Proxies for daily volatility

Robin de Vilder, Marcel P. Visser

PARIS SCHOOL OF ECONOMICS
ÉCOLE D'ÉCONOMIE DE PARIS

WORKING PAPER N° 2007 - 11

Proxies for daily volatility

Robin G. de Vilder

Marcel P. Visser

JEL Codes:
Keywords: volatility proxy, realized volatility, continuous time embedding, intraday periodicity

PARIS-JOURDAN SCIENCES ECONOMIQUES
LABORATOIRE D'ECONOMIE APPLIQUÉE - INRA

48, BD JOURDAN – E.N.S. – 75014 PARIS
TÉL. : 33(0) 1 43 13 63 00 – FAX : 33 (0) 1 43 13 63 10
www.pse.ens.fr

CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE – ÉCOLE DES HAUTES ÉTUDES EN SCIENCES SOCIALES
ÉCOLE NATIONALE DES PONTS ET CHAUSSÉES – ÉCOLE NORMALE SUPÉRIEURE

Proxies for Daily Volatility

Robin G. de Vilder* Marcel P. Visser †

March 14, 2007

Abstract

High frequency data are often used to construct proxies for the daily volatility in discrete time volatility models. This paper introduces a calculus for such proxies, making it possible to compare and optimize them. The two distinguishing features of the approach are (1) a simple continuous time extension of discrete time volatility models and (2) an abstract definition of volatility proxy. The theory is applied to eighteen years worth of S&P 500 index data. It is used to construct a proxy that outperforms realized volatility.

Keywords: volatility proxy, realized volatility, continuous time embedding, intraday periodicity.

*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

The volatility process is central to financial asset return modelling. Volatility is unobservable, so one often has to rely on proxies when specifying, estimating and evaluating volatility models.

The most common proxy for the square of volatility is the squared close-to-close return. Realized volatility, which is based on intraday data, is nowadays considered a benchmark proxy. A theoretical motivation for constructing realized volatility is that for a large class of semimartingales *quadratic variation* is an unbiased estimator of conditional return variance (ex post), see, for example, Barndorff-Nielsen and Shephard (2002), and Andersen, Bollerslev, Diebold, and Labys (2003a). See also Peters and de Vilder (2006), where realized volatility is used to test the hypothesis of a semimartingale. By using realized volatility to evaluate forecasts, Andersen and Bollerslev (1998) show that GARCH models perform better than expected. High frequency data create additional opportunities. Indeed, some intraday based proxies have superior forecasting power compared to daily returns, see, for example, Ghysels, Santa-Clara, and Valkanov (2006), and Engle and Gallo (2006). In general, the search for a good proxy has suffered from the absence of a straightforward way of comparing proxies. The goal of this paper is to compare and optimize proxies for daily volatility.

Discrete time models, such as GARCH and stochastic volatility models, often have a product structure

$$r_n = v_n Z_n, \tag{1}$$

where r_n is the close-to-close return, v_n the daily volatility, and the Z_n are iid innovations. We shall provide proxies for the factors v_n, based on intraday data, and show that it is possible to rank and to optimize proxies without knowing the scaling factors v_n themselves.

The first step is to introduce a continuous time extension of the discrete time model (1). The continuous time model also has a product structure. We assume that for each business day the intraday log return process is a randomly scaled, independent copy of a stochastic process Ψ on the unit interval. The iid innovations (Z_n) in (1) are replaced by independent copies of Ψ. The continuous time model is consistent with both the persistence of volatility captured by discrete time volatility models, and with intraday periodicity. There are several papers on the embedding of discrete time models into *stationary* continuous time models, see, for instance, Drost and Werker (1996), Meddahi and Renault (2004), and Ghysels, Harvey, and Renault (1996). For other approaches, related to embeddability, see, for example, Nelson (1990), Nelson and Foster (1994), and Klüppelberg, Lindner, and Maller (2004).

We proceed by defining a large class of volatility proxies. In formal terms one obtains

a proxy by applying a positively homogeneous functional to the intraday return process. Realized volatility, high minus low, and absolute returns, all are special cases of such proxies. We show that proxies may be compared by estimating the variance of their logarithm. In the associated ordering, good proxies are highly correlated with daily volatility. It is possible to improve proxies by a technique similar to the determination of the minimal variance portfolio in a portfolio selection problem, as introduced by Markowitz (1952). Optimal proxies for the scaling factors exist. On a data set of intraday S&P 500 index futures market data from January 1988 to mid 2006, the additional use of high-lows on intraday intervals yields proxies substantially better than realized volatility. For the S&P data, the empirical ranking of proxies is stable over time.

The remainder of the paper is organized as follows. Section 2 introduces the continuous time model, and relates it to models of daily volatility. Section 3 defines proxies, shows how to compare them, and how they may be improved. It also contains an existence theorem for optimal proxies. Section 4 is devoted to empirical analyses. We construct a good proxy for the S&P 500 data. Section 5 contains our conclusions. Appendix A describes the data used.

2 Model

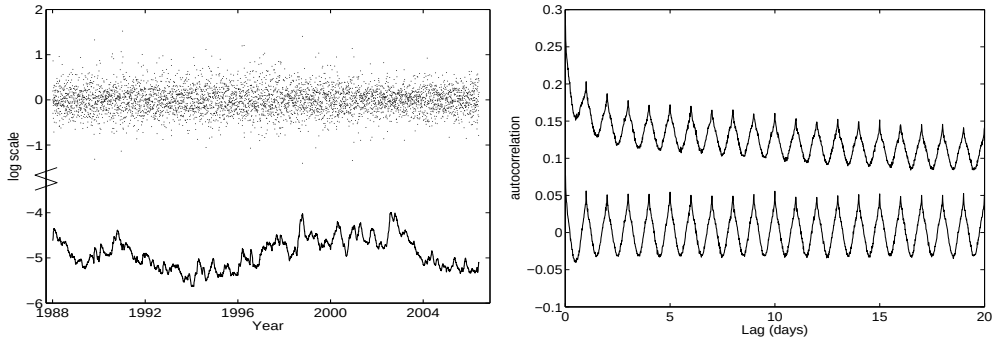
This section proposes a continuous time model of asset returns. The model embeds many discrete time, end-of-day models, such as GARCH and stochastic volatility models. The business day is our unit of time.

2.1 Empirical Considerations

Figures 1 and 2 show some properties of the S&P 500 price process at different time scales. Our S&P 500 futures data consist of 4575 days starting on 1988-01-04 and ending on 2006-05-31. The S&P 500 futures are traded from 8:30 A.M. to 3:15 P.M. Central Standard Time. See Appendix A for more details.

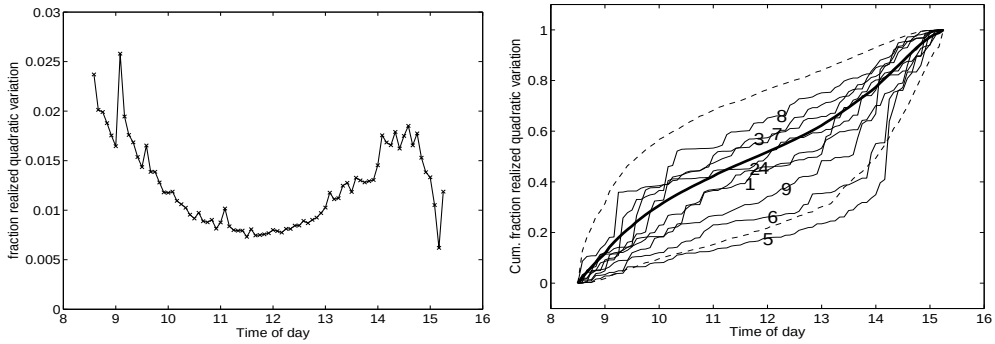
Financial price processes typically exhibit periods of small variation and periods of large variation. These are referred to as low and high volatility periods, respectively. Let us, for the moment, use realized volatility as a measure of the price fluctuations over the trading day. Realized volatility is an estimator of *actual volatility*, where actual volatility is defined as the square root of the daily increment in quadratic variation.[1] We use five-minute returns to define *realized quadratic variation* (RQV_n). This divides the day into 81 five-minute intervals,

[1] We use the term volatility in the context of quantities that scale like the returns, and not, for example, like the square of the returns as quadratic variation does.



(a) Daily log realized volatility

(b) Periodicity



(c) Average per-interval proportion of RQV

(d) Cumulative intraday proportion of RQV

Figure 1: (a) Daily log realized volatility. Difference (dots). Exponentially weighted moving average (solid line). (b) Periodicity. Lower line: autocorrelations of absolute five-minute returns, standardized by daily RV_n. Lags vary from 5 minutes up to 20 days. Upper line: non standardized absolute five-minute returns. Based on 370575 five-minute returns. (c) Average per-interval proportion of RQV. For each of the 81 intraday intervals, the average is taken over 4575 days. (d) Cumulative intraday proportion of RQV. Bold line: average of sample (4575 days). Numbered lines: nine consecutive days starting at 1997-02-12.

and we define daily *realized volatility* by $RV_n = \sqrt{RQV_n}$. For a particular day n, given the partition above, RQV_n is the sum of the 81 squared five-minute returns.

In Figure 1(a) the dots in the top half are the differences of the logs of realized volatility on consecutive days: $\log(RV_n) - \log(RV_{n-1})$. From this figure one may observe that realized volatility varies substantially from day to day. The standard deviation of the logarithmic difference is 0.29. This roughly means that realized volatility changes by about one third of its size a day. A more stable way of describing the fluctuations is to plot an exponentially weighted moving average (EWMA), given by the recursion $\text{EWMA}_n = \beta \cdot \text{EWMA}_{n-1} + (1 - \beta)\log(RV_{n-1})$ with smoothing parameter $\beta = 0.95$; this is the solid line in figure 1(a). It shows that realized volatility may be high for several years, and that the size of fluctuations

in periods of high volatility may differ by approximately a factor of four from the size in periods of low volatility.

Figure 1(b) shows two autocorrelation functions. The upper curve is the autocorrelation of the absolute five-minute returns. The lags vary from five minutes to twenty days. There is a clear periodicity of one day.[2] The lower curve depicts the autocorrelation of five-minute absolute returns standardized by the realized volatility RV_n of the day. This removes the persistence in the daily realized volatility.

Figure 1(c) shows that the hour of the day matters. For each five-minute interval the figure displays the average proportion of daily realized quadratic variation attributable to this interval, resulting in a standardized variation curve. For each interval, this average is taken over 4575 days. There is a distinct U-shape: on average a relatively large proportion of quadratic variation is realized in the period after opening and towards the end of the day. The peak at 9:00 A.M. might be explained by the regularly scheduled releases of important macroeconomic figures, such as Consumer Confidence, Chicago Purchasing Manager, and Leading Indicators. Intraday periodicities and volatility patterns have been widely reported, across markets and across time. For references, see McMillan and Speight (2004); for information about scheduled news release times, see Andersen, Bollerslev, Diebold, and Vega (2003b).

Although these patterns exist *on average*, the shape of a particular day generally differs considerably from this average and may take many forms. Figure 1(d) depicts the per-day departure from the U-shape in Figure 1(c) for the nine consecutive days in the middle of our sample. The bold line shows the average intraday increase in the cumulative proportion of daily realized quadratic variation (RQV) for the full data set; at the start of the day it is zero, and at the end of the day it is one. It is the integral of Figure 1(c). We observe a smooth S-shaped curve, which is steep in the morning and towards the end of the day. The numbered curves represent the *RQV proportion processes* for the nine consecutive days mentioned above. Individual shapes over the day vary substantially: on the fifth day 70% of the day's variation occurs after 2:00 P.M., while on the eighth day 50% of the day's variation has already occurred just after 10:00 A.M. Figure 2 displays the intraday return processes for these nine days, for visual reference. It shows the daily return processes, with respect to the previous day's close: each day the process starts around zero with the overnight return, and at the end of the day one may read off the close-to-close return. The relatively large afternoon volatility on day number 5 in Figure 1(d) corresponds to a sudden downward price movement on day number 5 in Figure 2. Figure 1(d) also shows two quantile lines. The

[2]Although the empirical autocorrelation function is an inappropriate tool for analyzing a nonstationary time series, it is useful for detecting periodic frequencies.

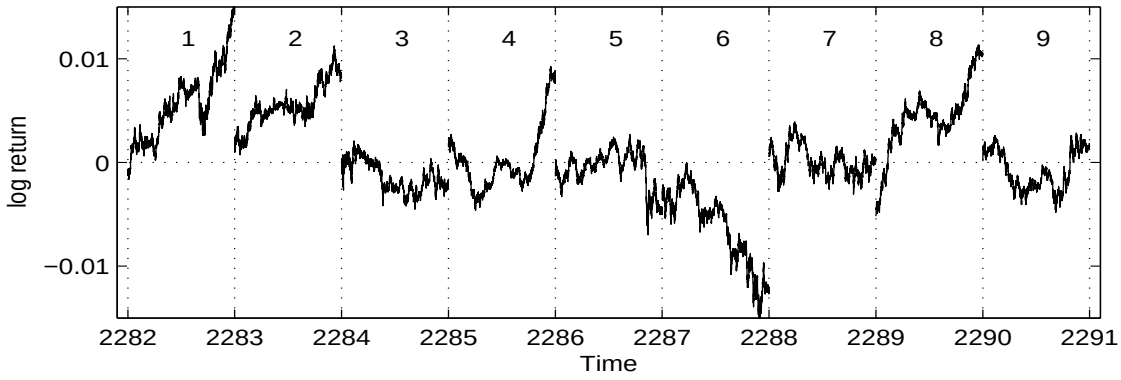


Figure 2: Nine intraday return processes, with respect to the previous day's close. Starting at 1997-02-12.

upper dashed line connects the per-interval 0.975 sample quantiles of the RQV-proportion process for all 81 intervals. The lower dashed line connects the 0.025 quantiles. The sample of the nine RQV-proportion processes is not atypical.

Summarizing, daily realized volatility may be high over periods of years, but varies substantially from day to day. The autocorrelations indicate the presence of a day structure. The average intraday volatility pattern shows a clear U-shape. For individual days the shape may differ substantially from the U-shape.

We shall now introduce a model for the log return process that allows for these characteristics. Let us first look at the discrete time models.

2.2 End-of-Day Models

Let (R_t), $t \in \mathbb{R}$, be the logarithmic cumulative return process of a financial price process. The time interval $[n-1, n)$ describes the process during the n-th business day. Note that R_n represents the *opening* value on day $n+1$. We consider discrete time models that describe the close-to-close returns, given by $r_n = R_n^C - R_{n-1}^C$ for $n \in \mathbb{Z}$. Here R_n^C is the closing value of day n, with the convention $R_0^C \equiv 0$. We adopt the standard assumption that sample paths are right continuous and have left limits (cadlag). We introduce a discrete time filtration $(\mathcal{F}_n^C)$. The σ-field $\mathcal{F}_n^C$ represents the observable information *at the close* of day n, including intraday information: we assume that R_t is $\mathcal{F}_n^C$-measurable for $t < n$. Information $\mathcal{F}_n^C$ may also contain, for example, past news events. The minimal choice for $\mathcal{F}_n^C$ is the sigma-field $\sigma(R_t, t < n)$.

The discrete time models have the form

$$r_n = v_n Z_n, \tag{1}$$

with (v_n) a sequence of strictly positive random variables, and (Z_n) a mean zero, variance one, iid sequence. Moreover, it is assumed that Z_n is independent of v_n. More specifically, Z_{n+1} is independent of the model information, $\mathcal{G}_n^C = \{\mathcal{F}_n^C,\ (v_i)_{i\leq n+1}\}$, for all n. Note that the model information $\mathcal{G}_n^C$ extends the observable information $\mathcal{F}_n^C$ with information from the process (v_k) up to day $k = n+1$. By construction, the innovation Z_n is $\mathcal{G}_n^C$-measurable. The process (v_n) is referred to as the volatility process and is used to describe the dependence structure in the returns (r_n). Distinctive asset price features may be captured in models of this form: uncorrelated returns, heavy tails, volatility persistence. In this paper we will not impose model assumptions on the process (v_n).

2.3 Continuous Time Embedding

Let time advance only during trading hours and normalize the length of the interval of intraday trading to one. This means, that time is a real valued index, expressed in terms of business days. The business day is a natural unit of time in financial applications. We assume a product structure as in (1), with the random variable Z_n replaced by a stochastic process Ψ_n on the unit time interval: for each day the return process over that day is a random *standard day process* multiplied by a random *scaling factor*. Let Ψ be a cadlag process on the closed interval $[0,1]$, left continuous in 1. The standard day processes (Ψ_n) are independent copies of the process Ψ. The scaling factors are strictly positive random variables s_n. The *day n return process* is given by $s_n\Psi_n$. Define the discrete time model filtration $(\mathcal{G}_n^C)$ as above, by $\mathcal{G}_n^C = \sigma\{\mathcal{F}_n^C,\ (s_i)_{i\leq n+1}\}$. As above, s_{n+1} is adapted to $\mathcal{G}_n^C$. The information set $\mathcal{G}_n^C$ contains all the information needed to start generating a sample path for the return process on day $n+1$. Observable information $\mathcal{F}_n^C \subset \mathcal{G}_n^C$ is extended with scaling factors s_n, which ensures that Ψ_n is $\mathcal{G}_n^C$-measurable. For each day n, the process Ψ_n is assumed independent of the scaling factor s_n.

The continuous time process (R_t) is built up from the previous day's close R_{n-1}^C. Recall the convention, $R_0^C \equiv 0$.

Model. *Consider the n-th day, $[n-1, n)$. Let $\vartheta \in [0,1)$ denote the time of day. Set $t = n-1+\vartheta$. The cumulative return process (R_t) satisfies*

$$R_{n-1+\vartheta} = R_{n-1}^C + s_n\Psi_n(\vartheta), \qquad 0 \leq \vartheta < 1. \tag{2}$$

The standard day process Ψ_{n+1} is independent of $\mathcal{G}_n^C$, and the processes (Ψ_n) are identically distributed.

Remark 1. *The overnight return, from day $n-1$ to day n, is given by $R_{n-1} - R_{n-1}^C =$*

$R_{n-1} - R_{n-1-0} = s_n \Psi_n(0)$.

Definition 2.1. *A process that satisfies the conditions of model (2) is said to satisfy the* scaling hypothesis.

Proposition 2.2. *Suppose* (R_t) *satisfies the scaling hypothesis. Moreover assume* $\mathbb{E}\Psi(1) = 0$ *and* $\mathrm{var}(\Psi(1)) = 1$. *Then the close-to-close returns* $(r_n) = (s_n \Psi_n(1))$ *satisfy the discrete time model (1), with* $v_n = s_n$ *and* $Z_n = \Psi_n(1)$.

Proof. The sequence $(Z_n) = (\Psi_n(1))$ is iid, with mean and variance one. The innovation Z_{n+1} is independent of $\mathcal{G}_n^C$, since Ψ_{n+1} is independent of $\mathcal{G}_n^C$, by definition. *Q.E.D.*

If $\Psi(1)$ is not standardized, then the discrete time returns satisfy $r_n = a v_n + v_n Z_n$. This means that the continuous time model (2) also embeds a simple version of the ARCH-M model introduced by Engle, Lilien, and Robins (1987).

The following example shows that one may use Brownian motion to interpolate.

Example 2.3.1. *Let* $(v_n)_{n \geq 1}$ *be a sequence of scaling factors from a discrete time model of the form (1) and let the innovations be standard Gaussian. Take* $s_n = v_n$ *and set* $R_0 = 0$. *Let* Ψ *be Brownian motion on* $[0, 1]$. *The process* $(R_t), t \in R$, *constructed according to the recursion (2) satisfies the scaling hypothesis. For* $Z_n = \Psi_n(1)$, *the discrete time returns* $r_n = s_n \Psi_n(1) = v_n Z_n$ *satisfy the discrete time model (1). One may use the same construction for a sequence* $(v_n)_{n \in \mathbb{Z}}$.

Proposition 2.3. *For any discrete time model of close-to-close returns* (r_n) *of the form (1) there exists a continuous time extension* (R_t) *that satisfies the scaling hypothesis.*

Proof. Consider a discrete time model $r_n = v_n Z_n$ of the form (1). Define $\Psi_n(\vartheta) = \vartheta Z_n$. *Q.E.D.*

Remark 2. *There are more realistic possibilities for* Ψ_n *than the one in the proof of proposition 2.3. For example, one may use the Brownian bridge to interpolate. Let* β *be a Brownian bridge on* $[0, 1]$. *Then* $\beta(0) = \beta(1) = 0$. *Define* $\Psi_n(\vartheta) = \vartheta Z_n + \beta_n(\vartheta)$, *where the* β_n *are Brownian bridges, and* β_{n+1} *is independent of* $\mathcal{G}_n^C$, *for all* n.

Corollary 2.4. *Assume the sequence* (v_n, Z_n) *for the discrete time model (1) is stationary. Then the process* (s_n, Ψ_n) *associated with the extension* (R_t) *in the proof of Proposition 2.3 is stationary.*

Proof. Since (s_n, Z_n) is a stationary sequence, and $\Psi_n = \vartheta Z_n$, the sequence (s_n, Ψ_n) is stationary. *Q.E.D.*

Remark 3. *In general, even if the process* (s_n, Ψ_n) *is stationary, the process* (R_t) *will not be stationary.*

Remark 4. *Corollary 2.4 also applies to the process* (s_n, Ψ_n) *of Remark 2, since* (s_n, Z_n, β_n) *is stationary.*

The scaling hypothesis yields a continuous time process with certain properties. The scaling factors s_n reflect the characteristics of the embedded discrete time model. The scaling factor s_n on day n may be thought of as the state of the market. This state may change from day to day. It is the only source of dependence between increments of (R_t) on *different* days. A change in the state may be driven by the return process over the past days. It may also include an external source. A realization of the standard day process Ψ determines the type, or shape of the intraday return process, such as up or down days. The process Ψ captures the intraday properties of the process (R_t). There are no restrictions on the day process Ψ. There may for example be deterministic or random diurnal effects, jumps, autocorrelation in intraday returns, stochastic spot volatility, leverage effects.

Let us now turn to the issue of identification. Processes that satisfy the scaling hypothesis do not have a unique representation.

Example 2.3.2. *Consider a representation* (s_n, Ψ_n) *for* (R_t) *that satisfies the scaling hypothesis. Set* $s'_n = s_n/2$ *and* $\Psi'_n = 2\Psi_n$. *Then* $(R'_t) \equiv (R_t)$.

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

3 Proxies

In statistics the term proxy is used for a variable that is not of prime interest itself, but is closely connected to an object of interest. It is related to the concept of estimator, but proxies tend to *replace* unobservable variables of interest, whereas an estimator is a recipe to estimate a parameter. A good estimator of a parameter will have small mean squared error. A good proxy to a variable will have large correlation with that variable.

3.1 Definitions and Basic Results

Recall that $\mathbb{D}[0,1]$ is the Skorohod space of cadlag functions on $[0,1]$. We assume that the elements of $\mathbb{D}[0,1]$ 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 process (R_t). Recall that Ψ_n denotes the standard day process, s_n the scaling factor, and ϑ the time of day. Our *proxies* are *homogeneous volatility proxies*. A proxy is a random variable, serving as a proxy for the scaling factor. It is the result of applying a certain estimator, the proxy functional, to the day n return process, $s_n\Psi_n$.

Recall that a functional $H : D \to [0, \infty)$, acting on a linear space D, is *positively homogeneous* if

$$H(\alpha f) = \alpha H(f), \qquad \alpha \in [0, \infty),\ f \in D.$$

Our *proxy functionals* are *proxy functionals for* Ψ.

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*

$$\Pi_n = H(s_n\Psi_n)$$

is a proxy.

Remark 5. *For a realization ψ_n of Ψ, with $\psi_n \notin D$, one may define $\Pi_n = 0$. A proxy functional is not linear: $H(\Psi) + H(-\Psi) > 0$ a.s., but $H(\Psi - \Psi) = H(0) = 0$. A proxy functional need not be symmetric: $H(\Psi)$ need not equal $H(-\Psi)$.*

Proxies are linear in the scaling factor s_n. By definition $\Pi_n = H(s_n\Psi_n)$. Homogeneity of H implies

$$\Pi_n = s_n V_n, \tag{3}$$

where $(V_n) = (H(\Psi_n))$ is a sequence of strictly positive iid innovations. Moreover, s_n and V_n are independent since s_n and Ψ_n are independent. Thus the proxy gives information about the scaling factor. Note that (3) has the same structure as the discrete time asset return model (1). Because of positivity one may take logarithms,

$$\log(\Pi_n) = \log(s_n) + \log(V_n).$$

This decomposition as a sum of independent random variables underlies many of the results below.

Example 3.1.1. *Here are some examples of proxies: absolute daily return; absolute overnight return; high minus low; realized volatility for any grid; maximal absolute two-minute return; actual volatility, defined by the square root of the daily increment in quadratic variation. If* $\Psi(1)$ *has a density, then the almost sure positivity of* $H(\Psi)$ *is ensured.*

The proxies that we consider suffer from a serious limitation: any construction that uses information from outside of the interval $[n, n-1)$ to estimate s_n is not a proxy. For example, the volatility forecast of a daily GARCH model is not a proxy. The restriction to positively homogeneous volatility proxies makes it possible to compare and optimize proxies, without imposing model assumptions on the scaling sequence (s_n). Moreover, a good proxy for s_n will be a good proxy under all models for s_n.

There are various recipes for creating additional proxy functionals.

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

Proof. Since G vanishes only at the origin, and every $H^{(i)}(\Psi) > 0$ a.s. we have $H(\Psi) > 0$ a.s. Homogeneity holds, since $H(\alpha f) = G(H^{(1)}(\alpha f), \ldots, H^{(d)}(\alpha f)) = G(\alpha H^{(1)}(f), \ldots, \alpha H^{(d)}(f)) = \alpha H(f)$. *Q.E.D.*

Example 3.1.2. *Let* $H^{(1)}$ *and* $H^{(2)}$ *be proxy functionals. The functionals* $f \mapsto H(f)$ *below define more proxy functionals.*

1. $H(f) \equiv aH^{(1)}(f)$, *for* $a > 0$.
2. $H(f) \equiv w_1 H^{(1)}(f) + w_2 H^{(2)}(f)$, *for* $w_1, w_2 \in (0, \infty)$.
3. $H(f) \equiv (H^{(1)}(f))^{w_1}\ (H^{(2)}(f))^{w_2}$, *for* $w_1, w_2 \in \mathbb{R},\ w_1 + w_2 = 1$.
4. $H(f) \equiv \max\{H^{(1)}(f), H^{(2)}(f)\}$ *and* $H(f) \equiv \min\{H^{(1)}(f), H^{(2)}(f)\}$.

Let us now define the nuisance proxy.

Definition 3.3. *The* nuisance proxy V_n *for a proxy functional* H *is*

$$V_n = H(\Psi_n).$$

If $\operatorname{var}\big(\log(H(\Psi))\big)$ *exists then* nuisance variation *is* $\lambda^2 = \operatorname{var}\big(\log(H(\Psi))\big)$, *else* $\lambda^2 \equiv \infty$.

Remark 6. *So far we have suppressed the proxy functional* H *in the notation for* Π_n, V_n, λ. *Whenever we need to distinguish between proxy functionals, this will be expressed in the notation.*

The nuisance proxy is unobservable. If one thinks of Π_n as a measurement of s_n, then the nuisance proxy V_n is the multiplicative measurement error, see (3). We use the nuisance proxy for comparing proxies and defining optimality.

Definition 3.4. *Let $H^{(1)}$ and $H^{(2)}$ be two proxy functionals. Define $(\lambda^{(i)})^2 = \mathrm{var}\big(\log(H^{(i)}(\Psi))\big)$, $i = 1, 2$. Then $H^{(1)}$ is* better than *$H^{(2)}$ if $\lambda^{(1)} \leq \lambda^{(2)}$. A proxy functional H is* optimal *if it has finite nuisance variation and is better than all other proxy functionals.*

Remark 7. *Let H be a proxy functional and $a > 0$. Then $G(f) \equiv aH(f)$ is an equally good proxy functional. This follows from the equalities* $\mathrm{var}(\log(aH(\Psi))) = \mathrm{var}(\log(a)) + \mathrm{var}(\log(H(\Psi))) = \mathrm{var}(\log(H(\Psi)))$.

Proposition 3.5 below underpins the use of the word proxies in the terminology: it shows that a good proxy has large correlation with the scaling factor s_n.

Proposition 3.5. *If $0 < \mathrm{var}(\log(s_n)) < \infty$ then*

$$\mathrm{corr}(\log(\Pi_n), \log(s_n)) = \big(1 + \frac{\lambda^2}{\mathrm{var}(\log(s_n))}\big)^{-1/2} \tag{4}$$

Moreover, assume $\lambda^{(1)} \leq \lambda^{(2)} < \infty$. Then (4) implies

$$\mathrm{corr}(\log(\Pi_n^{(1)}), \log(s_n)) \geq \mathrm{corr}(\log(\Pi_n^{(2)}), \log(s_n)).$$

If $\lambda^2 = 0$ then $\mathrm{corr}(\log(\Pi_n), \log(s_n)) = 1$.

Proof. Recall that $\mathrm{var}(\log(V_n)) = \lambda^2$. Since $\log(\Pi_n) = \log(s_n) + \log(V_n)$ we have,

$$\begin{aligned}\mathrm{corr}(\log(\Pi_n), \log(s_n)) &= \mathrm{cov}(\log(s_n), \log(s_n))/\big(\sqrt{\mathrm{var}(\log(s_n)) + \mathrm{var}(\log(V_n))}\sqrt{\mathrm{var}(\log(s_n))}\ \big) \\ &= \big(1 + \frac{\lambda^2}{\mathrm{var}(\log(s_n))}\big)^{-1/2}\end{aligned}$$

This correlation equals 1 if $\lambda^2 = 0$. *Q.E.D.*

Good proxies may suffer from a bias: if $\mathbb{E}H(\Psi) \neq 1$ then $\mathbb{E}\Pi_n \neq \mathbb{E}s_n$. The proxy is biased by a constant scaling factor $\mathbb{E}H(\Psi)$. One may rescale the proxy to obtain an unbiased version. The rescaled proxy is as good as the original one, see Remark 7.

The task of finding a good proxy amounts to finding a proxy functional H such that $H(\Psi)$ has small stochastic variation or, ideally, is a constant. The product structure $\Pi_n = s_n V_n$ of equation (3) and the definition of λ ensure that proxies with zero nuisance variation have perfect correlation with s_n. Perfect proxies do exist, in special cases. For example, take Ψ the standard Brownian motion. The quadratic variation of Ψ then equals one, so actual volatility

AV_n has zero nuisance variation. If a proxy has zero nuisance variation, then one knows the value s_n, without knowing the model for the sequence (s_n). However, actual volatility is not always the best proxy for s_n.

Example 3.1.3. *Consider a standard day process Ψ that is highly fluctuating after opening and has moderate fluctuation for the rest of the day. More specifically, let B denote standard Brownian motion for the interval [0,1] and suppose σ is a positive, discrete random variable, independent of B that takes the values σ_1 and σ_2 with probability $1/2$, where $\sigma_1 < 1,\ \sigma_2 < 1,\ \sigma_1 \neq \sigma_2$. Let $\vartheta_0 \in (0,1)$. Define the process*

$$\Psi(\vartheta) = B(\vartheta)\ 1_{\vartheta \leq \vartheta_0} + B(\vartheta_0)\ 1_{\vartheta > \vartheta_0} + \sigma\big(B(\vartheta) - B(\vartheta_0)\big)\ 1_{\vartheta > \vartheta_0}, \qquad 0 \leq \vartheta \leq 1.$$

The truncated actual volatility constructed for the interval $[0, \vartheta_0]$ is a constant, and as such has zero nuisance variation. Actual volatility over the interval $[0,1]$ is not a constant.

We emphasize that estimating the scaling factor s_n is a different exercise from estimating actual volatility AV_n. Actual volatility is a measure for the price fluctuations that actually occurred during the whole trading day. As an estimator of the volatility factor s_n, actual volatility is merely one out of many possible candidates.

Even when actual volatility is an optimal proxy for the scaling factor, and one is restricted to a finite time grid, it is sometimes possible to improve upon realized volatility.

Example 3.1.4. *Assume Ψ is standard Brownian motion on $[0,1]$. The quadratic variation over $[0,1]$ is 1. Realized quadratic variation sums the squared intraday returns $(\Delta R_i)^2 = (B(\vartheta_i) - B(\vartheta_{i-1}))^2$. Realized quadratic variation of Ψ has expectation 1 and vanishing variance when the mesh of the grid vanishes. Hence it converges in probability to 1.*
Consider the sum of squared per-interval high-lows. Replace the sum of squared returns $\sum(\Delta R_i)^2 = \sum(\vartheta_i - \vartheta_{i-1})(\Delta R_i)^2/\mathbb{E}(\Delta R_i)^2$, by the sum of squared, standardized high-lows $\sum(\vartheta_i - \vartheta_{i-1})hl_i^2/\mathbb{E}hl_i^2$. The new sum has expectation 1 and vanishing variance. It is well known that for standard Brownian motion on the unit time interval the estimator $hl^2/\mathbb{E}hl^2$ has approximately 5 times smaller variance than $B(1)^2$, see, for example, Parkinson (1980). Hence, the use of high-lows results in a higher rate of convergence to the quadratic variation.

Let us return to the issue of identification, which has been addressed by Example 2.3.2. Proposition 3.6 shows that different representations (s_n, Ψ_n) for the scaling hypothesis result in the same ordering for proxy functionals.

Proposition 3.6. *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_t). 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 $\mathrm{var}(\log(H^{(1)}(\Psi'_n))) \leq \mathrm{var}(\log(H^{(2)}(\Psi'_n)))$. *Q.E.D.*

3.2 Optimal Proxies

Theorem 3.7. *If there exists a proxy functional with finite nuisance variation, then there exists an optimal proxy functional.*

Proof. We need 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.

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, see Example 3.1.2.

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 parallellogram 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 above conditions. 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. *Q.E.D.*

Remark 8. *All proxy functionals of the form aH^*, $a > 0$, are optimal, see Remark 7. Any optimal proxy functional H satisfies $H(\Psi) \overset{a.s.}{=} aH^*(\Psi)$ for certain $a > 0$, see Proposition 3.9.*

Lemma 3.8. *If H^* is an optimal proxy functional, and H is a proxy functional, then* $\mathrm{cov}\big(\log(H^*(\Psi)),\, \log(H(\Psi))\big) = (\lambda^*)^2$.

Proof. Consider the proxy functional $H(f) \equiv \big(H^*(f)\big)^w \big(H(f)\big)^{1-w}$, with nuisance variation $\lambda_w^2 = w^2(\lambda^*)^2 + 2w(1-w)\ \mathrm{cov}\big(\log(H^*(\Psi)), \log(H(\Psi))\big) + (1-w)^2\lambda^2$. Since H^* is optimal, $\partial\lambda_w^2/\partial w \mid_{w=1} = 0$. Hence $\mathrm{cov}\big(\log(H^*(\Psi)), \log(H(\Psi))\big) = (\lambda^*)^2$. *Q.E.D.*

Proposition 3.9. *Suppose $H^{(1)}$ and $H^{(2)}$ are two optimal proxy functionals. Then there exists a constant $a > 0$, such that $H^{(1)}(\Psi) \overset{a.s.}{=} aH^{(2)}(\Psi)$.*

Proof. Both proxy functionals have nuisance variation $(\lambda^*)^2$. 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 3.8 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 $a > 0$. *Q.E.D.*

3.3 Empirically Ranking Proxies

Theorem 3.7 guarantees that optimal proxies for the scaling factor exist, it is however of no help in empirical applications. In empirical applications there is the problem that the process Ψ is of unknown form. As a first practical step, this section provides two propositions that are useful for comparing proxies in an applied setting.

Let $H^{(1)}$ and $H^{(2)}$ be two proxy functionals. The nuisance variations $(\lambda^{(i)})^2$ are used to compare these proxy functionals. This procedure is infeasible in an empirical situation, since the standard day processes (Ψ_n) are not observed. However, by the multiplicative structure of equation (3) and independence of s_n and Ψ_n,

$$\text{var}(\log(\Pi_n)) = \text{var}(\log(s_n)) + \lambda^2. \tag{5}$$

Hence it suffices to compare the variances of the log proxies.

Proposition 3.10. *Let $H^{(1)}$ and $H^{(2)}$ be two proxy functionals. Assume* $\text{var}(\log(H^{(1)}(\Psi))) < \infty$ *and* $\text{var}(\log(H^{(2)}(\Psi))) < \infty$. *Assume (s_n, Ψ_n) satisfies the scaling hypothesis. Moreover, assume* $\text{var}(\log(s_n)) < \infty$. *Then*

$$\text{var}(\log(\Pi_n^{(1)})) - \text{var}(\log(\Pi_n^{(2)})) \quad = \quad (\lambda^{(1)})^2 - (\lambda^{(2)})^2. \tag{6}$$

Proof. The common term $\text{var}(\log(s_n))$ drops out by equation (5). *Q.E.D.*

Remark 9. *Although the variance of the log proxy of equation (5) is useful for comparing proxies, it cannot be used as an absolute measure for the quality of a proxy, due to the unknown term* $\text{var}(\log(s_n))$.

Remark 10. *In empirical applications one uses estimates of the variances. In order to reduce estimation error, we shall use the technique of prescaling, see Section 3.5.*

The practical conclusion that follows from the following proposition is that good proxies have large autocorrelations: one may compare proxies by inspecting the autocorrelation function of their logarithms. It confirms the following intuition. Proxies are noisy measurements of the scaling factors (s_n). The less noisy a proxy, the more of the structure of (s_n) is visible in the autocorrelation of the proxy sequence. Condition (7) in the proposition below stipulates that the autocorrelation for distant scaling factors dominates the impact of distant innovations. Let $\rho_\cdot(j)$ denote j-th order autocorrelation. Let $(a_j) \sim (b_j)$ denote asymptotic equivalence: $a_j/b_j \to 1$ for $j \to \infty$.

Proposition 3.11. *Assume the conditions of Proposition 3.10 hold. Assume $\rho_{\log(s_n)}(j) \neq 0$ for all $j > 0$. Moreover, assume $V_n^{(i)} = H^{(i)}(\Psi_n),\ i = 1, 2$, satisfies*

$$\text{corr}(\log(V_{n-j}^{(i)}), \log(s_n))/\text{corr}(\log(s_{n-j}), \log(s_n)) \to 0 \qquad j \to \infty, \tag{7}$$

and $\text{var}(\log(s_n)) > \epsilon$ *for all n, for certain ϵ. Then the following asymptotic equivalence holds*

for all n,

$$\frac{\rho_{\log(\Pi_n^{(1)})}(j)}{\rho_{\log(\Pi_n^{(2)})}(j)} \sim \sqrt{\frac{\left(\mathrm{var}(\log(s_{n-j})) + (\lambda^{(2)})^2\right)\left(\mathrm{var}(\log(s_n)) + (\lambda^{(2)})^2\right)}{\left(\mathrm{var}(\log(s_{n-j})) + (\lambda^{(1)})^2\right)\left(\mathrm{var}(\log(s_n)) + (\lambda^{(1)})^2\right)}}, \qquad j \to \infty. \quad (8)$$

Proof. Form the left hand side fraction in (8) and decompose to obtain:

$$\sqrt{\frac{\left(\mathrm{var}(\log(s_{n-j})) + (\lambda^{(2)})^2\right)\left(\mathrm{var}(\log(s_n)) + (\lambda^{(2)})^2\right)}{\left(\mathrm{var}(\log(s_{n-j})) + (\lambda^{(1)})^2\right)\left(\mathrm{var}(\log(s_n)) + (\lambda^{(1)})^2\right)}} \cdot \frac{\mathrm{cov}(\log(s_{n-j}) + \log(V_{n-j}^{(1)}), \log(s_n))}{\mathrm{cov}(\log(s_{n-j}) + \log(V_{n-j}^{(2)}), \log(s_n))}.$$

The term to the right of the multiplication dot is equivalent to

$$\frac{\rho_{\log(s_n)}(j) + \mathrm{corr}(\log(V_{n-j}^{(1)}), \log(s_n)) \cdot \lambda^{(1)} / \sqrt{\mathrm{var}(\log(s_{n-j}))}}{\rho_{\log(s_n)}(j) + \mathrm{corr}(\log(V_{n-j}^{(2)}), \log(s_n)) \cdot \lambda^{(2)} / \sqrt{\mathrm{var}(\log(s_{n-j}))}},$$

which is asymptotically equivalent to 1: divide the numerator and the denominator by $\rho_{\log(s_n)}(j)$, take the limit $j \to \infty$ and apply (7). *Q.E.D.*

Remark 11. *If the process (s_n) is stationary, then the relation in (8) simplifies to*

$$\frac{\rho_{\log(\Pi_n^{(1)})}(j)}{\rho_{\log(\Pi_n^{(2)})}(j)} \sim \frac{\mathrm{var}(\log(s_0)) + (\lambda^{(2)})^2}{\mathrm{var}(\log(s_0)) + (\lambda^{(1)})^2}, \qquad j \to \infty.$$

3.4 Improving Proxies

Let $H^{(1)}, \ldots, H^{(d)}$ be given proxy functionals. In this section we study all proxy functionals of the form $f \mapsto H^{(w)}(f) = \prod_{i=1}^d (H^{(i)}(f))^{w_i}$, $w_1 + \ldots + w_d = 1$, and $w_i \in \mathbb{R}$ for all i. Here, the column vector w is the d-dimensional weight vector. Such a functional is a *geometric proxy functional*, based on $H^{(1)}, \ldots, H^{(d)}$. The restriction $\sum w_i = 1$ is needed to obtain a proxy functional, but the weights are not restricted to the interval $[0, 1]$. For this situation there exists an optimal geometric proxy functional that is simple to compute.

Let Λ denote the covariance matrix of the logarithm of the nuisance proxies and let λ_w^2 be the variance of the logarithm of the geometric proxy nuisance variation. Let ι be a d-dimensional column vector of ones. We then have $\lambda_w^2 = w'\Lambda w$ and, similarly to the minimal variance portfolio in Markowitz portfolio theory, λ_w^2 is optimal for $w^* = \frac{\Lambda^{-1}\iota}{\iota'\Lambda^{-1}\iota}$, with $\lambda_{w^*}^2 = \frac{1}{\iota'\Lambda^{-1}\iota}$. This solution is infeasible: it is impossible to estimate the variance matrix Λ, since the nuisance proxies $V_n^{(j)} = H^{(j)}(\Psi_n)$ are not observed. Theorem 3.12 shows that it suffices to know the covariance structure $\Lambda_{p,n}$ of the logarithm of the proxies at day n to

obtain the optimal weights. We have

$$\Lambda_{p,n} = \text{var}(\log(s_n))\ \iota\iota' + \Lambda. \tag{9}$$

The covariance matrix $\Lambda_{p,n}$ equals the covariance matrix Λ with a common noise term $\text{var}(\log(s_n))$ added to each element. The covariance matrix $\Lambda_{p,n}$ may be estimated.

Theorem 3.12. *Let (s_n, Ψ_n) satisfy the scaling hypothesis. Assume $\text{var}(\log(H^{(i)}(\Psi))) < \infty$ for $i = 1, \ldots, d$. Let $\Lambda_{p,n}$ be the covariance matrix of the log of the day n proxies, and let Λ denote the covariance matrix of the log nuisance proxies. The optimal weight vector w that minimizes $\lambda_w^2 = w'\Lambda w$ does not depend on the form of the process (s_n) and equals*

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

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

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

Proof. The optimal weight w^* does not depend on (s_n): it follows from equation (9) that

$$\arg\min_w\ w'\Lambda_{p,n}w = \arg\min_w\ \text{var}(\log(s_n)) + w'\Lambda w. \tag{10}$$

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 we have a unique solution to the first order condition, we have found the optimum.

Using (10) we obtain the equalities $w^* = \Lambda_{p,n}^{-1}\iota/\iota'\Lambda_{p,n}^{-1}\iota = \Lambda^{-1}\iota/\iota'\Lambda^{-1}\iota$, which imply $\text{var}(\log(\Pi_n^{(w^*)})) = \text{var}(\log(s_n))\ + \lambda_{w^*}^2$.

Q.E.D.

Remark 12. *In practice it may be useful to apply prescaling, see Section 3.5 below.*

Example 3.4.1. *Consider the two dimensional case, $d = 2$. Let $\Lambda = (\lambda_{i,j})$ denote the 2×2 covariance matrix $\text{cov}\big(\log(H^{(1)}(\Psi)),\ \log(H^{(2)}(\Psi))\big)$, with $\lambda_{i,i} = (\lambda^{(i)})^2$.*

Then,

$$w^* = \begin{pmatrix} w_1^* \\ w_2^* \end{pmatrix} = \frac{1}{\lambda_{1,1} + \lambda_{2,2} - 2\lambda_{1,2}} \begin{pmatrix} \lambda_{2,2} - \lambda_{1,2} \\ \lambda_{1,1} - \lambda_{1,2} \end{pmatrix}.$$

This illustrates that an optimal weight may turn negative or larger than 1: if $\lambda_{1,1} < \lambda_{1,2}$ *then* $w_2^* < 0$. *This occurs if* $\rho_{\log\ H^{(1)}(\Psi),\log\ H^{(2)}(\Psi)} > \sqrt{\lambda_{1,1}/\lambda_{2,2}}$, *that is if correlations are positive and large. If, as in Equation (9), there is an additional noise term* $\mathrm{var}(\log(s_n))$, *then*

$$\Lambda_{p,n} = \mathrm{var}(\log(s_n)) \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} + \begin{pmatrix} \lambda_{1,1} & \lambda_{1,2} \\ \lambda_{1,2} & \lambda_{2,2} \end{pmatrix}.$$

This leads to the same w^*.

3.5 Prescaling

The method of comparing proxies in Proposition 3.10 and the method of improving proxies in Theorem 3.12 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 scaling 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).

Definition 3.13. *A* prescaling sequence (p_n) *is an* $(\mathcal{F}_{n-1}^C)$ *adapted sequence of strictly positive random variables.*

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

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

Proposition 3.14. *Assume the process* (s_n, Ψ_n) *satisfies the scaling hypothesis. Prescale the scaling factors* (s_n) *to obtain the sequence* $(\tilde{s}_n)$. *The corresponding continuous time process* $(\tilde{R}_t)$ *constructed from* $(\tilde{s}_n, \Psi_n)$ *according to the relation (2) of the continuous time model, satisfies the scaling hypothesis.*

Proof. The variables $(\tilde{s}_n)$ are positive. Both p_{n+1} and s_{n+1} are $\mathcal{G}_n^C$-measurable, hence so is $\tilde{s}_{n+1}$. Therefore $(\tilde{R}_t)$ satisfies the scaling hypothesis. *Q.E.D.*

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

$$\tilde{\Pi}_n = \Pi_n/p_n.$$

This means that one may replace s_n by $\tilde{s}_n$ and Π_n by $\tilde{\Pi}_n$ in Proposition 3.10 and Theorem 3.12. As a consequence, the population value of the noise term $\mathrm{var}(\log(s_n))$ of equation (9)

changes into

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

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

Let us say a few words on the role of stationarity. Proposition 3.10 for comparing proxies and Theorem 3.12 for optimizing proxies do not assume stationarity of the process (s_n, Ψ_n). This is reflected by the covariance matrix $\Lambda_{p,n}$ in Theorem 3.12 depending on n. Suppose the process (s_n, Ψ_n) is stationary. Then the covariance matrix Λ_p is consistently estimated by the sample covariance matrix of the log of the proxies, thereby providing weights $\hat{w}$ that are consistent for w^*. Preliminary results indicate that it is possible to show that $\hat{w}$ is a consistent estimator for the optimal weight w^*, without assuming stationarity for the scaling factors (s_n).

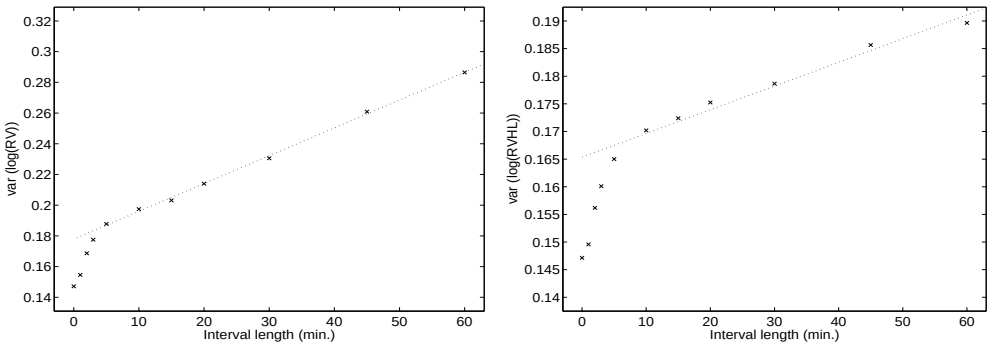
4 An Empirical Application

This section applies the theory of Sections 2 and 3 to the S&P 500 futures tick data. Appendix A describes the data.

4.1 Microstructure Noise Barrier

It is well known that 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 (2006). A commonly used way of avoiding such microstructure effects is to sample at sufficiently wide intervals.

In this paper the measure of comparison is the variance of the logarithm. The proxies RV and $RVHL$ (see Table 1) depend on the sampling interval $\Delta\vartheta$. Figure 3 shows the graph of $\Delta\vartheta \to \widehat{\text{var}}(\log(\Pi_n^{\Delta\vartheta}))$, 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 3: 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.

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 first observation cannot be prescaled and is left out of the variance computations. The measure of comparison is PV (prescaled variance), which is the variance of the logarithm of a proxy after prescaling, $PV = \text{var}(\log(\tilde{\Pi}_n))$; see Sections 3.3 and 3.5. Smaller variances correspond to better proxies. For the prescaling sequence (p_n) we take an exponentially weighted moving average filter 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 set the smoothing parameter so that the sample variance of the logarithm of prescaled five-minute realized volatility is minimal. Recall that the prescaled variance equals nuisance variation plus the unknown constant $\text{var}(\log(\tilde{s}_n))$. The prescaled variance is useful for ranking proxies, but it is not an absolute measure for the quality of a proxy.

Table 1 shows that the quality of realized volatility improves if one samples more frequently. The prescaled variance is largest for the absolute close-to-close returns.[3] 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

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

	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) filter 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.

to better proxies than sums of squared values. This 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-lows. 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 $maxar2$.

4.3 Improving Proxies

Let us now explore to what extent the proxies in Table 1 may be improved. We use the technique of Section 3.4.

Proxy	RV5	RV10	RV15	RV20	RV30	hl
weight	0.315	0.082	-0.041	-0.071	-0.036	-0.085
Proxy	maxar2	RAV5	RAV10	RVHL10	RAVHL10	
weight	0.020	-0.423	-0.608	-0.838	2.69	

Table 2: Optimal weights for the geometric proxy, based on the proxies of Table 1, for the full sample.

Table 2 provides the optimal weights for the proxies of Table 1. Note that $RAVHL10$ is the most important component. The prescaled variance of the optimized proxy is $PV = 0.039$, which outperforms all proxies in Table 1. We refer to this proxy as the optimized proxy $\Pi^{(\hat{w})}$. If one extrapolates the prescaled variances of the realized volatilities of Table 1 to a time interval of length zero, one obtains a value between 0.050 and 0.060. The value $PV = 0.039$ for the optimized proxy is well below those values.

Observe from Table 2 that weights may be negative. In geometrical terms this may be restated as follows. The log proxies are vectors in an affine space. The proxies are highly related, since they all approximate the same scaling factor s_n, cf. (3). The optimal proxy is not in the convex hull of the original proxies. The original proxies do not completely reflect the direction of the optimal proxy. The weights outside $[0, 1]$ correct for this.

	full	1st	2nd	3rd	4th
name	PV	PV	PV	PV	PV
$\Pi^{(\hat{w})}$	0.039	0.039	0.042	0.044	0.029
$\Pi^{(\hat{w},1)}$	0.039	0.038	0.043	0.045	0.030
$\Pi^{(\hat{w},2)}$	0.040	0.040	0.041	0.045	0.031
$\Pi^{(\hat{w},3)}$	0.040	0.041	0.043	0.043	0.030
$\Pi^{(\hat{w},4)}$	0.040	0.041	0.045	0.046	0.028

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

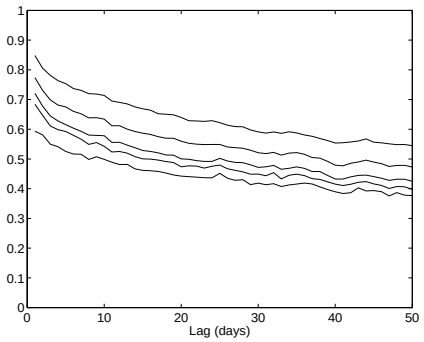


Figure 4: Autocorrelations of log proxies. Lags 1 to 50 days. From bottom to top: RV30, RV15, RV10, RV5, $\Pi^{(\hat{w})}$.

Table 3 investigates the stability of the optimized proxy $\Pi^{(\hat{w})}$. As in Table 1 we report performance measures for the full sample and for four subsamples. By comparing the performance of $\Pi^{(\hat{w})}$ in the different subsamples to Table 1, it is clear that $\Pi^{(\hat{w})}$ outperforms all those proxies in every subsample. The proxy $\Pi^{(\hat{w},i)}$ is constructed using weights based on the i-th subsample. In the first subsample the performance of the globally optimized $\Pi^{(\hat{w})}$ is not substantially improved by $\Pi^{(\hat{w},1)}$. A similar statement holds for the other subsamples. Moreover, proxies based on a particular subsample are close to optimality in all other subsamples. We conclude that the optimality of $\Pi^{(\hat{w})}$ is stable.

Figure 4 shows the autocorrelations of $\log(\Pi_n^{(\hat{w})})$ 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 figure also shows that the autocorrelations for the optimized proxy are substantially larger than those for five-minute

realized volatility. Hence, Proposition 3.11 provides an additional indication that $\Pi^{(\hat{w})}$ 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. A large R^2 in a particular column means that the proxy in that column is largely predictable, suggesting that it is a good proxy for daily volatility. A large R^2 in a particular row means that the proxy in that row is a good predictor. The optimized geometric proxy $\Pi^{(\hat{w})}$ is maximal in all rows and all columns.

	RV30	RV20	RV15	RV10	RV5	$\Pi^{(\hat{w})}$
RV30(-1)	0.35	0.39	0.42	0.46	0.50	0.59
RV20(-1)	0.38	0.42	0.45	0.49	0.54	0.62
RV15(-1)	0.39	0.44	0.47	0.50	0.55	0.64
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	0.70
$\Pi^{(\hat{w})}(-1)$	0.46	0.50	0.52	0.56	0.61	0.73

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

We end this section by supplying the simplified proxy

$$\Pi_n = (RAVHL10_n)^{1.82}(RAV10_n)^{-0.82}. \tag{11}$$

The proxy (11) is almost as good as the optimized proxy $\Pi_n^{(\hat{w})}$: it has prescaled variance $PV = 0.041$.

5 Conclusions

This paper provides a theoretical basis for the comparison and optimization of volatility proxies, based on intraday data. The theory is founded on a continuous time model that embeds discrete time, daily volatility models.

In this paper a volatility proxy is the result of applying a positively homogeneous functional to the intraday return process. This is a limitation that rules out, for instance, volatility predictors. On the other hand, it offers the possibility of developing a simple theory for comparing and optimizing proxies. By definition, the log of a good proxy has small variance. Equivalently, the correlation with daily volatility is large. We show that optimal proxies exist. An optimal proxy for the scaling factor v_n is an optimal proxy under all possible discrete time models of the form $r_n = v_n Z_n$, where the Z_n are iid innovations.

For the S&P 500 data a combination of the high-lows over ten-minute intervals, and the absolute returns over ten-minute intervals yields a good proxy.

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. (this occurs 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. This leaves 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 this 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 this 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.

References

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. (2003a). Modeling and forecasting realized volatility. *Econometrica*, **71,** number 2, 579–625.

Andersen, T.G., Bollerslev, T., Diebold, F.X. and Vega, C. (2003b). Micro effects of macro

announcements: Real-time price discovery in Foreign Exchange. *American Economic Review*, **93,** number 1, 38–62.

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

Engle, R.F. and Gallo, G.M. (2006). A multiple indicators model for volatility using intra-daily data. *Journal of Econometrics*, **131,** number 1-2, 2–27.

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.

Ghysels, E., Harvey, A.C. and Renault, E. (1996). Stochastic volatility. In *Statistical Methods in Finance, Amsterdam, North-Holland* (eds C.R. Rao and G.S. Maddala), Handbook of Statistics, volume 14, chapter 5, pp. 119–191. Elsevier.

Ghysels, E., Santa-Clara, P. and Valkanov, R. (2006). Predicting volatility: getting the most out of return data sampled at different frequencies. *Journal of Econometrics*, **131,** number 1-2, 59–95.

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

Klüppelberg, C., Lindner, A. and Maller, R. (2004). A continuous-time GARCH process driven by a Lévy process: Stationarity and second-order behaviour. *Journal of Applied Probability*, **41,** number 3, 601–622.

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

McMillan, D.G. and Speight, A.E.H. (2004). Intra-day periodicity, temporal aggregation and time-to-maturity in FTSE-100 index futures volatility. *Applied Financial Economics*, **14,** 253–263.

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

Nelson, D.B. (1990). ARCH models as diffusion approximations. *Journal of Econometrics*, **45,** number 1-2, 7–38.

Nelson, D.B. and Foster, D.P. (1994). Asymptotic filtering theory for univariate ARCH models. *Econometrica*, **62,** number 1, 1–41.

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.

Peters, R.T. and de Vilder, R.G. (2006). Testing the Continuous Semimartingale Hypothesis for the S&P 500. *Journal of Business & Economic Statistics*, **24,** number 4, 444–454.

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.