The Error of Statistical Volatility of Intra-daily Quoted Price Changes Observed over a Time Interval

ULRICH A. MÜLLER, MICHEL M. DACOROGNA,
OLIVIER V. PICTET

UAM.1995-07-31

August 8, 1995

A discussion paper by the O&A Research Group

The purpose of this paper is to promote discussions and the exchange of ideas. Comments on the concepts presented are gratefully received.

Contents

Abstract

For a proper computation of the scaling law exponent from intra-daily quoted prices, it is essential to discuss and estimate the different type of errors present when computing the volatility.

We present here the derivation of the error of a statistical average of an absolute or squared price change observed over a certain time interval. The sources of error are of two types: the conventional statistical error due to the number of observations and a measurement error due to the definition of the middle price which contains a fundamental uncertainty caused by the spread.

1 Motivation

We have shown in (Müller et al., 1990) that the average absolute price change, $\overline{|\Delta x|}$, as well as the root of the mean squared price changes, $(\overline{\Delta x^2})^{1/2}$, follows a scaling law as a function of the time interval on which this quantity is measured. The parameters of this law seem very stable (see (Guillaume et al., 1994)) but depend on the way the statistical quantities are computed and on the errors that enter the evaluation through the least square fit of the scaling law parameters. In (Müller et al., 1990), we briefly mention the problem but in order to help people reproduce our results, we give here the full derivation of the error.

When making statistical studies of price changes or return, people usually consider only one source of errors: the usual statistical error due to the limited number of observation. This error is clearly dominant when the return is measured over time intervals of a day or more. When the time interval is reduced to few minutes, however, the uncertainty on the price definition due to the spread must be also considered. The market makers are biased towards one of the two prices, either the bid or the ask, thus introducing a bouncing effect that reflects in a negative autocorrelation of the price changes in the very short term (Goodhart and Figliuoli, 1991; Guillaume et al., 1994). The true market price is between the bid and the ask quotes but not necessarily in their exact midpoint (Müller et al., 1990; Bollerslev and Domowitz, 1993). This uncertainty can be assessed to a considerable fraction of the nominal bid-ask spread. For short horizons, the amplitudes of price movements become comparable to the size of the spread. The uncertainty can thus imply a important measurement error.

The purpose of this short note is to derive the error on the statistical quantity entering the scaling law computation when the measurement error is also taken into account.

2 Few definitions and assumptions

We start from few definitions of the important variables (some of them can be found in (Guillaume et al., 1994) together with a discussion of their meaning):

- the price

$$x(t_j) \equiv [\log p_{ask}(t_j) + \log p_{bid}(t_j)] / 2 \tag{2.1}$$

where t_j is the sequence of the tick recording times which is unequally spaced. An alternative notation is

$$x(t_i) \equiv x(\Delta t, t_i) \equiv x_i \equiv [\log p_{ask}(t_i) + \log p_{bid}(t_i)] / 2 \tag{2.2}$$

where t_i is the sequence of the regular spaced in time data separated by the time interval Δt. To obtain the price at time t_i an interpolation method is needed. We usually linearly interpolate between two neighboring prices (Müller et al., 1990).

- the price change or return

$$r(t_i) \equiv r(\Delta t; t_i) \equiv r_i \equiv x(t_i) - x(t_i - \Delta t) = x_i - x_{i-1} \tag{2.3}$$

where x_{i-1} and x_i are two consecutive elements in the homogeneous time series.

- the average absolute value of the return which we also often call volatility:

$$v(t_i) \equiv v(\Delta t, S; t_i) \equiv \overline{|\Delta x|} \equiv \frac{1}{n} \sum_{k=0}^{n-1} |r(\Delta t; t_{i-k})| \tag{2.4}$$

where $S = n\Delta t$ is the sample period on which the volatility is computed.

- the relative spread

$$s(t_i) \equiv \log p_{ask}(t_i) - \log p_{bid}(t_i) \tag{2.5}$$

Using these definitions, we can write the empirical scaling law (Müller et al., 1990) as follows,

$$v(\Delta t, S; t_i) = \overline{|\Delta x|} = \left(\frac{\Delta t}{\Delta T} \right)^{\frac{1}{E}} \tag{2.6}$$

This empirical law is well fulfilled for sufficiently large samples S. ΔT and the drift exponent E are constants depending on the FX rate. The law relates the volatility over a time interval Δt to the size of this interval in time.

The scaling law is empirically computed by fitting its logarithmic form,

$$\log \overline{|\Delta x|} = \frac{1}{E} (\log \Delta t - \log \Delta T) \tag{2.7}$$

The law becomes linear in this form. For the linear regression, we need to know the errors of $\log \overline{|\Delta x|}$.

A similar scaling law is valid for $(\overline{\Delta x^2})^{1/2}$ instead of $\overline{|\Delta x|}$. We define

$$\overline{\Delta x^2} \equiv \frac{1}{n} \sum_{k=0}^{n-1} r_{i-k}^2 \tag{2.8}$$

analogous to eq. 2.4. The scaling law for $(\overline{\Delta x^2})^{1/2}$ is

$$\log \overline{\Delta x^2}^{1/2} = \frac{1}{E'} (\log \Delta t - \log \Delta T') \tag{2.9}$$

3 The error of mean squared price changes

The problem is to find the error on $\overline{|\Delta x|}$ knowing that we have an uncertainty related to the price definition and to the spread. Expressions with absolute values such as $\overline{|\Delta x|}$ are known for their poor analytical tractability. Therefore, the whole error computation is done for the analogous case of $(\overline{\Delta x^2})^{1/2}$.

Following the arguments given in the introduction, let us assume that x_i^* is the series of true logarithmic market prices whereas the observed middle values x_i as defined by eq. 2.2 are subject to an additional market maker bias ε_i:

$$x_i \;=\; x_i^* + \varepsilon_i \tag{3.1}$$

The true return is defined analogous to eq. 2.3:

$$r_i^* \;\equiv\; r^*(t_i) \;\equiv\; x_i^* - x_{i-1}^* \tag{3.2}$$

Its relation to the observed return follows from eqs. 2.3 and 3.1:

$$r_i \;=\; r_i^* + \varepsilon_i - \varepsilon_{i-1} \tag{3.3}$$

To compute the error, a minimum knowledge on the distribution of the stochastic quantities is required. We know that the returns r_i and r_i^* follow a Gaussian distribution only as a very crude approximation, see (Guillaume et al., 1994), and the market maker bias ε_i might also be non-normally distributed. Nevertheless, we shall assume Gaussian distributions as approximations to make the problem analytically tractable:

$$r_i^* \;\in\; N(0, \varrho^{*2}) \tag{3.4}$$

$$\varepsilon_i \;\in\; N(0, \frac{\eta^2}{2}) \tag{3.5}$$

We approximate the standard deviation of the market maker bias by 1/4 of a typical value of the relative spread (equation 2.5); this means that η^2 is assumed to 1/8 of the squared relative spread[1]. Studies with transaction prices have shown that the “true” spread is very different from the quoted spread (Goodhart et al., 1994). This quantity is a kind of convention, the market maker is really interested in one of the bid or ask price and adds or subtracts a canonical value to the price he wants to use. In normal market conditions, the price is settled with an equal distribution of buyers and sellers. Thus, in a first approximation, the two random variables, the true returns and the market maker bias as it appears in quoted prices, can well be assumed to be independent.

Now, we are ready to compute the expectation of r_i^2 from eq. 3.3, using eqs. 3.4 and 3.5 and the independence of r_i^* and ε_i:

$$\varrho^2 \;\equiv\; \mathsf{E}(r_i^2) \;=\; \mathsf{E}(\overline{\Delta x^2}) \;=\; \varrho^{*2} + \eta^2 \tag{3.6}$$

[1] We neglect spread changes as our assumptions are anyway too fuzzy to take more than an estimation for the measurement error.

The squared observed returns are thus biased by the positive amount of η^2.

Empirical measures of $\overline{\Delta x^2}$ are not only biased but contain also a stochastic error which is defined as the deviation of $\overline{\Delta x^2}$ from its expectation ϱ^2. The variance of this stochastic error can be formulated:

$$\sigma^2 \equiv \mathrm{E}[(\overline{\Delta x^2} - \varrho^2)^2] = \mathrm{E}(\overline{\Delta x^2}^2 - 2\,\overline{\Delta x^2}\,\varrho^2 + \varrho^4) \tag{3.7}$$

The last form of this equation has the expanded terms of the square. The first term, $\overline{\Delta x^2}^2$, can be explicitly written by inserting eqs. 2.8 and 3.3; the other two terms can be simplified by inserting eq. 3.6. We obtain

$$\sigma^2 = \mathrm{E}\{[\frac{1}{n}\sum_{i=1}^{n}(r_i^* + \varepsilon_i - \varepsilon_{i-1})^2]^2\} - \varrho^4 \tag{3.8}$$

The first term is somewhat tedious to compute because of the two squares and the sum. We expand the squares to get many terms for which we have to compute the expectation values. All those terms that contain r^* or ε to an odd power have a zero expectation due to the symmetry of the normal distribution and the independence of r^* and ε. The expectations of r^{*2} and ε^2 can be taken from eqs. 3.4 and 3.5. The expectations of the fourth moments of normal distribution are

$$\mathrm{E}(r_i^{*4}) = 3\,[\mathrm{E}(r_i^{*2})]^2 = 3\,\varrho^{*4} \tag{3.9}$$

$$\mathrm{E}(\varepsilon_i^4) = \mathrm{E}(\varepsilon_{i-1}^4) = 3\,[\mathrm{E}(\varepsilon_i^2)]^2 = \frac{3}{4}\,\eta^4 \tag{3.10}$$

as to be found in (Kendall et al., 1987) (p. 321 and 338), for example. By inserting this and carefully evaluating all the terms, we obtain

$$\sigma^2 = \frac{n+2}{n}\,\varrho^{*4} + \frac{2n+2}{n}\,\varrho^{*2}\,\eta^2 + \frac{n+2+\frac{n-1}{n}}{n}\,\eta^4 - \varrho^4 \tag{3.11}$$

By inserting eq. 3.6, we can express the resulting stochastic error variance either in terms of ϱ^*,

$$\sigma^2 = \frac{2}{n}\,\varrho^{*4} + \frac{2}{n}\,\varrho^{*2}\,\eta^2 + (\frac{3}{n} - \frac{1}{n^2})\,\eta^4 \tag{3.12}$$

or in terms of ϱ,

$$\sigma^2 = \frac{2}{n}\,\varrho^4 - \frac{2}{n}\,\varrho^2\,\eta^2 + (\frac{3}{n} - \frac{1}{n^2})\,\eta^4 \tag{3.13}$$

Now, we know both the bias η^2 of an empirically measured $\overline{\Delta x^2}$ and the variance of its stochastic error. For reporting the results and using them in the scaling law computation, two alternative approaches are possible:

1. We can subtract the bias η^2 from the observed $\overline{\Delta x^2}$ and take the result with a stochastic error of a variance following eq. 3.13, approximating ϱ^2 by $\overline{\Delta x^2}$. We do not recommend this because η^2 is only approximately known and thus contains another unknown error.

2. We can take the originally obtained value of $\overline{\Delta x^2}$ and regard the bias η^2 as a separate error component in addition to the stochastic error. This is the appropriate way to go, given the uncertainty of η^2.

Following the second approach, we formulate a total error with variance σ^2_{total}, containing the bias and the stochastic error. The stochastic error is independent of the bias by definition, so the total error variance is the sum of the stochastic variance and the squared bias:

$$\sigma^2_{\text{total}} \;=\; \sigma^2 + \eta^4 \;=\; \frac{2}{n}\,\varrho^4 - \frac{2}{n}\,\varrho^2\,\eta^2 + (1+\frac{3}{n}-\frac{1}{n^2})\,\eta^4 \tag{3.14}$$

This is the final, resulting variance of the total error of $\overline{\Delta x^2}$.

For the application in the scaling law, we can use a good approximation for large values of n, as to be shown. For this, we introduce the auxiliary variable

$$c \;=\; \sqrt{\frac{n}{2}}\,\frac{\eta^2}{\varrho^2} \tag{3.15}$$

and reformulate eq. 3.14:

$$\sigma^2_{\text{total}} \;=\; \frac{2}{n}\,\varrho^4\,[1 - \sqrt{\frac{2}{n}}\,c + (1+\frac{3}{n}-\frac{1}{n^2})\,c^2] \tag{3.16}$$

Using the relation

$$c \;<\; 1 + c^2 \tag{3.17}$$

we obtain

$$\sigma^2_{\text{total}} \;=\; \frac{2}{n}\,\varrho^4\,\{1+O(n^{-1/2}) + [1+O(n^{-1/2})]\,c^2\} \tag{3.18}$$

For large time intervals of a scaling law study, c is very small, so we can anyway neglect all the terms of eq. 3.16 that contain c. For small time intervals on the other hand, the number n of observations becomes large and we can neglect all higher order terms of the asymptotic expansion against n. By dropping the higher order terms and re-inserting eq. 3.15, the error variance becomes

$$\sigma^2_{\text{total}} \approx \frac{2}{n}\,\varrho^4 + \eta^4 \;\approx\; \frac{2}{n}\,\overline{\Delta x^2}^2 + \eta^4 \tag{3.19}$$

In the last form, the theoretical constant ϱ^2 has been replaced by its estimator $\overline{\Delta x^2}$, see eq. 3.6.

The mean squared return with error can be formulated as follows:

$$\overline{\Delta x^2}_{\text{with error}} \;=\; \overline{\Delta x^2} \pm \sqrt{\eta^4 + \frac{2}{n}\,\overline{\Delta x^2}^2} \tag{3.20}$$

where the second term is the standard deviation of the error according to eq. 3.19.

The scaling law is usually formulated for $(\overline{\Delta x^2})^{1/2}$ rather than $\overline{\Delta x^2}$. Applying the law of error propagation, we obtain:

$$\overline{\Delta x^2}^{1/2}_{\text{with error}} = \overline{\Delta x^2}^{1/2} \pm \frac{\mathrm{d}\overline{\Delta x^2}^{1/2}}{\mathrm{d}\overline{\Delta x^2}} \sqrt{\eta^4 + \frac{2}{n}\,\overline{\Delta x^2}^2} = \overline{\Delta x^2}^{1/2} \pm \sqrt{\frac{\eta^4}{4\,\overline{\Delta x^2}} + \frac{\overline{\Delta x^2}}{2\,n}} \quad (3.21)$$

The scaling law fitting is done in the linear form obtained for $\log(\overline{\Delta x^2})^{1/2}$, see eq. 2.9. Again applying the law of error propagation, we obtain:

$$\begin{aligned} \log \overline{\Delta x^2}^{1/2}_{\text{with error}} &= \log \overline{\Delta x^2}^{1/2} \pm \frac{\mathrm{d}\log \overline{\Delta x^2}^{1/2}}{\mathrm{d}\overline{\Delta x^2}^{1/2}} \sqrt{\frac{\eta^4}{4\,\overline{\Delta x^2}} + \frac{\overline{\Delta x^2}}{2\,n}} \\ &= \log \overline{\Delta x^2}^{1/2} \pm \sqrt{\frac{\eta^4}{4\,\overline{\Delta x^2}^2} + \frac{1}{2\,n}} \end{aligned} \quad (3.22)$$

which gives rise to the following expression for the error variance of this quantity:

$$\mathrm{Var}(\log \overline{\Delta x^2}^{1/2}) = \frac{\eta^4}{4\,\overline{\Delta x^2}^2} + \frac{1}{2\,n} \quad (3.23)$$

4 The error of mean absolute price changes

The assumption is now that the variance of the error for $\log \overline{|\Delta x|}$ is approximately the same as that of $\log(\overline{\Delta x^2})^{1/2}$ in equation (3.23) and that we only need to replace there the empirically obtained $(\overline{\Delta x^2})^{1/2}$ by the empirically obtained $\overline{|\Delta x|}$. This approximation is justified by the similar sizes and behaviors of both quantities. We obtain

$$\mathrm{Var}(\log \overline{|\Delta x|}) = \frac{\eta^4}{4\,\overline{|\Delta x|}^4} + \frac{1}{2\,n} \quad (4.1)$$

This expression has interesting properties. In the case of long time intervals, $\overline{|\Delta x|} \gg \eta$ and the term $1/(2n)$ becomes the essential cause of errors. In the case of short time intervals, n is very big but $\overline{|\Delta x|}$ is of the same order as η and the first term of the right hand side of the equation plays the central role.

References

Bollerslev T. and Domowitz I., 1993, Trading patterns and prices in the interbank foreign exchange market, The Journal of Finance, 48, 1421–1443.

Goodhart C., Ito T., and Payne R., 1994, One day in june, 1993: a study of the working of the reuters dealing 2000-2 electronic foreign exchange trading system, mimeo, London School of Economics, 1–36.

Goodhart C. A. E. and Figliuoli L., 1991, Every minute counts in financial markets, Journal of International Money and Finance, 10, 23–52.

Guillaume D.M., Dacorogna M.M., Davé R.D., Müller U.A., Olsen R.B., and Pictet O.V., 1994, From the bird's eye to the microscope: A survey of new stylized facts of the intra-daily foreign exchange markets, Internal document DMG.1994-04-06, Olsen & Associates, Seefeldstrasse 233, 8008 Zürich, Switzerland.

Kendall M., Stuart A., and Ord J.K., 1987, Advanced Theory of Statistics, volume 1, Charles Griffin & Company Limited, London, fifth edition.

Müller U.A., Dacorogna M.M., Olsen R.B., Pictet O.V., Schwarz M., and Morgenegg C., 1990, Statistical study of foreign exchange rates, empirical evidence of a price change scaling law, and intraday analysis, Journal of Banking and Finance, 14, 1189–1208.