Chapter 2

Ranking and Combining Volatility Proxies for GARCH and Stochastic Volatility Models

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

GARCH and stochastic volatility models are standard tools used for the time series analysis of daily volatility. This chapter is the first research that analyses proxies in relation to these discrete time models. It addresses the problem of optimizing volatility proxies when intraday high-frequency data are available. In an ideal world, with a continuously observed asset price process in a frictionless market, an obvious candidate for a proxy would be the (square root of the) quadratic variation. The daily quadratic variation is the limit of the realized variance\(^1\) as the lengths of the sampling intervals approach zero, see for instance Andersen, Bollerslev, Diebold, and Labys (2001). However, in discrete time models the volatility is a scale factor, and the square root of the quadratic variation is generally not a perfect estimator of this scale factor. Moreover, the quadratic variation does not always lead to an optimal estimator of the scale factor, as shall be clear from a simple example. So, even in ideal circumstances, finding good proxies for the discrete time scale factor is not a trivial task.

Discrete time volatility models were developed before high-frequency data became

\(^1\) One obtains the realized variance by summing the squared intraday financial returns over, for instance, five-minute intervals.
readily available, and are typically applied to daily, or lower frequency returns. Most discrete time models for the daily financial return \( r_n \) satisfy the canonical product structure:

\[
r_n = \sigma_n Z_n. \tag{2.1}
\]

Here the observed financial return \( r_n \) is modelled as the product of an iid innovation \( Z_n \) and a positive scale factor \( \sigma_n > 0 \). One usually assumes that \( Z_n \) has mean zero and, for standardization, unit variance. Specific models differ in their specification of the scale factors; an example is the stationary GARCH(1,1) recursion

\[
\sigma_n^2 = \kappa + \alpha r_{n-1}^2 + \beta \sigma_{n-1}^2, \tag{2.2}
\]

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

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

\[
\lambda^2 = \text{var}(\log(H_n/\sigma_n)). \tag{2.3}
\]

The smaller \( \lambda^2 \), the smaller the standard errors of the parameter estimators. This result holds for surprisingly general intraday price processes. It also holds irrespective of the particular model for the scale factors \( (\sigma_n) \). The criterion \( \lambda^2 \) is valid within a large class of volatility proxies \( H \), including commonly applied proxies such as the intraday high-low range, the realized volatility and realized power variation.

The main result of this chapter is a procedure for combining volatility proxies into a single, highly efficient proxy. The combined proxy has minimal measurement variance \( \lambda^2 \), as given by equation (2.3). The chapter takes a model-free approach: it develops a theory for ranking and combining proxies without assuming a particular model for the sequence of scale factors \( (\sigma_n) \), and without making strong model assumptions for the intraday price process. The resulting tools developed in the chapter are easy to apply.

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

**Related Literature**

This chapter is the first research that specifically addresses the problem of optimizing volatility proxies from the perspective of discrete time models, but its underlying theme, dealing with high-frequency data in daily volatility modelling, is shared with two other branches of the literature. First, there is the temporal aggregation literature, see for instance Drost and Nijman (1993), Drost and Werker (1996), and Meddahi and Renault (2004); second, there is the literature based on the framework of continuous time semimartingales. An important part of the semimartingale literature is concerned with estimating the quadratic variation, and dealing with microstructure noise (e.g. Barndorff-Nielsen, Hansen, Lunde, and Shephard, 2008). In that literature, as in our empirical analysis, the use of high-low ranges over five or ten-minute intervals has received attention, see for instance Martens and van Dijk (2007) and Christensen and Podolskij (2007). One may also improve quadratic variation estimators by subsampling and averaging realized variances, see Zhang, Mykland, and Aït-Sahalia (2005). Hansen and Lunde (2005b) discussed how to combine unbiased estimators of the integrated variance, in particular how to combine the realized variance and the squared overnight return. Patton (2008) studied ranking estimators of the quadratic variation when the sequence of daily quadratic variations is a random walk or an AR-process, and Patton and Sheppard (2009b) extended these results to obtain optimal combinations of estimators of quadratic variation. For overviews see McAleer and Medeiros (2008), and Andersen, Bollerslev and Diebold (2009). We shall see that the measures proposed by the quadratic variation literature may be used as proxies for $\sigma_n$. They can be used as input to a combination of proxies for $\sigma_n$.

The remainder of the chapter is organized as follows. Section 2.1 discusses the model for the intraday return process $R_n(\cdot)$. Section 2.2 introduces proxies, and develops tools for ranking and optimizing them. Section 2.3 constructs a good volatility proxy for the S&P 500 data. Section 2.4 sets out the most important conclusions. The appendices 2.A, 2.B, 2.C, and 2.D contain a description of the data, a discussion of microstructure effects, introduce the empirical technique of prescaling, and provide mathematical details.


2.1 Model

For each trading day \( n \) we observe a process \( R_n(\cdot) \), being the continuous time log-return process within that day. It starts with the overnight return at time \( u = 0 \), and at the end of the day, at time \( u = 1 \), we obtain the close-to-close return \( r_n = R_n(1) \). Figure 2.1 depicts five actual intraday log-return processes \( R_n \), for the S&P 500, for \( n = 2285, \ldots, 2289 \). The second day in the figure (\( n = 2286 \) in our sample), for example, starts with a small positive overnight return, and the value of the index increases towards the end of the day to arrive at a plus of \( r_n = R_n(1) \approx 0.01 \), or +1% at the close of the day.

Now, the standard framework for intraday high-frequency data is to assume that \( R_n(\cdot) \) is a semimartingale on the unit time interval (i.e. the trading day). Although the model that the chapter introduces is not at odds with the semimartingale approach, for our purposes more insight is provided by first looking at a simple example, the scaled Brownian motion:

\[
R_n(\cdot) = \sigma_n W_n(\cdot).
\]

Here, \( W_n(\cdot) \) is a standard Brownian motion on the unit time interval independent of \( \sigma_n \), and one may think of \( \sigma_n \) as the scale factor of a daily GARCH process. So the scale factor \( \sigma_n \) is a measure for the size of the intraday price fluctuations and is constant over the day, whereas the Brownian motion \( W_n(\cdot) \) captures the intraday price movements. Estimating \( \sigma_n \) is an easy task in this model. The daily quadratic variation \( QV \) now yields an exact relationship:

\[
QV(R_n) = \sigma_n^2,
\]

so the square root of the quadratic variation is a perfect estimator of \( \sigma_n \). Our model for
2.1 Model

$R_n(\cdot)$ is a generalization of the scaled Brownian motion:

$$R_n(\cdot) = \sigma_n \Psi_n(\cdot), \quad (2.4)$$

where $\Psi_n(\cdot)$ is an arbitrary process on the unit time interval, again independent of $\sigma_n$. For different days the standard processes $\Psi_k(\cdot)$ and $\Psi_n(\cdot)$, $k \neq n$, are assumed to be independent, and to have the same probability distribution. Moreover, there are no model assumptions for the sequence $(\sigma_n)$. To the best of our knowledge, the general form that the model (2.4) takes is new; we refer to it as the scaling model. For our purposes, it is extremely interesting because of its relation to discrete time models: the model yields daily close-to-close returns $r_n$ that satisfy

$$r_n = \sigma_n Z_n,$$

(setting $Z_n = \Psi_n(1)$) which is the canonical discrete time model structure, see equation (2.1). In general the quadratic variation is not a perfect estimator of $\sigma_n^2$: if $\Psi_n(\cdot)$ has nondeterministic quadratic variation, then in general

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

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

$$R_n(u) = \sigma_n \Psi_n(u), \quad 0 \leq u \leq 1,$$

and

M1. The scale factors $\sigma_n$ are strictly positive,

M2. The standard processes $\Psi_n(\cdot)$ are cadlag\(^3\) on the closed interval $[0, 1]$,

M3. The standard processes $\Psi_n(\cdot)$ have the same probability distribution for all $n$,

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

Conditions (M1) and (M2) are technical and do not have any practical limitations. By (M3) and (M4), the sequence of processes $\Psi_n(\cdot)$ is iid. The process $\Psi_n(\cdot)$ may be

\(^{2}\)The statistician only observes the processes $R_n(\cdot)$.

\(^{3}\)The sample paths are right-continuous and have left limits.
any process representing the intraday price pattern. The sequence of scale factors \((\sigma_n)\) may be any strictly positive stochastic process, as long as the process \(\Psi_n(\cdot)\) is independent of current and past scale factors \(\sigma_k\). So the factors \((\sigma_n)\) may satisfy a GARCH model, or a stochastic volatility model. They may also contain structural breaks, so be nonstationary. The actual fluctuations in the process \(\Psi_n(\cdot)\) determine the pattern of the intraday return process, such as up or down days, quiet or hectic days. The scaling model is not a model with constant intraday volatility: depending on \(\Psi_n(\cdot)\) the day may be hectic (a large quadratic variation) when \(\sigma_n\) is low, and vice versa. The process \(\Psi_n(\cdot)\) allows for time-of-day effects. It may also have, for instance, leverage effects, jumps (e.g. a Lévy process), a stochastic diffusion coefficient, a non-zero mean process.

Conditions (M1) to (M4) ensure that volatility proxies satisfy an additive measurement equation involving the scale factor and an independent measurement error, see equation (2.7) in Section 2.2.1. Note that \(\Psi(1)\) is not standardized; identification of \(\sigma_n\) and \(\Psi_n(\cdot)\) are not necessary for the study of proxies, see Section 2.2.1.

### 2.2 Proxies

The scale factors \(\sigma_n\) are latent. They have to be estimated from the data. This section discusses proxies for \(\sigma_n\). Section 2.2.1 discusses how proxies may be compared. Section 2.2.2 shows how proxies may be combined into a superior one.

Let us first address how the theory in the chapter relates to the quadratic variation \(QV_n\) of the intraday log price process, and why \(\sqrt{QV_n}\) is in general not an ideal estimator of \(\sigma_n\). The daily quadratic variation \((QV_n)\) is the limit of the realized variance\(^4\) as the lengths of the sampling intervals approach zero. In the scaling model the scale factor \(\sigma_n\) determines \(\sqrt{QV_n}\) up to multiplicative noise,

\[
\sqrt{QV_n} = \sigma_n \sqrt{QV(\Psi)}.
\]

If \(\Psi\) is a standard Brownian motion then \(QV(\Psi) = 1\), so \(\sqrt{QV_n} = \sigma_n\). In this case the square root of the quadratic variation over the day is a perfect estimator of \(\sigma_n\). In general \(\sqrt{QV_n}\) is not an optimal estimator; if the standard process has continuous sample paths of bounded variation, then \(\sqrt{QV_n} \equiv 0\) is clearly not a good estimator of \(\sigma_n\). For another example with \(\sqrt{QV_n} > 0\), see Example 2.D.2.1 in Appendix 2.D.

We can say more. Suppose \(\Psi_n\) has the standardization \(\mathbb{E}QV(\Psi_n) = 1\). One could achieve this standardization through dividing \(\Psi_n\) by \(\sqrt{\mathbb{E}QV(\Psi_n)}\) (if \(0 < \mathbb{E}QV(\Psi_n) < \infty\)). Also, if \(\Psi_n\) is a martingale with \(\mathbb{E}\Psi_n^2(1) = 1\), then \(\mathbb{E}QV(\Psi_n) = 1\). Take for \(\sigma_n\) a GARCH scale factor (which is \(\mathcal{F}_{n-1}\)-measurable; here \(\mathcal{F}_{n-1}\) denotes information observed until

\(^4\)The realized variance may be obtained as the sum of the squared five-minute returns over the trading day.
2.2 Proxies

yesterday). The relation \( QV_n = \sigma_n^2 QV(\Psi_n) \) now yields

\[
\mathbb{E}(QV_n | \mathcal{F}_{n-1}) = \sigma_n^2.
\]

A common interpretation (e.g., Andersen, Bollerslev, Diebold, and Labys, following Corollary 1, 2003) is that quadratic variation is a natural (ex-post) estimator for the squared scale factor (or conditional return variance). Another insightful interpretation is that the **squared scale factor is the best forecaster (in terms of mean squared error) of quadratic variation**. Quadratic variation and scale factor are two different concepts for price fluctuations. Quadratic variation is a measure pertaining to the sample path fluctuations during the trading day. The scale factor is a **latent variable** that influences the size of the fluctuations during the trading day, and that predicts \( \sqrt{QV_n} \) (if \( QV_n \) exists).

### 2.2.1 Ranking Proxies

Since the square root of the quadratic variation is not necessarily an optimal estimator of \( \sigma_n \), it is natural to consider other estimators. We shall consider a general class of statistics \( H \) based on intraday data. Per day these statistics are applied to the sample path of the intraday return process. These statistics act as stand-ins for \( \sigma_n \), and are called proxies. They have a particular form. Proxies may be seen as estimators. We prefer to use the term “proxy” rather than “estimator”: the proxies do not estimate a fixed population parameter, but instead approximate a sequence of unobservable random variables.

A number of volatility proxies have appeared in the literature: the intraday high-low range (e.g. Parkinson, 1980), the realized volatility (e.g. Barndorff-Nielsen and Shephard, 2002b, and Andersen, Bollerslev, Diebold, and Labys, 2003), the sum of absolute returns (more generally the \( 1/p \)-power of the realized \( p \)-power variation, see Barndorff-Nielsen and Shephard, 2003, 2004), the square root of bipower variation (Barndorff-Nielsen and Shephard, 2004), the square root of the realized range (e.g. Martens and van Dijk, 2007, and Christensen and Podolskij, 2007). Many authors have motivated these statistics with a Brownian motion or a semimartingale in mind. These proxies all have the property of **positive homogeneity**: if the intraday return process \( R_n(\cdot) \) is multiplied by a factor \( \alpha \geq 0 \), then so is the proxy:

\[
H(\alpha R_n) = \alpha H(R_n), \quad \alpha \geq 0. \tag{2.5}
\]

A **proxy**\(^5\) is a positive statistic that satisfies the positive homogeneity (2.5). We shall refer to both the random variable \( H_n \),

\[
H_n \equiv H(R_n),
\]

\(^5\)For quadratic proxies we refer to the final paragraph of this section.
and the functional $H$ as proxies. The property of positive homogeneity fits naturally into the scaling model. Homogeneity implies $H(\sigma_n \Psi_n) = \sigma_n H(\Psi_n)$, and hence

$$H_n = \sigma_n H(\Psi_n).$$  \hfill (2.6)

Here, for each day $n$, $H(\Psi_n)$ is independent of $\sigma_n$. So in the scaling model a proxy equals $\sigma_n$ up to multiplicative noise. The sequence of random variables $H(\Psi_n)$ is iid. Equation (2.6) shows that a proxy provides information on the scale factor $\sigma_n$. If $\sigma_n$ is large the proxy $H_n$ tends to be large too.

The following decomposition (2.7) is central to the results of the chapter. Applying logarithms leads to an additive measurement equation:

$$\log(H_n) = \log(\sigma_n) + U_n.$$  \hfill (2.7)

So the log of a proxy consists of the sum of two independent terms, the log of the scale factor $\sigma_n$ and a measurement error $U_n \equiv \log(H_n/\sigma_n)$,

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

The measurement errors $U_n$ form an iid sequence, and importantly do not depend on the scale factors $\sigma_n$.

Let us look at the quality of a proxy. Consider the Mean Squared Error (MSE) for $\log(\sigma_n)$,

$$\mathbb{E}(\log(H_n) - \log(\sigma_n))^2 = \mathbb{E}U_n^2 = \text{var}(U_n) + (\mathbb{E}U_n)^2.$$  

In general the proxies $H_n$ are biased:

$$\mathbb{E}U_n \neq 0.$$  

To adjust for the bias, one may consider the proxy $H_n$ rescaled by a constant $c > 0$, which gives another proxy $\tilde{H}_n = c H_n$. So $\tilde{U}_n = U_n + \log(c)$. This yields the MSE:

$$\mathbb{E}(\log(\tilde{H}_n) - \log(\sigma_n))^2 = \text{var}(\tilde{U}_n) + (\mathbb{E}\tilde{U}_n)^2,$$

$$= \text{var}(U_n) + (\mathbb{E}U_n + \log(c))^2.$$  

The constant $c$ affects the MSE by changing the bias, but it does not affect the measurement variance $\text{var}(U_n)$. The constant $c = c^*$ that minimizes MSE is obtained by setting $\mathbb{E}\tilde{U}_n = 0$, yielding

$$\mathbb{E}(\log(\tilde{H}_n) - \log(\sigma_n))^2 = \text{var}(U_n).$$
So the problem of finding a positively homogeneous proxy $H$ with minimal MSE comes down to first minimizing measurement variance, and in a second and trivial step, removing the bias. It is therefore natural to compare the quality of two proxies by comparing their measurement variances. For a proxy $H$, let $\lambda^2$ denote measurement variance,

$$\lambda^2 = \text{var}(U_n).$$ \hspace{1cm} (2.8)

We say that a proxy $H^{(1)}$ is better than $H^{(2)}$ if it has a smaller measurement variance:

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

For this ranking to make sense, it has to be the same for all possible representations of $R_n(\cdot) = \sigma_n \Psi_n(\cdot)$. This is confirmed by Proposition 2.4 in Appendix 2.D.2.

In certain applications the bias $\mathbb{E}U_n \neq 0$ does not matter; in quasi maximum likelihood estimation of the autoregression parameters in a classical GARCH(1,1) recursion, the bias in $H_n$ plays no role and asymptotic efficiency is determined by $\lambda^2$ (see the log-Gaussian QMLE in Visser, 2008b); in obtaining the coefficient of determination $R^2$ for predictive ability of a volatility model, again the bias plays no role. This chapter focuses on the principal problem in finding good proxies, that is on obtaining proxies that have small measurement variance. Proxies with small measurement variance constitute a time series that resembles the time series of scale factors ($\sigma_n$), see Figure 2.2 (this is made precise by the theoretical fact that small measurement variance corresponds to large correlation, see equation (2.20) in Appendix 2.D.1). It is often practical to restrict attention to proxies that have finite measurement variance $\lambda^2 < \infty$, as does Theorem 2.1. For additional discussion of the measurement variance see Appendix 2.D.1.

An optimal proxy $H^\ast$ satisfies

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

For a proxy $H$, the measurement error $U_n$ only depends on the process $\Psi_n(\cdot)$. So the optimality of a proxy $H$ is independent of the particular discrete time model for the scale factors ($\sigma_n$). Appendix 2.D.3 provides a mathematically rigorous definition of proxy. Optimal proxies exist and they can be shown to be unique up to a constant factor, see Appendix 2.D.2, yet there does not seem to be a concrete way of determining this optimal proxy, nor of calculating its measurement variance.

A first practical step is to achieve a data-based ranking: how can one tell from the time series of realizations $H_n$ of several proxies, which one is the best? Taking variances
on both sides of the decomposition (2.7) gives
\[ \text{var}(\log(H_n^{(i)})) = (\lambda^{(i)})^2 + \text{var}(\log(\sigma_n)). \] (2.9)

There is no covariance term due to the independence of \( \sigma_n \) and \( \Psi_n(\cdot) \). Equation (2.9) shows that the variances of the proxies all have the common term \( \text{var}(\log(\sigma_n)) \). It follows that if the variance of the log proxy is smaller, then the measurement variance is smaller. Assuming \( \text{var}(\log(\sigma_n)) < \infty \) results in the equivalence
\[ (\lambda^{(1)})^2 \leq (\lambda^{(2)})^2 \iff \text{var}(\log(H_n^{(1)})) \leq \text{var}(\log(H_n^{(2)})). \] (2.10)

So, in empirical applications proxies may be simply ranked by estimating the variances of their logarithms. See Section 2.3.1 for ranking proxies for the S&P 500 index.

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

### 2.2.2 Combining Proxies

We now come to the main result of the chapter. To find a good proxy one first needs to think up some simple proxies. The procedure below can then be used for combining these into a single, more efficient proxy. Suppose we are supplied with the proxies \( H^{(1)}, \ldots, H^{(d)} \). Consider the geometric combination of these proxies,
\[ H_n^{(w)} = \prod_{i=1}^{d} (H_n^{(i)})^{w_i}, \quad w_1 + \ldots + w_d = 1, \ w_i \in \mathbb{R}. \] (2.11)

Here, the column vector \( w \) is the \( d \)-dimensional coefficient vector. The condition \( \sum w_i = 1 \) is needed to obtain a proxy, though the coefficients are not restricted to the interval \([0, 1]\). It is natural to have the coefficients \( w_i \) acting as exponents in equation (2.11), since taking logarithms now yields an additive problem. Let \( \Lambda \) denote the covariance matrix of the measurement errors \( U^{(i)} = \log(H^{(i)}(\Psi)) \):
\[ \Lambda = \text{cov}([U^{(1)}, \ldots, U^{(d)}]') \] (2.12)
2.2 Proxies

The measurement error $U^{(w)} \equiv \log(H^{(w)}(\Psi))$ of the geometric combination in (2.11) equals

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

which has variance $\lambda_w^2 = w' \Lambda w$ and, as for the global minimal variance portfolio in Markowitz portfolio theory (using straightforward Lagrangian optimization, as in the proof of Theorem 2.1 below), $\lambda_w^2$ is minimal for

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

(2.13)

with optimal variance $\lambda_{w^*}^2 = \frac{1}{\iota' \Lambda^{-1} \iota}$. This solution is empirically infeasible since the measurement errors $U_n^{(i)}$ are not observed. However, by equation (2.10) one may equivalently minimize $\text{var}(\log(H_n^{(w)}))$ to obtain $w^*$. Now, let $\Lambda_{p,n}$ denote the covariance matrix of the log of the simple proxies:

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

(2.14)

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

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

(2.15)

The optimal coefficients $w^*$ may now be obtained upon replacing $\Lambda$ by $\Lambda_{p,n}$ in equation (2.13), see formula (2.16) below. In empirical applications one may want to assume stationarity for $(\sigma_n)$, so that the covariance matrix

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

may simply be estimated by the sample covariance matrix.

Theorem 2.1 summarizes. Its main point is that one merely needs the covariance matrix for the log-proxies (and not for the log measurement errors) to obtain the weights $w^*$ in (2.13).

**Theorem 2.1** Let $R_n(\cdot)$ satisfy the scaling model. Assume $(\lambda^{(i)})^2 = \text{var}(\log(H^{(i)}(\Psi))) < \infty$ for $i = 1, \ldots, d$, and $\text{var}(\log(\sigma_n)) < \infty$. Let the covariance matrices $\Lambda$ and $\Lambda_{p,n}$ be defined by (2.12) and (2.14). The optimal coefficient vector $w^*$ in (2.13) does not depend on the form of the process $(\sigma_n)$ and may be expressed as

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

(2.16)
The variance of the logarithm of the optimal geometric combination is
\[
\text{var}(\log(H_n^{(w)})) = \lambda_w^2 + \text{var}(\log(\sigma_n)),
\]
where \(\lambda_w^2 = \frac{1}{\iota' \Lambda_{p,n}^{-1} \iota} \) is its measurement variance.

**Proof.** The optimal coefficient \(w^*\) does not depend on \((\sigma_n)\): by equation (2.10)
\[
\arg \min_w \text{var}(\log(H_n^{(w)})) = \arg \min_w \lambda_w^2 \equiv \arg \min_w \text{var}(\log(H_n^{(w)}(\Psi_n))),
\]
so
\[
\arg \min_w w' \Lambda_{p,n} w = \arg \min_w w' \Lambda w. \tag{2.17}
\]

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 (2.17) 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
\[
\text{var}(\log(H_n^{(w^*)})) = \lambda_{w^*}^2 + \text{var}(\log(\sigma_n)).
\]

**Remark 2.1** In empirical applications one uses estimates of the covariance matrix. To reduce estimation error, we shall use the technique of prescaling, see Appendix 2.C.

We end this section with a few words regarding empirical implementation. Assuming stationarity for the process \((\sigma_n, \Psi_n)\), the covariance matrix \(\Lambda_{p,n} = \Lambda_{p}\) is consistently estimated by the sample covariance matrix of the log of the proxies, thereby providing coefficients \(\hat{w}\) that are consistent for \(w^*\). More generally, this estimator for \(w^*\) may remain consistent while allowing, for example, for structural breaks in the scale factors \((\sigma_n)\). See the consistency condition (2.21) in Appendix 2.D.3 for details.

### 2.3 A Good Volatility Proxy for the S&P 500 Index

This section constructs a good volatility proxy for the S&P 500 index, applying the techniques from Section 2.2 to the S&P 500 futures tick data over the years 1988–2006, a total of 4575 trading days. Appendix 2.A describes the data. The proxies below are constructed in such a way as to avoid microstructure noise, see Appendix 2.B.
2.3 A Good Volatility Proxy for the S&P 500 Index

2.3.1 Ranking Proxies

There are many possible proxies for the scale factor $\sigma_n$. Table 2.1 compares twelve simple proxies constructed from the data.

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

Table 2.1: Performance of twelve proxies. The table gives the PV: the variance of the logarithm after prescaling by $\text{EWMA}(0.7)$ predictor for RV5. The full sample is split into four subsamples. The following proxies are included. We abbreviate square root to sqrt.: RV5: sqrt. of sum of squared 5-min. returns; RV10: sqrt. of sum of squared 10-min. returns; RV15: sqrt. of sum of squared 15-min. returns; RV20: sqrt. of sum of squared 20-min. returns; RV30: sqrt. of sum of squared 30-min. returns; abs-r: absolute close-to-close return; hl: high-low of the intraday return process; maxar2: maximum of the absolute 2-min. returns; RAV5: sum of absolute 5-min. intraday returns; RAV10: sum of absolute 10-min. intraday returns; RVHL10: sqrt. of sum of 10-min. squared high-lows; RAVHL10: sum of 10-min. high-lows;

For each proxy a measure of comparison is given for five samples: first the full sample (days 2 to 4575) and then for four subsamples spanning the full sample (2:1144, 1145:2287, 2288:3431, 3432:4575). The measure of comparison is $PV$ (prescaled variance), which is the variance of the logarithm of a proxy $H$ after prescaling by a suitable series $p_n$ of $F_{n-1}$-measurable positive random variables,

$$PV(H^{(i)}) = \text{var}(\log(H_n^{(i)}/p_n))$$

$$= \text{var}(\log(\sigma_n/p_n)) + (\lambda^{(i)})^2.$$ 

Smaller $PV$s correspond to more efficient proxies. Prescaling is an empirical technique that does not affect the theoretical ranking of proxies, but is used for diminishing statistical noise, see Appendix 2.C. The better the prescaling sequence (the better $p_n$ predicts $\sigma_n$), the smaller statistical noise. The first observation cannot be prescaled and is left out of the variance calculations. 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 $p_n = 0.7p_{n-1} + 0.3 RV5_{n-1}$. We set the smoothing parameter so that the sample variance of the logarithm of prescaled five-minute
Table 2.2: Performance of upward/downward decomposed proxies. The table splits proxies from Table 2.1 according to upward and downward price movements. For example, RV5-up is \( \sqrt{\text{sum of squared 5-min positive returns}} \), RAV10HIGH is the sum of 10-min highs (taking the start of the 10-min interval as the origin), and RAV10LOW is the sum of 10-min absolute lows (taking the start of the 10-min interval as the origin).

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

Realized volatility is minimal. A smoothing parameter around \( \beta = 0.7 \) for a realized volatility filter was found to fit well in earlier research, see for instance Engle (2002). Recall that the prescaled variance ranks proxies, but the measurement variance \( \lambda^2 \) itself remains unknown.

The first column of Table 2.1 gives the prescaled variances over the full sample. The standard five-minute realized volatility RV5 has \( PV = 0.064 \). The first column shows that the quality of the realized volatility RV improves if one increases the sampling frequency from 30 minutes to 5 minutes. The prescaled variance is maximal for the absolute close-to-close returns, confirming that absolute or squared daily returns are poor proxies. Note that the maximal absolute two-minute return outperforms the intraday high-low range, which tends to use returns based on much longer time spans. Overall, we find that sums of absolute values lead to more efficient proxies than sums of squared values. This observation relates to a finding of Barndorff-Nielsen and Shephard (2003), whose simulations indicated that absolute power variation, based on the sum of absolute returns, has better finite-sample behaviour than the realized variance. The best performing proxy in Table 2.1 is RAVHLL10, the sum of the ten-minute high-low ranges. The remaining columns of Table 2.1 show that the ranking of the different proxies is stable: the ranking in the subsamples is the same as in the full sample, with one exception in the second subsample for RV30 and maxar2. Though the measurement variances are not observed, from the full sample column we may infer that the measurement variance \( \lambda^2 \) of RAVHLL10 is at least 25% smaller than the measurement variance of RV5, by \( (0.064 - 0.047)/0.064 \approx 0.27 \).

\( ^6 \)We use the absolute returns larger than 0.001, or 10 basis points, to avoid taking the log of zero. This leaves 4079 daily returns.
Table 2.2 splits proxies from Table 2.1 into upward and downward components. For instance, the five-minute realized volatility is decomposed according to upward and downward price movements:

\[
(RV5)^2 = \sum_{i=1}^{m} r_{n,i}^2
= \sum_{i=1}^{m} r_{n,i}^2 I\{r_{n,i}>0\} + \sum_{i=1}^{m} r_{n,i}^2 I\{r_{n,i}<0\}
= (RV5-up)^2 + (RV5-down)^2.
\]

Here, \(r_{n,i}\) denotes the return over the \(i\)-th intraday five-minute interval on day \(n\). There are \(m = 81\) five-minute intervals each day. As one would expect \(RV5 (PV=0.064)\) is better than \(RV5-up (PV=0.066)\) and \(RV5-down (PV=0.094)\). Note that the upward proxies are consistently more efficient than their downward counterparts. This difference suggests that, when proxying the scale factor \(\sigma_n\), more weight should be put on the upward movements.

### 2.3.2 Optimized Combination

The proxies in Tables 2.1 and 2.2 may each be of value for measuring the scale factor \(\sigma_n\), but certain proxies may be more useful than others. This section combines the proxies in Tables 2.1 and 2.2 into a more efficient one using the combination formula from Section 2.2.2. We also conduct a thorough stability analysis, and discuss properties of the optimized proxy.

The five-minute realized volatility \(RV5\) is a standard proxy, and has \(PV=0.064\). Let us improve upon this value. First, by using the (square root of the) sum of the squared high-low ranges over intraday intervals, \(RVHL10\) has \(PV = 0.053\), see Table 2.1. It is better to use absolute values: \(RAVHL10\) has the smaller \(PV = 0.047\). Now use the theory from Section 2.2.2 to combine the high-low ranges in \(RAVHL10\) with the absolute returns in \(RAV10\): inserting the covariance matrix \(\hat{\Lambda}_{p,n}\) of the log of these two prescaled proxies into formula (2.16), yields the proxy

\[
H_n = (RAVHL10_n)^{1.82}(RAV10_n)^{-0.82}, \quad (PV = 0.041).
\]

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

\[
H_n^{(\hat{w})} = (RAV10HIGH_n)^{1.04}(RAV10LOW_n)^{0.72}(RAV10n)^{-0.76}, \quad (PV = 0.038). (2.18)
\]

Of course, one may also apply the optimal coefficient formula (2.16) to all twenty-one
proxies in Tables 2.1 and 2.2 at once. The full combination yields a proxy with $PV = 0.037$, which only marginally outperforms the proxy obtained in (2.18). We prefer to work with the simpler proxy in (2.18) and we refer to it as the optimized proxy $H_{n}^{\hat{w}_{1}}$.

The optimized proxy easily outperforms all proxies in Tables 2.1 and 2.2. If one extrapolates the full sample prescaled variances of the realized volatilities of Table 2.1 to a time interval of length zero (corresponding to the limiting case of the quadratic variation), this results in a value between 0.050 and 0.060. The value $PV = 0.038$ for the optimized proxy is well below these values, suggesting that this proxy for the daily scale factor is more efficient than the square root of the quadratic variation. Indeed, the optimized proxy has a measurement variance $\lambda^2$ which is at least 40% smaller than the five-minute realized volatility, by $(0.064 - 0.038)/0.064 \approx 0.41$, and may be 25% smaller than the measurement variance of the square root of the quadratic variation, by $(0.05 - 0.038)/0.05 \approx 0.24$.

Note 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 similar, since all proxies approximate the same daily scale factor $\sigma_n$. The optimized proxy is not in the convex hull of the proxies in Tables 2.1 and 2.2. The original proxies do not completely reflect the direction of the optimal proxy. The coefficients outside $[0, 1]$ correct the direction.

Table 2.3 investigates the stability of the optimized proxy $H_{n}^{\hat{w}_{1}}$. As in Table 2.1 it lists performance measures for the full sample and for four subsamples. The first row gives the performance of $H_{n}^{\hat{w}_{1}}$ in the different subsamples; comparison with Tables 2.1 and 2.2 shows that $H_{n}^{\hat{w}_{1}}$ outperforms all those proxies in every subsample. The proxy $H_{n}^{\hat{w}_{i}}$ is constructed using the coefficients that are optimal for the $i$-th subsample. In the first subsample the performance of the globally optimized $H_{n}^{\hat{w}_{1}}$ (PV=0.038) is not substantially outperformed by $H_{n}^{\hat{w}_{1}}$ (PV=0.039). A similar statement holds for the other subsamples. Moreover, proxies based on any particular subsample are close to optimality in all other subsamples. For instance, the proxy optimized for the first subsample (the years 1988–1992) is nearly optimal for the years 2002–2006. We conclude therefore that

---

7We exclude the absolute close-to-close return in performing this calculation.
the optimality of $H_n^{(w)}$ is stable.

Proxies are important for volatility forecast evaluation. Good proxies help to distinguish better from poorer forecasts, see for instance Hansen and Lunde (2006a) and Patton and Shephard (2009a). Table 2.4 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 $H^{(j)}$ on the logarithm of another proxy $H^{(i)}$ lagged by one day:

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

Large $R^2$s in a particular column mean that the proxy in that column is largely predictable, suggesting that it is a good proxy to evaluate volatility forecasts. The $R^2$s are greatest for the optimized proxy, in the rightmost column.\footnote{The optimized proxy is constructed as an optimized proxy for $\sigma_n$, not as an optimal predictor for $\sigma_{n+1}$. Even so, the $R^2$s attained in the bottom row $H^{(w)}(-1)$ are large.}

<table>
<thead>
<tr>
<th></th>
<th>RV30</th>
<th>RV20</th>
<th>RV15</th>
<th>RV10</th>
<th>RV5</th>
<th>$H^{(w)}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>RV30(-1)</td>
<td>0.35</td>
<td>0.39</td>
<td>0.42</td>
<td>0.46</td>
<td>0.50</td>
<td>0.58</td>
</tr>
<tr>
<td>RV20(-1)</td>
<td>0.38</td>
<td>0.42</td>
<td>0.45</td>
<td>0.49</td>
<td>0.54</td>
<td>0.61</td>
</tr>
<tr>
<td>RV15(-1)</td>
<td>0.39</td>
<td>0.44</td>
<td>0.47</td>
<td>0.50</td>
<td>0.55</td>
<td>0.63</td>
</tr>
<tr>
<td>RV10(-1)</td>
<td>0.41</td>
<td>0.45</td>
<td>0.48</td>
<td>0.52</td>
<td>0.57</td>
<td>0.66</td>
</tr>
<tr>
<td>RV5(-1)</td>
<td>0.43</td>
<td>0.48</td>
<td>0.51</td>
<td>0.55</td>
<td>0.60</td>
<td>0.69</td>
</tr>
<tr>
<td>$H^{(w)}(-1)$</td>
<td>0.44</td>
<td>0.48</td>
<td>0.51</td>
<td>0.54</td>
<td>0.60</td>
<td>0.71</td>
</tr>
</tbody>
</table>

Table 2.4: $R^2$ of the regression $\log(H_n^{(j)}) = \alpha + \beta \log(H_{n-1}^{(i)}) + \varepsilon_n$, for $i, j = 1, \ldots, 6$, and $n = 2, \ldots, 4575$.

Table 2.5 gives an impression of the distributions of the measurement errors of the five-minute realized volatility and of the optimized proxy. It provides descriptive statistics based on the logarithm of the prescaled proxies. We do not provide the sample averages since the quality of a proxy is insensitive to scaling, see Appendix 2.D.1. It appears that putting less weight on the sum of the lows, and more on the sum of the highs, helps to diminish the skewness. The optimized proxy is more symmetrical and more concentrated than the five-minute realized volatility.

<table>
<thead>
<tr>
<th></th>
<th>log($RV5_n/p_n$)</th>
<th>log($H_n^{(w)}/p_n$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>st. dev.</td>
<td>0.25</td>
<td>0.20</td>
</tr>
<tr>
<td>skewness</td>
<td>0.62</td>
<td>0.00</td>
</tr>
<tr>
<td>kurtosis</td>
<td>6.00</td>
<td>4.05</td>
</tr>
</tbody>
</table>

Table 2.5: Proxy distributions. For the full sample, $n = 2, \ldots, 4575$: standard deviation, skewness, and kurtosis of the logarithm of the prescaled proxy, log($H_n/p_n$). Prescaling filter $p_n$ is EWMA($\beta = 0.7$) based on RV5.

Finally, Figure 2.2 shows the time series graphs of four different proxies. The proxies were standardized to have mean one, by dividing them by their mean. From top to bottom
the curves become “less erratic”, suggesting a decrease in the variance of the measurement errors $U_n$. Each step shows a marked improvement.

2.4 Conclusions

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

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

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

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

Figure 2.2: Time series of four standardized proxies, $H_n/\bar{H}_n$. 
Appendices

2.A Data

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

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

An advantage of using future data rather than the S&P 500 cash index is the absence of non-synchronous trading effects which cause positive autocorrelation between successive observations, see Dacorogna et al. (2001). As in the cash index there may be bid-ask effects in the future prices which induce negative autocorrelation between successive observations. We deal with these effects by taking large enough time intervals, see Appendix 2.B. Since we study a very liquid asset the error term due to microstructures is relatively small.

2.B Microstructure Noise Barrier

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

In this chapter the measure of comparison is the variance of the logarithm. The standard realized volatility $RV$ and the realized range $RVHL$ (see Table 2.1) depend on the sampling interval $\Delta u$. Figure 2.3 shows the graph of $\Delta u \rightarrow \hat{\text{var}}(\log(H^{\Delta u}(R_n)))$, for $\Delta u$ ranging from zero to sixty minutes. These curves suggest that a qualitative change of behaviour occurs for $\Delta u \approx$ five minutes for realized volatility, and $\Delta u \approx$ eight minutes for
realized range. The realized volatilities in the chapter are based on five-minute sampling intervals or larger. For realized range our minimal sampling interval is ten minutes.

Figure 2.3: Plots of the sample variance of the log of a proxy with $\Delta u$ ranging from zero to 60 minutes (zero is tick per tick). (a) Realized volatility, $RV$. (b) Realized range, $RVHL$.

2.C Prescaling

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

Definition 2.2 A prescaling sequence $(p_n)$ is an $(F_{n-1})$ adapted sequence of strictly positive random variables.

The prescaling factors $p_n$ are used for defining adjusted scale factors

$$\tilde{\sigma}_n = \sigma_n/p_n.$$

Proposition 2.3 Assume that the processes $(R_n)$ satisfy the scaling model. Prescale the scale factors $(\sigma_n)$ to obtain the sequence $(\tilde{\sigma}_n)$ above. The corresponding processes $(\tilde{R}_n)$, where $\tilde{R}_n = \tilde{\sigma}_n \Psi_n$, satisfy the scaling model.

Proof. The variables $(\tilde{\sigma}_n)$ are positive. Both $p_{n+1}$ and $\sigma_{n+1}$ are $G_n$-measurable, hence so is $\tilde{\sigma}_{n+1}$. Therefore $\tilde{R}_n$ satisfies the scaling model. $\blacksquare$
As a result one obtains proxies for $\tilde{\sigma}_n$. These are prescaled proxies:

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

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

$$\tilde{U}_n = \log(\tilde{H}_n) - \log(\tilde{\sigma}_n) = \log(H_n) - \log(p_n) - (\log(\sigma_n) - \log(p_n)) = U_n.$$  

So, ranking and optimizing proxies before and after prescaling are equivalent in terms of population statistics; therefore one may replace $\sigma_n$ by $\tilde{\sigma}_n$, and $H_n$ by $\tilde{H}_n$ in the chapter. As a consequence, the population value of the term $\text{var}(\log(\sigma_n))$ in equations (2.9) and (2.15) changes into

$$\text{var}(\log(\tilde{\sigma}_n)) = \text{var}(\log(\sigma_n/p_n)).$$  

A perfect predictor $p_n$ of $\sigma_n$ results in $\text{var}(\log(\tilde{\sigma}_n)) = 0$.

2.D Mathematical Details

2.D.1 Properties of the Measurement Variance

Let us here provide additional intuition on the measurement variance $\lambda^2$. Proxies are the product of scale factor and noise. Since proxies are positive and may have a heavy tail, it is natural to apply logarithms, yielding (as in Equation (2.7)),

$$\log(H_n) = \log(\sigma_n) + U_n.$$  

In the ideal situation of a perfect proxy, i.e. zero measurement variance, a proxy gives the scale factor up to a constant factor $c > 0$:

$$H_n = c \sigma_n.$$  

This underlines that the measurement variance $\lambda^2$ is central to the quality of a proxy. Changing the measurement units of a proxy by a positive factor $c > 0$ does not change

---

9In practice proxies are strictly positive. An exception is the absolute close-to-close return $|r_n|$, for which one could either set the measurement variance to infinity, or exclude the zeros from the sample. In either case $|r_n|$ proves a poor proxy.
the measurement variance $\lambda^2$:

$$\text{var}(\log(cH(\Psi))) = \text{var}(\log(H(\Psi))) = \lambda^2. \quad (2.19)$$

From a practical point of view, it is an advantage that the criterion $\lambda^2$ is insensitive to changes in scale: one does not need to rule out potentially biased proxies, which bypasses the difficulty of telling a priori whether a proxy is unbiased, and the difficulty of finding a priori a suitably rescaled version of each proxy. A good proxy is such that its time series $(H_n)$ has a high degree of comovement with the series $(\sigma_n)$. This is confirmed by looking at correlations. Assume $0 < \text{var}(\log(\sigma_n)) < \infty$. It then follows, using the decomposition (2.7), that

$$\text{corr}(\log(H_n), \log(\sigma_n)) = \left(1 + \frac{\lambda^2}{\text{var}(\log(\sigma_n))}\right)^{-1/2}. \quad (2.20)$$

So log proxies with smaller measurement variance $\lambda^2$ have larger correlation with $\log(\sigma_n)$. The ideal situation of zero measurement variance gives perfect correlation.

### 2.D.2 Identification and Optimality

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

**Proposition 2.4** Suppose $H^{(1)}$ and $H^{(2)}$ are proxies. Moreover, assume $(\sigma_n, \Psi_n)$ and $(\sigma_n', \Psi_n')$ both satisfy the scaling model for $R_n(\cdot)$. If $H^{(1)}$ is better than $H^{(2)}$ for $\Psi$, then $H^{(1)}$ is also better than $H^{(2)}$ for $\Psi'$.

**Proof.** By assumption $\sigma_n'\Psi_n' = \sigma_n\Psi_n$. Independence of $\sigma_n$ and $\Psi_n$ implies

$$\text{var}(\log(\sigma_n')) + \text{var}(\log(H^{(1)}(\Psi_n'))) = \text{var}(\log(\sigma_n)) + \text{var}(\log(H^{(1)}(\Psi_n)))$$

$$\leq \text{var}(\log(\sigma_n)) + \text{var}(\log(H^{(2)}(\Psi_n)))$$

$$= \text{var}(\log(\sigma_n')) + \text{var}(\log(H^{(2)}(\Psi_n')))$$

Hence $\text{var}(\log(H^{(1)}(\Psi_n'))) \leq \text{var}(\log(H^{(2)}(\Psi_n')))$. ■

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

**Example 2.D.2.1** Take $\Psi(\cdot)$ a diffusion:

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

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

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

Theorem 2.5 If there is a proxy with finite measurement variance, then there exists an optimal proxy.


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

Proposition 2.6 Suppose $H^{(1)}$ and $H^{(2)}$ are two optimal proxies. Then there is a constant $a > 0$, such that $H^{(1)}(\Psi) \overset{a.s.}{=} aH^{(2)}(\Psi)$.


2.D.3 Proof of Existence of Optimal Proxies

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

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

Definition 2.7 Let $H$ be a measurable, positively homogeneous functional $D \to [0, \infty)$, on a linear subspace $D$ of $\mathbb{D}[0, 1]$. Assume $\Psi \in D$ a.s., and $H(\Psi) > 0$ a.s. Then $H$ is a proxy functional. The random variable $H_n = H(R_n)$ is a proxy.

Usually there is no danger of misunderstanding, and we refer to both $H$ and $H_n$ as proxies. The condition $H(\Psi) > 0$ ensures that the measurement errors $U_n = \log(H(\Psi_n))$ are
well-defined almost surely. It can sometimes be practical to allow for a proxy that is zero with positive probability. One example of such a proxy is the absolute close-to-close return $|r|$. One then defines the measurement variance $\lambda^2_{|r|} = \infty$. Notice that Theorem 2.5 assumes the existence of a proxy with $\lambda^2 < \infty$.

Proof of Theorem 2.5. 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 $H$. The proof uses standard Hilbert space arguments.

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.

Define $\lambda_{in}^2 = \inf_{H: \log(H) \in \mathcal{U}^0} \{\lambda_H^2\}$. Consider the space $L^2(D, \mathcal{B})$, of equivalence classes $[U]$ of log proxy functionals $U$, with inner product $<[U]^{(1)}, [U]^{(2)}> = \mathbb{E}(U^{(1)}(\Psi)U^{(2)}(\Psi))$. Here, $\mathcal{B}$ denotes the Borel sigma-field for $D$. Notice that $\mathcal{U}^0$ is a subset of $L^2$ and that $\lambda$ 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 $||U_i|| \to \lambda_{in}$. 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 $||\frac{U_m + U_n}{2}||^2 \geq \lambda_{in}^2$. Therefore $||U_m - U_n||^2 \leq -4\lambda_{in}^2 + 2\lambda_m^2 + 2\lambda_n^2 \to 0$ 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 $\lambda_0^2 = \lambda_{in}^2$. 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 $D_0$ almost everywhere in $D$. Define on the convergence set $D_0$: $H^*(f) = \lim H^{(i_k)}(f)$. For $\{\alpha f : f \in D_0, \alpha \notin D_0, \alpha \in [0, \infty), \}$ 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 $f_1, f_2 \in D_0$, $\alpha_1, \alpha_2 > 0$, and $f = \alpha_1 f_1 = \alpha_2 f_2$. Then $H^*(\alpha_1 f_1) \equiv \alpha_1 H^*(f_1) = H^*(\alpha_2/\alpha_1 f_2)$. By homogeneity of $H^*$ on $D_0$ 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 
\[ U_0(\Psi) \overset{a.s.}= \log(H^*(\Psi)) \text{ and } \operatorname{var}(U_0(\Psi)) = \lambda_0^2 < \infty. \] Finally, \( \operatorname{var}(\log(H^*(\Psi))) = \lambda_{in}^2 \leq \lambda_H^2 \) for all \( H \).

**Lemma 2.8** Consider a proxy \( H \) and an optimal proxy \( H^* \). Then \( \operatorname{cov}(\log(H^*(\Psi)), \log(H(\Psi))) = (\lambda^*)^2. \)

**Proof of Lemma 2.8.** Consider the proxy functional \( H(f) \equiv (H^*(f))^w (H(f))^{1-w} \), with measurement variance \( \lambda_w^2 = w^2(\lambda^*)^2 + 2w(1-w) \operatorname{cov}(\log(H^*(\Psi)), \log(H(\Psi))) + (1-w)^2 \lambda^2. \) Since \( H^* \) is optimal, \( \partial \lambda_w^2/\partial w \big|_{w=1} = 0. \) Hence \( \operatorname{cov}(\log(H^*(\Psi)), \log(H(\Psi))) = (\lambda^*)^2. \)

**Proof of Proposition 2.6.** Both proxies have measurement variance \( (\lambda^*)^2. \) Let \( H_0 \) denote the centered proxy: \( H_0 = \exp(-E \log(H(\Psi))) \) \( H \), with \( E \log(H(\Psi)) = 0. \) Consider the covariance of the centered log proxies: \( \operatorname{cov}(\log(H_0^{(1)}(\Psi)), \log(H_0^{(2)}(\Psi))). \) By Lemma 2.8 this covariance equals \( (\lambda^*)^2. \) Applying 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 \( a > 0. \)

2.D.4 Consistency Condition for the Coefficients \( \hat{w} \)

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

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

\[ \widehat{\operatorname{var}}(\log(\mathbb{H}(R_n))) = \widehat{\operatorname{var}}(\log(\sigma_n)) \cdot \iota' + \widehat{\operatorname{var}}(\log(U_n)) + 2 \cdot \widehat{\operatorname{cov}}(U_n, \log(\sigma_n) \cdot \iota). \]

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

\[ \hat{w} = \arg \min_w w' \left( \widehat{\operatorname{var}}(\log(U_n)) + 2 \cdot \widehat{\operatorname{cov}}(U_n, \log(\sigma_n) \cdot \iota) \right) \cdot w. \]

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

\[
\hat{\text{cov}}(U_n, \log(\sigma_n) \cdot \iota) \xrightarrow{P} 0, \quad N \to \infty.
\]  

(2.21)

In addition to existence of second moments and the independence of \( U_n \) and \( \log(\sigma_n) \), the stationarity for \((\sigma_n, \Psi_n)\) is a sufficient, but not necessary, condition that ensures that the consistency condition (2.21) holds.