Volatility Proxies for Discrete Time Models

de Vilder, Robin G. and Visser, Marcel P.

Korteweg-de Vries Instute for Mathematics, University of Amsterdam

14 September 2007

Volatility proxies for discrete time models

Robin G. de Vilder* Marcel P. Visser †

September 14, 2007

Abstract

Discrete time volatility models typically employ a latent scale factor to represent volatility. High frequency data may be used to construct proxies for these scale factors. Examples are the intraday high-low range and the realized volatility. This paper develops a method for ranking and optimizing volatility proxies. It is possible to outperform the quadratic variation as a proxy for the discrete time scale factor. For the S&P 500 index data over the years 1988-2006 this is achieved by a proxy which puts, among other things, more weight on the highs than on the lows over intraday intervals.

JEL classification: C22,C52,C65.

Keywords: volatility proxy, realized volatility, quadratic variation, scale factor, arch/garch/stochastic volatility, intraday seasonality.

*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.

1 Introduction

Volatility refers to the degree to which financial prices tend to fluctuate. It changes over time and financial time series display periods of low and high volatility. A large and widely used body of discrete time models for dealing with time varying volatility has a product structure

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

The observed financial return r_n is modelled as the product of an iid innovation Z_n and a positive scale factor s_n. The scale factor is generally called the volatility. Attention typically centers on the volatility process (s_n), specifications for which include Arch/Garch, stochastic volatility, long memory, and Markov switching.

Since only the returns r_n are observed, the values of the scale factors s_n have to be estimated. High frequency data may be used to construct proxies for the scale factor s_n, such as the intraday high-low range or the realized volatility. Proxies serve as measurements of the scale factor. Good proxies are needed for efficient parameter estimation (see, e.g., Alizadeh, Brandt, and Diebold (2002)), they are essential for volatility forecast evaluation (see, for instance, Andersen and Bollerslev (1998), and Hansen and Lunde (2006a)), and good proxies will help to improve the economic understanding of the volatility process.

There is no straightforward way to tell, for any two proxies, which one is better, since the scale factors s_n are unobservable. Much of the current research focuses on efficiently estimating the quadratic variation, which is (the limit of) the sums of squared intraday returns, see, for instance, Barndorff-Nielsen and Shephard (2002), and Andersen, Bollerslev, Diebold and Labys (2003). However, we shall see that the quadratic variation does not necessarily lead to the most efficient proxy for the discrete time scale factor s_n.

This paper proposes a way to rank and optimize proxies for the *daily scale factor* s_n, based on intraday high frequency data. We take three steps: extend the discrete time models (1) to a generic continuous time model with intraday seasonality; fix a large class of volatility proxies; and develop tools for ranking and optimizing those proxies.

We integrate high frequency data into the discrete time models (1), by replacing the iid innovations (Z_n) by iid stochastic processes (Ψ_n) on the unit time interval:

$$R_n(\vartheta) = s_n \Psi_n(\vartheta) \qquad \vartheta \in [0,1]. \tag{2}$$

The continuous time models describe, for each day, the daily log return process R_n from the opening until closing, starting from the overnight return at the opening. The process Ψ_n

may be *any* stochastic process, and is scaled by s_n resulting in a bridge between the discrete time, close-to-close returns r_n in (1). These continuous time models are consistent with both the persistence of volatility captured by discrete time volatility models, and with widely observed empirical facts such as intraday seasonalities, leverage effects, and jumps. As in the discrete time model, s_n is a state of the market representing the level of the *day-specific price sensitivity*. The *actual fluctuations* in the day n return process R_n are the result of the local fluctuations in the process Ψ_n, scaled by the day-specific factor s_n. The structure in (2) is *not* a model of constant intraday volatility: depending on Ψ_n it is possible to have a hectic day (for instance, a large quadratic variation) when s_n is low, and vice versa. For discrete time models the term volatility tends to have a slightly different meaning than it has for continuous time models. To avoid confusion we will, from now on, refer to s_n as the daily scale factor.

We proceed by formalizing the notion of a proxy in the spirit of Alizadeh, Brandt, and Diebold (2002). The definition covers popular proxies such as the absolute return, the intraday high-low range, and the realized volatility. For many applications it is helpful to have proxies with small measurement variance. We develop some easy-to-implement tools for ranking and improving proxies. In particular, proxies may be ranked by the variance of their logarithm and they may be combined into a more efficient proxy by a technique similar to the Markowitz (1952) optimization procedure. By the end of Section 3, we will have at our disposal a largely model free methodology for ranking and improving proxies: optimality of a proxy for s_n does not depend on the particular discrete time model, as long as (1) holds.

Empirical analysis of intraday S&P 500 index futures market data from January 1988 to mid 2006 shows that the techniques provide a good proxy for the S&P 500 data. In terms of the variance of the logarithm, the additional use of the high-low range over intraday intervals yields proxies that substantially outperform the realized volatility, which uses only squared intraday returns. Martens and van Dijk (2007) arrive at a similar finding in the context of estimating the quadratic variation. Moreover, our empirical results suggest that the optimized proxy is more efficient than the (square root of) the quadratic variation. Interestingly, the optimized proxy based on the highs, the lows, and 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, when proxying volatility.

The discrete time model class (1) provides a flexible way for dealing with time varying volatility. These models are important tools for the practice of risk management and asset allocation, as well enable academic researchers to investigate fundamental questions in

finance, such as the tradeoff between risk and return, see, for instance, Engle, Lilien, and Robins (1987).

Many discrete time volatility models were developed when high frequency data were not readily available, and were initially applied to daily, or lower frequency returns. The continuous time models in (2) offer a natural way to introduce high frequency data into the discrete time models and to construct proxies for the scale factors.

There have been several ways in the literature to deal with high frequency data in discrete time volatility models. The first one is to apply the model to a five-minute grid, or even study the continuous time limit and to infer the implications for the daily or weekly sampling frequency, see, for example, Drost and Nijman (1993), Drost and Werker (1996), and Meddahi and Renault (2004). In empirical applications this approach leads to serious time aggregation problems: the parameters estimated using very short intervals, typically turn out to be internally inconsistent with those obtained from daily or lower frequencies. These problems have often led to the impression that the discrete time model structure (1) is fundamentally flawed and cannot be made consistent with the data. Andersen and Bollerslev (1997) argue, on the other hand, that intraday seasonality causes a daily barrier below which stationary models should not be applied. Their daily barrier agrees with our choice for the business day as a unit of time in the models (1) and (2). They propose a model for five-minute returns, which explicitly takes into account intraday periodicities on top of the daily volatility s_n. Alizadeh, Brandt, and Diebold (2002) assume an intraday Brownian motion, and use the daily high-low range for quasi maximum likelihood estimation of a discrete time stochastic volatility model. They obtain improved estimators due to the high-low range being superior to absolute or squared daily returns as a proxy for the volatility s_n.

The remainder of the paper is organized as follows. Section 2 provides a continuous time extension of the discrete time models. Section 3 introduces proxies, and develops tools for ranking and optimizing them. Section 4 constructs a good proxy for the S&P 500 data. In Section 5 we conclude. Appendices A, B, and C contain a description of the data, the empirical technique of prescaling and a number of proofs, respectively.

2 The Continuous Time Models

In the classical Black Scholes world the log return process is a continuous time Brownian motion σB_t with constant volatility $\sigma > 0$. This leads to discrete time returns

$$r_n = \sigma Z_n, \tag{3}$$

where the Z_n are iid standard Gaussian random variables. The Black Scholes model has been extended to deal with time varying volatility. In the continuous time extensions the term volatility refers to either the instantaneous diffusion coefficient, or the (square root of) the quadratic variation over a given time period. The *discrete time* generalizations replace the constant σ in (3) by a sequence (s_n) of strictly positive random variables:

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

The innovations (Z_n) are iid, and the innovation Z_n is independent of s_n[1]. The term volatility now refers to the scale factor s_n, which generally differs from the continuous time concept of daily volatility.

In order to study intraday based proxies for the scale factor s_n, we integrate high frequency data into the discrete time models by replacing the iid innovations (Z_n) by iid stochastic processes (Ψ_n) :

$$R_n(\vartheta) = s_n \Psi_n(\vartheta), \qquad \vartheta \in [0,1]. \tag{2}$$

The *day n return process* R_n is the log return process with respect to the previous day's close. The return process R_n is the product of the daily scale factor s_n and a stochastic process Ψ_n. The process Ψ_n is a cadlag process[2] representing the intraday price pattern. The *time of day* ϑ is in $[0,1]$. The processes Ψ_n are iid. The sequence of scale factors (s_n) may be any strictly positive stochastic process, as long as the scale factor s_n is independent of all (Ψ_k) for $k = n$ up to infinity. In particular, the scale factor s_n and the process Ψ_n are independent, for each day n. The continuous time model structure (2) will be referred to as the *scaling hypothesis*. A model that satisfies the scaling hypothesis leads to the well-known product structure (1) for the daily close-to-close returns,

$$r_n = R_n(1) = s_n \Psi_n(1).$$

The continuous time models have the following interpretation. The discrete time scale factor s_n describes the state of the market on day n, and represents the level of overall price sensitivity. The factor s_n may be determined by the process over the past days, (s_{n-k}, Ψ_{n-k}), $k \geq 1$, for instance like a Garch model. It may also include an external source of randomness. The actual fluctuations in the process Ψ_n determine the pattern of the intraday return

[1]For identification, the innovations are typically assumed mean zero, unit variance. Such identification assumptions will not be necessary in this paper.

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

process, such as up or down days, quiet or hectic days. The process Ψ_n offers a flexible way of taking into account intraday seasonality. It may also have, for instance, leverage effects, jumps, stochastic spot volatility.

As an example, consider the case that Ψ is a diffusion with stochastic instantaneous volatility $v(\vartheta)$,

$$d\Psi(\vartheta) = v(\vartheta)dB(\vartheta),$$

which implies for the n-th day return process R_n,

$$dR_n(\vartheta) = s_n v_n(\vartheta)\ dB_n(\vartheta), \qquad \vartheta \in [0,1]. \tag{4}$$

So, under the scaling hypothesis, in the often used diffusive model for the log price process $p(t)$,

$$dp(t) = \sigma(t)dB(t),$$

the instantaneous volatility $\sigma(t)$ allows a decomposition into a daily scale factor s_n and an independent intraday stochastic volatility component $v_n(\vartheta)$, modelling intraday effects such as seasonalities. The intraday volatility processes (v_n) are iid and independent of events up to the close of the n-1$^{\text{st}}$ day[3].

Let us provide the remaining formal details for the continuous time process R_n. The trading hours are normalized to cover the interval $[0,1]$. Time advances only during trading hours. Let Ψ be a cadlag process on the closed interval $[0,1]$, left continuous in 1. The processes (Ψ_n) are independent copies of the process Ψ. In order to describe the dependence structure, it is convenient to introduce the discrete time *model filtration* $(\mathcal{G}_n)$. The σ-field $\mathcal{G}_n$ includes the history of (s_n, Ψ_n) extended with s_{n+1}. So, $\mathcal{G}_n = \sigma\{(\Psi_i)_{i\le n},\ (s_i)_{i\le n+1}\}$. The σ-field $\mathcal{G}_n$ represents the model information at the start of day $n+1$. The model information includes the observable day n return process, as $R_n = s_n\Psi_n$. *The process Ψ_{n+1} is independent of $\mathcal{G}_n$, and the processes (Ψ_n) are identically distributed.*

Note that $\Psi(1)$ is not standardized[4]. More generally, processes that satisfy the scaling

[3]Andersen and Bollerslev (1997) propose to model the seasonal intraday volatility in five-minute returns r_{n,ϑ_i} on day n as the product of a daily scale factor s_n and deterministic intraday components (v_{ϑ_i}) : $r_{n,\vartheta_i} = s_n \cdot v_{\vartheta_i} \cdot \varepsilon_{n,\vartheta_i}, \quad \vartheta_i = i/K, \quad i = 1,\dots,K$, where $\varepsilon_{n,\vartheta_i}$ is an iid mean zero, unit variance error. Their model is a special case of the continuous time model (4), discretized to five-minute returns.

[4]If $\Psi(1)$ is not standardized, then the discrete time returns satisfy $r_n = a\sigma_n + \sigma_n\varepsilon_n$, where the ε_n are iid(0,1). So the continuous time model (2) also embeds a simple version of the ARCH-M model introduced by Engle, Lilien, and Robins (1987).

hypothesis do not have a unique representation. Consider a representation (s_n, Ψ_n) for (R_n) that satisfies the scaling hypothesis. Set $s'_n = s_n/2$ and $\Psi'_n = 2\Psi_n$. Then

$$R'_n \equiv R_n. \tag{5}$$

As we will see later, identification of s_n and Ψ_n is not necessary for the study of proxies.

3 Proxies

The daily scale factor s_n in the continuous time model (2) is difficult to estimate, even if we observe the full sample path of the asset price. This stands in contrast to the situation of obtaining a perfect estimate for the *diffusion coefficient* using the full sample path, as Merton (1980) discusses. We will work with (ex post) statistics of the sample path to estimate s_n. Let us introduce the concept of volatility proxy in the spirit of Alizadeh, Brandt, and Diebold (2002).

Consider volatility proxies that are positive functionals $H(R_n)$ of the day n return process R_n. If the functional H is *positively homogeneous*,

$$H(\alpha R_n) = \alpha H(R_n), \qquad \alpha \in [0, \infty),$$

then a *proxy* $H(R_n)$ is linear in s_n:

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

Applying logarithms leads to a measurement equation: the log of a proxy is the sum of two independent terms, the log of the scale factor and a *measurement error* $U_n = \log(H(\Psi_n))$:

$$\log(H(R_n)) = \log(s_n) + U_n. \tag{6}$$

The measurement errors U_n form an iid sequence. Naturally the quality of the proxy is related to the bias and the variance of the measurement error U_n. Rewriting equation (6),

$$\log(H(R_n)) = \log(s_n) + \mathbb{E}U_n + \big(U_n - \mathbb{E}U_n\big),$$

makes clear that for a given functional H the measurement error introduces a *constant* bias, $\mathbb{E}U_n = \mu$. In applications where the proxy is used as a variable in a regression, the bias will be corrected by the regression parameters. It is also possible to rescale the proxy, replacing H by aH, in order to obtain a bias-corrected version. So, the key determinant of the quality

is the *measurement variance* λ^2,

$$\lambda^2 = \mathrm{var}(U_n). \tag{7}$$

An optimal proxy functional H^* will satisfy

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

For a proxy functional H, the measurement error U_n only depends on the process Ψ_n. This means that the optimality of a proxy functional is independent of the particular discrete time model for the scale factors (s_n).

3.1 Definition and Basic Properties

Let us now provide the formal details for volatility proxies. Recall from Section 2 that the process Ψ is cadlag on $[0,1]$. Let $\mathbb{D}[0,1]$ denote the Skorohod space of cadlag functions on $[0,1]$, which are left continuous in 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 $\mathbb{D}[0,1]$.

From now on we assume the scaling hypothesis for the day n return process: $R_n(\vartheta) = s_n\Psi_n(\vartheta)$. A proxy is the result of applying a certain estimator, the proxy functional H, to the day n return process R_n. We restrict attention to positively homogeneous functionals in order to ensure that the decomposition in (6) holds.

Definition 3.1. *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 $H(R_n)$ is a* proxy.

Remark 1. *Proxy functionals are nonlinear: $H(\Psi) + H(-\Psi) > 0$ a.s., but $H(\Psi - \Psi) = H(0) = 0$; proxy functionals need not be symmetric: $H(\Psi) \neq H(-\Psi)$, in general.*

Example 3.1.1. *Here are some examples of proxies: the absolute daily return; the absolute overnight return; the high-low range; realized volatility, defined for any grid as the square root of the sum of the squared returns over the grid; maximal absolute two-minute return; the square root of the daily increment in quadratic variation. In these cases, if $\Psi(1)$ has a density, the almost sure positivity of $H(\Psi)$ is ensured.*

Combining different proxies through a positively homogeneous function, generates a new proxy. Write $x >> 0$ if $\min_i x_i > 0$.

Proposition 3.2. *Let $H^{(i)},\ i = 1, \ldots, d$, be proxy functionals on D. Let $G : [0, \infty)^d \to [0, \infty)$ be a measurable, positively homogeneous function. Moreover, assume $G(x) > 0$ for $x >> 0$. Then the functional $H : f \mapsto G(H^{(1)}(f), \ldots, H^{(d)}(f))$ is a proxy functional on D.*

Proof. Since every $H^{(i)}(\Psi) > 0$ a.s. we have $(H^{(1)}(\Psi), \ldots, H^{(d)}(\Psi)) >> 0$ a.s., so $H(\Psi) > 0$ a.s. Homogeneity holds, since

$$H(\alpha f) = G(H^{(1)}(\alpha f), \ldots, H^{(d)}(\alpha f)) = G(\alpha H^{(1)}(f), \ldots, \alpha H^{(d)}(f)) = \alpha H(f).$$

□

As a consequence, given two proxy functionals $H^{(1)}$ and $H^{(2)}$, it is possible to obtain a new proxy functional, for example, by scaling one of them with a positive number, by adding them, or by taking the maximum or the minimum. One may also take a geometric combination:

$$H(R_n) \equiv (H^{(1)}(R_n))^{w_1}\ (H^{(2)}(R_n))^{w_2}, \qquad w_1, w_2 \in \mathbb{R},\ w_1 + w_2 = 1. \tag{8}$$

Scaling a proxy with a positive number $a > 0$ does not change the measurement variance λ^2 :

$$\mathrm{var}(\log(aH(\Psi))) = \mathrm{var}(\log(H(\Psi))) = \lambda^2. \tag{9}$$

Assume $0 < \mathrm{var}(\log(s_n)) < \infty$. It then follows, using the decomposition (6), that

$$\mathrm{corr}(\log(H(R_n)), \log(s_n)) = \big(1 + \frac{\lambda^2}{\mathrm{var}(\log(s_n))}\big)^{-1/2}. \tag{10}$$

This means that proxies with smaller measurement variance λ^2 have larger correlation with $\log(s_n)$. The ideal situation of zero measurement variance gives perfect correlation. The proxy is then a perfect proxy:

$$H(R_n) = cs_n,$$

for a certain constant $c > 0$. Such perfect proxies do exist, in special cases. For example, if Ψ is the standard Brownian motion on $[0, 1]$. The quadratic variation of Ψ then equals one, and s_n is the square root of the quadratic variation: $\sqrt{QV_n} = s_n$. If a proxy has zero measurement variance, then one knows the value of s_n, without knowing the model for the sequence (s_n). The following example shows that the square root of the quadratic variation is not necessarily the most efficient proxy for s_n.

Example 3.1.2. *Consider the case that Ψ is a diffusion as in (4):*

$$d\Psi(\vartheta) = v(\vartheta)dB(\vartheta),$$

where B denotes standard Brownian motion. Let the instantaneous volatility process $v(\vartheta)$ be deterministic at the opening and stochastic for the rest of the day. More specifically, suppose $v(\vartheta)$ equals 1 before time of day $\vartheta_0 = 1/2$, and $v(\vartheta)$ equals either c_1 or c_2 after ϑ_0, both with probability $1/2$. The square root of the truncated quadratic variation over $[0, 1/2]$ equals s_n times a constant, and as such is perfect. The square root of the quadratic variation of R_n is the product of s_n and a random variable with positive variance

Estimating the daily scale factor s_n is a different exercise from estimating (the square root of) the quadratic variation. The quadratic variation is a measure for the price fluctuations that actually occurred during the whole trading day. As an estimator of the daily scale factor s_n, the square root of the quadratic variation is merely one out of many possible candidates.

Let us return to the issue of identification. The following proposition states that different representations (s_n, Ψ_n) for the scaling hypothesis result in the same ordering for proxy functionals. So, for our purposes identification plays no role.

Proposition 3.3. *Suppose $H^{(1)}$ and $H^{(2)}$ are proxy functionals. Moreover, assume (s_n, Ψ_n) and (s'_n, Ψ'_n) both satisfy the scaling hypothesis for R_n. If $H^{(1)}$ is better than $H^{(2)}$ for Ψ, then $H^{(1)}$ is also better than $H^{(2)}$ for Ψ'.*

Proof. See appendix C. □

3.2 Existence of Optimal Proxies

Recall that an optimal proxy functional H^* satisfies

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

Theorem 3.4. *If there exists a proxy functional with finite measurement variance, then there exists an optimal proxy functional.*

Proof. See appendix C. □

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

Proposition 3.5. *Suppose $H^{(1)}$ and $H^{(2)}$ are two optimal proxy functionals. Then there exists a constant $a > 0$, such that $H^{(1)}(\Psi) \stackrel{a.s.}{=} aH^{(2)}(\Psi)$.*

Proof. See appendix C. □

3.3 Empirically Ranking Proxies

This section provides a practical way to compare proxies. By definition, a proxy functional $H^{(1)}$ is more efficient than $H^{(2)}$ if it has smaller measurement variance:

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

Comparison by measurement variances is infeasible in an empirical situation: the measurement variances cannot be estimated, since the processes (Ψ_n) are not observed. However, taking variances on both sides of the decomposition (6) gives

$$\text{var}(\log(H^{(i)}(R_n))) = (\lambda^{(i)})^2 + \text{var}(\log(s_n)). \tag{11}$$

There is no covariance term by the independence of s_n and Ψ_n. Equation (11) shows that the variances of the proxies all have the common term $\text{var}(\log(s_n))$. It follows that if the variance of the log proxy is smaller, then the measurement variance must be smaller. So, proxies may be ranked by the variances of the log proxies. Let us summarize this in a proposition.

Proposition 3.6. *Let $H^{(1)}$ and $H^{(2)}$ be two proxy functionals. Assume $U^{(i)} = \log(H^{(i)}(\Psi))$, $i = 1, 2$, and $\log(s_n)$ have finite variances. Assume R_n satisfies the scaling hypothesis. Then*

$$\text{var}(\log(H^{(1)}(R_n))) - \text{var}(\log(H^{(2)}(R_n))) \;=\; (\lambda^{(1)})^2 - (\lambda^{(2)})^2.$$

The autocorrelation function is an additional indicator of the quality of a proxy, see Figure 2 below. Let $\rho_\cdot(j)$ denote j-th order autocorrelation. If the process (s_n, Ψ_n) is stationary, then, using (6),

$$\begin{aligned} \rho_{\log(H(R_0))}(j) &= a_\lambda \cdot \rho_{\log(s_0)}(j) + e_\lambda(j) \\ &\approx a_\lambda \cdot \rho_{\log(s_0)}(j). \end{aligned} \tag{12}$$

where $a_\lambda \in (0, 1]$, see Appendix C. The proportion factor a_λ is close to unity if the measurement variance λ^2 is small. In practice, the correction term $e_\lambda(j)$ vanishes for $j \to \infty$, and the autocorrelations $\rho_{\log(s_0)}(j)$ are large and decay slowly. Under suitable assumptions the

ratio of the autocorrelation functions for different proxies converges,

$$\frac{\rho_{\log(H^{(1)}(R_0))}(j)}{\rho_{\log(H^{(2)}(R_0))}(j)} \quad \to \quad \frac{\mathrm{var}(\log(s_0)) + (\lambda^{(2)})^2}{\mathrm{var}(\log(s_0)) + (\lambda^{(1)})^2}, \qquad j \to \infty. \tag{13}$$

This means that better proxies have *larger* autocorrelations. See Appendix C for details.

3.4 Improving Proxies

This section provides a way to combine given proxy functionals $H^{(1)}, \ldots, H^{(d)}$ into a more efficient proxy functional. Consider the geometric combination of these functionals,

$$H^{(w)}(R_n) = \prod_{i=1}^{d} (H^{(i)}(R_n))^{w_i}, \qquad w_1 + \ldots + w_d = 1,\ w_i \in \mathbb{R}. \tag{14}$$

Here, the column vector w is the d-dimensional coefficient vector. The restriction $\sum w_i = 1$ is needed to obtain a proxy functional, but the coefficients are not restricted to the interval $[0, 1]$. Let Λ denote the covariance matrix of the measurement errors $U^{(i)} = \log(H^{(i)}(\Psi))$:

$$\Lambda = \mathrm{cov}([U^{(1)}, \ldots, U^{(d)}]'). \tag{15}$$

The measurement variance λ_w^2 of the geometric combination $U^{(w)}$ in (14) is $\lambda_w^2 = w'\Lambda w$ and, as for the minimal variance portfolio in Markowitz portfolio theory, λ_w^2 is minimal for

$$w^* = \frac{\Lambda^{-1}\iota}{\iota'\Lambda^{-1}\iota}, \qquad \iota = (1, \ldots, 1)', \tag{16}$$

with optimal variance $\lambda_{w^*}^2 = \frac{1}{\iota'\Lambda^{-1}\iota}$.

This solution is empirically infeasible: it is impossible to estimate the variance matrix Λ, since the measurement errors $U_n^{(i)}$ are not observed. Let $\Lambda_{p,n}$ denote the covariance matrix of the log proxies:

$$\Lambda_{p,n} = \mathrm{cov}([\log(H^{(1)}(R_n)) \ldots \log(H^{(d)}(R_n))]'). \tag{17}$$

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

$$\Lambda_{p,n} = \Lambda + \mathrm{var}(\log(s_n))\ \iota\iota'. \tag{18}$$

Assuming stationarity for (s_n), the covariance matrix $\Lambda_p = \Lambda_{p,n}$ may be estimated from the data. Formula (19) in Theorem 3.7 below shows how to obtain the optimal coefficients w^* in (16) from the covariance matrix $\Lambda_{p,n}$.

Theorem 3.7. *Let R_n satisfy the scaling hypothesis. Assume* $\mathrm{var}(\log(H^{(i)}(\Psi))) < \infty$ *for $i = 1, \ldots, d$. Let the covariance matrices Λ and $\Lambda_{p,n}$ be defined by (15) and (17), respectively. The optimal coefficient vector w^* in (16) does not depend on the form of the process (s_n) and may be expressed as*

$$w^* = \frac{\Lambda_{p,n}^{-1}\iota}{\iota'\Lambda_{p,n}^{-1}\iota}. \tag{19}$$

Let $\lambda_{w^*}^2 = \frac{1}{\iota'\Lambda^{-1}\iota}$. The variance of the logarithm of the optimal geometric proxy is

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

Proof. The optimal coefficient w^* does not depend on (s_n): it follows from equation (18) that

$$\arg\min_w \; w'\Lambda_{p,n}w = \arg\min_w \; \big(w'\Lambda w + \mathrm{var}(\log(s_n))\big) = \arg\min_w \; w'\Lambda w. \tag{20}$$

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 (20) 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

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

□

Remark 2. *In empirical applications one uses estimates of the variances. In order to reduce estimation error, we shall use the technique of prescaling, see Appendix B.*

Assuming stationarity for the process (s_n, Ψ_n), the covariance matrix Λ_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^*. However, this estimator for w^* may remain consistent while allowing, for example, for structural breaks in the scale factors (s_n). See the consistency condition (22) in Appendix C for more details.

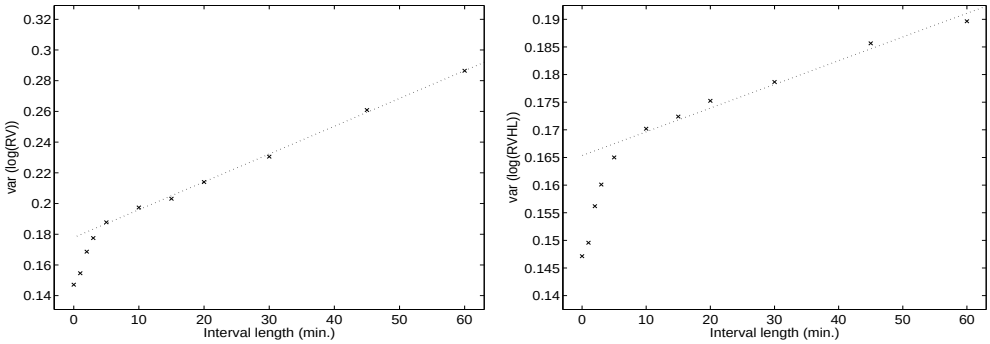
4 A good proxy for the S&P 500

This section applies the techniques of Section 3 to the S&P 500 futures tick data. Appendix A describes the data.

4.1 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 (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 this paper the measure of comparison is the variance of the logarithm. The standard realized volatility RV and the high-low based realized volatility $RVHL$ (see Table 1) depend on the sampling interval $\Delta\vartheta$. Figure 1 shows the graph of $\Delta\vartheta \to \widehat{\text{var}}(\log(H^{\Delta\vartheta}(R_n)))$, for $\Delta\vartheta$ ranging from zero to sixty minutes. These curves suggest that a qualitative change of behaviour occurs for $\Delta\vartheta \approx$ five minutes for realized volatility, and $\Delta\vartheta \approx$ eight minutes for realized high-low. From now on realized volatility is based on five-minute sampling intervals or larger. For realized high-low our minimal sampling interval will be ten minutes.



(a) Realized volatility (b) Realized high-low

Figure 1: Plots of the sample variance of the log of a proxy with $\Delta\vartheta$ ranging from zero to 60 minutes (zero is tick per tick). (a) Realized volatility. (b) Realized high-low.

4.2 Ranking Proxies

Table 1 compares twelve simple proxies constructed from the data. We emphasize that these proxies are a set of twelve out of many possible proxies. They are of no special importance

themselves.

	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

Table 1: Performance of twelve proxies. The full sample is split into four subsamples. Prescaling by EWMA(0.7) predictor for RV5. The following proxies are included. RV5: root of sum of squared 5 min returns; RV10: root of sum of squared 10 min returns; RV15: root of sum of squared 15 min returns; RV20: root of sum of squared 20 min returns; RV30: root 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-minute returns; RAV5: sum of absolute 5-minute intraday returns; RAV10: sum of absolute 10-minute intraday returns; RVHL10: root of sum of 10-minute squared high-lows; RAVHL10: sum of 10-minute high-lows;

	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

Table 2: Performance of upward/downward decomposed proxies. The table contains proxies from Table 1 splitted according to upward and downward price movements. For example, RV5-up is root of sum of squared 5 min *positive* returns, RAV10HIGH is the sum of 10-minute highs, and RAV10LOW is the sum of 10-minute absolute 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 after prescaling, $PV = \mathrm{var}(\log\big(H(R_n)/p_n\big))$; see Section 3.3 and Appendix B. The first observation cannot be prescaled and is left out of the variance computations. Smaller variances correspond to more efficient proxies. 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. The prescaled variance ranks proxies, but its value is not a measure for the quality of a proxy.

The first column of Table 1 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 the fact that absolute or squared daily returns are poor proxies[5]. Note that the maximal absolute two-minute return is better than high minus low, 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 realized quadratic variation. 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 in the various subsamples is the same as in the full sample, with one exception in the second subsample for $RV30$ and $maxar2$.

Table 2 provides the prescaled variances for the *upward* and *downward* components of a number of proxies. For instance, the five-minute realized volatility is decomposed according to upward and downward price movements:

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

Note that RV5 (PV=0.064) is better than RV5-up (PV=0.066) and RV5-down (PV=0.094). 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 *positive* returns.

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

4.3 Optimized Proxy

Let us now combine the proxies in Tables 1 and 2 into a more efficient one using the technique of Section 3.4. The five-minute realized volatility RV5 has PV=0.064. Let us improve upon this value. First, by using the high-low range over intraday intervals, RVHL10 has $PV = 0.053$, see Table 1. It is even better to use absolute values: RAVHL10 has $PV = 0.047$. Now use the theory of Section 3.4 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 (19), yields the proxy

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

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

$$H^{(\hat{w})}(R_n) = (RAV10HIGH_n)^{1.04}(RAV10LOW_n)^{0.72}(RAV10_n)^{-0.76} \qquad (PV = 0.038). \quad (21)$$

Of course, one may also apply the optimal coefficient formula (19) to all twenty-one proxies in Tables 1 and 2 at once[6]. The full combination yields a proxy with $PV = 0.037$, which only marginally outperforms the proxy obtained in (21). We prefer to work with the simpler proxy in (21) and we will refer to it as the *optimized proxy* $H^{(\hat{w})}(R_n)$.

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 it is a more efficient proxy for the daily scale factor than the square root of the quadratic variation.

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 they all 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 $H^{(\hat{w})}(R_n)$. Similarly to Table 1 it reports performance measures for the full sample and for four subsamples. The first row reports the performance of $H^{(\hat{w})}(R_n)$ in the different subsamples; comparison with Tables 1 and 2 shows that $H^{(\hat{w})}(R_n)$ outperforms all those proxies in every subsample. The proxy $H^{(\hat{w},i)}(R_n)$ is constructed using the coefficients that are optimal for the i-th subsample. In

[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

Table 3: Geometric proxies, optimized for different subsamples: performance and stability.

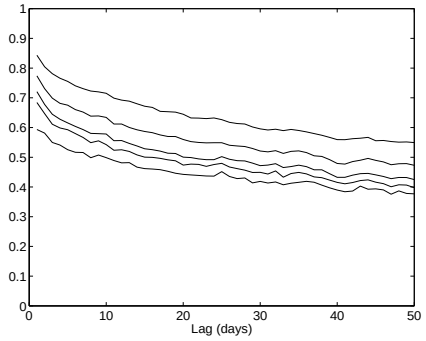


Figure 2: Autocorrelations of log proxies. Lags 1 to 50 days. From bottom to top: RV30, RV15, RV10, RV5, $H^{(\hat{w})}$.

the first subsample the performance of the globally optimized $H^{(\hat{w})}(R_n)$ (PV=0.039) is not substantially improved by $H^{(\hat{w},1)}(R_n)$ (PV=0.039). A similar statement holds for the other subsamples. Moreover, proxies based on a 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 $H^{(\hat{w})}(R_n)$ is stable.

Figure 2 shows the autocorrelations of $\log(H^{(\hat{w})}(R_n))$ and the log of four different realized volatilities, RV30, RV15, RV10, and RV5. The autocorrelations of the different realized volatilities increase as the sampling interval decreases. The autocorrelations for the optimized proxy are substantially larger than those for five-minute realized volatility. Hence, formula (13) provides an additional indication that $H^{(\hat{w})}$ is the best.

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:

$$\log(H^{(j)}(R_n)) = \alpha + \beta \log(H^{(i)}(R_{n-1})) + \varepsilon_n.$$

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 for daily volatility. The R^2's attain their maximum at the optimized proxy, in the most right column.[7]

Finally, Figure 3 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

[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 row $H^{(\hat{w})}(-1)$ are large.

	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 $\log(H^{(j)}(R_n)) = \alpha + \beta \log(H^{(i)}(R_{n-1})) + \varepsilon_n$, for $i, j = 1, \ldots, 6$, and $n = 2, \ldots, 4575$.

curves become 'less erratic', suggesting a decrease in the measurement errors U_n. Each step shows a marked improvement.

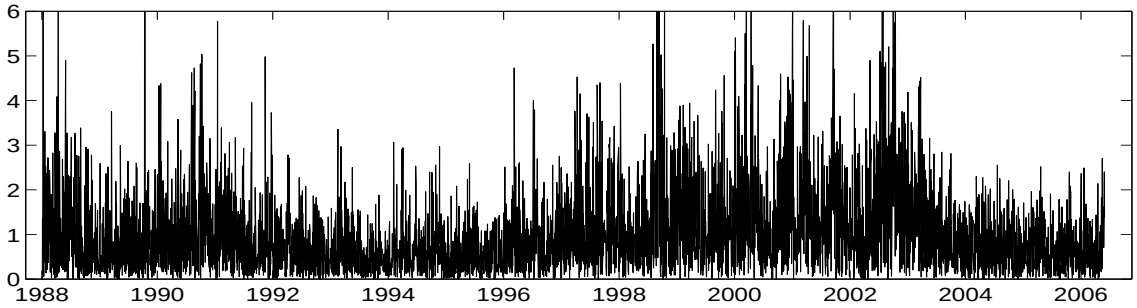
5 Conclusions

This paper provides a theoretical basis for ranking and optimizing volatility proxies, based on intraday data. The theory is founded on a natural class of continuous time extensions of discrete time, daily volatility models, taking into account intraday seasonality. Good proxies are needed for parameter estimation and forecast evaluation of discrete time volatility models.

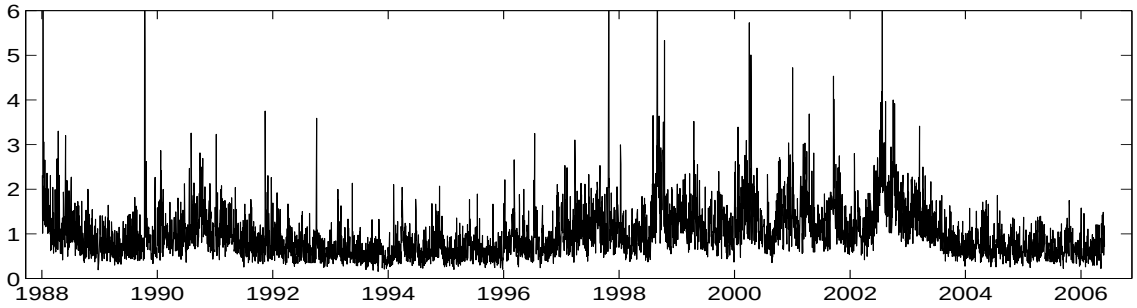
In this paper a volatility proxy is the result of applying a positively homogeneous functional to the intraday return process. By definition, a good proxy has small measurement variance. We show that optimal proxies exist and provide easy-to-implement tools for ranking and improving proxies. The approach is to a large extent model free: an optimal proxy for the scaling factor s_n is an optimal proxy under all possible discrete time models of the form $r_n = s_n Z_n$, where the Z_n are iid innovations, and the innovation Z_n is independent of the scale factor s_n.

For the S&P 500 data a combination of the highs, the lows, and the absolute returns over ten-minute intervals yields a good proxy. One should put more weight on the highs than on 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 continuous sampling.

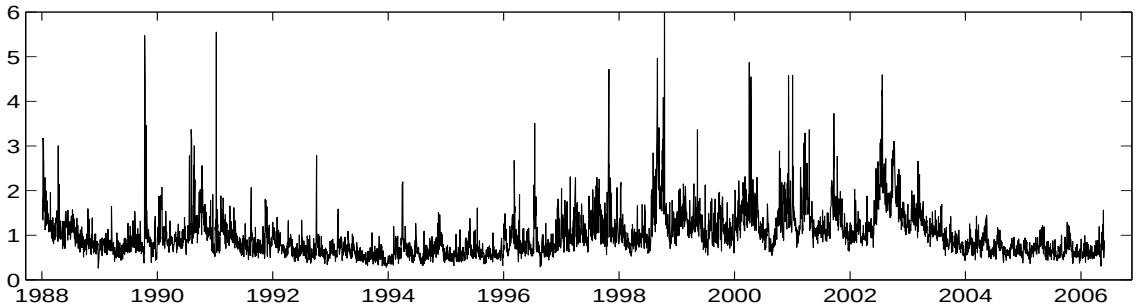
This paper has addressed the problem of ranking and optimizing proxies for today's scale factor s_n. We see opportunities to use proxies for the specification of the daily volatility process (s_n), and accordingly use proxies to forecast future volatility. We aim to pursue these ideas in future research.



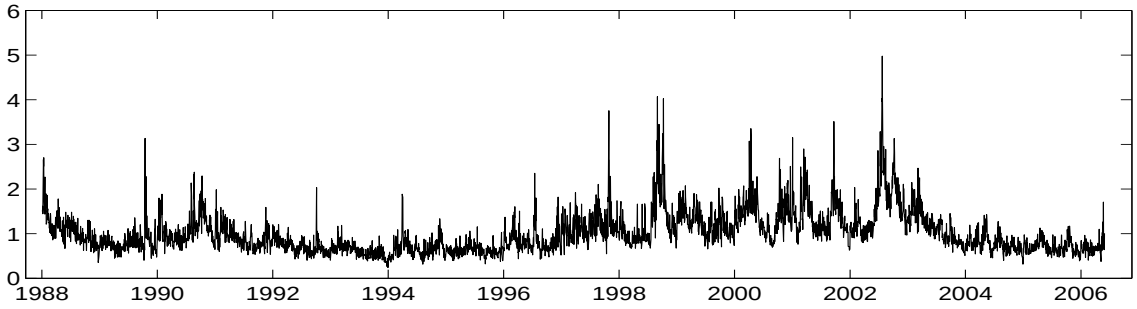
(a) Absolute close-to-close returns



(b) Intraday high-low range



(c) Five-minute realized volatility RV5



(d) Optimized proxy from formula (21)

Figure 3: Time series of four standardized proxies, $H(R_n)/\bar{H}(R_n)$.

A Data

Our data set is the U.S. Standard & Poor's 500 stock index future, traded on 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 when 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 are 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 Section 4.1. Since we study a very liquid asset the error term due to microstructures is relatively small.

B Prescaling

The methods of comparing proxies in Proposition 3.6 and of improving proxies in Theorem 3.7 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 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 B.1. *A* prescaling sequence (p_n) *is an* $(\mathcal{F}_{n-1})$ *adapted sequence of strictly positive random variables.*

The prescaling factors p_n will be used to define adjusted scale factors

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

Proposition B.2. *Assume the processes* (R_n) *satisfy the scaling hypothesis. 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 hypothesis.*

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 hypothesis. □

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

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

The proxy $H(\tilde{R}_n)$ for $\tilde{s}_n$ has the same measurement error as the proxy $H(R_n)$ for s_n :

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

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 $H(R_n)$ by $H(\tilde{R}_n)$ in Proposition 3.6 and Theorem 3.7. As a consequence, the population value of the noise term $\mathrm{var}(\log(s_n))$ in equations (11) and (18) changes into

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

A good predictor p_n of s_n results in a small term $\mathrm{var}(\log(\tilde{s}_n))$.

C Proofs

Proof of Proposition 3.3. 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 $\mathrm{var}(\log(H^{(1)}(\Psi'_n))) \leq \mathrm{var}(\log(H^{(2)}(\Psi'_n)))$. □

Proof of Theorem 3.4. We have to show that there exists a measurable, positively homogeneous functional $H^* : D \to [0, \infty)$, with $H^*(\Psi) > 0$ *a.s.*, and $\text{var}(\log(H^*(\Psi))) \leq \text{var}(\log(H(\Psi)))$ for all proxy functionals H.

For a proxy functional H, write $U = \log(H)$. Define $\lambda_H^2 = \text{var}(\log(H(\Psi)))$. 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, see equation (8).

Define $\lambda_{inf}^2 = \inf_{H:\log(H)\in\mathcal{U}^0}\{\lambda_H^2\}$. Consider the space $\text{L}^2(D, \mathcal{B})$, 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 $U_0(\Psi) \stackrel{a.s.}{=} \log(H^*(\Psi))$ and $\text{var}(U_0(\Psi)) = \lambda_0^2 < \infty$. Finally, $\text{var}(\log(H^*(\Psi))) = \lambda_{inf}^2 \leq \lambda_H^2$ for all H. □

Lemma C.1. *If H^* is an optimal proxy functional, and H is a proxy functional, then* $\text{cov}\big(\log(H^*(\Psi)), \log(H(\Psi))\big) = (\lambda^*)^2$.

Proof of Lemma C.1. 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 3.5. Both proxy functionals 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 C.1 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 $a > 0$. □

Autocorrelation formulas (12) and (13). Correlation calculations yield for a_λ and $e_\lambda(j)$:

$$a_\lambda = \frac{\mathrm{var}(\log(s_0))}{\mathrm{var}(\log(s_0)) + \lambda^2},\ \text{and}\ e_\lambda(j) = \mathrm{corr}(\log(s_0), U_{-j}) \cdot \frac{\sqrt{\mathrm{var}(\log(s_0))}\ \lambda}{\mathrm{var}(\log(s_0)) + \lambda^2}.$$

The proportion factor $a_\lambda \in (0, 1]$ is large when the measurement variance λ^2 is small. The correction term $e_\lambda(j)$ depends on the impact of $U_{-j} = \log(H(\Psi_{-j}))$ on $\log(s_0)$. The factor to the right hand side of the multiplication dot is in $[0, 1]$. The ratio of two autocorrelation functions equals

$$\frac{\rho_{\log(H^{(1)}(R_0))}(j)}{\rho_{\log(H^{(2)}(R_0))}(j)} = \frac{a_{\lambda^{(1)}} \cdot \rho_{\log(s_0)}(j) + e_{\lambda^{(1)}}(j)}{a_{\lambda^{(2)}} \cdot \rho_{\log(s_0)}(j) + e_{\lambda^{(2)}}(j)}.$$

Assuming that the impact of U_{-j} is small compared to the autocorrelation in $\log(s_0)$:

$$\mathrm{corr}(U_{-j}^{(i)}, \log(s_0))/\mathrm{corr}(\log(s_{-j}), \log(s_0)) \to 0 \qquad j \to \infty,$$

gives, upon dividing both numerator and denominator by $\rho_{\log(s_0)}(j)$ (assuming $\rho_{\log(s_0)}(j) \neq 0$ for all $j > 0$),

$$\frac{\rho_{\log(H^{(1)}(R_0))}(j)}{\rho_{\log(H^{(2)}(R_0))}(j)} \to \frac{a_{\lambda^{(1)}}}{a_{\lambda^{(2)}}} = \frac{\mathrm{var}(\log(s_0)) + (\lambda^{(2)})^2}{\mathrm{var}(\log(s_0)) + (\lambda^{(1)})^2}, \qquad j \to \infty. \tag{13}$$

□

Consistently estimating the optimal coefficients. 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), respectively. 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:

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

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 3.7, 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 converges to zero in probability:

$$\widehat{\text{cov}}(\mathbb{U}_n, \log(s_n) \cdot \iota) \xrightarrow{P} 0, \qquad N \to \infty. \tag{22}$$

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 which ensures that the consistency condition (22) holds. □

References

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

Andersen, T.G. and Bollerslev, T. (1997). Intraday periodicity and volatility persistence in financial markets. *Journal of Empirical Finance*, **4,** 115–158.

Andersen, T.G. and Bollerslev, T. (1998). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review*, **39,** number 4, 885–905.

Andersen, T.G., Bollerslev, T., Diebold, F.X. 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.

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

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., Lilien, D.M. and Robins, R.P. (1987). Estimating time varying risk premia in the term structure: The ARCH-M model. *Econometrica*, **55,** number 2, 391–407.

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.

Markowitz, H. (1952). Portfolio selection. *The Journal of Finance*, **7,** number 1, 77–91.

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

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

Merton, R.C. (1980). On estimating the expected return on the market: an exploratory investigation. *Journal of Financial Economics*, **8,** 323–361.

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

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.