Brownian motion, both $d(\cdot)$ and $\tau(\cdot)$ are assumed deterministic functions, $\tau(\cdot)$ has positive increments, and ε_t is the mean zero random noise independent of the Brownian motion $B(\cdot)$. The noise, ε_t , is a combination of several sources that were mentioned previously. No distribution assumptions have been made for this noise. It can be a very general stochastic process.

The negative autocorrelation found in high-frequency data is attributed to the noise term in (2). Let $X(s, t) = S(t) - S(s)$, the return of an exchange rate in interval $[s, t]$. Then, following Equation (2),

$$X(s, t) = \mu(s, t) + \sigma(s, t)Z_t + \varepsilon_t - \varepsilon_s, \quad (3)$$

where Z_t is a standard normal random variable, $\sigma^2(s, t) = \tau(t) - \tau(s)$, and $\mu(s, t) = d(t) - d(s)$. The variance of the return is $\text{var}(X(s, t)) = \sigma^2(s, t) + \eta_t^2 + \eta_s^2 - 2c(s, t)$, where $\eta_t^2 = \text{var}(\varepsilon_t)$ and $c(s, t) = \text{cov}(\varepsilon_s, \varepsilon_t)$. When $|s - t|$ increases, $\sigma^2(s, t)$ increases as well. For large $|s - t|$, the variances of the noises become negligible, and so does the noise. $X(s, t)$ behaves just like a random walk. When $|s - t|$ decreases to near 0, $\sigma^2(s, t)$ diminishes. The return, $X(s, t)$, is mainly the difference of the two noises. The sample first-order autocorrelation of such series is about -50% . This indicates that, when we study high-frequency data, the noise component is no longer negligible. An autocorrelation of -45% for the DEM/USD and -42% for JPY/DEM indicates that tick-by-tick price changes for both exchange rates are mostly the difference of the two noise components. JPY/DEM has less negative autocorrelation because it is observed less frequently. The negative autocorrelation increases as observation frequency increases.

There are several difficulties in analyzing the process (2). One of the difficulties is the lack of information about $\tau(t)$, which I call the *cumulative volatility*. $\sigma^2(t - \delta, t) = \tau(t) - \tau(t - \delta)$ is called the δ increment of volatility or simply δ volatility. In Section 2, I will devote my attention to estimating the volatility increment.

Table 2. Estimation of Signal and Noise Level

Series	n	Yearly volume	η^2	Opt. n^*	n/n^*
DEM/USD	1,472,032	.0102	3.21e-8	275,398	5.35
JPY/USD	570,689	.0118	5.88e-8	174,123	3.28
JPY/DEM	158,958	.0153	1.30e-8	1,017,299	0.16

2. VOLATILITY ESTIMATION

In this section, I concentrate on estimating the volatility of a given time interval $[a, b]$ from a series of observations within the interval. The interval can be one day or one hour. The data are the observations within that day or that hour. Let $\sigma^2 = \tau(b) - \tau(a)$, the volatility of the interval, and $\{S(t_i), i = 0, 1, \dots, n\}$ be a series of observations within the interval. Because the mean of tick-by-tick return is negligible compared to the variance, I assume that the observation comes from process:

$$S(t) = B(\tau(t)) + \varepsilon_t. \quad (4)$$

Assuming that $\varepsilon_t, i = 1, \dots, n$, are independent, I propose the following unbiased estimator of the volatility:

$$\hat{\sigma}_U^2 = \sum_{i=1}^n (X_i^2 + X_i X_{i-1} + X_{i+1} X_i), \quad (5)$$

where $X_i = S(t_i) - S(t_{i-1})$.

It is easy to prove that the estimator is unbiased because $E(X_i^2 + X_i X_{i-1} + X_{i+1} X_i) = \tau(t_i) - \tau(t_{i-1})$. The variances of X_i do not have to be equal. The noise variable ε_i is only assumed to be uncorrelated. One may challenge the assumption of uncorrelated noises. Although only bid prices are used here, the fighting-screen effect can cause negative autocorrelation among noise. Such correlation dies down quickly, however. The correlation can be avoided by skipping a few tick points if it poses a problem.

The variance of the estimator (5) is a long formula. To simplify the formula, I assume all variances of X_i to be equal and all variances of ε_i to be equal. These assumptions are used to simplify the formula of the variance. It is not essential to the estimator.

Theorem 1. Let $\text{var}(X_i) = \sigma^2/n$ and $\text{var}(\varepsilon_i) = \eta^2$. The variance of the estimator (5) is

$$\text{var}(\hat{\sigma}_U^2) = \sigma^4 \left(\frac{6}{n} + 8 \frac{\eta^2}{\sigma^2} + 8 \frac{n\eta^4}{\sigma^4} \right) - 2 \frac{\sigma^4}{n^2} - 4\eta^4. \quad (6)$$

Proof.

$$\begin{aligned} \text{var}(\hat{\sigma}_U^2) &= \text{var} \sum_{i=1}^n (X_i^2 + X_{i+1} X_i + X_i X_{i-1}) \\ &= \text{var} \sum_{i=1}^n \left[\frac{\sigma^2}{n} (Z_i^2 + Z_i Z_{i-1} + Z_{i+1} Z_i) \right. \\ &\quad + \frac{\sigma}{\sqrt{n}} (Z_i \varepsilon_{i+1} + Z_i \varepsilon_i - Z_i \varepsilon_{i-1} - Z_i \varepsilon_{i-2} \\ &\quad + Z_{i+1} \varepsilon_i - Z_{i+1} \varepsilon_{i-1} + Z_{i-1} \varepsilon_i - Z_{i-1} \varepsilon_{i-1}) \\ &\quad \times \left. \varepsilon_{i+1} \varepsilon_i - \varepsilon_{i+1} \varepsilon_{i-1} - \varepsilon_i \varepsilon_{i-2} + \varepsilon_{i-1} \varepsilon_{i-2} \right] \\ &= \sigma^4 \left(\frac{6}{n} + 8 \frac{\eta^2}{\sigma^2} + 8 \frac{n\eta^4}{\sigma^4} \right) - 2 \frac{\sigma^4}{n^2} - 4\eta^4. \end{aligned}$$

From (6), it is easy to see that increasing observation frequency infinitely does not improve accuracy of the volatility estimation. There is an optimal observation frequency

Table 3. Estimated One-Year Volatility With Different k

Series	$k = 1$	$k = 2$	$k = 3$	$k = 4$	$k = 5$	$k = 6$	$k = 7$
DEM/USD	.0102	.0144	.0145	.0148	.0148	.0149	.0147
JPY/USD	.0118	.0158	.0153	.0150	.0145	.0142	.0101
JPY/DEM	.0153	.0161	.0165	.0165	.0163	.0162	.0160

that optimizes the variance of $\hat{\sigma}_U^2$. Because σ^2 is the total variance of the price change over interval $[a, b]$, it does not change when the sample frequency n changes. η^2 is also

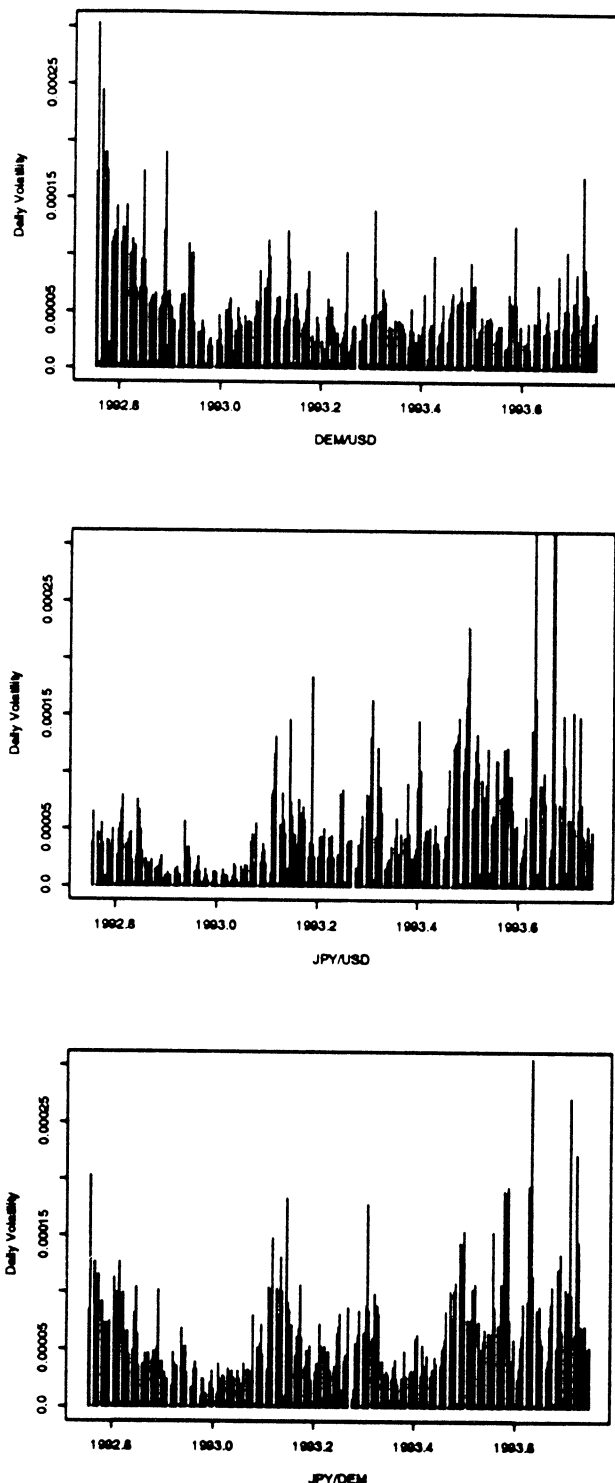


Figure 2. Daily Volatility Estimates of Three Exchange Rates.

independent from observation frequency. From (6), the optimal number of observations n is near $\sqrt{3}\sigma^2/(2\eta^2)$ when η^2/σ^2 is small. The optimal number of observations is proportional to σ^2/η^2 , the signal-to-noise ratio. When there is significant noise in the data, fewer data sometimes are better than more data. To examine if there are too many data in our data set for volatility estimation, I use (6) to give an initial estimate of one year's total volatility. Sample covariance $\text{cov}(X_i, X_{i-1})$ is used as an estimator of $-\eta^2$. The results are listed in Table 2. The results show that both DEM/USD and JPY/USD have more data than we needed in (5). Of course, it is easy to reduce the sample frequency. I can either resample observations at a fixed time interval, say every minute, or every k th tick. Because the price recorded in the fixed time interval is often more heteroscedastic than one recorded in tick flow, I choose the latter.

Let $X_{i,k} = S(i) - S(i-k)$, $i = k, 2k, \dots, n$. Adjusting the number k , one can achieve the desired number of observations needed in (5). I can further improve the estimator (5) by using overlapping returns. This leads to the following volatility estimator:

$$\hat{\sigma}_{U,k}^2 = \frac{1}{k} \sum_{i=1}^n (X_{i,k}^2 + X_{i+k,k}X_{i,k} + X_{i-k,k}X_{i,k}). \quad (7)$$

If $\text{cov}(\varepsilon_i, \varepsilon_{i-k}) = 0$,

$$\begin{aligned} E\hat{\sigma}_{U,k}^2 &= (1/k) \sum_{i=1}^n \sigma_{i,k}^2 \\ &= \sigma^2 + \sum_{i=1}^{k-1} (i/k)(\sigma_{i-k}^2 - \sigma_{n-i+1}^2), \end{aligned}$$

where $\sigma_i^2 = \text{var}(X_i)$. Because k is often much smaller than n , the bias is negligible. This estimator is nearly unbiased. The variance of the estimator is also improved. In fact, it can be proved that

$$\text{var}(\hat{\sigma}^2) \leq \sigma^4 \left(6 \frac{k}{n} + 8 \frac{\eta^2}{k\sigma^2} + 8n \frac{\eta^4}{k^2\sigma^4} \right). \quad (8)$$

The assumption of autocorrelation is now relaxed to $\text{cov}(\varepsilon_i, \varepsilon_{i-k}) = 0$ that, even for moderate k , is more likely to be satisfied.

The volatility estimator presented here has its advantages and disadvantages. It is simple and easy to update when new information comes. It has few assumptions about the noise variables. The disadvantage is that it cannot take full advantage of high-frequency data with optimal observation frequency. The estimator can be improved if more assumptions about the structure of the noise are added.

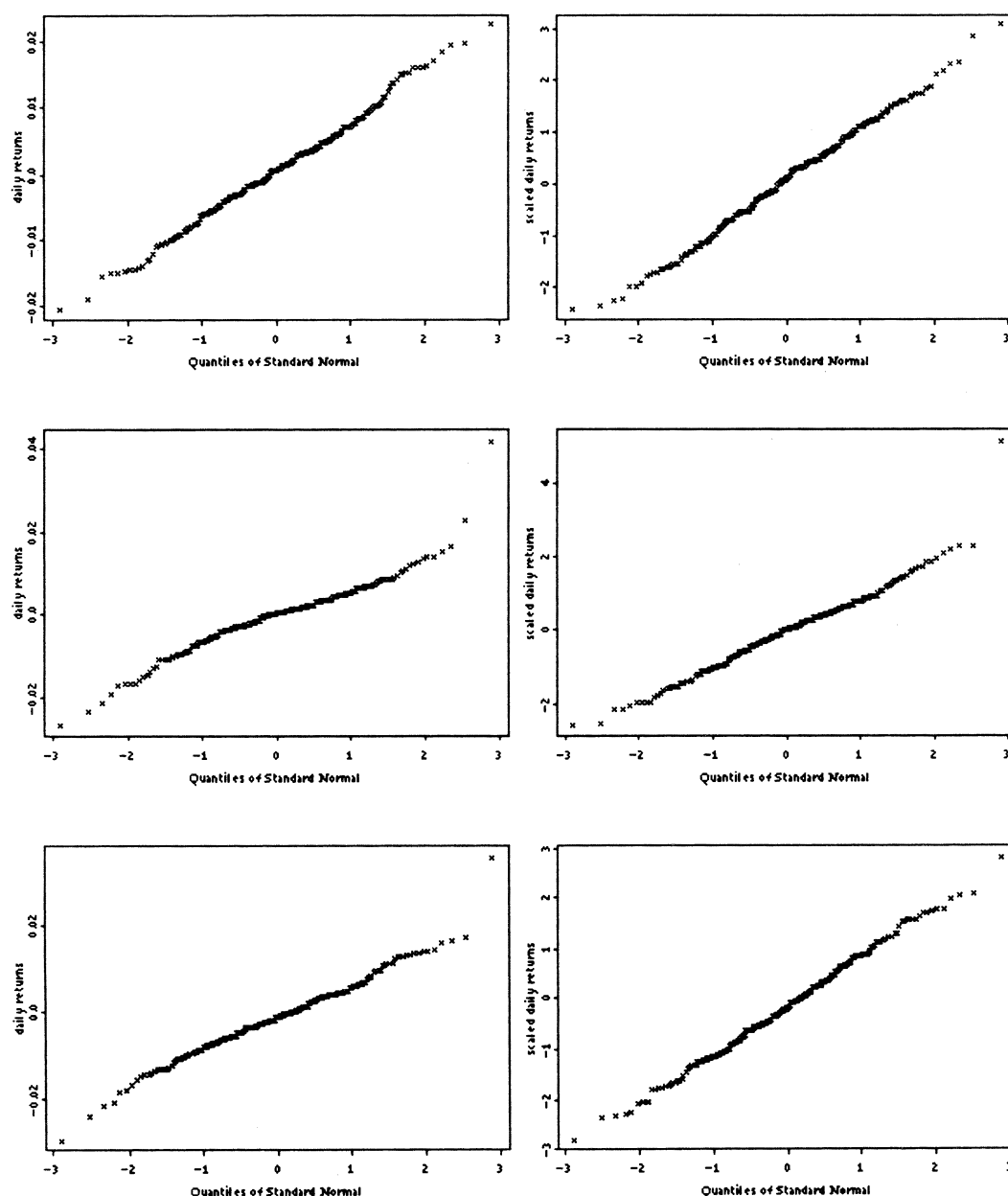


Figure 3. Normal Score Plot of Daily and Standardized Daily Returns.

3. ESTIMATING VOLATILITY OF EXCHANGE RATES

Using volatility estimator (7) and intraday high-frequency data, I now can estimate the volatility of any given period—one year, one day, or one hour. From Table 2, optimal k is around 5 for DEM/USD, 3 for JPY/USD, and 1 for JPY/DEM. However, k also has to be chosen to satisfy the correlation condition $\text{cov}(\varepsilon_i, \varepsilon_{i-k}) = 0$. To determine the value of k , I calculate the volatility of an entire year using different k values. The results are listed in Table 3.

From the proof of Theorem 1, it is not difficult to see that negative autocorrelation causes underestimation. Therefore, from Table 3, DEM/USD and JPY/USD both have very strong negative correlation in their noises. One major reason for that is the fighting-screen effect. To be safe, I choose $k = 5$ for DEM/USD and $k = 3$ for both JPY/USD

and JPY/DEM for the following analysis.

Several times I was asked why the volatility estimate is always higher when high-frequency data is used. The answer is the noise component. The conventional volatility estimator, the sample variance, does not adjust the noise component. For currency, one has to pay attention to the noise component when the frequency of observation is less than an hour.

Next I want to explore the daily volatilities. A day is defined as 24 hours from 0:00 Greenwich mean time. There are, on average, more than 5,000 ticks per day on weekdays. There are far fewer ticks on weekends or holidays. The daily volatility estimates of the three currencies are plotted in Figure 2.

Because the true volatility is unknown, one indirect way to verify the volatility estimates is to test the normality of

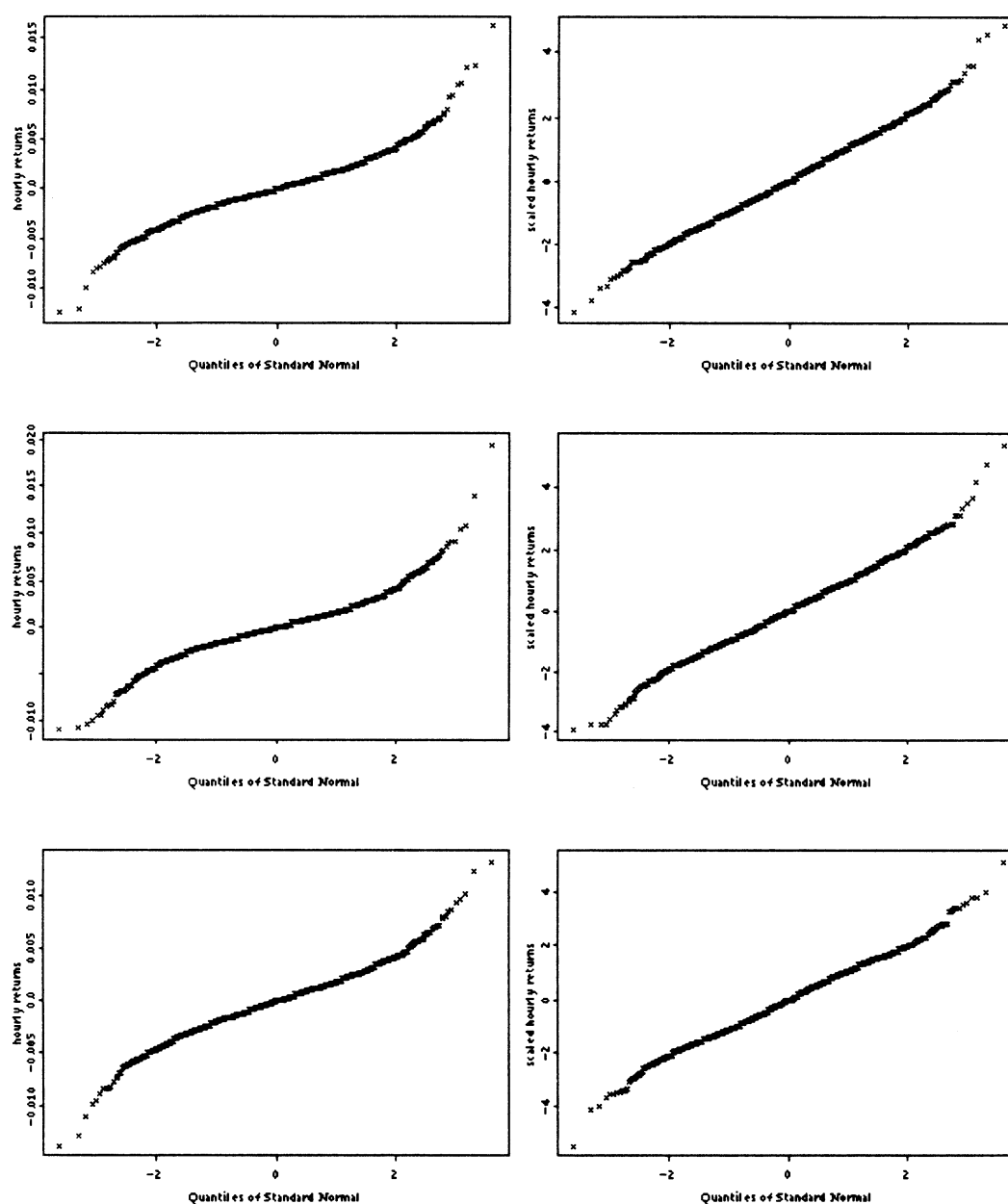


Figure 4. Normal Score Plot of Hourly and Standardized Hourly Returns.

daily returns normalized by its volatility. From process (2), the daily return normalized by the square root of its volatility is the standard normal random variable. Let $Y_i = X_i/\sigma_i$, where X_i and σ_i^2 are the daily return and daily volatility. Excluding all zero volatility estimates (all but one on Saturdays), the normal score plots for both return X_i and standardized return Y_i are given in Figure 3.

Because the daily return for this set of data is itself fairly normally distributed, the comparison is not dramatic. I therefore repeat the same comparison for hourly returns. Excluding all low-volatility estimates (less than $1e-6$), the normal score plots for both hourly returns X_i and standardized hourly returns Y_i are given in Figure 4. The basic statistics of both X_i and Y_i are shown in Table 4.

Table 4. Basic Statistics of Daily Returns

Series		<i>n</i>	Mean	Var.	Skewness	Kurtosis
DEM/USD	<i>X</i>	3,524	4.10e-5	3.99e-6	.272	7.24
	<i>X/σ</i>	3,524	1.80e-2	1.02e+0	.096	3.28
JPY/USD	<i>X</i>	3,285	-3.82e-5	4.13e-6	.314	9.15
	<i>X/σ</i>	3,285	-1.55e-2	1.03e+0	.114	3.63
JPY/DEM	<i>X</i>	3,452	-7.63e-5	4.65e-6	-.064	5.72
	<i>X/σ</i>	3,452	-3.05e-2	1.15e+0	-.014	3.27

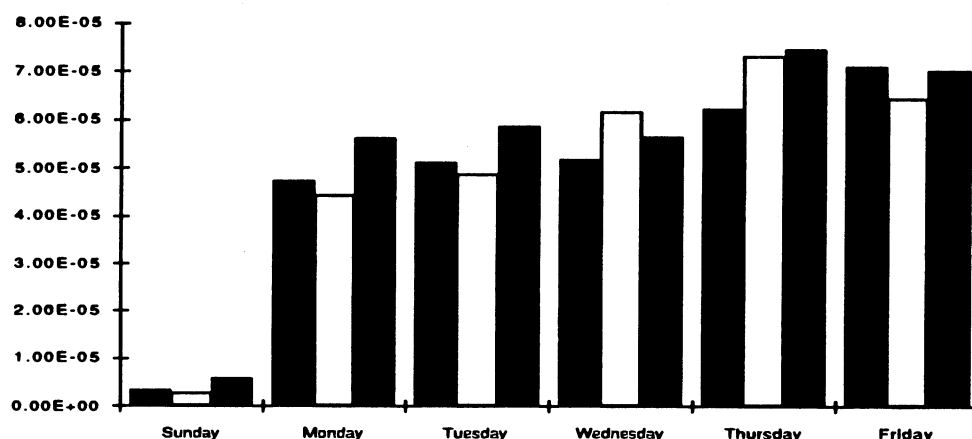


Figure 5. Average Daily Volatilities: First Black Column, DEM/USD; White Column, JPY/USD; Second Black Column, JPY/DEM.

All statistics in Table 4 indicate that the normalized return is close to a standard normal random variable. Although this does not prove the accuracy of the volatility estimation, this is a feature that cannot be achieved by using other estimators with low-frequency data.

There are many explanations for heavy-tailed distribution of returns. This part of the analysis indicates that the heavy-tailed distribution is mainly caused by changing volatility. Therefore, I have more confidence in using Process (2) to describe the financial time series.

To study the day-of-the-week effect and the hour-of-day effect, average daily volatility for the week and average hourly volatility for the day are plotted in Figures 5 and 6. All three exchange rates have rising volatility as the week progresses. This pattern has also been found in other years. The reason for this pattern needs to be further investigated. Hourly volatility has the most understandable pattern. The Deutsche mark and the U.S. dollar are most active during New York morning hours, and the Japanese yen is most active during Tokyo morning hours. New York late hours and the Tokyo lunch hour have low volatility. Every day has three waves of volatility. The first one peaks in Tokyo

early morning. The second one peaks at the opening hour of the London market, and the last one peaks at the opening hour of the New York market. On the whole, the New York market is the most active market.

4. CONCLUSION AND DISCUSSION

High-frequency data are extremely noisy. The noise causes a strong negative first-order autocorrelation in tick-by-tick returns. The autocorrelation decreases as frequency decreases because the role of the noise is reduced. A volatility estimator using high-frequency data has to adjust to such noises. The advantage of using high-frequency data to estimate volatility is that volatility of any reasonable length can be estimated without delay. The volatility estimator presented in this article needs few distribution assumptions. It can be used in tracking market volatility dynamically because it is easy to update. One disadvantage is that the estimator is not consistent when sample frequency increases and time span is fixed. Further improvement can be made.

When daily and hourly returns are standardized by the volatility estimate (square root), they are very close to the standard normal random variable. This indicates that the

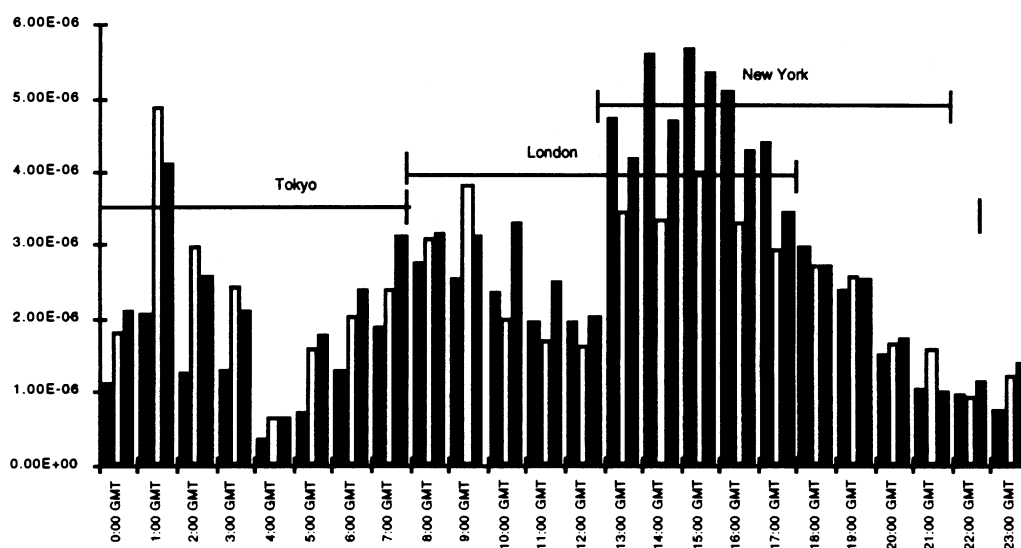


Figure 6. Average Hourly Volatilities: First Black Column, DEM/USD; White Column, JPY/USD; Second Black Column, JPY/DEM.

heavy-tailed distribution of price change is mostly caused by changing volatility. Process (2) can reasonably well describe the high-frequency exchange rates.

The methodology introduced here can independently estimate daily or hourly volatilities. This allows one to study the volatility process itself further. For example, one can study the distribution of the daily volatility, the dynamic structure of the daily process. One study (Xi 1993) indicated that daily volatility is approximately log-normally distributed with strong positive autocorrelation. One can further apply autoregressive integrated moving average (ARIMA) analysis to daily volatilities and possibly forecast the volatility. Compared to estimating autoregressive conditional heteroscedasticity coefficients, estimating ARIMA coefficients using daily volatility estimates is much easier and possibly more fruitful. The volatility estimates can also be used in many other ways such as studying relations between the market volatility and market price movement.

ACKNOWLEDGMENTS

I thank Professor David Donoho for his advice and helpful comments. I also thank anonymous referees who provided helpful suggestions. This research is partially supported by MIT International Financial Service Research Center, and the data was provided by Olsen & Associates.

APPENDIX: VALIDATION PROGRAM

There are many reasons to have outliers in the high-frequency exchange-rate quotes. The quotes are nonbinding and typed in by people. There are unavoidable noises and keying errors. The validation program removes outliers but

does not change any recorded quote. A quote is validated by comparing it to the medians of three neighboring points on both sides. The left three points have been validated and the right three have not yet been validated. When a quote is 25 points above or below both neighboring medians, it is removed. One point is equal to .0001 in DEM/USD and .01 in JPY/USD and JPY/DEM.

[Received March 1993. Revised July 1995.]

REFERENCES

- Clark, P. K. (1973), "A Subordinate Stochastic Process Model With Finite Variance for Speculative Price," *Econometrica*, 41, 135-155.
- Diebold, F. X. (1988), *Empirical Modeling of Exchange Rate Dynamics*, New York: Springer-Verlag.
- Friedman, D., and Vandersteel, S. (1982), "Short-run Fluctuation in Foreign Exchange Rates," *Journal of International Economics*, 13, 171-186.
- Goodhart, C. A. E., and Figliuoli, L. (1991), "Every Minute Counts in Financial Markets," *Journal of International Money and Finance*, 10, 23-52.
- Hsieh, D. A. (1988), "The Statistical Properties of Daily Foreign Exchange Rates: 1974-1983," *Journal of International Economics*, 24, 129-145.
- Mandelbrot, B., and Taylor, H. (1969), "On the Distribution of Stock Price Differences," *Operations Research*, 15, 1057-1062.
- Meese, R. A., and Rogoff, K. (1983a), "Empirical Exchange Rate Models of the Seventies: Do They Fit Out of Sample?" *Journal of International Economics*, 14, 3-24.
- (1983b), "The Out of Sample Failure of Empirical Exchange Rate Models: Sampling Error or Misspecification?" in *Exchange Rates And International Microeconomics*, ed. J. Frenkel, Chicago: University of Chicago Press, pp. 67-112.
- Praetz, P. D. (1972), "The Distribution of Share Price Changes," *Journal of Business*, 45, 49-55.
- Roll, R. (1984), "A Simple Implicit Measure of the Effective Bid-Ask Spread in an Efficient Market," *Journal of Finance*, 39, 1127-1159.
- Xi, G. (1993), "An Exploratory Study of Stock Price Behavior and Volatility Estimation Using High Frequency Data," unpublished thesis, Massachusetts Institute of Technology, Sloan School of Management.