Ranking and Combining Volatility Proxies for Garch and Stochastic Volatility Models

Visser, Marcel P.

Korteweg-de Vries Instute for Mathematics, University of Amsterdam

9 October 2008

Ranking and Combining Volatility Proxies for Garch and Stochastic Volatility Models

Robin G. de Vilder* Marcel P. Visser †

October 9, 2008

*PSE (CNRS-EHESS-ENPC-ENS), Paris, France and Korteweg-de Vries Institute for Mathematics, University of Amsterdam, The Netherlands.

†Corresponding author, Korteweg-de Vries Institute for Mathematics, University of Amsterdam. Plantage Muidergracht 24, 1018 TV Amsterdam, The Netherlands. Tel. +31 20 5255861. Email: m.p.visser@uva.nl.

Abstract

Daily volatility proxies based on intraday data, such as the high-low range and the realized volatility, are important to the specification of discrete time volatility models, and to the quality of their parameter estimation. The main result of this paper is a simple procedure for combining such proxies into a single, highly efficient volatility proxy. The approach is novel in optimizing proxies in relation to the scale factor (the volatility) in discrete time models, rather than optimizing proxies as estimators of the quadratic variation. For the S&P 500 index tick data over the years 1988–2006 the procedure yields a proxy which puts, among other things, more weight on the sum of the highs than on the sum of the lows over ten-minute intervals. The empirical analysis indicates that this finite-grid optimized proxy outperforms the standard five-minute realized volatility by at least 40%, and the limiting case of the square root of the quadratic variation by 25%.

JEL classification: C22, C52, C65, G1.

Key Words: volatility proxy, realized volatility, quadratic variation, scale factor, arch/garch/stochastic volatility, intraday seasonality, variance of logarithm.

Much of the understanding of financial asset price volatility has to be deduced from volatility proxies, as volatility itself is inherently unobservable. Proxies such as the intraday high-low range or the realized volatility are important objects for modelling financial asset prices and volatility. Good proxies increase forecast accuracy and improve parameter estimation for discrete time volatility models. So the search for optimal proxies is beneficial to topics central to financial economics, such as portfolio allocation, pricing financial instruments, and risk management.

Garch and stochastic volatility models are standard tools for the time series analysis of daily volatility. This paper is novel in analyzing proxies in relation to these *discrete time* models. It addresses the problem of optimizing volatility proxies when intraday high-frequency data are available. In an ideal world, with a continuously observed asset price process in a frictionless market, a first natural candidate for a proxy would be the (square root of the) quadratic variation. The daily quadratic variation is the limit of the realized variance[1] as the lengths of the sampling intervals approach zero, see for instance Andersen, Bollerslev, Diebold, and Labys (2001). However, in discrete time models the volatility is

[1]One obtains the realized variance by summing the squared intraday financial returns over, for instance, five-minute intervals.

a scale factor, and the square root of the quadratic variation is generally not a perfect estimator of this scale factor. Moreover, the quadratic variation does not always lead to an optimal estimator of the scale factor, as shall be clear from a simple example. So, even in ideal circumstances, finding good proxies for discrete time volatility is not a trivial task.

Discrete time volatility models were developed before high-frequency data became readily available, and are typically applied to daily, or lower frequency returns. Most discrete time models for the *daily* financial return r_n satisfy the canonical product structure:

$$r_n = s_n Z_n. \tag{1}$$

Here the observed financial return r_n is modelled as the product of an iid innovation Z_n and a positive scale factor s_n, called the volatility. One usually assumes that Z_n has mean zero and, for standardization, unit variance. Specific models differ in their specification of the volatility process; an example is the stationary Garch(1,1) recursion

$$s_n^2 = \kappa + \alpha r_{n-1}^2 + \beta s_{n-1}^2, \tag{2}$$

where $\kappa, \alpha, \beta > 0$ and $\alpha + \beta < 1$.

The scale factors (s_n) are not observed, and one may use the daily close-to-close returns r_n to estimate and evaluate models for s_n. An early paper that makes use of intraday data to obtain a daily volatility proxy is Parkinson (1980): under the assumption that the log price process is a Brownian motion within the day, the intraday high-low range provides a superior volatility estimator compared with the daily close-to-close return. See also Alizadeh, Brandt, and Diebold (2002). For more general intraday price processes, Visser (2008) develops a quasi maximum likelihood estimator (QMLE) for daily volatility models. This QMLE makes use of intraday based volatility proxies H_n, and yields a precise criterion for the quality of a proxy by looking at the relative errors $\log(H_n/s_n)$. The quality of the parameter estimators is then determined by the measurement variance λ^2 of the relative error,

$$\lambda^2 = \mathrm{var}(\log(H_n/s_n)). \tag{3}$$

The smaller λ^2, the smaller the standard errors of the parameter estimators. This result holds for surprisingly general intraday price processes. It also holds irrespective of the particular volatility model for s_n. The criterion λ^2 is valid within a large class of volatility proxies H, including popular proxies such as the intraday high-low range, the realized volatility and realized power variation.

The main result of the present paper is a procedure for combining volatility proxies into a single, highly efficient proxy. The combined proxy has minimal measurement variance λ^2, as given by equation (3). The paper takes a model free approach: it develops a theory for ranking and combining proxies without assuming a particular model for the sequence of volatilities (s_n), and without making strong model assumptions for the intraday price process. The resulting tools developed in the paper are straightforward to apply.

Empirical analysis of S&P 500 index futures market tick data from January 1988 to mid 2006 shows that the techniques in this paper enable one to construct a good proxy for the S&P 500 volatility. Moreover, our empirical results suggest that indeed in practice the quadratic variation is not optimal for the scale factor s_n: the analysis indicates that our finite-grid optimized proxy is more efficient than the (square root of the) quadratic variation. Interestingly, the optimized proxy based on the sum of the highs, the sum of the lows, and the sum of the absolute returns over ten-minute intervals, puts more weight on the highs than on the lows. From this point of view, the *upward* price movements are more informative than the downward movements.

Related Literature

The present paper is first to specifically address the problem of optimizing volatility proxies from the perspective of discrete time volatility models, but its underlying theme, dealing with high-frequency data in daily volatility modelling, is shared with two other branches of the literature. On the one hand there is the temporal aggregation literature, see for instance Drost and Nijman (1993), Drost and Werker (1996), and Meddahi and Renault (2004); on the other hand there is the literature cast in the framework of continuous time semimartingales. An important part of the semimartingale literature is concerned with estimation of the quadratic variation, and dealing with microstructure noise. In that literature, as in our empirical analysis, the use of high-low ranges over five or ten-minute intervals has received attention, see for instance Martens and van Dijk (2007) and Christensen and Podolskij (2007). One may also improve quadratic variation estimators by subsampling and averaging realized variances, see Zhang, Mykland, and Aït-Sahalia (2005). Hansen and Lunde (2005) discuss how to combine unbiased estimators of the integrated variance, in particular how to combine the realized variance and the squared overnight return. For overviews see Andersen, Bollerslev and Diebold (2008) and McAleer and Medeiros (2008). Of course, the measures proposed by the quadratic variation literature may be used as proxies for s_n, and prove valuable as input to a combination of proxies for s_n.

The remainder of the paper is organized as follows. Section 1 discusses the model for the intraday return process $R_n(\cdot)$. Section 2 introduces proxies, and develops tools for ranking and optimizing them. Section 3 constructs a good volatility proxy for the S&P 500 data. In Section 4 we give the most important conclusions. The appendices A, B, C, and D contain a description of the data, a discussion of microstructure effects, introduce the empirical technique of prescaling, and provide mathematical details.

1 Model

We observe for each trading day n a process $R_n(\cdot)$, being the continuous time log-return process within that day. It starts with the overnight return at time $u = 0$, and at the end of the day, at time $u = 1$, we obtain the close-to-close return $r_n = R_n(1)$. Figure 1 depicts five actual intraday log-return processes R_n, for the S&P 500, for $n = 2285, \ldots, 2289$. The

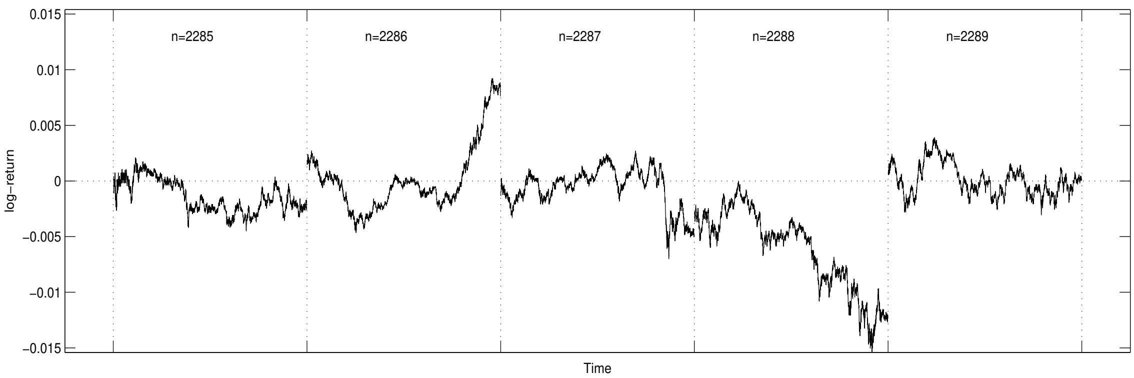


Figure 1: Five intraday return processes $R_n(\cdot)$, with respect to the previous day's close, $n = 2285, \ldots, 2289$, for the S&P 500. Starting at 1997–02–14.

second day in the figure ($n = 2286$ in our sample), for example, starts with a small positive overnight return, and the value of the index increases towards the end of the day to arrive at a plus of $r_n = R_n(1) \approx 0.01$, or +1% at the close of the day.

Now, the standard framework for intraday high-frequency data would be to assume that $R_n(\cdot)$ is a semimartingale on the unit time interval (i.e. the trading day). Although the model that the paper proposes is not at odds with the semimartingale approach, it is for our purposes perhaps more insightful to first have a look at a simple example, the scaled Brownian motion:

$$R_n(\cdot) = s_n W_n(\cdot).$$

Here, $W_n(\cdot)$ is a standard Brownian motion on the unit time interval independent of s_n, and one may think of s_n as the scale factor of a daily Garch process. So the scale factor

s_n represents daily volatility and is constant within the day, whereas the Brownian motion $W_n(\cdot)$ captures the intraday price movements. Estimation of s_n is an easy task in this model. Indeed, the daily quadratic variation QV now yields an exact relationship:

$$QV(R_n) = s_n^2,$$

so the square root of the quadratic variation is a perfect estimator of s_n. Our model for $R_n(\cdot)$ is a generalization of the scaled Brownian motion:

$$R_n(\cdot) = s_n \Psi_n(\cdot), \tag{4}$$

where $\Psi_n(\cdot)$ is an *arbitrary* process on the unit time interval, again independent of s_n. The processes $\Psi_n(\cdot)$ over different days are assumed independent. Moreover, there are no model assumptions for the sequence (s_n). To the best of our knowledge, the general form that the model (4) takes is new; we refer to it as the *scaling model*. It is highly interesting for our purposes because of its relation to discrete time volatility models: the model yields daily close-to-close returns r_n that satisfy

$$r_n = s_n Z_n,$$

(setting $Z_n = \Psi_n(1)$) which is the canonical discrete time model structure, see equation (1). In general the quadratic variation is not a perfect estimator of s_n^2: if $\Psi_n(\cdot)$ has nondeterministic quadratic variation, then

$$QV(R_n) \neq s_n^2.$$

Let us be precise on the model assumptions for the scaling model. To this purpose we introduce the discrete time *model filtration* $(\mathcal{G}_n)$, which includes the history of (s_n, Ψ_n) extended with s_{n+1}. So, $\mathcal{G}_n = \sigma\{(\Psi_i)_{i\leq n},\ (s_i)_{i\leq n+1}\}$. The σ-field $\mathcal{G}_n$ represents the model information[2] at the start of day $n+1$. The intraday return processes $R_n(\cdot)$ satisfy the scaling model whenever

$$R_n(u) = s_n \Psi_n(u), \qquad 0 \leq u \leq 1,$$

and

M1. The daily scale factors s_n are strictly positive,

[2]The statistician only observes the processes $R_n(\cdot)$.

M2. $\Psi_n(\cdot)$ is a cadlag[3] process on the closed interval $[0, 1]$,

M3. The processes $\Psi_n(\cdot)$ are identically distributed,

M4. The process $\Psi_n(\cdot)$ is independent of $\mathcal{G}_{n-1}$, for all n.

Conditions (M1) and (M2) are technical and do not lead to practical limitations. By (M3) and (M4), the processes $\Psi_n(\cdot)$ are iid over different days. The process $\Psi_n(\cdot)$ may be any process representing the intraday price pattern. The sequence of scale factors (s_n) may be any strictly positive stochastic process, as long as the process $\Psi_n(\cdot)$ is independent of current and past scale factors s_k. So the factors (s_n) may satisfy a Garch model, or a stochastic volatility model. They may also contain structural breaks, so be nonstationary. The actual fluctuations in the process $\Psi_n(\cdot)$ determine the pattern of the intraday return process, such as up or down days, quiet or hectic days. The scaling model is *not* a model of constant intraday volatility: depending on $\Psi_n(\cdot)$ the day may be hectic (for instance, a large quadratic variation) when s_n is low, and vice versa. The process $\Psi_n(\cdot)$ allows for intraday seasonality. It may also have, for instance, leverage effects, jumps, stochastic spot volatility, a non-zero mean process.

Conditions (M1) to (M4) ensure that daily volatility proxies may be decomposed into a scale factor and an independent measurement error, see equation (5) in Section 2.1. Note that $\Psi(1)$ is not standardized; identification of s_n and $\Psi_n(\cdot)$ shall not be necessary for the study of proxies, see Section 2.1.

2 Proxies

This section discusses proxies for s_n. Section 2.1 discusses how proxies may be compared. Section 2.2 shows how proxies may be combined into a superior one.

Let us first address how the theory in the paper relates to the quadratic variation of the intraday log price process. In particular, what is the relation between s_n and the quadratic variation? In recent years the quadratic variation of financial processes has received attention as a way of dealing with high-frequency data. The daily *quadratic variation (QV)* is the limit of the realized variance as the lengths of the sampling intervals approach zero. The standard framework is to assume that the intraday log price process is a semimartingale. Under fairly mild regularity conditions the quadratic variation then is an unbiased estimator of the conditional variance of the daily close-to-close return r_n, see for instance Andersen, Bollerslev, Diebold and Labys (2003). So, if s_n satisfies a Garch model (hence is

[3]The sample paths are right-continuous and have left limits.

$\mathcal{F}_{n-1}$-measurable), then the quadratic variation is an unbiased estimator of s_n^2:

$$\mathbb{E}(QV_n|\mathcal{F}_{n-1}) = \mathrm{var}(r_n|\mathcal{F}_{n-1}) = s_n^2,$$

where in general the scale factor s_n is a truly latent variable.

2.1 Ranking Proxies

A number of alternative volatility proxies have appeared in the literature: the intraday high-low range (e.g. Parkinson, 1980), the realized volatility (e.g. Barndorff-Nielsen and Shephard, 2002, and Andersen *et al.*, 2003), the sum of absolute returns (more generally the square root of the realized power variation, see Barndorff-Nielsen and Shephard, 2003, 2004), the square root of bipower variation (Barndorff-Nielsen and Shephard, 2004), the square root of the realized range (e.g. Martens and van Dijk, 2007, and Christensen and Podolskij, 2007).

All these proxies have the property of positive homogeneity: if the intraday process $R_n(\cdot)$ is multiplied by a factor $\alpha \geq 0$, then so is the proxy:

$$H(\alpha R_n) = \alpha H(R_n), \qquad \alpha \geq 0.$$

The present paper allows any positive and positively homogeneous proxy. We shall refer to both the random variable H_n,

$$H_n \equiv H(R_n),$$

and the functional H as *proxies*. The proxy H_n is linear in s_n:

$$H_n = s_n H(\Psi_n).$$

For quadratic proxies we refer to the final paragraph of this section.

The following decomposition (5) is central to the results of the paper. Applying logarithms leads to an additive measurement equation:

$$\log(H_n) = \log(s_n) + U_n. \tag{5}$$

So the log of a proxy consists of the sum of two independent terms, the log of the scale factor s_n and a *measurement error*

$$U_n \equiv \log(H(\Psi_n)).$$

The measurement errors U_n form an iid sequence. Our criterion for the quality of proxy is given by the measurement variance $\lambda^2 = \mathrm{var}(\log(H_n/s_n))$,

$$\lambda^2 = \mathrm{var}(U_n). \tag{6}$$

The smaller λ^2, the more efficient the proxy is for QML parameter estimation of discrete time volatility models.[4] For additional discussion of the measurement variance see Appendix D.1. A proxy $H^{(1)}$ is better than $H^{(2)}$ if it has smaller measurement variance:

$$(\lambda^{(1)})^2 \leq (\lambda^{(2)})^2.$$

For this ranking to make sense, it has to be the same for all possible representations of $R_n(\cdot) = s_n \Psi_n(\cdot)$. This is confirmed by Proposition D.1 in Appendix D.2. An optimal proxy H^* satisfies

$$\mathrm{var}(\log(H^*(\Psi))) = \inf_H \mathrm{var}(\log(H(\Psi))).$$

For a proxy H, the measurement error U_n only depends on the process $\Psi_n(\cdot)$. So *the optimality of a proxy H is independent of the particular discrete time model for the scale factors* (s_n). Optimal proxies exist and they can be shown to be unique up to a constant factor, see Appendix D.2, yet there does not seem to be a concrete way of determining this optimal proxy, or for computing its measurement variance. The appendix also provides a simple example that shows that the square root of the quadratic variation need not be an optimal proxy, Example D.2.1.

A first *practical* step is to achieve a data-based ranking: how can one tell from the time series of realizations H_n of several proxies, which one is the best? Taking variances on both sides of the decomposition (5) gives

$$\mathrm{var}(\log(H_n^{(i)})) = (\lambda^{(i)})^2 + \mathrm{var}(\log(s_n)). \tag{7}$$

There is no covariance term by the independence of s_n and $\Psi_n(\cdot)$. Equation (7) shows that the variances of the proxies all have the common term $\mathrm{var}(\log(s_n))$. It follows that if the variance of the log proxy is smaller, then the measurement variance must be smaller.

[4] See the theory on the log-Gaussian QMLE in Visser (2008).

Assuming $\mathrm{var}(\log(s_n)) < \infty$, one has the equivalence

$$(\lambda^{(1)})^2 \leq (\lambda^{(2)})^2 \quad \Leftrightarrow \quad \mathrm{var}(\log(H_n^{(1)})) \leq \mathrm{var}(\log(H_n^{(2)})). \tag{8}$$

So, in empirical applications one may simply rank proxies by estimating the variance of their logarithm. See Section 3.1 for ranking proxies for the S&P 500 index.

We end this section with a remark on quadratic proxies. One may be interested in proxies that are homogeneous of a degree $p \neq 1$, for instance proxies that are quadratic in nature. These proxies satisfy $\tilde{H}(\alpha R_n) = \alpha^2 \tilde{H}(R_n)$, and are proxies for s_n^2. The theory of the paper directly applies to quadratic proxies, since $\tilde{H}$ is linear in s_n^2. The ranking of a quadratic proxy corresponds to the ranking of its homogeneous version (which one obtains by taking the square root). So the optimal quadratic proxy is the square of our optimal proxy: $\tilde{H}^* = (H^*)^2$, and one may restrict attention to proxies that are homogeneous of degree $p = 1$.

2.2 Combining Proxies

We are now ready for the main result of the paper. For finding a good proxy one first needs to think up some simple proxies. The procedure below will then combine these into a single, more efficient proxy. Suppose we are supplied with the proxies $H^{(1)}, \ldots, H^{(d)}$. Consider the geometric combination of these proxies,

$$H_n^{(w)} \equiv \prod_{i=1}^{d} (H_n^{(i)})^{w_i}, \qquad w_1 + \ldots + w_d = 1, \; w_i \in \mathbb{R}. \tag{9}$$

Here, the column vector w is the d-dimensional coefficient vector. The restriction $\sum w_i = 1$ is needed to obtain a proxy, though the coefficients are not restricted to the interval $[0, 1]$. It is natural to have the coefficients w_i acting as exponents in equation (9), since taking logarithms now yields an additive problem. Let Λ denote the covariance matrix of the measurement errors $U^{(i)} = \log(H^{(i)}(\Psi))$:

$$\Lambda = \mathrm{cov}([U^{(1)}, \ldots, U^{(d)}]'). \tag{10}$$

The measurement error $U^{(w)} \equiv \log(H^{(w)}(\Psi))$ of the geometric combination in (9) equals

$$U^{(w)} = \sum_{i=1}^{d} w_i U^{(i)},$$

which has variance $\lambda_w^2 = w'\Lambda w$ and, as for the *global* minimal variance portfolio in Markowitz portfolio theory, λ_w^2 is minimal for

$$w^* = \frac{\Lambda^{-1}\iota}{\iota'\Lambda^{-1}\iota}, \qquad \iota = (1, \ldots, 1)', \tag{11}$$

with optimal variance $\lambda_{w^*}^2 = \frac{1}{\iota'\Lambda^{-1}\iota}$. This solution is empirically infeasible since the measurement errors $U_n^{(i)}$ are not observed. However, by equation (8) one may equivalently minimize $\mathrm{var}(\log(H_n^{(w)}))$. Now, let $\Lambda_{p,n}$ denote the covariance matrix of the log of the simple proxies:

$$\Lambda_{p,n} = \mathrm{cov}([\log(H_n^{(1)}) \ldots \log(H_n^{(d)})]'). \tag{12}$$

The covariance matrix $\Lambda_{p,n}$ is the covariance matrix Λ with a common term $\mathrm{var}(\log(s_n))$ added to each element:

$$\Lambda_{p,n} = \Lambda + \mathrm{var}(\log(s_n))\ \iota\iota'. \tag{13}$$

The optimal coefficients w^* may now be obtained upon replacing Λ by $\Lambda_{p,n}$ in equation (11), see formula (14) below. In empirical applications one may want to assume stationarity for (s_n), so that the covariance matrix

$$\Lambda_{p,n} = \Lambda_p,$$

may simply be estimated by the sample covariance matrix.

Theorem 2.1. *Let $R_n(\cdot)$ satisfy the scaling model. Assume $\mathrm{var}(\log(H^{(i)}(\Psi))) < \infty$ for $i = 1, \ldots, d$, and $\mathrm{var}(\log(s_n)) < \infty$. Let the covariance matrices Λ and $\Lambda_{p,n}$ be defined by (10) and (12). The optimal coefficient vector w^* in (11) does not depend on the form of the process (s_n) and may be expressed as*

$$w^* = \frac{\Lambda_{p,n}^{-1}\iota}{\iota'\Lambda_{p,n}^{-1}\iota}. \tag{14}$$

The variance of the logarithm of the optimal geometric combination is

$$\mathrm{var}(\log(H_n^{(w^*)})) \quad = \quad \lambda_{w^*}^2 + \mathrm{var}(\log(s_n)),$$

where $\lambda_{w^*}^2 = \frac{1}{\iota'\Lambda^{-1}\iota}$ is its measurement variance.

Proof. The optimal coefficient w^* does not depend on (s_n): by equation (8)

$$\arg\min_w \mathrm{var}(\log(H_n^{(w)})) = \arg\min_w \mathrm{var}(\log(H^{(w)}(\Psi_n))),$$

so

$$\arg\min_w\ w'\Lambda_{p,n}w = \arg\min_w\ w'\Lambda w. \tag{15}$$

Define the Lagrangian $w'\Lambda_{p,n}w + \mu\ (1 - w'\iota)$. Differentiating the Lagrangian with respect to w yields $2\Lambda_{p,n}w - \mu\iota = 0$, hence $w = 1/2\ \Lambda_{p,n}^{-1}\mu\iota$. By $\iota'w = 1$, this yields $\mu = 2/\iota'\Lambda_{p,n}^{-1}\iota$ and $w = \Lambda_{p,n}^{-1}\iota/\iota'\Lambda_{p,n}^{-1}\iota$. Since $w'\Lambda_{p,n}w$ is convex in w and there is a unique solution to the first order condition, it is the optimum.

Use (15) to obtain the equalities $w^* = \Lambda_{p,n}^{-1}\iota/\iota'\Lambda_{p,n}^{-1}\iota = \Lambda^{-1}\iota/\iota'\Lambda^{-1}\iota$, which imply

$$\mathrm{var}(\log(H_n^{(w^*)})) = \lambda_{w^*}^2 +\ \mathrm{var}(\log(s_n)).$$

□

Remark 1. *In empirical applications one uses estimates of the covariance matrix. To reduce estimation error, we shall use the technique of prescaling, see Appendix C.*

We end this section with a few words relevant to empirical implementation. Assuming stationarity for the process (s_n, Ψ_n), the covariance matrix $\Lambda_{p,n} = \Lambda_p$ is consistently estimated by the sample covariance matrix of the log of the proxies, thereby providing coefficients $\hat{w}$ that are consistent for w^*. More generally, this estimator for w^* may remain consistent while allowing, for example, for structural breaks in the scale factors (s_n). See the consistency condition (19) in Appendix D.3 for details.

3 A Good Proxy for S&P 500 Volatility

This section constructs a good proxy for volatility, applying the techniques of Section 2 to the S&P 500 futures tick data over the years 1988–2006, a total of 4575 trading days. Appendix A describes the data. The proxies below are constructed taking care of microstructure noise, see Appendix B.

3.1 Ranking Proxies

One can think of many proxies for the daily volatility s_n. Table 1 compares twelve simple proxies constructed from the data.

	full	1st	2nd	3rd	4th
name	PV	PV	PV	PV	PV
RV5	0.064	0.070	0.070	0.073	0.042
RV10	0.080	0.085	0.093	0.090	0.052
RV15	0.089	0.096	0.105	0.093	0.061
RV20	0.100	0.110	0.117	0.103	0.071
RV30	0.117	0.133	0.134	0.113	0.087
abs-r	0.611	0.683	0.550	0.635	0.568
hl	0.161	0.179	0.176	0.160	0.130
maxar2	0.118	0.134	0.124	0.118	0.088
RAV5	0.058	0.060	0.065	0.066	0.040
RAV10	0.072	0.072	0.085	0.082	0.049
RVHL10	0.053	0.057	0.061	0.061	0.034
RAVHL10	0.047	0.048	0.055	0.054	0.031
minimal PV	0.047	0.048	0.055	0.054	0.031

Table 1: Performance of twelve proxies. The table gives the PV: the variance of the logarithm after prescaling by EWMA(0.7) predictor for RV5. The full sample is split into four subsamples. The following proxies are included. We abbreviate square root to sqrt.: RV5: sqrt. of sum of squared 5-min. returns; RV10: sqrt. of sum of squared 10-min. returns; RV15: sqrt. of sum of squared 15-min. returns; RV20: sqrt. of sum of squared 20-min. returns; RV30: sqrt. of sum of squared 30-min. returns; abs-r: absolute close-to-close return; hl: high-low of the intraday return process; maxar2: maximum of the absolute 2-min. returns; RAV5: sum of absolute 5-min. intraday returns; RAV10: sum of absolute 10-min. intraday returns; RVHL10: sqrt. of sum of 10-min. squared high-lows; RAVHL10: sum of 10-min. high-lows;

For each proxy a measure of comparison is given for five samples: first the full sample (days 2 to 4575) and then for four subsamples spanning the full sample (2:1144, 1145:2287, 2288:3431, 3432:4575). The measure of comparison is PV (prescaled variance), which is the variance of the logarithm of a proxy H after prescaling by a suitable series p_n,

$$\begin{aligned} PV(H) &= \mathrm{var}(\log\big(H_n/p_n\big)) \\ &= \mathrm{var}(\log(s_n/p_n)) + \lambda_H^2. \end{aligned}$$

Smaller PV's correspond to more efficient proxies. Prescaling does not effect the theoretical ranking of proxies, but helps to diminish statistical noise, see Appendix C. The first observation cannot be prescaled and is left out of the variance computations. For the prescaling sequence (p_n) we take an exponentially weighted moving average predictor of five-minute realized volatility with smoothing parameter $\beta = 0.7$, yielding a prescaling sequence $p_n = 0.7\ p_{n-1} + 0.3\ RV5_{n-1}$. We have set the smoothing parameter so that the sample variance of the logarithm of prescaled five-minute realized volatility is minimal. A smoothing parameter around $\beta = 0.7$ for a realized volatility filter was found to fit well in earlier research, see for instance Engle (2002). Recall that the prescaled variance ranks proxies, but the measurement variance λ^2 itself remains unknown.

The first column of Table 1 gives the prescaled variances over the full sample. The

	full	1st	2nd	3rd	4th
name	PV	PV	PV	PV	PV
RV5-up	0.066	0.070	0.070	0.073	0.051
RV5-down	0.094	0.103	0.101	0.104	0.068
RV10-up	0.091	0.094	0.101	0.095	0.074
RV10-down	0.133	0.138	0.146	0.149	0.100
RAV5-up	0.064	0.066	0.070	0.067	0.051
RAV5-down	0.097	0.099	0.101	0.110	0.077
RAV10-up	0.093	0.093	0.102	0.097	0.081
RAV10-down	0.147	0.145	0.155	0.168	0.120
RAV10HIGH	0.053	0.055	0.061	0.054	0.041
RAV10LOW	0.081	0.082	0.086	0.090	0.064
minimal PV	0.053	0.055	0.061	0.054	0.041

Table 2: Performance of upward/downward decomposed proxies. The table splits proxies from Table 1 according to upward and downward price movements. For example, RV5-up is sqrt. of sum of squared 5-min. *positive* returns, RAV10HIGH is the sum of 10-min. highs, and RAV10LOW is the sum of 10-min. absolute lows.

standard five-minute realized volatility RV5 has $PV = 0.064$. The first column shows that the quality of the realized volatility RV improves if one increases the sampling frequency from 30 minutes to 5 minutes. The prescaled variance is maximal for the absolute close-to-close returns, confirming that absolute or squared daily returns are poor proxies.[5] Note that the maximal absolute two-minute return outperforms the intraday high-low range, which tends to use returns based on much longer time spans. Overall, we find that sums of absolute values lead to more efficient proxies than sums of squared values. This observation relates to a finding of Barndorff-Nielsen and Shephard (2003), whose simulations indicate that absolute power variation, based on the sum of absolute returns, has better finite-sample behaviour than the realized variance. The best performing proxy in Table 1 is $RAVHL10$, the sum of the ten-minute high-low ranges. The remaining columns of Table 1 show that the ranking of the different proxies is stable: the ranking in the subsamples is the same as in the full sample, with one exception in the second subsample for $RV30$ and $maxar2$. Though the measurement variances are not observed, from the full sample column we may infer that the measurement variance λ^2 of $RAVHL10$ is at least 25% smaller than the measurement variance of $RV5$, by $(0.064 - 0.047)/0.064 \approx 0.27$.

Table 2 splits proxies from Table 1 into *upward* and *downward* components. For instance, the five-minute realized volatility is decomposed according to upward and downward price

[5] We use the absolute returns larger than 0.001, or 10 basis points, to avoid taking the log of zero. This leaves 4079 daily returns.

movements:

$$\begin{aligned}
(\text{RV5})^2 &= \sum_{i=1}^{K} r_{n,i}^2 \\
&= \sum_{i=1}^{K} r_{n,i}^2 \; I_{\{r_{n,i}>0\}} + \sum_{i=1}^{K} r_{n,i}^2 \; I_{\{r_{n,i}<0\}} \\
&= (\text{RV5-up})^2 + (\text{RV5-down})^2.
\end{aligned}$$

Here, $r_{n,i}$ denotes the return over the i-th intraday five-minute interval on day n. As one would expect RV5 (PV=0.064) is better than RV5-up (PV=0.066) and RV5-down (PV=0.094). Note that the upward proxies are consistently more efficient than their downward counterparts. This difference suggests that, when proxying the scale factor s_n, one should put more weight on the *upward* movements.

3.2 Optimized Combination

The proxies in Tables 1 and 2 may each be of value for measuring the scale factor s_n, but certain proxies may be more useful than others. This section combines the proxies in Tables 1 and 2 into a more efficient one using the combination formula of Section 2.2. We also conduct a thorough stability analysis, and discuss properties of the optimized proxy.

The five-minute realized volatility RV5 is a standard proxy, and has PV=0.064. Let us improve upon this value. First, by using the (square root of the) sum of the squared high-low ranges over intraday intervals, RVHL10 has $PV = 0.053$, see Table 1. It is better to use absolute values: RAVHL10 has $PV = 0.047$. Now use the theory of Section 2.2 to combine the high-low ranges in RAVHL10 with the absolute returns in RAV10: inserting the covariance matrix $\tilde{\Lambda}_{p,n}$ of the log of these two prescaled proxies into formula (14), yields the proxy

$$H_n = (RAVHL10_n)^{1.82}(RAV10_n)^{-0.82}, \qquad (PV = 0.041).$$

Decomposing RAVHL10 into its upward and downward components, RAV10HIGH and RAV10LOW, we obtain the proxy

$$H_n^{(\hat{w})} = (RAV10HIGH_n)^{1.04}(RAV10LOW_n)^{0.72}(RAV10_n)^{-0.76}, \quad (PV = 0.038). \tag{16}$$

Of course, one may also apply the optimal coefficient formula (14) to all twenty-one proxies in Tables 1 and 2 at once.[6] The full combination yields a proxy with $PV = 0.037$, which

[6]We exclude the absolute close-to-close return in performing this calculation.

	full	1st	2nd	3rd	4th
name	PV	PV	PV	PV	PV
$H^{(\hat{w})}$	0.038	0.039	0.043	0.043	0.028
$H^{(\hat{w},1)}$	0.038	0.039	0.043	0.043	0.028
$H^{(\hat{w},2)}$	0.039	0.039	0.043	0.043	0.029
$H^{(\hat{w},3)}$	0.039	0.040	0.044	0.042	0.029
$H^{(\hat{w},4)}$	0.039	0.040	0.045	0.044	0.027
minimal PV	0.038	0.039	0.043	0.042	0.027

Table 3: Combined proxy, optimized for different subsamples: performance and stability.

only marginally outperforms the proxy obtained in (16). We prefer to work with the simpler proxy in (16) and we refer to it as the *optimized proxy* $H_n^{(\hat{w})}$.

The optimized proxy easily outperforms all proxies in Tables 1 and 2. If one extrapolates the full sample prescaled variances of the realized volatilities of Table 1 to a time interval of length zero (corresponding to the limiting case of the quadratic variation), one obtains a value between 0.050 and 0.060. The value $PV = 0.038$ for the optimized proxy is well below these values, suggesting that this proxy for the daily scale factor is more efficient than the square root of the quadratic variation. Indeed, the optimized proxy has a measurement variance λ^2 which is at least 40% smaller than the five-minute realized volatility, by $(0.064-0.038)/0.064 \approx 0.41$, and may be 25% smaller than the measurement variance of the square root of the quadratic variation, by $(0.05-0.038)/0.05 \approx 0.24$.

Observe that the coefficient for RAV10 in the optimized proxy is negative ($\hat{w}_3 = -0.76$). In geometrical terms this negative coefficient may be explained as follows. The log proxies are vectors in an affine space. The proxies are highly related, since all proxies approximate the same daily scale factor s_n. The optimal proxy is not in the convex hull of the proxies in Tables 1 and 2. The original proxies do not completely reflect the direction of the optimal proxy. The coefficients outside $[0,1]$ correct the direction.

Table 3 investigates the stability of the optimized proxy $H_n^{(\hat{w})}$. Similarly to Table 1 it lists performance measures for the full sample and for four subsamples. The first row gives the performance of $H_n^{(\hat{w})}$ in the different subsamples; comparison with Tables 1 and 2 shows that $H_n^{(\hat{w})}$ outperforms all those proxies in every subsample. The proxy $H_n^{(\hat{w},i)}$ is constructed using the coefficients that are optimal for the i-th subsample. In the first subsample the performance of the globally optimized $H_n^{(\hat{w})}$ (PV=0.039) is not substantially outperformed by $H_n^{(\hat{w},1)}$ (PV=0.039). A similar statement holds for the other subsamples. Moreover, proxies based on *any* particular subsample are close to optimality in all other subsamples. For instance, the proxy optimized for the first subsample (the years 1988–1992) is nearly optimal for the years 2002–2006. We conclude that the optimality of $H_n^{(\hat{w})}$ is stable.

Proxies are important for volatility forecast evaluation. Good proxies help in distinguishing better from poor forecasts, see for instance Hansen and Lunde (2006a) and Patton and Shephard (2008). Table 4 explores the quality of the optimized proxy in a heuristic way. It gives the coefficient of determination, R^2, of a linear regression of the logarithm of a proxy on the logarithm of another proxy lagged one day:

$$\log(H_n^{(j)}) = \alpha + \beta \log(H_{n-1}^{(i)}) + \varepsilon_n.$$

Large R^2's in a particular column mean that the proxy in that column is largely predictable, suggesting that it is a good proxy to use for volatility forecast evaluation. The R^2's attain their maximum at the optimized proxy, in the most right column.[7]

	RV30	RV20	RV15	RV10	RV5	$H^{(\hat{w})}$
RV30(-1)	0.35	0.39	0.42	0.46	0.50	0.58
RV20(-1)	0.38	0.42	0.45	0.49	0.54	0.61
RV15(-1)	0.39	0.44	0.47	0.50	0.55	0.63
RV10(-1)	0.41	0.45	0.48	0.52	0.57	0.66
RV5(-1)	0.43	0.48	0.51	0.55	0.60	0.69
$H^{(\hat{w})}(-1)$	0.44	0.48	0.51	0.54	0.60	0.71

Table 4: R^2 of the regression $\log(H_n^{(j)}) = \alpha + \beta \log(H_{n-1}^{(i)}) + \varepsilon_n$, for $i, j = 1, \ldots, 6$, and $n = 2, \ldots, 4575$.

Table 5 gives an impression of the distributions of the measurement errors of the five-minute realized volatility and of the optimized proxy. It gives descriptive statistics based on the logarithm of the prescaled proxies. We do not provide the sample averages since the quality of a proxy is insensitive to scaling, see Appendix D.1. It appears that putting less weight on the sum of the lows, and more on the sum of the highs, helps to diminish the skewness. The optimized proxy is more symmetrical and more concentrated than the five-minute realized volatility.

	$\log(RV5_n/p_n)$	$\log(H_n^{(\hat{w})}/p_n)$
st. dev.	0.25	0.20
skewness	0.62	0.00
kurtosis	6.00	4.05

Table 5: Proxy distributions. For the full sample, $n = 2, \ldots, 4575$: standard deviation, skewness, and kurtosis of the logarithm of the prescaled proxy, $\log(H_n/p_n)$. Prescaling filter p_n is EWMA($\beta = 0.7$) based on RV5.

Finally, Figure 2 shows the time series graphs of four different proxies. The proxies were standardized to have mean one, by dividing them by their mean. From top to bottom the curves become 'less erratic', suggesting a decrease in the variance of the measurement errors U_n. Each step shows a marked improvement.

[7] The optimized proxy is constructed as an optimized proxy for s_n, not as an optimal predictor for s_{n+1}. Even so, the R^2's attained in the bottom row $H^{(\hat{w})}(-1)$ are large.

4 Conclusions

This is the first paper to address the problem of optimizing volatility proxies in relation to discrete time volatility models, such as Garch and stochastic volatility models. The results of the paper should be of use to researchers that aim at improving discrete time models: proxies are important for the specification of these models; they are also an essential input to parameter estimation and volatility forecast evaluation. Most of these models satisfy the canonical product structure $r_n = s_n Z_n$, where r_n is the daily financial return, s_n the volatility, and Z_n an iid innovation. The problem of finding good proxies for discrete time models differs from the problem of finding good estimators of the quadratic variation for continuous time semimartingales. Our theory is founded on three distinctive elements:

- The continuous time model for the intraday return process $R_n(\cdot)$ is new, but yields the canonical discrete time model if one samples daily close-to-close returns. Moreover, it requires only minimal assumptions.
- The class of volatility proxies H for s_n is new, but encompasses most well-known proxies, such as the realized volatility, the high-low range, and the absolute return.
- The criterion of ranking proxies by the variance of the relative measurement error $U_n = \log(H_n/s_n)$ is new. It is natural because of the multiplicative role of the scale factor, and it is consistent with optimal QML parameter estimation for discrete time models.

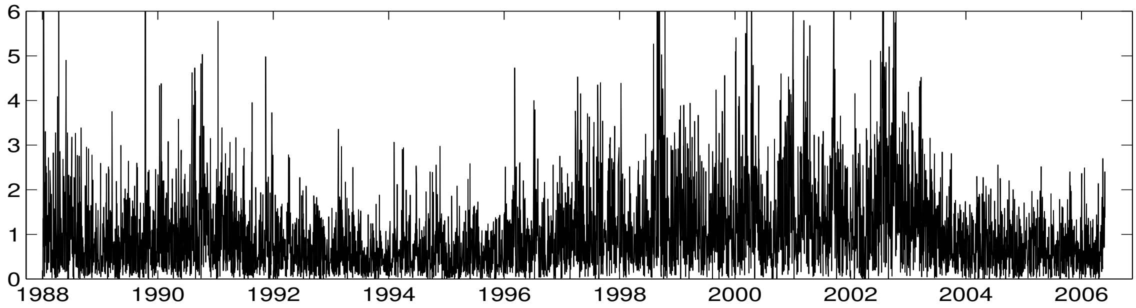
In this paper a volatility proxy is a positive, and positively homogeneous statistic of the intraday log-return process. We provide easy-to-implement tools for ranking proxies and combining them into a highly efficient proxy. The approach is to a large extent model free: an optimal proxy for the scale factor s_n is an optimal proxy under all possible discrete time models of the form $r_n = s_n Z_n$.

For the S&P 500 data a combination of the sum of the highs, the sum of the lows, and the sum of the absolute returns over ten-minute intervals yields a good proxy. One should put more weight on the sum of the highs than on the sum of the lows, when proxying volatility. The empirical results indicate that the optimized proxy, although it uses only a finite sampling grid, is more efficient for the scale factor s_n than (the square root of) the quadratic variation, which is based on the limiting case of continuous sampling. Our optimized proxy outperforms the five-minute realized volatility by at least 40%, and the square root of the quadratic variation by 25%.

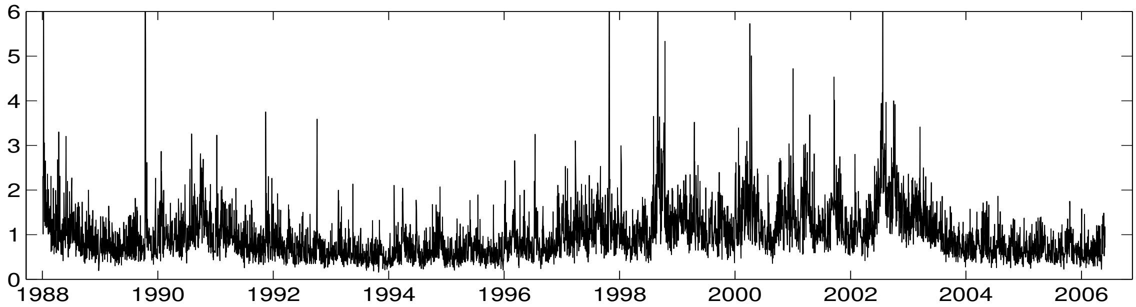
This paper has addressed the problem of ranking and optimizing proxies for today's scale factor s_n. We see opportunities for future research to use proxies for the specification of the daily volatility process (s_n), and accordingly use proxies to forecast future volatility.

Acknowledgment

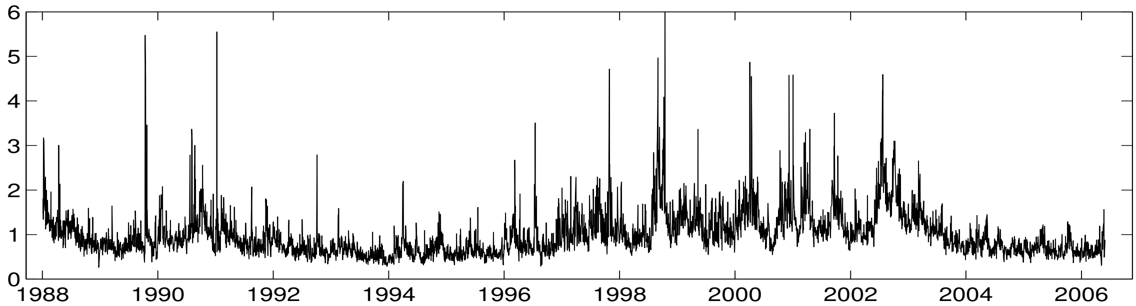
The authors are indebted to Guus Balkema, Peter Boswijk, Chris Klaassen, Oleg Kozlovski, Remco Peters, and Duncan Sands for detailed comments and suggestions.



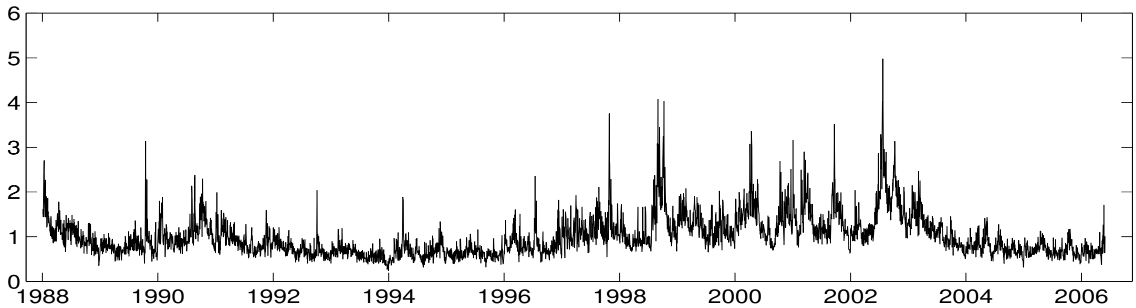
(a) Absolute close-to-close returns



(b) Intraday high-low range



(c) Five-minute realized volatility RV5



(d) Optimized proxy from formula (16)

Figure 2: Time series of four standardized proxies, $H_n/\bar{H}_n$.

Appendices

A Data

Our data set is the U.S. Standard & Poor's 500 stock index future, traded at the Chicago Mercantile Exchange (CME), for the period 1st of January, 1988 until May 31st, 2006. The data were obtained from Nexa Technologies Inc. (www.tickdata.com). The futures trade from 8:30 A.M. until 15:15 P.M. Central Standard Time. Each record in the set contains a timestamp (with one second precision) and a transaction price. The tick size is \$0.05 for the first part of the data and \$0.10 from 1997–11–01. The data set consists of 4655 trading days. We removed sixty four days for which the closing hour was 12:15 P.M. (early closing hours occur on days before a holiday). Sixteen more days were removed, either because of too late first ticks, too early last ticks, or a suspiciously long intraday no-tick period. These removals leave us with a data set of 4575 days with nearly 14 million price ticks, on average more than 3 thousand price ticks per day, or 7.5 price ticks per minute.

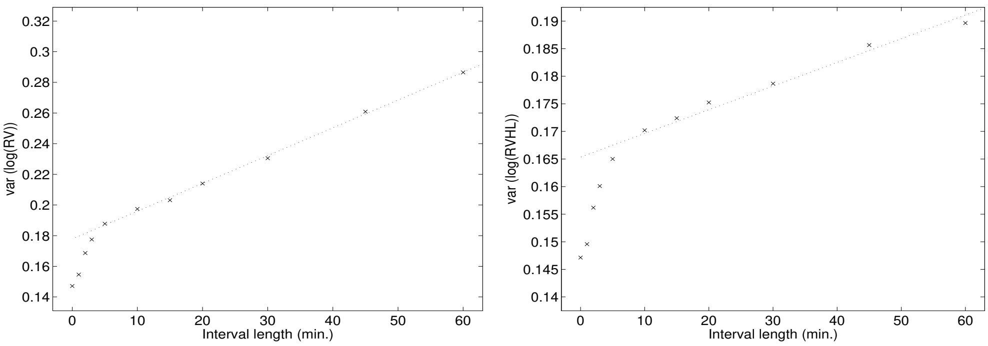
There are four expiration months: March, June, September, and December. We use the most actively-traded contract: we roll to a next expiration as soon as the tick volume for the next expiration is larger than for the current expiration.

An advantage of using future data rather than the S&P 500 cash index is the absence of non-synchronous trading effects which cause positive autocorrelation between successive observations, see Dacorogna *et al.* (2001). As in the cash index there may be bid-ask effects in the future prices which induce negative autocorrelation between successive observations. We deal with these effects by taking large enough time intervals, see Appendix B. Since we study a very liquid asset the error term due to microstructures is relatively small.

B Microstructure Noise Barrier

On small time scales financial prices are subject to market microstructure effects, such as the bid-ask bounce, price discreteness, and asynchronous trading, see, for instance, Zhang, Mykland, and Aït-Sahalia (2005), Oomen (2005,2006), and Hansen and Lunde (2006b). These effects may invalidate the model assumptions. Microstructure effects may be avoided by sampling at sufficiently wide intervals.

In the present paper the measure of comparison is the variance of the logarithm. The standard realized volatility RV and the realized range $RVHL$ (see Table 1) depend on the sampling interval Δu. Figure 3 shows the graph of $\Delta u \to \widehat{\mathrm{var}}(\log(H^{\Delta u}(R_n)))$, for Δu ranging from zero to sixty minutes. These curves suggest that a qualitative change of behaviour occurs for $\Delta u \approx$ five minutes for realized volatility, and $\Delta u \approx$ eight minutes for realized range. The realized volatilities in the paper are based on five-minute sampling intervals or larger. For realized range our minimal sampling interval is ten minutes.



(a) Realized volatility

(b) Realized range

Figure 3: Plots of the sample variance of the log of a proxy with Δu ranging from zero to 60 minutes (zero is tick per tick). (a) Realized volatility, RV. (b) Realized range, $RVHL$.

C Prescaling

The methods of comparing proxies by the variance of the logarithm and of combining proxies in Theorem 2.1 are formulated in terms of population variances and covariances. In practical situations, one has to work with the sample counterparts of these quantities, which introduces sampling error. To reduce the part of the sampling error caused by the scale factors (s_n), we propose the technique of prescaling. The idea is to stabilize the sequence (s_n), by scaling it by a predictable sequence of random variables (p_n). Let $\mathcal{F}_n = \sigma(R_i, i \leq n)$ denote the observable information up until day n.

Definition C.1. *A* prescaling sequence (p_n) *is an* $(\mathcal{F}_{n-1})$ *adapted sequence of strictly positive random variables.*

The prescaling factors p_n are used to define adjusted scale factors

$$\tilde{s}_n = s_n / p_n.$$

Proposition C.2. *Assume the processes* (R_n) *satisfy the scaling model. Prescale the scale factors* (s_n) *to obtain the sequence* $(\tilde{s}_n)$ *above. The corresponding processes* $(\tilde{R}_n)$, *where* $\tilde{R}_n = \tilde{s}_n \Psi_n$, *satisfy the scaling model.*

Proof. The variables $(\tilde{s}_n)$ are positive. Both p_{n+1} and s_{n+1} are $\mathcal{G}_n$-measurable, hence so is $\tilde{s}_{n+1}$. Therefore $\tilde{R}_n$ satisfies the scaling model. □

As a result one obtains proxies for $\tilde{s}_n$. These are prescaled proxies:

$$\tilde{H}_n = H(\tilde{R}_n) = H(R_n)/p_n.$$

The proxy $\tilde{H}_n$ for $\tilde{s}_n$ has the same measurement error as the proxy H_n for s_n :

$$\begin{aligned}\tilde{U}_n &= \log(\tilde{H}_n) - \log(\tilde{s}_n) \\ &= \log(H_n) - \log(p_n) - (\log(s_n) - \log(p_n)) \\ &= U_n.\end{aligned}$$

So, ranking and optimizing proxies before and after prescaling are equivalent in terms of population statistics; therefore one may replace s_n by $\tilde{s}_n$, and H_n by $\tilde{H}_n$ in the paper. As a consequence, the population value of the term $\mathrm{var}(\log(s_n))$ in equations (7) and (13) changes into

$$\mathrm{var}(\log(\tilde{s}_n)) = \mathrm{var}(\log(s_n/p_n)).$$

A perfect predictor p_n of s_n results in $\mathrm{var}(\log(\tilde{s}_n)) = 0$.

D Mathematical Details

D.1 Properties of the Measurement Variance

A smaller measurement variance is better for QML parameter estimation for volatility models, see Section 2.1. Let us here provide some additional intuition on the measurement variance λ^2. Since proxies are positive and may have a heavy tail, it is natural to apply logarithms.[8] Let us have a look at the bias and the variance of the measurement error U_n. Equation (5),

$$\log(H_n) = \log(s_n) + U_n,$$

makes clear that for a given proxy H the measurement error introduces a *constant* bias, $\mathbb{E}U_n = \mu$, independent of s_n. In applications where the proxy is used as a variable in a

[8] In practice proxies are strictly positive. An exception is the absolute close-to-close return $|r_n|$, for which one could either set the measurement variance to infinity, or exclude the zeros from the sample. In either case $|r_n|$ proves a poor proxy.

regression, the bias is corrected by the regression parameters. It is also possible to rescale the proxy, replacing H by aH, to obtain a bias-corrected version. Consider for instance the ideal situation of a *perfect proxy*, a proxy with zero measurement variance. Such a proxy gives the volatility up to a constant factor $c > 0$:

$$H_n = c\, s_n.$$

So, the key determinant of the quality[9] is the *measurement variance* λ^2. Moreover, changing the measurement units of a proxy by a positive factor $a > 0$ does not change the measurement variance λ^2 :

$$\operatorname{var}(\log(aH(\Psi))) = \operatorname{var}(\log(H(\Psi))) = \lambda^2. \tag{17}$$

From a practical point of view, it is an advantage that the criterion λ^2 is insensitive to changes in scale: one does not need to rule out potentially biased proxies, which bypasses the difficulty of telling a priori whether a proxy is unbiased, and the difficulty of finding a priori a suitable rescaled version of each proxy. A good proxy is such that its time series (H_n) has a high degree of comovement with the series (s_n). This is confirmed by looking at correlations. Assume $0 < \operatorname{var}(\log(s_n)) < \infty$. It then follows, using the decomposition (5), that

$$\operatorname{corr}(\log(H_n), \log(s_n)) = \big(1 + \frac{\lambda^2}{\operatorname{var}(\log(s_n))}\big)^{-1/2}. \tag{18}$$

So log proxies with smaller measurement variance λ^2 have larger correlation with $\log(s_n)$. The ideal situation of zero measurement variance gives perfect correlation. For example, if $\Psi(\cdot)$ is the standard Brownian motion on $[0, 1]$, then the square root of the quadratic variation yields $\sqrt{QV_n} = s_n$.

D.2 Identification and Optimality

Let us first address the issue of identification. The following proposition states that different representations (s_n, Ψ_n) for $R_n(\cdot)$ result in the same ordering for proxies. So, for our purposes identification of s_n and $\Psi_n(\cdot)$ plays no role.

Proposition D.1. *Suppose $H^{(1)}$ and $H^{(2)}$ are proxies. Moreover, assume (s_n, Ψ_n) and*

[9]One may observe that minimizing the measurement variance λ^2 is closely related to minimizing the mean squared error $\mathbb{E}(\log(H_n) - \log(s_n))^2 = \lambda^2 + (\mathbb{E}U_n)^2$, since the bias-corrected version aH will yield $\mathbb{E}(\log(aH_n) - \log(s_n))^2 = \lambda^2$ by equation (17).

(s'_n, Ψ'_n) *both satisfy the scaling model for* $R_n(\cdot)$*. If* $H^{(1)}$ *is better than* $H^{(2)}$ *for* Ψ*, then* $H^{(1)}$ *is also better than* $H^{(2)}$ *for* Ψ'*.*

Proof. By assumption $s'_n \Psi'_n = s_n \Psi_n$. Independence of s_n and Ψ_n implies

$$\begin{aligned} \mathrm{var}(\log(s'_n)) + \mathrm{var}(\log(H^{(1)}(\Psi'_n))) &= \mathrm{var}(\log(s_n)) + \mathrm{var}(\log(H^{(1)}(\Psi_n))) \\ &\leq \mathrm{var}(\log(s_n)) + \mathrm{var}(\log(H^{(2)}(\Psi_n))) \\ &= \mathrm{var}(\log(s'_n)) + \mathrm{var}(\log(H^{(2)}(\Psi'_n))). \end{aligned}$$

Hence $\mathrm{var}(\log(H^{(1)}(\Psi'_n))) \leq \mathrm{var}(\log(H^{(2)}(\Psi'_n)))$. □

The following example shows that the square root of the quadratic variation is not necessarily the most efficient proxy for s_n.

Example D.2.1. *Consider the case that* $\Psi(\cdot)$ *is a diffusion:*

$$d\Psi(u) = \sigma(u)\, dB(u), \qquad 0 \leq u \leq 1,$$

where $B(\cdot)$ *denotes standard Brownian motion. Let the instantaneous volatility process* $\sigma(u)$ *be deterministic at the opening and stochastic for the rest of the day. More specifically, suppose* $\sigma(u)$ *equals* 1 *before time of day* $u_0 = 1/2$*, and* $\sigma(u)$ *equals either* c_1 *or* c_2 *after* u_0*, both with probability* $1/2$*. The square root of the truncated quadratic variation over* $[0, 1/2]$ *equals* s_n *times a constant, hence has zero measurement variance. The square root of the quadratic variation of* $R_n(\cdot)$ *is the product of* s_n *and a random variable with positive variance.*

One may wonder whether optimal proxies exist in general. Recall that an optimal proxy H^* satisfies

$$\mathrm{var}(\log(H^*(\Psi))) = \inf_H \mathrm{var}(\log(H(\Psi))).$$

Theorem D.2. *If there is a proxy with finite measurement variance, then there exists an optimal proxy.*

Proof. See appendix D.3. □

The next proposition states that optimal proxies are scaled versions of one another, except possibly on a set of measure zero.

Proposition D.3. *Suppose* $H^{(1)}$ *and* $H^{(2)}$ *are two optimal proxies. Then there is a constant* $a > 0$*, such that* $H^{(1)}(\Psi) \stackrel{a.s.}{=} aH^{(2)}(\Psi)$*.*

Proof. See appendix D.3. □

D.3 Proof of Existence of Optimal Proxies

To prove the existence of optimal proxies we need a rigorous definition of proxy. Recall that the process $\Psi(\cdot)$ is cadlag on $[0,1]$. Let $\mathbb{D}[0,1]$ denote the Skorohod space of cadlag functions on $[0,1]$. Endow $\mathbb{D}[0,1]$ with the Skorohod topology. The space $\mathbb{D}[0,1]$ is a separable, complete metric space (see Billingsley (1999)). The space $C[0,1]$ of continuous functions on the unit interval is a linear subspace of $\mathbb{D}[0,1]$.

A proxy is the result of applying a certain estimator, the functional H, to the day n intraday return process $R_n(\cdot)$. Our proxies are positive, and positively homogeneous.

Definition D.4. *Let H be a measurable, positively homogeneous functional $D \to [0,\infty)$, on a linear subspace D of $\mathbb{D}[0,1]$. Assume $\Psi \in D$ a.s., and $H(\Psi) > 0$ a.s. Then H is a* proxy functional. *The random variable $H_n = H(R_n)$ is a* proxy.

Usually there is no danger of misunderstanding, and we refer to both H and H_n as proxies.

Proof of Theorem D.2. We have to show that there exists a measurable, positively homogeneous functional $H^* : D \to [0,\infty)$, with $H^*(\Psi) > 0$ *a.s.*, and $\mathrm{var}(\log(H^*(\Psi))) \leq \mathrm{var}(\log(H(\Psi)))$ for all proxy functionals H. The proof uses standard Hilbert space arguments.

For a proxy functional H, write $U = \log(H)$. Define $\lambda_H^2 = \mathrm{var}(\log(H(\Psi)))$. Let $\mathcal{U}$ denote the space of all log proxy functionals with $\lambda_H^2 < \infty$. The space $\mathcal{U}$ is not empty, by assumption. If $\mathbb{E}U(\Psi) = a \neq 0$, then $H' = e^{-a}H$ is an equally good proxy functional for which $\mathbb{E}\log(H'(\Psi)) = 0$. Therefore we may restrict attention to the subspace $\mathcal{U}^0$ of $\mathcal{U}$ of centered functionals. The space $\mathcal{U}^0$ is affine: if $U_1, U_2 \in \mathcal{U}^0$, and $w \in \mathbb{R}$, then $wU_1 + (1-w)U_2 \in \mathcal{U}^0$, since $(H^{(1)})^w (H^{(2)})^{(1-w)}$ is a proxy functional.

Define $\lambda_{inf}^2 = \inf_{H:\log(H)\in\mathcal{U}^0}\{\lambda_H^2\}$. Consider the space $\mathrm{L}^2(D,\mathcal{B})$, of equivalence classes $[U]$ of log proxy functionals U, with inner product $< [U^{(1)}],[U^{(2)}] >= \mathbb{E}\big(U^{(1)}(\Psi)U^{(2)}(\Psi)\big)$. Here, $\mathcal{B}$ denotes the Borel sigma-field for D. Notice that $\mathcal{U}^0$ is a subset of L^2 and that λ coincides with the L^2-norm $||.||$ on $\mathcal{U}^0$. Let $U_1, U_2, \ldots \in \mathcal{U}^0$ be a sequence for which $||U_i|| \to \lambda_{inf}$. Then $[U_1],[U_2],\ldots$ is a Cauchy sequence in L^2 : apply the parallelogram law to obtain

$$0 \leq ||U_m - U_n||^2 \leq -4||\frac{U_m + U_n}{2}||^2 + 2||U_m||^2 + 2||U_n||^2.$$

Since $\mathcal{U}^0$ is affine, $(U_m + U_n)/2 \in \mathcal{U}^0$, hence $||\frac{U_m+U_n}{2}||^2 \geq \lambda_{inf}^2$. Therefore $||U_m - U_n||^2 \leq -4\lambda_{inf}^2 + 2\lambda_m^2 + 2\lambda_n^2 \to 0$ for $m,n \to \infty$.

By completeness of L^2 the sequence $[U_1], [U_2], \ldots$ converges to an element $[U_0]$ in L^2 and by continuity of the norm $\lambda_0^2 = \lambda_{inf}^2$. Pick a functional $U_0 \in \mathcal{U}^0$ from $[U_0]$. Let us use U_0 to construct a functional H^* that satisfies the conditions stated at the start of the proof. For every L^2 convergent sequence there exists a subsequence that converges almost surely. Let $U_{i_k} = \log(H^{(i_k)}(f)) \to U_0(f)$ on a set C almost everywhere in D. Define on the convergence set C: $H^*(f) = \lim H^{(i_k)}(f)$. For $\{\alpha f : f \in C, \alpha f \notin C, \alpha \in [0, \infty)\}$, define $H^*(\alpha f) = \alpha H^*(f)$. For remaining $f \in D$ define $H^*(f) \equiv 0$. The functional H^* assigns a single value to each $f \in D$: consider $f_1, f_2 \in C$, $\alpha_1, \alpha_2 > 0$, and $f = \alpha_1 f_1 = \alpha_2 f_2$. Then $H^*(\alpha_1 f_1) \equiv \alpha_1 H^*(f_1) = \alpha_1 H^*(\alpha_2/\alpha_1\, f_2)$. By homogeneity of H^* on C this equals $\alpha_2 H^*(f_2) \equiv H^*(\alpha_2 f_2)$. Being the result of a limit, the functional H^* is measurable. Positive homogeneity follows by construction. Moreover, $H^*(\Psi) > 0$ almost surely, since $U_0(\Psi) \overset{a.s.}{=} \log(H^*(\Psi))$ and $\mathrm{var}(U_0(\Psi)) = \lambda_0^2 < \infty$. Finally, $\mathrm{var}(\log(H^*(\Psi))) = \lambda_{inf}^2 \leq \lambda_H^2$ for all H. □

Lemma D.5. *If H^* is an optimal proxy, and H is a proxy, then* $\mathrm{cov}\big(\log(H^*(\Psi)), \log(H(\Psi))\big) = (\lambda^*)^2$.

Proof of Lemma D.5. Consider the proxy functional $H(f) \equiv \big(H^*(f)\big)^w \big(H(f)\big)^{1-w}$, with measurement variance $\lambda_w^2 = w^2(\lambda^*)^2 + 2w(1-w)\,\mathrm{cov}\big(\log(H^*(\Psi)), \log(H(\Psi))\big) + (1-w)^2\lambda^2$. Since H^* is optimal, $\partial\lambda_w^2/\partial w\ |_{w=1} = 0$. Hence $\mathrm{cov}\big(\log(H^*(\Psi)), \log(H(\Psi))\big) = (\lambda^*)^2$. □

Proof of Proposition D.3. Both proxies have measurement variance $(\lambda^*)^2$. Let H_0 denote the centered proxy: $H_0 = \exp(-\mathbb{E}\log(H(\Psi)))\,H$, with $\mathbb{E}\log(H_0(\Psi)) = 0$. Consider the covariance of the centered log proxies: $\mathrm{cov}\big(\log(H_0^{(1)}(\Psi)), \log(H_0^{(2)}(\Psi))\big)$. By Lemma D.5 this covariance equals $(\lambda^*)^2$. By Cauchy-Schwarz this equality holds if and only if $H_0^{(1)}(\Psi) \overset{a.s.}{=} H_0^{(2)}(\Psi)$. In other words, if and only if $H^{(1)}(\Psi) \overset{a.s.}{=} aH^{(2)}(\Psi)$, for certain $a > 0$. □

D.4 Consistency Condition for the Coefficients $\hat{w}$

We provide additional discussion on consistent estimation of the optimal coefficients w^*. First some notation. Let $(X_n)_{n\in 1\ldots N}$ be a series of vectors. Let $\widehat{\mathrm{var}}(X_n)$ and $\widehat{\mathrm{cov}}(X_n)$ denote the standard empirical variance and covariance matrices of the series (X_n). Let $\mathbb{H}(R_n)$ be shorthand for the d-dimensional column vector of proxies $H^{(i)}(R_n)$, and let $\mathbb{U}_n$ denote the accompanying measurement errors. Let $\log(\mathbb{H}(R_n))$ denote the element wise logarithms. So, $\log(\mathbb{H}(R_n)) = \log(s_n) \cdot \iota + \mathbb{U}_n$.

The standard formula for the sample variance of the sum of random vectors gives:

$$\widehat{\mathrm{var}}(\log((\mathbb{H}(R_n)) = \widehat{\mathrm{var}}(\log(s_n))\iota\iota' + \widehat{\mathrm{var}}(\log(\mathbb{U}_n)) + 2 \cdot \widehat{\mathrm{cov}}(\mathbb{U}_n, \log(s_n) \cdot \iota).$$

The estimator $\hat{w}$ is given by $\hat{w} = \arg\min_w w'\widehat{\text{var}}\big(\log(\mathbb{H}(R_n))\big)w$. As in the proof of Theorem 2.1, the variance of $\log(s_n)$ drops out:

$$\hat{w} = \arg\min_w w' \left(\widehat{\text{var}}(\log(\mathbb{U}_n)) + 2 \cdot \widehat{\text{cov}}(\mathbb{U}_n, \log(s_n) \cdot \iota)\right) w.$$

If $\hat{w}$ is consistent for w^*, then asymptotically it should solve $\arg\min_w w'\Lambda w$. The term $\widehat{\text{var}}(\log(\mathbb{U}_n))$ converges to Λ for increasing sample sizes, since the measurement error vectors $\mathbb{U}_n$ are iid. So, the consistency of $\hat{w}$ comes down to the *consistency condition* that the 'sample covariance' of $\log(s_n)$ and $\mathbb{U}_n$ (recall that both $\mathbb{U}_n$ and s_n are not observed) converges to zero in probability:

$$\widehat{\text{cov}}(\mathbb{U}_n, \log(s_n) \cdot \iota) \xrightarrow{P} 0, \qquad N \to \infty. \tag{19}$$

In addition to existence of second moments and the independence of $\mathbb{U}_n$ and $\log(s_n)$, the stationarity for (s_n, Ψ_n) is a sufficient, but not necessary, condition that ensures that the consistency condition (19) holds.

References

Alizadeh, S., Brandt, M.W. and Diebold, F.X. (2002). Range-based estimation of stochastic volatility models. *The Journal of Finance*, **57,** number 3, 1047–1091.

Andersen, T.G., Bollerslev, T. and Diebold, F.X. (2008). Parametric and Nonparametric Volatility Measurement. In *Handbook of Financial Econometrics* (ed. Y. Aït-Sahalia). Forthcoming.

Andersen, T.G., Bollerslev, T., F.X., Diebold and Labys, P. (2001). The distribution of realized exchange rate volatility. *Journal of the American Statistical Association*, **96,** number 453, 42–55.

Andersen, T.G., Bollerslev, T., F.X., Diebold and Labys, P. (2003). Modeling and forecasting realized volatility. *Econometrica*, **71,** number 2, 579–625.

Barndorff-Nielsen, O.E. and Shephard, N. (2002). Estimating quadratic variation using realized variance. *Journal of Applied Econometrics*, **17,** number 5, 457–477.

Barndorff-Nielsen, O.E. and Shephard, N. (2003). Realized power variation and stochastic volatility models. *Bernoulli*, **9,** number 2, 243–65 and 1109–1111.

Barndorff-Nielsen, O.E. and Shephard, N. (2004). Power and Bipower Variation with Stochastic Volatility and Jumps. *Journal of Financial Econometrics*, **2,** number 1, 1–37.

Billingsley, P. (1999). *Convergence of Probability Measures*, 2nd edn. New York: John Wiley & Sons, Inc.

Christensen, K. and Podolskij, M. (2007). Realized range-based estimation of integrated variance. *Journal of Econometrics*.

Dacorogna, M.M., Gençay, R., Müller, U., Olsen, R.B. and Pictet, O.V. (2001). *An Introduction to High-Frequency Finance*. London: Academic Press.

Drost, F.C. and Nijman, T.E. (1993). Temporal aggregation of Garch processes. *Econometrica*, **61,** number 4, 909–927.

Drost, F.C. and Werker, B.J.M. (1996). Closing the GARCH gap: Continuous time GARCH modeling. *Journal of Econometrics*, **74,** number 1, 31–57.

Engle, R.F. (2002). New frontiers for arch models. *Journal of Applied Econometrics*, **17,** number 5, 425–446.

Hansen, P.R. and Lunde, A. (2005). A Realized Variance for the Whole Day Based on Intermittent High-Frequency Data. *Journal of Financial Econometrics*, **3,** number 4, 525–554.

Hansen, P.R. and Lunde, A. (2006a). Consistent ranking of volatility models. *Journal of Econometrics*, **131,** number 1-2, 97–121.

Hansen, P.R. and Lunde, A. (2006b). Realized variance and market microstructure noise. *Journal of Business & Economic Statistics*, **24,** number 2, 127–161.

Martens, M. and van Dijk, D. (2007). Measuring volatility with the realized range. *Journal of Econometrics*, **138,** number 1, 181–207.

McAleer, M. and Medeiros, M.C. (2008). Realized Volatility: A Review. *Econometric Reviews*, **27,** number 1, 10–45.

Meddahi, N. and Renault, E. (2004). Temporal aggregation of volatility models. *Journal of Econometrics*, **119,** number 2, 355–377.

Oomen, R.C.A. (2005). Properties of bias-corrected realized variance under alternative sampling schemes. *Journal of Financial Econometrics*, **3,** number 4, 555–577.

Oomen, R.C.A. (2006). Properties of realized variance under alternative sampling schemes. *Journal of Business & Economic Statistics*, **24,** number 2, 219–237.

Parkinson, M. (1980). The extreme value method for estimating the variance of the rate of return. *Journal of Business*, **53,** 61–65.

Patton, A.J. and Sheppard, K. (2008). Evaluating Volatility Forecasts. In *Handbook of Financial Time Series* (eds T.G. Andersen, R.A. Davis, J.P. Kreiss and T. Mikosch). Springer. Forthcoming.

Visser, M.P. (2008). Garch parameter estimation using high-frequency data. MPRA paper no. 9076.

Zhang, L., Mykland, P.A. and Aït-Sahalia, Y. (2005). A tale of two time scales: Determining integrated volatility with noisy high-frequency data. *Journal of the American Statistical Association*, **100,** number 472, 1394–1411.