Estimation and Forecast Evaluation of Risk Measures with High Frequency Financial Data

Colin T. Bowers

Department of Economics, Macquarie University

October 2014

Contents

List of Figures

List of Tables

Statement

This thesis is my own work, and none of the material has been previously submitted as part of a higher degree at any other university or institution.

Colin T. Bowers

Acknowledgements

I thank my principal supervisor, Christopher Heaton, and in particular, would like to emphasize the enormous amount of time he has dedicated to furthering my education and research abilities - far beyond what any doctoral candidate could reasonably expect. I also thank my associate supervisor, Roselyne Joyeux, for her valuable contribution.

I thank my parents, Margaret and John Bowers, and sister, Karen Bowers, for their support throughout this period, and all the periods that came before. To this day, my father remains the most insightful economist I have ever encountered. I treasure our weekly discussions on this subject. My mother's contribution to my mathematical education cannot be overstated. She is to be particularly thanked for her patience when I was a less willing student than I am now.

Last, but not least, I thank my wife, Samantha Bowers, and son, Flynn Bowers. My wife, for her patience, love, and perhaps most importantly, sense of humour. I give her less credit than I should for her insistence that I explain my research in simple terms a non-mathematical audience can appreciate. It brings to mind a quote variously attributed to both Einstein and Feynmann: "If you can't explain it to a six year old, you don't really understand it.". I thank my son for being a source of joy to all those around him. Time will tell if I am able to explain my thesis to him when he is six.

Abstract

This thesis contributes to the financial econometric literature in the areas of estimation, forecasting, and forecast evaluation, using high frequency financial data. The thesis focuses on the use of this data to estimate risk parameters commonly described at lower frequencies, e.g. using *intraday* data to estimate *daily* variance or *daily* value-at-risk. A common theme in all chapters is that the use of high frequency data can dramatically improve the solutions to common financial problems.

Chapter 2 demonstrates how a dependent bootstrap can be used to consistently estimate a wide range of risk measures associated with a daily return, given a sequence of intraday returns. Estimable parameters include variance, value-at-risk, expected shortfall, semi-variance, skewness, kurtosis, and robust risk measures. Excluding the case of variance, all estimators are, to the best of my knowledge, the first of their kind in the literature: non-parametric, and consistent, in the presence of market microstructure noise. The theory also contains a new result on the convergence of bootstrapped parameters that is more generally applicable in the theoretical literature on dependent bootstraps.

Chapter 2 also demonstrates an application of the proposed estimation methodology. The method is used to construct a consistent proxy for value-at-risk which is used to rank value-at-risk forecast models in a dual-asymptotic framework. The approach is shown via both simulation and empirical work to exhibit much greater power to distinguish between competing value-at-risk forecasts than other tests in the literature.

Chapter 3 extends the empirical application in Chapter 2 to 351 value-at-risk forecast models, and to a larger dataset which spans two exchanges and two distinct forecasting intervals. A new class of value-at-risk forecast models based on the estimation methodology from Chapter 2 are proposed, and are shown to provide more accurate forecasts than all other models under consideration. More generally, the results strongly suggest that value-at-risk forecasts that utilise simple time series models of proxies based on intraday data significantly outperform forecasts which utilise daily data exclusively.

Chapter 4 proposes a data-based method for ranking variance estimators constructed from intraday data. This paper draws from the literature on loss-based forecast evaluation, but accounts for the inevitable dependencies that occur when ranking *estimators* as opposed to *forecasts*. Under certain conditions, the method is shown via simulation to exhibit greater power than other methods in the literature. The chapter also contains a new technical result on the product of near epoch dependent processes that is widely applicable in the time-series literature.

Chapter 1

Introduction

Many important financial decisions are made precisely once per day. Examples include the allocation of capital cushions at financial institutions, and the portfolio re-balancing decisions of a fund manager. These daily decisions are typically based on some measure of the risk of a financial position. In the above examples, value-at-risk and variance are the most common measures, respectively.

The daily cycle of decisions is usually taken to imply that these risk measures, or parameters, are studied at a daily frequency. That is, that they are parameters of the distribution of a daily return on a risky asset. This distribution is unknown, and so the risk parameters are not directly observable. Further complicating the analysis of these parameters is the fact that they appear to change from day to day. Despite these problems, the importance of these parameters for financial decision-makers has ensured that a significant portion of the existing financial econometrics literature is devoted to their estimation and forecast. Building more accurate *forecasts* is useful as it allows economic agents to make better-informed financial decisions. Continuing with the above two examples, improved accuracy in value-at-risk forecasts enables financial institutions to allocate capital cushions that are large enough to afford sufficient protection from adverse market movements, yet not so large as to be wasteful. Improved accuracy in variance forecasts enables fund managers to more accurately estimate the optimal portfolio weights on any given day. Building more accurate *estimates* is useful for many reasons, although two stand out in particular. First, under certain modelling assumptions, these estimates can be used as proxies for the true risk parameters in *ex post* forecast evaluation procedures. Second, accurate estimates themselves are frequently useful as predictive variables in forecast models.

Analysis of daily risk parameters was significantly altered with the introduction of high frequency, or intraday, data. Under certain modelling assumptions, consistent estimation of *daily* variance[1] is now available. The first such estimator, commonly referred to as "realised variance", is the sum of squared intraday returns, and was shown

[1]Application of intraday data to value-at-risk has been less common in the literature, although is central to this thesis.

to converge to daily variance in a continuous-time framework by Merton (1980).[2] The consistency of the estimator is achieved as the number of intraday transactions grows. The benefit of using realised variance as a proxy for true daily variance in a forecast evaluation context was made very clear by Andersen & Bollerslev (1998), who debunked the findings in several earlier studies that relied on squared daily returns as a proxy. Further, there is growing evidence that volatility forecasts themselves can be greatly improved by employing these consistent estimators as predictors.[3] The statistical gains provided by realised variance have also been shown to have a measurable economic benefit. Fleming, Kirby & Ostdiek (2003) extend the results in an earlier paper[4] on this subject and find that a typical risk-averse investor would be willing to pay 50 to 200 basis points per year to use forecasts based on realised variance, instead of forecasts based on daily returns, in a volatility timing strategy.

In recent years, the availability of high frequency financial data has massively increased. For many markets, the limit has been reached: every transaction and quotation is now available for the interested empiricist. This has spurred a large collection of econometric papers analysing estimates and forecasts of risk measures using this high frequency data. The material in this thesis is taken from three papers that sit firmly within this collection. All three papers are available on the *Social Science Research Network* (see footnotes 5, 6, and 7) and are co-authored with my doctoral supervisor, Christopher Heaton. However, I am the sole author of this thesis, and any mistakes or omissions are mine alone.

To date, most of the existing literature on high frequency data has focused on intraday data-based estimators of daily variance. Intraday data-based estimation of other daily risk parameters has, thus far, been largely overlooked in the literature. Chapters 2 and 3 address this gap, with a particular focus on value-at-risk. In Chapter 4, I provide a method for empirically ranking the large number of intraday data-based variance estimators in the literature.

In the remainder of this introduction, I provide some more technical details on the contents of these chapters.

Across all three chapters, it is worth making a careful distinction between an *estimate* and a *forecast*. For the purposes of this thesis, an estimate of a parameter on day t is assumed to be constructed using data, intraday or otherwise, from day t, and only day t. A forecast of a parameter on day t is assumed to use data from any day up to and including day $t-1$.

A paper titled *Bootstrapping Daily Returns*[5] provides the material for Chapter 2. This chapter is about the construction of estimators using intraday data. Specifically, it is demonstrated that a stationary bootstrap can be used to estimate a wide range

[2] Specifically, in the appendix of that paper.
[3] See Section 3.4.
[4] Fleming, Kirby & Ostdiek (2001).
[5] Bowers & Heaton (2014*a*).

of risk measures associated with the distribution of a daily return. Consistency is achieved as the number of intraday observations grows. Examples of parameters which can be estimated by this procedure include variance, quantiles (e.g. value-at-risk), expected shortfall, semi-variance, skewness, kurtosis, and L-estimators. Except for the case of variance, which has been studied extensively, the proposed estimators of these parameters are, to my knowledge, the first of their kind in the literature: consistent, non-parametric, intraday data-based estimators that are robust to market microstructure effects. Chapter 2 also contains a new result on the convergence of bootstrapped parameters. Applications for this class of estimators include parameter estimation, forecast, and forecast evaluation. Regarding parameter estimation, the methods discussed in Patton (2011*a*) are used to demonstrate that the proposed variance estimator has accuracy comparable to the popular realised kernels estimator of Barndorff-Nielsen, Hansen, Lunde & Shephard (2008*b*), and that it significantly outperforms 5-minute realised variance. Regarding forecast evaluation, Chapter 2 demonstrates how the estimators can be used, in combination with the framework in Patton & Li (2013), to construct a single coherent framework for the evaluation of forecast models of any of the aforementioned risk measures. For the specific case of value-at-risk, this framework is shown to exhibit much greater power to distinguish between competing forecast models than other methods in the literature.

A paper titled *An Empirical Analysis of Value-at-Risk Forecasting Models*[6] provides the material for Chapter 3. This chapter extends the value-at-risk application in Chapter 2 to a much wider range of models and across a much larger dataset. Chapter 3 also introduces a new class of value-at-risk forecasting models based on simple time series models of the value-at-risk estimator proposed in Chapter 2. These new models are shown to yield accurate forecasts. More generally, this chapter provides strong evidence to support the use of high frequency data in value-at-risk forecasts.

A paper titled *Ranking Intraday Volatility Estimators Using Empirical Criteria*[7] provides the material for Chapter 4, which considers the problem of ranking intraday variance estimators, such as realised variance,[8] purely via empirical criteria. This topic was first considered in Patton (2011*a*), and then extended in Patton & Sheppard (2009). These authors overcame the difficulties associated with this problem by requiring a specific time series model for the true volatility dynamics. In contrast, the method in Chapter 4 allows for a wide variety of popular time series models, as well as generalizing some other technical assumptions. Further, this chapter also contains a new result on the product of near epoch dependent processes which applies generally in the time-series literature.

Chapters 2 and 3 concern the construction of estimators, and the application of

[6] Bowers & Heaton (2014*b*).
[7] Bowers & Heaton (2014*c*).
[8] See Andersen & Bollerslev (1998) or Barndorff-Nielsen & Shephard (2002*a*).

these estimators to forecast models and forecast evaluation. In contrast, Chapter 4 concerns empirical ranking of estimators. Because of this, I do not provide a literature review for the thesis as a whole, but rather refer the interested reader to the smaller reviews contained within each chapter. I also employ variable definitions in these chapters which are consistent with the prior literature on which each chapter draws, rather than using a single set of notation for the entire thesis.

I conclude this introduction with a description of the data sources and code, as these are common to all chapters. All data are sourced from the Thomson Reuters Tick History database, via an Application Programming Interface provided by the Securities Industry Research Centre of Asia-Pacific (SIRCA)[9] and written in the R programming language.[10] All high frequency data were cleaned following the procedures recommended in Barndorff-Nielsen, Hansen, Lunde & Shephard (2009).[11] Some additional cleaning was deemed necessary for the empirical work in Chapter 3. This is described in detail within that chapter. All code was written and executed in Matlab. Extensive use was made of Kevin Sheppard's Oxford MFE Matlab toolbox, which is publicly available from http://www.kevinsheppard.com/MFE_Toolbox as of 2014-04-01. Additional code for estimating the optimal block length of a stationary bootstrap, written by Andrew Patton and Kevin Sheppard, was also used and is publicly available from http://public.econ.duke.edu/~ap172/code.html as of 2014-04-01. Source code is available upon request.

[9] http://www.sirca.org.au/.

[10] R Core Team (2014).

[11] Some minor tweaks were necessary when working with Australian exchange data due to differing exchange rules.

Chapter 2

Bootstrapping Daily Returns

2.1 Introduction

In this chapter I propose the use of standard bootstrapping techniques to estimate a class of interesting functions of the distribution of daily asset returns using intraday data. This class of functions includes the variance, quantiles such as value-at-risk (VaR), expected shortfall or conditional VaR, semi-variance, kurtosis, L-estimators, and other moments (if they exist). In what follows, this approach to estimation will be referred to as the Bootstrapped Return Method (BRM). Consistency is achieved as the number of intraday observations grows. The method proposed is robust to market microstructure effects.

The BRM has a range of possible applications including parameter estimation, forecasting, and forecast evaluation. In the present chapter, the ability of the BRM to equal or exceed competing methods in the topics of parameter estimation and forecast evaluation is demonstrated, while Chapter 3 focuses on the ability of the BRM to produce superior forecasts.

An intuitive description of the BRM follows: it is well-known that under mild regularity conditions, the distribution of a sample mean can be consistently and non-parametrically estimated using a dependent bootstrap. Given a sequence of intraday returns, the daily return is a re-scaled sample mean, and so its distribution function can be consistently estimated. In this chapter, this result is extended to cover Riemann-Stieltjes integrals of uniformly integrable functions over the daily return distribution, allowing for estimation of a variety of interesting parameters. To the best of my knowledge, this result is new in the literature. In this chapter I focus on the stationary bootstrap of Politis & Romano (1994*b*), but the result holds for any re-sampling procedure that estimates the distribution with uniform consistency.

It is worth emphasizing that this approach is an unusual application for a bootstrap. Typically, a bootstrap is used to perform inference on a test statistic that itself is an estimator of some unknown parameter of interest. For example, Goncalves &

Meddahi (2009) use a bootstrap to estimate confidence intervals for the popular realised variance estimator, where the unknown parameter of interest is quadratic variation. In the present paper, the "test statistic" is a daily return, the probability limit of which is not of direct interest. However functions of the distribution of this "test statistic" are of interest, and so the application of a bootstrap is appropriate.

The chapter proceeds as follows: In Section 2.2 the modelling assumptions are described and the main theoretical statement is presented and proven. This section also contains a proposition regarding convergence of integrals that, to the best of my knowledge, is new in the literature. In Section 2.3 I provide examples where the theory applies. In Section 2.4 the empirical ranking methods proposed in Patton (2011*a*) are used to demonstrate that the BRM variance estimator strongly outperforms 5-minute realised variance and has performance comparable to realised kernels.[12] These conclusions are also supported by simulations. In Section 2.5 I describe a unified framework for evaluating forecast models with the BRM, and then consider its application to VaR. Both simulation and empirical work are used to demonstrate that the framework has much greater power to distinguish between competing forecast models than other methods in the literature. In Section 2.6 I summarize the results and discuss other possible applications and extensions of the BRM. Appendix 2.B discusses some important differences between discrete-time and continuous-time modelling, while Appendix 2.C provides a brief review of the stationary bootstrap.

2.2 Modelling Assumptions and Theory

This is not the first paper to explore the idea of resampling intraday random variables. Goncalves & Meddahi (2009) employ the *iid* bootstrap of Efron (1979) and the Wild bootstrap of Wu (1986) to estimate confidence bounds for the popular realised variance estimator.[13] Note that both these bootstrap methodologies require an independent sequence.

The present chapter differs significantly in two ways. First, as discussed in Section 2.1, I am not concerned with the estimation of confidence bounds for an estimator. Rather, I am interested in direct estimation of a range of parameters. This is an unusual application for resampling methods, although it poses no additional theoretical challenges once one accepts that a daily return can be conceptualized as a test statistic. Second, I aim to use as much information as possible in the construction of the estimators. Thus it is apparent that a modelling framework that can accommodate dependence, and preferably heterogeneity, in the intraday sequence is needed.[14]

[12] Barndorff-Nielsen, Hansen, Lunde & Shephard (2008*a*).

[13] See also Dovonon, Goncalves & Meddahi (2013) and Goncalves, Hounyo & Meddahi (Forthcoming).

[14] Hwang & Shin (2013) have independently proposed using a dependent bootstrap with intraday

Let $Y_{t,n}$, $n = 1, ..., N$ denote sequential, non-overlapping increments in some observable intraday process from any day t, such that $\sum_n Y_{t,n} = R_{t,N}$, where $R_{t,N}$ denotes a daily return. Let $\Delta = N^{-1}$, and assume:

$$Y_{t,n} = \sqrt{\Delta} X_{t,n}. \tag{2.1}$$

Consider the following modelling assumptions:

Assumptions 2.1

1. *for $r > 2$ and $\delta > 0$,*
   (a) *$X_{t,n}$ is $L_{2+\delta}$ near epoch dependent (NED) of size -1 on a strong mixing base of size $-(2+\delta)(r+\delta)/(r-2)$, for any N,*
   (b) *$\mathbb{E}|X_{t,n}|^{r+\delta} < \infty$ and $\mathbb{E}X_{t,n}^2 > 0$, $\forall n$,*
2. *$\mathbb{E}X_{t,n}$ obeys the homogeneity condition described in Goncalves & White (2002)*[15] *and additionally $\mathbb{E}\sum_n X_{t,n} = 0$, and*
3. *$p_N \to 0$ and $Np_N^2 \to \infty$, as $N \to \infty$, where p_N denotes the parameter of the geometric distribution used in construction of the stationary bootstrap indices.*

Remark 2.1 Assumption 2.1.1a allows for intraday increment sequences that exhibit weak dependence. This is necessary since microstructure effects typically induce serial correlation at high frequencies. Assumptions 2.1.1a, 2.1.1b, and 2.1.2 allow for heterogeneity in the intraday sequence, as well as infinite higher order moments. This is important since the variance of intraday sequences typically varies throughout the day. Both assumptions are needed for the theorems that follow since they allow a bootstrap central limit theorem to be invoked for each daily return.

Remark 2.2 The literature on high frequency financial data typically employs a continuous-time modelling framework. Specifically, the most popular assumption for the intraday price process is that of a continuous-time semi-martingale plus microstructure noise term.[16] In contrast, Assumption set 2.1 is a discrete-time framework. A discussion of how Assumption set 2.1 and Equation (2.1) compare to the commonly employed continuous-time framework is interesting, although not necessary for the theory to follow. The interested reader is referred to Appendix 2.B at the end of this chapter.

data. However, their results pertain only to the construction of confidence intervals for a specific bias-corrected realised variance estimator and only under very specific modelling assumptions for microstructure effects.

[15] Let $\mu_n = \mathbb{E}X_n$ and let $\bar{\mu}_N = \sum_{n=1}^N \mu_n$. Then the homogeneity condition is satisfied if $N^{-1}\sum_{n=1}^N (\mu_n - \bar{\mu}_N)^2$ is $o(N^{\frac{1}{2}})$.

[16] See, for example, all references in Footnote 30, excluding Merton (1980).

Remark 2.3 Assumptions 2.1.1a and 2.1.1b, along with Equation (2.1), imply the weak dependence in the intraday sequence is modelled in observation time, *not* calendar time. This is useful as it implies the assumption set is robust to intraday sequences that exhibit non-uniform partitions in calendar time, as long as the vanishing variance implied by Equation (2.1) holds. However, there is an additional subtlety worth emphasizing: Assumption 2.1.1a allows for arbitrary dependence between $Y_{t,n}$ and $Y_{t,n+s}$ for finite integers s, but it implies asymptotic independence as $s \to \infty$. This means that if the sequence $Y_{t,1}, ..., Y_{t,N}$ is mapped onto a finite calendar interval - such as is implicitly done when employing continuous time mathematics - then the degree of dependence between any two distinct (in calendar time) increments will vanish as $N \to \infty$. This is a typical approach to modelling microstructure effects. For example, Barndorff-Nielsen et al. (2008*a*) similarly tie the degree of dependence to observation time such that the dependence between any two distinct (in calendar time) noise terms vanishes as the sampling frequency increases. Further, intraday data support this assertion. When N is small, as it is for illiquid equities, one can observe significant sample correlations between intraday returns separated by hours of calendar time. In contrast, when N is large, 5 minutes is typically a sufficient span for sample correlations between intraday returns to be statistically insignificant. The downside to this modelling assumption is that it implies that asymptotically, the dependence between $|Y_{t,n}|^p$ and $|Y_{t,n+s}|^p$ vanishes as $s \to \infty$. There is some evidence against this within the data, e.g. the autocorrelation function of some squared intraday return sequences is persistent. Interestingly, the simulations in Section 2.4 suggest the method proposed in this chapter may be robust to such effects. A more thorough theoretical treatment of this issue is left for future work.

Remark 2.4 Assumption 2.1.2 explicitly assumes daily returns are mean zero, although it allows for some heterogeneity in the first moment within the day. $\mathbb{E}R_{t,N} = 0$ is probably violated in practice. However, typically the location parameter of a daily return is so close to zero relative to the dispersion parameter that assuming a zero mean is of little consequence. In the present case, the assumption is necessary for identification because the scaling term in Equation (2.1) prevents consistent estimation of $\mathbb{E}R_{t,N}$. It is worth emphasizing that assuming $\mathbb{E}R_{t,N} = 0$ is common for these types of problems.[17] A discussion on why this assumption is preferable to employing a noisy estimator of $\mathbb{E}R_{t,N}$ can be found in Ait-Sahalia et al. (2005). It is also worth pointing out that for estimation of naturally centred characteristics, such as variance, the zero-mean assumption is irrelevant. It is, of course, relevant for estimation of non-centred parameters, e.g. quantiles.

Remark 2.5 Equation (2.1) implies the variance of each intraday increment vanishes in N. It might be argued that the existence of an exchange regulated minimum

[17] See Merton (1980) or Ait-Sahalia, Mykland & Zhang (2005).

tick size implies that placing a strictly positive lower bound on the variance of an increment would be sensible. In practice, a close examination of real-world equity data suggests that as N increases to ultra-high frequencies, the probability of a new increment being exactly equal to zero increases towards unity. This implies that a framework in which the variance of increments vanishes asymptotically is appropriate. Further, vanishing variance is appropriate for sequences that are not affected (or minimally affected) by the exchange regulated minimum tick size; e.g. a volume-weighted average of the best bid and best ask series.

A second argument against vanishing variance is the possibility of jumps in the intraday sequence. Since I work in discrete time, rather than continuous, the definition of "jump" needs to be treated carefully. Consider the following two possibilities:

1. a jump is modelled as a spike in variance over an interval of time where the interval is bounded below by a strictly positive constant, or

2. a jump is modelled as a discrete price jump at a fixed *point* in time.

The first definition is explicitly allowed for by the modelling assumptions of this chapter since in discrete time the result is simply a heteroskedastic sequence. This model is not unreasonable, as jumps are typically caused by surprise announcements to which some market participants may be slower to respond than others. Further, recent research suggests that bursts of volatility, such as are described by the first definition, are often incorrectly identified as jumps as described by the second definition, and that true jumps are relatively rare.[18]

If the second definition obtains, then, given a jump, the variance of at least one intraday return will not vanish for any N. This violates the modelling assumptions since Assumption 2.1.1b implies $X_{t,n}$ must have non-zero, finite variance, and hence Equation (2.1) implies the variance of $Y_{t,n}$ must vanish as N grows.[19] Given this, consistency will not obtain. Interestingly, this does not appear to cause problems in practice. Empirically, the variance estimator proposed in this chapter is close in value to the popular realised kernels estimator of Barndorff-Nielsen et al. (2008*a*). This is true irrespective of the presence of jumps.[20] In contrast, the variance estimator proposed in this chapter deviates significantly from the bipower variation estimator of Barndorff-Nielsen & Shephard (2004) on days that contain jumps, and in a fashion similar to realised kernels. More detail, as well as evidence to support this claim, can be found in Appendix 2.A.

Next, let $Y^*_{t,n}$ denote a stationary bootstrap draw[21] from $Y_{t,n}, n = 1, ..., N$, and

[18] Bajgrowicz, Scaillet & Treccani (2013).

[19] Intuitively, the theory relies on no single intraday increment dominating (in terms of variance) the entire sequence of intraday increments.

[20] Jumps are detected following Barndorff-Nielsen & Shephard (2006).

[21] See Patton, Politis & White (2009) for a description of a data-driven method to estimate bandwidth.

let $R^*_{t,N} = \sum_n Y^*_{t,n}$. Also, for $r \in \mathbb{R}$, let $F_N(r) = \mathbb{P}(R_{t,N} \leq r)$ and $F^*_N(r) = \mathbb{P}^*(R^*_{t,N} - \mathbb{E}^* R^*_{t,N} \leq r)$ denote unconditional and conditional (respectively) cumulative distribution functions (cdf), where $\mathbb{E}^*$ and $\mathbb{P}^*$ denote expectation and probability conditional on the observable sequence $Y_{t,n}$, $n = 1, ..., N$, and, as is well known for the stationary bootstrap, $\mathbb{E}^* R^*_{t,N} \equiv R_{t,N}$.

Proposition 2.2.1 *Given Assumption set 2.1 and Equation (2.1), it follows that:*

$$\sup_{r \in R} |F^*_N(r) - F_N(r)| \xrightarrow{\mathbb{P}} 0, \ \ as\ N \to \infty. \tag{2.2}$$

Proof The proof of Proposition 2.2.1 is immediate since the modelling assumptions for $X_{t,n}$ match those of Theorem 2 of Goncalves & de Jong (2003), and, by construction, $R_{t,N} = N^{\frac{1}{2}} \bar{X}_N$, with $\mathbb{E}\bar{X}_N = 0$. $\square$

Remark 2.6 The proof of Proposition 2.2.1 relies on Central Limit Theorems. That is, both $F_N(r)$ and $F^*_N(r)$ are converging to the same Normal limiting distribution. This implies that *asymptotically*, the BRM is estimating parameters of Normal random variables. Nonetheless, I claim the BRM yields *nonparametric* estimates. The basis for this claim lies in *finite sample* arguments. Specifically, it has been established that the stationary bootstrap[22] of test statistics such as the sample mean is *second-order correct* in an Edgeworth expansion.[23] This implies that in finite sample, the BRM captures not only the Normal limiting function, but also the next term which is of order $O(N^{-1/2})$. The remaining terms in the expansion which are not captured are of order $o(N^{-1/2})$. Stated simply, the BRM is able to capture deviations from Normality in finite sample and so there is *no need to assume that daily returns are Normal.* In this sense, the BRM is nonparametric. In contrast, employing an existing consistent estimator of daily variance, such as realised kernels, in combination with a Normal assumption, would result in a fully parametric estimator of a characteristic of the daily return distribution. Such an estimator would be invalid for the application in Section 2.5, as it would unfairly advantage any forecasting models that employ the same parametric assumption. More to the point, such an estimator would very likely be less accurate than the BRM when the daily return violates Normality.

Proposition 2.2.1 states that the difference between the conditional bootstrap cdf and the true cdf of a daily return converges uniformly to 0 in probability. But much more is also true. Let $q(\lambda, F) = \inf\{r \in \mathbb{R} | \lambda \leq F(r)\}$ denote the quantile function of the cdf F:

Proposition 2.2.2 *Given Assumption set 2.1 and Equation (2.1), it follows that:*

$$|q(\lambda, F^*_N) - q(\lambda, F_N)| \xrightarrow{\mathbb{P}} 0, \ \ as\ N \to \infty, \tag{2.3}$$

[22] Along with several other block bootstrap procedures.
[23] Lahiri (2003) Chapter 6.

and for any Borel function $g : \mathbb{R}^K \to \mathbb{R}$ locally continuous at its argument:

$$|g(q(\lambda_1, F_N^*), ..., q(\lambda_K, F_N^*)) - g(q(\lambda_1, F_N), ..., q(\lambda_K, F_N))| \xrightarrow{\mathbb{P}} 0, \tag{2.4}$$

as $N \to \infty$.

Proof Discussion of Equation (2.3) is provided in Politis & Romano (1994*b*) and Goncalves & de Jong (2003), and proof can be found in Chapter 1 of Politis, Romano & Wolf (1999). Equation (2.4) follows from an application of Slutsky's Theorem. □

Proposition 2.2.2 motivates the use of the BRM to estimate many interesting characteristics of the daily return distribution, some of which are discussed below. However, it does *not* provide a theoretical justification for use of the BRM to estimate moments, or, more generally, integrals over the cdf. For that the following proposition is required which, to my knowledge, is new in the literature:

Proposition 2.2.3 *Consider a function $g(r) : \mathbb{R} \to \mathbb{R}$ that is bounded over any finite interval and has bounded total variation over any finite interval. If $g(r)$ is uniformly integrable with respect to $F_N(r)$ and $F_N^*(r)$, then Equation (2.2) is sufficient for:*

$$I(g) = \left| \int_{-\infty}^{\infty} g(r) dF_N^*(r) - \int_{-\infty}^{\infty} g(r) dF_N(r) \right| \xrightarrow{\mathbb{P}} 0, \text{ as } N \to \infty. \tag{2.5}$$

Proof : $\forall a \in \mathbb{R}$:

$$\begin{aligned} I(g) \leq & \left| \int_{-a}^{a} g(r) dF_N^*(r) - \int_{-a}^{a} g(r) dF_N(r) \right| + \left| \int_{-\infty}^{-a} g(r) dF_N^*(r) \right| \\ & + \left| \int_{-\infty}^{-a} g(r) dF_N(r) \right| + \left| \int_{a}^{\infty} g(r) dF_N^*(r) \right| + \left| \int_{a}^{\infty} g(r) dF_N(r) \right|. \end{aligned} \tag{2.6}$$

Applying integration by parts to the first right-hand-side term, and recalling the definition of uniform integrability, it follows from Equation (2.6) that $\forall \epsilon > 0$, $\exists a \in \mathbb{R}$ such that $0 < a < \infty$ and:

$$\begin{aligned} I(g) & \leq 2 \sup_{r \in [-a,a]} |g(r)| \sup_{r \in R} |F_N^*(r) - F_N(r)| + \left| \int_{-a}^{a} \left(F_N^*(r) - F_N(r)\right) dg(r) \right| + \epsilon \\ & \leq \left(2 \sup_{r \in [-a,a]} |g(r)| + V_{-a}^{a}(g) \right) \sup_{r \in R} |F_N^*(r) - F_N(r)| + \epsilon, \end{aligned} \tag{2.7}$$

where the second inequality follows from the definition of the Riemann-Stieltjes integral, and $V_{-a}^{a}(g)$ is the total variation of $g(r)$ over $[-a, a]$. Both $\sup_{r \in [-a,a]} |g(r)|$ and $V_{-a}^{a}(g)$ are bounded by assumption, so the required result follows from Equation (2.2). □

The following corollary will also prove useful:

Corollary 2.2.4 *Under the assumptions of Proposition 2.2.3,* $\forall (a, b) \in \mathbb{R}^2$

$$\left| \int_a^b g(r) dF_N^*(r) - \int_a^b g(r) dF_N(r) \right| \xrightarrow{\mathbb{P}} 0, \text{ as } N \to \infty. \tag{2.8}$$

Proof If $g(r)$ satisfies the assumptions of Proposition 2.2.3 then so must $g(r)\mathbb{I}\{a < r < b\}$ where $\mathbb{I}\{\cdot\}$ is the indicator function, and the result follows from Proposition 2.2.3. □

Choose $g(r) = |r|^q$. It is well known that the existence of moment number $q + \delta$ on the underlying process, for any $\delta > 0$, is sufficient for uniform integrability of $g(r)$.[24] It immediately follows that Assumption 2.1.1b is sufficient for the bootstrap to consistently estimate any absolute moment up to the second of the underlying distribution. For the special case of the variance of the underlying distribution, this result was already known, see Goncalves & de Jong (2003) Theorem 1. However note the difference in the assumption set: Goncalves & de Jong (2003) use assumptions 2.1.1a to 2.1.3 to obtain their result. In contrast, I require only the standard bootstrap distributional result, i.e. Equation (2.2). It follows that if the sufficient conditions for Equation (2.2) are weakened in future work, Proposition 2.2.3 need not be revisited.

2.3 Examples

In this section, several examples of parameters that can be estimated by the BRM are provided, and methods for building the corresponding estimators are described. All the estimators discussed below are both non-parametric and consistent as $N \to \infty$. To the best of my knowledge, this makes them the first of their kind in the literature, with the exception of variance (Example 3).

Example 1 (*quantiles*): As discussed in Proposition 2.2.2, quantiles can be estimated consistently. From a computational perspective, a quantile can be obtained by sorting $R^*_{t,N,b} - R_{t,N}, b = 1, ..., B$ in ascending order then choosing the $\lfloor \lambda B \rfloor$ element. In what follows, this is referred to as the BRM quantile estimator and, since VaR is just a quantile, it will be used extensively in Section 2.5 to proxy true VaR. Importantly, since the stationary bootstrap is second-order correct in an Edgeworth expansion, the BRM quantile estimator is non-parametric (as well as consistent), and so is the first estimator of its kind in the literature.[25] Prior to the BRM quantile estimator, a parametric assumption was necessary to consistently estimate VaR, e.g. one might employ a consistent and non-parametric estimator of daily variance and then (parametrically) transform it to VaR using the inverse Normal function.

Example 2 (*expected shortfall*): Choose $g(r) = r$ in Corollary 2.2.4 and set $a = -\infty$ and $b = q(\lambda, F_N)$, where, for simplicity, assume $q(\lambda, F_N)$ is locally continuous

[24] Davidson (1994) Theorem 12.10.
[25] See Remark 2.6 for more detail.

at λ. The uniform integrability requirement is guaranteed by Assumption 2.1.1b. Corollary 2.2.4 then guarantees consistent estimation of expected shortfall,[26] also known under these conditions as conditional-VaR or tail-VaR. From a computational perspective, one simply identifies the λ-quantile, as in Example 1, and then takes the average of all centred, re-sampled returns that lie below the quantile. Given the recent proposal of the Basel committee to switch metrics from VaR to expected shortfall, this example could prove important for future work.[27] It is worth adding that there is a prevalent view in industry that it is virtually impossible to obtain: 1) a consistent and non-parametric estimator of expected shortfall, and 2) a method of empirically verifying an expected shortfall forecast via a backtest.[28] As this chapter demonstrates, both views are incorrect.[29]

Example 3 (*variance*): Choose $g(r) = r^2$, $a = -\infty$, and $b = \infty$ in Proposition 2.2.3 to demonstrate that the BRM variance estimator is consistent. The uniform integrability requirement is guaranteed by Assumption 2.1.1b. From a computational perspective, apply the sample variance formula to $R^*_{t,N,b}, b = 1, ..., B$; note that the $\mathbb{E}R_{t,N} = 0$ assumption is not required for this estimator as the parameter is naturally centred. To the best of my knowledge, this is the only risk measure where other consistent and non-parametric estimators using intraday data already exist in the literature.[30]

Example 4 (*downside variance*): Choose $g(r) = r^2$, and set $a = -\infty$, and $b = 0$. The uniform integrability requirement of Corollary 2.2.4 is guaranteed by Assumption 2.1.1b. The corollary then guarantees consistent estimation of semivariance, also known as downside risk or downside variance. Harry Markowitz has been a consistent advocate of the semivariance as a measure of risk since 1959.[31] Taking the square root of downside variance yields a consistent estimator for semi-deviation (via Slutsky's Theorem). Note, if we instead choose $b \in \mathbb{R}$, then this is known as target semi-deviation. From a computational perspective, simply take the sample variance of all centred re-sampled daily returns less than b, and then take the square root if desired. Obviously, for upside risk, set $b = \infty$ and $a = 0$.

Example 5 (*higher moments*): Set $g(r) = r^3$ or $g(r) = r^4$, in Proposition 2.2.3. Assuming the uniform integrability assumption is satisfied, the BRM can then be used to estimate Skewness or Kurtosis consistently,[32] given appropriate scaling by

[26]For more detail on expected shortfall, see Acerbi & Tasche (2002).
[27]Basel Committee On Banking Supervision (2012).
[28]For example Rowe (2012).
[29]See Section 2.5 for more detail on how to empirically verify an expected shortfall forecast via backtest.
[30] A short-list of other papers proposing intraday data-based variance estimators includes Merton (1980), Zhou (1996), Andersen & Bollerslev (1998), Barndorff-Nielsen & Shephard (2002*a*), Zhang, Mykland & Ait-Sahalia (2005), Christensen & Podolskij (2007), Barndorff-Nielsen et al. (2008*a*), and Jacod, Li, Mykland, Podolskij & Vetter (2009).
[31]Markowitz (1991).
[32]To compute the statistic, simply apply the corresponding sample formula to the re-sampled

a consistent variance estimator - see Example 3. As discussed in Example 1, since the stationary bootstrap is second-order correct in an Edgeworth expansion, these estimates are still of interest, even though asymptotically $F_N(r)$ and $F_N^*(r)$ both converge to the Normal. Note, Skewness and Kurtosis are frequently provided as interesting summary statistics for a dataset of daily returns.[33]

Example 6 (*L-estimators*): Proposition 2.2.2 guarantees consistent estimation of, for example, linear combinations of order statistics; i.e. L-estimators.[34] Examples include the λ-quantile range, the median absolute deviation, or numerous other robust estimators of scale and asymmetry.[35] From a computational perspective, apply the appropriate transformation to the quantile estimators, which are constructed as in Example 1. These estimators may be of interest to investors wishing to investigate tail-thickness or asymmetry of the distribution of daily returns, when doubt exists over the existence of the third and fourth moments and therefore the uniform integrability requirement in Example 5 above.

2.4 Estimating Daily Variance

In this section, the performance of the BRM variance estimator is compared to realised variance and realised kernels[36] using both simulations and the empirical methods proposed in Patton (2011*a*). Given an appropriate choice of kernel function and a particular continuous-time semi-martingale plus noise modelling framework, realised kernels is known to be consistent and have asymptotic variance equal to the fully parametric case. The implication is that realised kernels likely provides an upper bound on the performance one might expect from a fully non-parametric estimator. 5-minute realised variance is considered due to its considerable popularity.

It is worth emphasizing that the purpose of this section is not to argue that the BRM variance estimator outperforms all other approaches.[37] This would be a difficult task indeed. Rather, the aim is to demonstrate that for a particular task that has received comprehensive treatment in the literature, the BRM approach has performance comparable to the best-case non-parametric estimator. Strong performance in this specific application of the BRM implies that we might expect to see strong performance when applying the BRM to other, less explored, characteristics of the distribution (such as quantiles - see Section 2.5).

Evaluation of the BRM variance estimator, realised variance, and realised kernels is done in two ways. First, in Section 2.4.1, I evaluate the estimators in simulated

daily returns.

[33] See for example Andersen, Bollerslev, Diebold & Ebens (2001) table 1.

[34] Huber (1981).

[35] Hogg (1974).

[36] Barndorff-Nielsen et al. (2008*a*).

[37] See Footnote 30 for a list of estimators.

environments that adhere to Assumption set 2.1 and Equation (2.1), as well as simulated environments that adhere to the popular continuous-time semi-martingale plus noise framework.[38] Second, in Section 2.4.2, I evaluate the estimators using the empirical methods proposed in Patton (2011*a*) across 20 of the largest (by market capitalization) equities listed on the New York Stock Exchange.

2.4.1 Simulations

Six environments are simulated to evaluate the proposed variance estimator. The first four environments satisfy Assumption set 2.1 and Equation (2.1), hence the notation from Section 2.2 is used to describe them (the t subscript is dropped as it is not relevant). The final two environments are a form of the continuous-time semi-martingale plus noise model, and so are misspecified for the methodology proposed in this chapter (see Appendix 2.B for more detail). However, I include them to demonstrate the surprising fact that the BRM still works well in the presence of this model when common specifications of the variance of the noise process are employed. I describe these last two environments using $\tilde{p}_s$ to denote the latent log-price and p_s to denote the observed log-price.

The environments are simulated for $N = \{390, 1170, 4680, 23400\}$, corresponding to 1 minute, 20 second, 5 second, and 1 second increments respectively on an exchange opened for $6\frac{1}{2}$ hours. Their definitions follow:

SE1 $X_n = \beta X_{n-1} + u_n$, where $\beta = -0.5$, and $u_n | W_n \overset{iid}{\backsim} \mathcal{N}(0, W_n)$, where W_n is a random walk independent of u_n, with the variance of each increment set to $0.01\Delta W_0^2$ and $W_0 = 0.0000897$,

SE2 identical to SE1 except $u_n = \sqrt{0.0000897}\tilde{u}_n$, where $\tilde{u}_n$ denotes *iid* draws from the Skewed-T distribution of Hansen (1994) with degrees of freedom 3 and skew parameter -0.5 (note, by construction this distribution has a zero mean and unit variance),

SE3 $X_n = u_n + \beta u_{n-1}$, where $\beta = -0.5$, and u_n is defined as in SE1, but now $W_0 = 0.000158$,

SE4 identical to SE3 except $u_n = \sqrt{0.000158}\tilde{u}_n$, where $\tilde{u}_n$ is defined as in SE2,

SE5 23401 observations[39] of p_s are simulated using an Euler discretization of the continuous-time model $d\tilde{p}_s = \frac{1}{100}\nu_s \left(\kappa_1 dW_{1,s} + \kappa_2 dW_{2,s} + \sqrt{1 - \kappa_1^2 - \kappa_2^2} dW_{3,s}\right)$, where $W_{i,s}$ denotes a Wiener process, with the GARCH diffusion from Patton (2011*a*), Goncalves & Meddahi (2009) and Andersen & Bollerslev (1998), i.e.

[38] See Appendix 2.B.

[39] The lower sampling frequency datasets are obtained by skipping the appropriate number of simulated observations.

$d\nu_s^2 = 0.035(0.636 - \nu_s^2)dt + 0.144\nu_s^2 dW_{1,s}$, where $\kappa_1 = -0.576$ and $\kappa_2 = 0$, and where $p_s = \tilde{p}_s + \epsilon_s$ with $\epsilon_s \overset{iid}{\backsim} \mathcal{N}(0, \sigma_\epsilon^2)$, where the variance of the microstructure noise term is calculated following Ait-Sahalia et al. (2005) and Huang & Tauchen (2005), see for example Patton (2011*a*) equation 20, and

SE6 identical to SE5 except with the two factor affine diffusion from Andersen, Bollerslev & Meddahi (2005) and Bollerslev & Zhou (2002): $\nu_t^2 = \nu_{1,t}^2 + \nu_{2,t}^2$, where $d\nu_{1,t}^2 = 0.5708(0.3257 - \nu_{1,t}^2)dt + 0.2286\nu_{1,t}dW_{1,t}$ and $d\nu_{2,t}^2 = 0.0757(0.1786 - \nu_{2,t}^2)dt + 0.1096\nu_{2,t}dW_{2,t}$, where $\kappa_1 = 0.9$ and $\kappa_2 = -0.4$; see also Chernov, Gallant, Ghysels & Tauchen (2003) .

SE1 is an AR(1) process with Normal errors. The presence of the random walk component ensures the variance of the errors is heterogenous. SE2, on the other hand, has homogenous variance, but replaces the Normal errors with a Skewed-T error. SE3 and SE4 replicate SE1 and SE2, except the AR(1) structure is replaced with an MA(1) structure. Interestingly, this implies that both SE1 and SE3 violate Assumption 1a, since the autocorrelation function of squared intraday returns does not vanish. This is a deliberate modelling choice designed to test how robust the BRM methodology is to long memory in the square of intraday returns, which is a valid concern for high frequency financial data.[40]

When simulating an AR(1), I "warm-up" using $\ell = 100$ observations, $u_{-99}, ..., u_0$, where for simplicity I assume $\mathbb{V}u_n = W_0, n \leq 0$, i.e. constant variance in the warm-up. Also, let:

$$\beta_{[a,b]}^2 = \left(\sum_{j=a}^{b} \beta^j \right)^2 . \tag{2.9}$$

It is straightforward to show that the true variance of a daily return for SE1 and SE2 is:

$$\mathbb{V}r_{t,N} = \Delta \left(\sum_{n=1}^{\ell} \beta_{[n,N+n-1]}^2 \mathbb{V}u_0 + \sum_{n=1}^{N} \beta_{[0,N-n]}^2 \mathbb{V}u_n \right), \tag{2.10}$$

while true variance of a daily return for SE3 and SE4 is:

$$\mathbb{V}r_{t,N} = \Delta \left(\mathbb{V}u_N + \beta^2 \mathbb{V}u_0 + (1+\beta)^2 \sum_{n=1}^{N-1} \mathbb{V}u_n \right). \tag{2.11}$$

The coefficients for SE1 to SE4 are chosen to provide a challenging sequence of heterogeneous increments that exhibits strong negative serial correlation, and (for the Skewed-T distribution) negatively-skewed, fat tails. The variances of the residuals of SE1 to SE4 are chosen to yield an annualized return volatility of approximately 10%.

For each of these simulated models Table 2.1 provides the Mean Absolute Error

[40] The continuous-time modelling assumptions for realised kernels explicitly allow for many kinds of dependence, including long-memory, in the volatility process.

(MAE) for the BRM variance estimator, realised kernels, and realised variance at the 5 minute sampling frequency. MAE is estimated using 5000 iterations over each of the above simulated environments, for each sampling frequency N. The results are scaled so the best performing estimator has unit MAE. The bias of each estimator is also investigated, and to this end, Table 2.2 provides the average percentage deviation of each estimator from the true variance, i.e. $100(\ln \hat{\theta} - ln\theta)$. Average percentage deviation is a more intuitive measure of bias in these simulations since true variance is small in magnitude and not constant.

As can be seen in Table 2.1, for SE1 to SE4, the BRM variance estimator performs almost as well as realised kernels, and given SE1 and SE2, actually outperforms it for large N, likely because the kernel function in realised kernels is optimized for a modelling framework where the variance of increments does not vanish in N. This is particularly encouraging as the stationary bootstrap is known to be less efficient than many other block bootstrap methods.[41] The similar performance is not surprising, as the BRM variance estimator, when using a stationary bootstrap, can itself be reformulated as a kernel estimator (asymptotically); albeit the kernel function exhibits different properties to those required by realised kernels.[42] Note that the BRM variance estimator does not appear to be negatively affected by the long memory in squared returns present in SE1 and SE3.

Table 2.2, for SE1 to SE4, demonstrates that bias is something of a problem for the BRM variance estimator, although it is mostly alleviated by large N. This is a well-known issue with the stationary bootstrap variance estimator, as the bias vanishes at rate $O(1/b)$, where b is the average block length.[43] The implication is that further reductions in the bias of the BRM variance estimator, and consequently the MAE (and MSE), are possible through the use of bootstrapping procedures with bias that vanishes at faster rates. For example, the tapered block bootstrap of Paparoditis & Politis (2002) is known to have a bias that vanishes at rate $O(1/b^2)$. However, I leave investigation of the performance of other block bootstrap procedures to future work.

As discussed, SE5 and SE6 are misspecified for the BRM variance estimator. Surprisingly, Table 2.1 indicates that its MAE is not that far off realised kernels, and Table 2.2 indicates that the impact of the model misspecification on the bias is only visible when $N = 23400$. It is interesting that the BRM variance estimator consistently outperforms 5-minute realised variance, even for large N, which, as discussed in Appendix 2.B, is the scenario in which the most errors accumulate. Admittedly, the main problem faced by 5-minute realised variance in SE5 and SE6 is the simulation design which, following Patton (2011*a*), ensures the bias for 5-minute realised

[41] Lahiri (1999) as well as the corrections provided in Nordman (2009).
[42] Politis & Romano (1994*b*) Lemma 1.
[43] Lahiri (1999).

Table 2.1: Mean Absolute Error of Daily Variance Estimators

SE1				**SE2**			
N	RV5	RK	BRMVE	N	RV5	RK	BRMVE
390	1.67	1.00	1.47	390	1.22	1.00	1.12
1170	1.15	1.00	1.26	1170	1.03	1.00	1.04
4680	1.35	1.00	1.04	4680	1.10	1.00	1.00
23400	1.99	1.09	1.00	23400	1.46	1.08	1.00
SE3				**SE4**			
N	RV5	RK	BRMVE	N	RV5	RK	BRMVE
390	4.11	1.00	1.41	390	2.19	1.00	1.11
1170	2.03	1.00	1.39	1170	1.29	1.00	1.08
4680	1.48	1.00	1.28	4680	1.15	1.00	1.06
23400	1.85	1.00	1.18	23400	1.39	1.00	1.02
SE5				**SE6**			
N	RV5	RK	BRMVE	N	RV5	RK	BRMVE
390	1.53	1.00	1.22	390	1.56	1.00	1.23
1170	1.91	1.00	1.36	1170	1.89	1.00	1.35
4680	2.57	1.00	1.75	4680	2.57	1.00	1.73
23400	3.62	1.00	2.53	23400	3.63	1.00	2.50

RV5 = 5-minute realised variance; RK = realised kernels; BRMVE = Bootstrap Return Method variance estimator. Mean Absolute Errors are scaled so that for any row the smallest is equal to unity.

variance is 20%, even when $N = 23400$. This value is reflected in Table 2.2.

2.4.2 Empirical

Next, I compare the performance of the BRM variance estimator, realised kernels, and 5-minute realised variance using empirical data from the New York Stock Exchange. A visual of how the BRM variance estimator compares to realised kernels is provided in Figure 2.1, which presents the time-series of both estimators at different levels of zoom. When viewing 4 years of data, the two time-series are visually indistinguishable. Even when zoomed in on the seven months of the global financial crisis period, the two series track each other closely. It is only at the highest level of zoom provided (a two month interval) that the differences become easily distinguishable. For the most part, the two series appear to track each other closely. This is promising, as it implies the BRM procedure, when used to estimate variance, closely tracks one of the most accurate and popular intraday data-based variance estimators (realised

Table 2.2: Average Percentage Deviation of Daily Variance Estimators

SE1				**SE2**			
N	RV5	RK	BRMVE	N	RV5	RK	BRMVE
390	22.7	-2.14	7.88	390	12.5	-11.7	-2.45
1170	6.97	-1.95	7.18	1170	1.74	-7.34	1.73
4680	0.64	-1.16	5.03	4680	-1.47	-3.39	2.86
23400	-1.02	-0.51	3.34	23400	-1.61	-1.42	2.25
SE3				**SE4**			
N	RV5	RK	BRMVE	N	RV5	RK	BRMVE
390	55.8	3.16	8.73	390	45.0	-5.21	-0.38
1170	22.0	0.50	8.99	1170	14.5	-6.38	1.83
4680	5.30	-0.70	6.47	4680	2.33	-3.01	4.13
23400	-0.21	-0.44	4.51	23400	-1.34	-1.91	2.99
SE5				**SE6**			
N	RV5	RK	BRMVE	N	RV5	RK	BRMVE
390	21.1	-0.52	8.03	390	20.8	-0.44	7.89
1170	20.8	-0.12	8.61	1170	20.5	-0.20	8.30
4680	20.7	0.56	8.50	4680	20.7	0.35	8.14
23400	20.8	0.88	11.6	23400	21.1	0.96	11.7

RV5 = 5-minute realised variance; RK = realised kernels; BRMVE = Bootstrap Return Method variance estimator. The table provides the average deviation of each estimator from true variance, expressed as a percentage.

kernels). This suggests we might expect the BRM procedure to similarly provide good estimates when applied to other parameters of the daily return distribution.

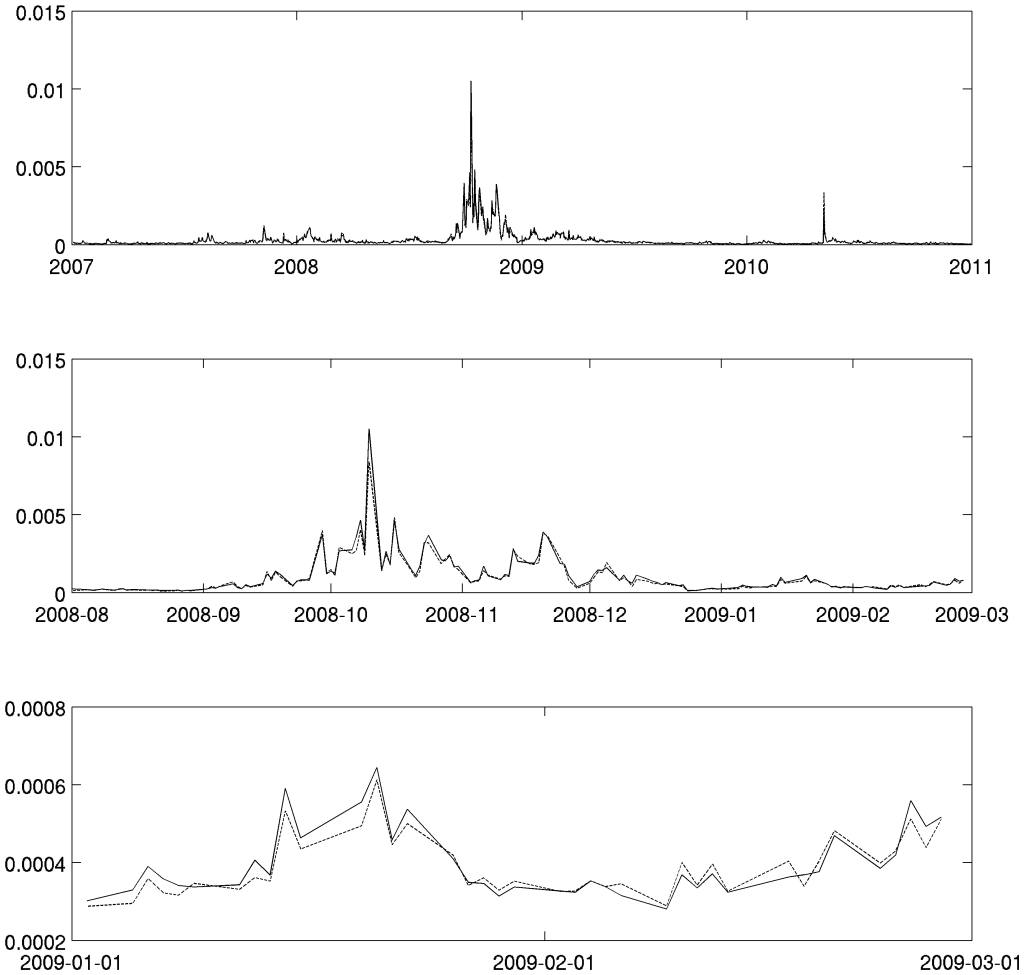


Figure 2.1: Three plots of the variance estimators at different levels of zoom. The solid line is the BRM variance estimator, the dashed line is realised kernels.

Although they closely resemble each other, it is still interesting to test whether one estimator is more accurate than another in a probabilistic framework. A theoretical comparison is not feasible due to differences in the modelling framework of the BRM variance estimator versus realised kernels and 5-minute realised variance. However, an empirical procedure for comparing intraday data-based variance estimators is proposed in Patton (2011*a*), and a large-scale application of this procedure is undertaken in Liu, Patton & Sheppard (2013). I employ essentially the same method as Liu et al. (2013). In particular, I utilise the Model Confidence Set[44] methodology of Hansen, Lunde & Nason (2011) with the QLIKE loss function, and employ one-

[44] A description of this procedure can be found in Chapter 3 Appendix 3.A.

Table 2.3: Estimators in the Model Confidence Set (QLIKE)

Estimator	*Total*
5-minute realised variance	2
realised kernels	16
BRM Variance Estimator	20

step-ahead 30-minute realised variance as a proxy for the true variance. Importantly, the theoretical results in Patton (2011*a*) demonstrate that in large sample, the rankings obtained using this proxy are consistent to those that would be obtained using true variance. For robustness, I also consider the MSE loss function.

Every intraday transaction from 1st January, 2004 to 31st December, 2011 is used, resulting in just over 2000 daily observations for each estimator. I analyse 20 of the largest ticker codes (by market capitalization) over this interval.[45] For these assets, one might expect anywhere between 3000 and 10000 distinct transactions[46] on any given day.

Table 2.3 contains the total number of times an estimator was in the Model Confidence Set, where the aggregation is performed across the 20 ticker codes, and the method used was that of Patton (2011*a*) with the QLIKE loss function. In all 20 cases, the Model Confidence Set contained the BRM variance estimator, while it contained realised kernels in 16 of the 20 cases. This suggests the performance of the two estimators is comparable, with the BRM variance estimator perhaps being slightly ahead. Importantly, we can be confident that the test has power to reject poor estimators, since 5-minute realised variance is only in the Model Confidence Set for 2 of the 20 cases. It is worth adding that the results for Realized Kernels and 5-minute realised variance are consistent with those in Liu et al. (2013), although that study does not include the BRM variance estimator.

Unfortunately, when the analysis was repeated with the MSE loss function, the Model Confidence Set algorithm exhibited no ability to distinguish between the three estimators. The problem is that a small number of observations during the global financial crisis period (especially September to October of 2008) dominate the sample and this effect is greatly exacerbated when the MSE loss function is employed. The intervals on either side of the global financial crisis period provide insufficient data to distinguish between the models at standard significance levels.

In summary, these results suggest that in practice there is little difference between the BRM variance estimator and the current best-case non-parametric estimator of daily variance; realised kernels. With this in mind, I now consider another possible

[45] The ticker codes are AXP, BAC, C, DIS, HD, JPM, KO, MCD, MRK, PEP, SLB, WFC, CVX, GE, HPQ, IBM, JNJ, PG, UNH, and WMT.

[46] That is, transactions occurring at distinct times.

application of the BRM: quantile estimation.

2.5 Evaluating Value-at-Risk Models

2.5.1 Forecast Evaluation Methods

In this section I discuss how the BRM can be combined with the framework in Patton & Li (2013) and many of the loss-based tests from the forecast evaluation literature to construct a single, unified framework for evaluating competing forecast models for almost any risk measure of the daily return distribution. The effectiveness of this framework is demonstrated for the specific case of VaR forecast evaluation.

By employing a loss-based test, one directly tests the accuracy of a forecast for a target sequence. In the present case the problem is particularly difficult since the target sequence is a characteristic of an unknown distribution and so is unobservable. One possible solution is to replace the latent target sequence with a conditionally unbiased proxy, as in Hansen & Lunde (2006*a*) and Patton (2011*b*). Unbiased proxies are easy to construct for some target sequences, e.g. variance,[47] but less obvious for other targets, such as quantiles.

More recently, Patton & Li (2013) construct an asymptotic framework in which, given a *consistent* and *non-parametric* proxy, one can apply a battery of tests from the loss-based forecast evaluation literature.[48] A sufficient condition is that the proxy and distance metric satisfy Patton & Li (2013) assumptions C2 and C3, which, roughly speaking, bound the proxy error, and bound the corresponding error in the loss differential caused by using the proxy in place of the unobservable target variable. Intuitively, the asymptotic theory relies on the proxy error on any given day vanishing at a sufficiently fast rate such that the total contribution of the proxy errors across all days is negligible. Further intuition can be found by examining Table 2 in Patton (2011*b*). From a practical perspective, the assumptions are mild if $N \gg T$, where T denotes the number of days used by the forecast evaluation procedure. It is worth adding that Laurent, Rombouts & Violante (2013) independently discuss these same ideas.

Patton & Li (2013) include a methodology for constructing proxies from intraday data for several characteristics of the distribution of a daily return, but do not extend their work to quantiles such as VaR. Further, their proxies are not robust to microstructure effects, except in the case where one can apply the subsampling and averaging approach of Zhang et al. (2005), an approach that is only (currently) well understood for estimation of variance.

[47] A squared daily return, or low sampling frequency realised variance are common proxies.

[48] A by no means complete list includes: Diebold & Mariano (1995), West (1996), White (2000), Hansen (2005), Romano & Wolf (2005), and Hansen et al. (2011).

Importantly, the BRM yields consistent and non-parametric proxies for almost any risk measure, and is robust to microstructure effects, so N can be maximized. It follows that given Assumptions C2 and C3 in Patton & Li (2013), the BRM in combination with the Patton and Li framework and many of the tests from the loss-based forecasting literature provide a single unified framework for evaluation of forecast models for almost any risk measure of the daily return distribution.

I emphasize: to the best of my knowledge, prior to the BRM, consistent and non-parametric proxies that are robust to microstructure effects are only available for the daily variance. Obviously, obtaining VaR from one of these variance estimators necessitates a parametric assumption (typically Normality). In contrast, the ability of dependent bootstraps to reflect departures from Normality is well understood based on Edgeworth expansion arguments.[49] Note that consistent fully parametric estimators of daily VaR have been proposed in the literature.[50] However, their use in the stated framework would unfairly advantage any forecasting model that utilizes the same parametric assumptions.

The effectiveness of this framework in evaluating VaR forecast models is now examined. Prior to this chapter, VaR forecast evaluation has primarily relied on an analysis of the coverage of a given forecast sequence; that is, an analysis of the number of VaR violations over a given interval, along with the manner in which they occur. Although several sophisticated tests have been proposed in this sphere, see in particular the Conditional Coverage (CC) test of Christoffersen (1998) and the Dynamic Quantile (DQ) test of Engle & Manganelli (2004)[51], they all suffer from the information loss inherent in analysing a transformation of a VaR forecast and corresponding actual return into a binary random variable. As pointed out in Engle & Manganelli (2004), this inevitably leads to low power at rejecting forecasting models that generate the correct number of VaR violations, but are nonetheless wildly inaccurate. For example, consider a forecast equal to an *iid* random variable that takes value $(-1)A$ with probability 0.95 and A with probability 0.05, $A > 0$. For arbitrarily large A, this forecast will generate the correct coverage for 5% VaR, even though it is massively inaccurate for the true VaR sequence.

The BRM proxy in combination with the Patton & Li (2013) framework should exhibit more power at rejecting misspecified VaR forecasting models than existing techniques based on coverage, since even small deviations from the forecast target can be penalized. In what follows, this is investigated using both simulations and empirical work to compare the BRM quantile estimator proxy approach to the CC and DQ test.[52]

[49]See Lahiri (2003) example 4.8 and chapter 6.

[50]See Giot (2005), Giot & Grammig (2006), and Dionne, Duchesne & Pacurar (2009).

[51]A finite sample improvement based on quantile regression has been proposed by Gaglianone, Lima, Linton & Smith (2011). Their method has identical asymptotic properties to the DQ test.

[52]Some previous work has been done to leverage loss-based forecast evaluation tests in the VaR

First, power is investigated using the BRM quantile estimator proxy with the test of Diebold & Mariano (1995) to estimate simulated expected loss differentials between misspecified VaR forecasting models and the true forecasting model. These simulated power curves are contrasted with the simulated power curves of the CC and DQ test to reject the same misspecified models.

Second, the BRM quantile estimator proxy is combined with the Model Confidence Set[53] of Hansen et al. (2011) to determine which of five popular VaR forecasting models are in the Model Confidence Set across twenty publicly listed stocks on the NYSE. I also investigate the recommendations of the CC and DQ tests. Particular attention is paid to how the results match *a priori* theoretical rankings.

2.5.2 Simulations

In the simulation, true volatility is generated using the popular GARCH(1,1) model of Bollerslev (1986), i.e. $\sigma_t^2 = (1-\alpha-\beta)+\alpha r_{t-1}^2+\beta\sigma_{t-1}^2$, with parameters following Christoffersen (1998), i.e. $\alpha = 0.1$ and $\beta = 0.85$. Given $\mathbb{V}[r_{t,N}|\mathcal{F}_{t-1}] = \sigma_t^2$, the concurrent daily return r_t is generated by summing N intraday returns simulated via a zero-mean AR(1) with coefficient of -0.5 and Normal residuals with constant intraday variance. Let ψ denote the AR(1) coefficient, and let:

$$\psi_{[a,b]}^2 = \left(\sum_{j=a}^{b} \psi^j\right)^2 \tag{2.12}$$

I "warm-up" the AR(1) with $\ell = 100$ observations. For any t, the variance of the AR(1) residual is obtained from the GARCH daily variance using

$$\mathbb{V}u_{t,n} = N\left(\sum_{n=1}^{\ell} \psi_{[n,N+n-1]}^2 + \sum_{n=1}^{N} \psi_{[0,N-n]}^2\right)^{-1} \mathbb{V}r_{t,N}, \forall n. \tag{2.13}$$

This guarantees the daily return will have the correct variance.

Since the residuals of the AR(1) process are Normal, it immediately follows that $VaR_{\lambda,t} = \sigma_t\Phi_\lambda^{-1}$, where $\Phi(\cdot)$ denotes the standard Normal cumulative distribution function. Next, a forecast of $VaR_{\lambda,t}$ is constructed using $VaR_{\lambda,t}^{[i]} = \sigma_t^{[i]}\Phi_\lambda^{-1}$, $i = 1, 2, 3$, where $\sigma_t^{[i]}$ is generated by one of three misspecified GJR-GARCH(1,1,1) models, ie $\sigma_t^2 = (1-\alpha-\beta-\delta)+\alpha r_{t-1}^2+\delta r_{t-1}^2\mathbb{I}\{r_{t-1} < 0\}+\beta\sigma_{t-1}^2$. The following parameterizations

sphere, see Lopez (1998), Bao, Lee & Saltoglu (2006), and Sener, Baronyan & Menguturk (2012). However, these approaches require ad hoc assumptions. Moreover, they exhibited poor performance relative to the DQ test in my early simulations and so were dropped from the final version of this chapter. The simulation results are available upon request.

[53] A description of this procedure can be found in Chapter 3 Appendix 3.A.

are utilized:

$$i = 1 \Rightarrow \beta = 0.85, \alpha = 0.08, \delta = 0.02, \tag{2.14a}$$

$$i = 2 \Rightarrow \beta = 0.85, \alpha = 0.06, \delta = 0.04, \text{ and} \tag{2.14b}$$

$$i = 3 \Rightarrow \beta = 0.85, \alpha = 0.04, \delta = 0.06. \tag{2.14c}$$

Note that $i = 1$ indexes a minor deviation from the true volatility process through the introduction of a small leverage effect. Thus whenever $r_{t-1} > 0$, $VaR_{\lambda,t}^{[1]}$ will be slightly biased for $VaR_{\lambda,t}$. Similarly, $VaR_{\lambda,t}^{[2]}$ will exhibit a medium bias, while $VaR_{\lambda,t}^{[3]}$ will exhibit a large bias.

I perform a Diebold & Mariano (1995) bivariate comparison[54] of each misspecified model with the true VaR sequence, where, utilizing the Patton & Li (2013) framework, the BRM quantile estimator is used to proxy the true latent VaR. Ideally, the test should reject the null hypothesis of equal expected loss in favour of the true model. I report results for the MAE loss function[55], although note that results for the MSE loss function were near-identical.

I also perform CC and DQ tests on each misspecified model. Ideally, the test will reject the null hypothesis of correct coverage for all three models, since each is misspecified. Power curves are constructed for $T = \{20, 50, 150, 400, 1100\}$, which closely correspond to the log-scale $\{3, 4, 5, 6, 7\}$.

First however, I evaluate the size of each statistical test. For the CC and DQ tests, this is simply done by testing the true VaR sequence. For the BRM quantile estimator approach, the size is evaluated by performing a bivariate comparison of $VaR_{\lambda,t} + Z_{1,t}$ and $VaR_{\lambda,t} + Z_{2,t}$, where $Z_{1,t}$ and $Z_{2,t}$ are both *iid* zero-mean Normal with variance 0.01.

For all tests, I consider $\lambda = \{0.05, 0.01\}$, i.e. the 5% and 1% VaR quantiles. Size and power rejection frequencies are obtained by averaging over 5000 iterations. To keep the simulation run-time feasible, I set $N = 1170$ for each day, regardless of T. Interestingly, this implies $N \approx T$ rather than $N \gg T$ when $T = 1100$. It is noteworthy that the BRM quantile estimator approach suffers no ill-effects in this situation.

As Table 2.4 shows, the BRM quantile estimator approach has near-perfect statistical size across both quantiles, even for very small T. In contrast, the CC and DQ tests both exhibit some size distortion for small samples. Of particular note, the CC and DQ tests have more difficulty with the 0.01 quantile. This makes sense, since

[54] Confidence bounds for the estimator of the expected loss differential are obtained by regressing the loss difference sequence on a constant and using Newey & West (1987) standard errors. I chose this approach instead of the more common bootstrap approach since it significantly reduces the run-time of the simulation while still providing consistent estimates of the confidence bounds.

[55] In cases where the proxy is noisy and unbiased, it is known that the use of the MAE loss function can lead to inconsistent rankings, see Hansen & Lunde (2006*a*) or Patton (2011*b*). However, in the dual-asymptotic framewok of Patton & Li (2013), this problem vanishes.

Table 2.4: Simulated Test Sizes

$T =$	20	50	150	400	1100
0.05 quantile					
CC	0.40	0.09	0.03	0.04	0.06
DQ	0.10	0.06	0.05	0.05	0.05
BRMQE/DM	0.06	0.06	0.05	0.05	0.05
0.01 quantile					
CC	0.85	0.63	0.24	0.03	0.02
DQ	0.11	0.08	0.06	0.04	0.04
BRMQE/DM	0.06	0.05	0.05	0.05	0.05

CC = Conditional Coverage; DQ = Dynamic Quantile; BRMQE/DM = Bootstrap Return Method quantile estimator combined with Diebold & Mariano (1995).

these tests both rely on VaR violations, which are rare events for small quantiles.

The real difference between the tests can be seen upon an analysis of the power to reject a misspecified model. Table 2.5 contains power curves for the 0.05 quantile, while Table 2.6 contains the corresponding numbers for the 0.01 quantile. For the CC and DQ tests, elements where Table 2.4 indicates significant size distortion have been replaced with a asterisk, since analysis of power in the presence of size distortion is not meaningful.

Across both quantiles and all three models, the BRM quantile estimator approach demonstrates an exceptional ability to reject a false null hypothesis. Of particular note: for the forecast model with minor misspecification ($i = 1$), the BRM quantile estimator approach is able to identify that the model is misspecified 70% of the time, using only 20 observations.

In contrast, the CC and DQ tests only demonstrate a reasonable amount of power for model $i = 3$, i.e. the most severely misspecified forecast model. For $i = 1$ and $i = 2$, i.e. minor and medium misspecification, the CC and DQ tests show little ability to detect the misspecification.

The simulation results suggest that employing the BRM quantile estimator in conjunction with loss-based forecast evaluation tests should yield much more powerful recommendations than existing, coverage-based, approaches, such as the CC and DQ tests. With this in mind, I now turn to an empirical application.

Table 2.5: Simulated Power Curves: 0.05 Quantile

$T =$	20	50	150	400	1100
Model $i = 1$					
CC	*	*	0.04	0.06	0.12
DQ	*	0.08	0.07	0.09	0.13
BRMQE/DM	0.65	0.76	0.81	0.85	0.88
Model $i = 2$					
CC	*	*	0.08	0.15	0.32
DQ	*	0.11	0.11	0.18	0.37
BRMQE/DM	0.73	0.85	0.93	0.99	1.00
Model $i = 3$					
CC	*	*	0.14	0.31	0.69
DQ	*	0.15	0.20	0.37	0.74
BRMQE/DM	0.80	0.92	0.99	1.00	1.00

CC = Conditional Coverage; DQ = Dynamic Quantile; BRMQE/DM = Bootstrap Return Method quantile estimator combined with Diebold & Mariano (1995).

Table 2.6: Simulated Power Curves: 0.01 Quantile

$T =$	20	50	150	400	1100
Model $i = 1$					
CC	*	*	*	0.04	0.06
DQ	*	0.10	0.08	0.09	0.12
BRMQE/DM	0.65	0.77	0.82	0.86	0.90
Model $i = 2$					
CC	*	*	*	0.09	0.23
DQ	*	0.13	0.13	0.20	0.34
BRMQE/DM	0.73	0.85	0.93	0.99	1.00
Model $i = 3$					
CC	*	*	*	0.23	0.56
DQ	*	0.18	0.22	0.39	0.67
BRMQE/DM	0.80	0.92	0.99	1.00	1.00

CC = Conditional Coverage; DQ = Dynamic Quantile; BRMQE/DM = Bootstrap Return Method quantile estimator combined with Diebold & Mariano (1995).

2.5.3 Empirical

In this section the BRM quantile estimator is used in combination with the Model Confidence Set[56] of Hansen et al. (2011), to evaluate five VaR forecasting models. Specifically, three conditional volatility models with Normal innovations and two naive moving window schemes are considered:

1. GJR-GARCH(1,1,1),
2. GARCH(1,1),
3. JP Morgan's RiskMetrics (RM),
4. Naive Historical Simulation with moving window of 250 observations (NHS250), and
5. a moving window of 500 observations (NHS500).

I chose these models in particular, not just for their simplicity and popularity, but also because intuition suggests an *a priori* ranking. Specifically, the naive historical simulations are expected to have the worst performance since they respond slowly to shifts in VaR. RiskMetrics should outperform the naive historical simulation since it will be quicker to respond to shifts in VaR (via the conditional variance). However, the innovation and conditional volatility parameters in RiskMetrics are bound to be 0.06 and 0.94 respectively, so it exhibits no parameter freedom. The GARCH(1,1) is essentially a RiskMetrics model, but with parameter freedom; thus, given a large estimation window, it is likely that GARCH(1,1) will outperform RiskMetrics. Finally, GJR-GARCH(1,1,1) introduces a leverage effect on top of the GARCH(1,1) specification, which is generally agreed to exist in equity markets, hence *a priori* it is expected to have the best overall performance. Support for this set of *a priori* rankings can be found in, for example, Glosten, Jagannathan & Runkle (1993), Berkowitz & O'Brien (2002), and Gaglianone et al. (2011).

The Model Confidence Set for twenty of the largest NYSE equities is obtained.[57] VaR forecasts are calculated from January 2010 to December 2011, so that $T \approx 500$. The length of this interval was limited to 2 years to ensure $N \gg T$. This interval in particular is chosen since it is the latest in my dataset and thus is the most likely to exhibit large N. For the stated assets, one typically observes $N \approx 5000$, which should be sufficient for the regularity conditions in Patton & Li (2013). Note, the parameters of the GARCH and GJR-GARCH models are estimated using the six years of data immediately prior to the forecast interval, i.e. January 2004 to December 2009, which yields approximately 1500 observations.

[56]A description of this procedure can be found in Chapter 3 Appendix 3.A. The MAE loss function is employed. Results with the MSE loss function were very similar.

[57]See Footnote 45 for a list of ticker codes.

Table 2.7: VaR Model Recommendations Across 20 Assets

0.05 quantile	GJR-GARCH	GARCH	RM	NHS250	NHS500
CC	17	16	19	18	9
DQ	18	15	18	11	1
BRMQE/MCS	19	18	14	6	0
0.01 quantile	GJR-GARCH	GARCH	RM	NHS250	NHS500
CC	18	17	18	17	19
DQ	13	15	10	15	13
BRMQE/MCS	19	18	14	6	0

CC = Conditional Coverage; DQ = Dynamic Quantile; BRMQE/MCS = Bootstrap Return Method quantile estimator combined with the model confidence set of Hansen et al. (2011); RM = RiskMetrics; NHS = Naive Historical Simulation

After performing the BRM quantile estimator combined with Model Confidence Set for each asset, a forecasting model is regarded as "recommended" if it is in the Model Confidence Set for that asset. I also perform CC and DQ tests for each forecasting model/asset combination and define a model to be recommended if the tests fails to reject the null hypothesis.

The results are contained in Table 2.7. Each number in the table depicts the number of times a given forecast model is recommended by a given test. Thus the maximum possible score is 20, i.e. the forecasting model is recommended for all 20 assets, and the minimum score is 0 (not recommended for any assets). Ideally, a statistical test will recommend the single best forecasting model in all 20 cases, and will not recommend any other forecasting model.

Table 2.7 demonstrates that for the 0.05 quantile, the BRM quantile estimator with model confidence set recommendations exactly match our *a priori* expectations; that is, GJR-GARCH is recommended for the largest number of assets, and as we move down the forecast model rankings, the number of recommendations drops monotonically. Of particular interest, the model that is expected to have the worst performance is rejected for all 20 assets. Interestingly, for the 0.05 quantile, the DQ test also has quite good performance, although it is not able to reject the historical simulation with window length of 250 as often as the approach advocated in this chapter. The CC test has poor performance, showing little ability to distinguish between the five forecasting models, except for a mild ability to single out the worst performer.

For the 0.01 quantile, the performance of the BRM quantile estimator approach is identical in aggregate to the 0.05 quantile, although there were some small differences in the individual results for each ticker code. In contrast, both the CC and DQ tests show no ability to distinguish between the five forecasting models. This result it not unexpected, since these tests will always have more difficulty with smaller quantiles since a VaR break is a rarer event. This contrast is striking, since the 0.01 quantile is favoured by most industry practitioners.

Figure 2.2 provides some intuition for the results in Table 2.7. For IBM on the New York Stock Exchange, the time-series of the BRM quantile estimator (the VaR proxy) is provided, along with the forecasts of the best and worst performing models (GJR-GARCH and NHS500), over the forecast interval of January 2010 to December 2011. As expected, GJR-GARCH responds quickly to changes in the VaR proxy, while NHS500 is much slower to respond (and is initially much too large in absolute value due to after-effects of the global financial crisis). This results in the NHS500 lying much further, on average, from the VaR proxy, which explains its poor performance in Table 2.7.

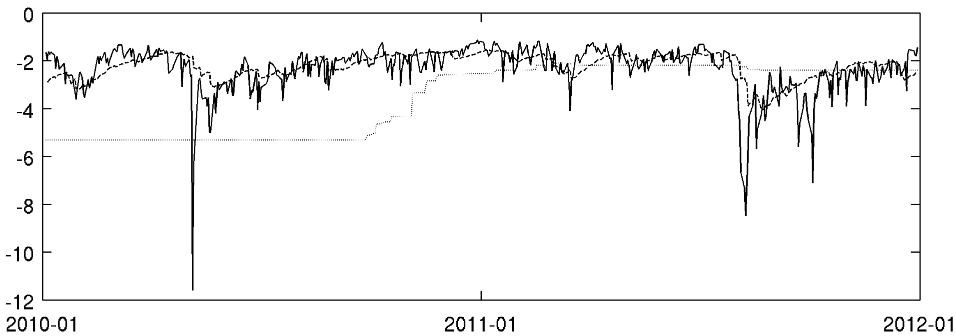


Figure 2.2: Two daily return VaR forecasts and the daily return VaR proxy over the forecast interval, January 2010 to December 2011, for IBM data, expressed as a percentage. The proxy (BRM quantile estimator) is the solid line, a GJR-GARCH forecast is the dashed line, and a naive historical forecast is the dotted line.

More generally, the time-series of the BRM proxy is also of independent interest as it is the first consistent, non-parametric proxy for true VaR available in the literature. To this end, Figure 2.3 provides the plot of this time-series for IBM on the New York Stock Exchange for both the estimation and forecast interval, i.e. January 2004 to December 2011.

In summary, combining the BRM with tests from the forecast evaluation literature is a novel and powerful approach to evaluating VaR forecasting models, especially for the industry standard 0.01 quantile. Under simulation, this approach has power to reject misspecified models far exceeding that of coverage-based evaluation tests. Empirically, this approach yields sensible results that match *a priori* expectations and provide firmer recommendations than coverage-based tests.

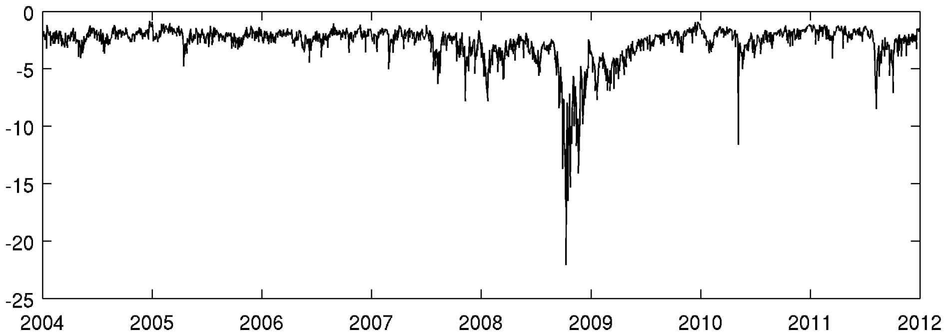


Figure 2.3: Time-series of the daily return BRM quantile estimator from January 2004 to December 2011 for IBM, expressed as a percentage.

2.6 Conclusion

In this chapter, a bootstrap-based estimator for a range of financial risk measures using intraday data is proposed. For the specific case of the variance, the estimator is shown to have performance comparable to realised kernels; arguably the best-case non-parametric estimator previously proposed in the literature.

It is also shown that this approach to estimation can be combined with the framework in Patton & Li (2013) in order to evaluate forecasting models of an important set of measures of risk. For the specific case of value-at-risk, this novel approach is shown to convincingly outperform existing popular forecast evaluation methods based on coverage, particularly for smaller quantiles, with respect to both size and power.

In this chapter, I focus on the stationary bootstrap. However, if the appropriate regularity conditions are satisfied, there is nothing in the theory to prevent researchers from using other resampling methods. For example, the tapered block bootstrap of Paparoditis & Politis (2001) or the extended tapered block bootstrap of Shao (2010) may yield faster rates of convergence for the bias of the estimator. Of particular interest, given the extraordinarily general regularity conditions, the subsampling approach of Politis & Romano (1994*a*) could potentially enable the analysis of infinite variance intraday sequences, paving the way for a robust test of the conditional Normality of daily returns.

Given that the framework of Patton & Li (2013) can accommodate multivariate intraday data, it would also be interesting to extend the methods in this chapter to a multivariate framework. Intuitively, if an intraday sequence on two assets is sampled at identical times throughout the day, then it is possible that application of an identical resample scheme to both intraday sequences will preserve the correlation structure of the assets. An immediate application is the consistent and non-parametric estimation of market betas.

Finally, in the present chapter, only consistency of the BRM was proven, since this is all that was needed for the application in Section 2.5. In particular, an asymptotic distribution theory for BRM estimators was not discussed. One possible approach to this problem is iterative application of the bootstrap.[58] I am currently investigating all of these applications and extensions.

Appendix 2.A Impact of Jumps on the Bootstrap Return Method

In this appendix, the output of the BRM variance estimator is contrasted with that of realised kernels[59] and bipower variation[60]. In a continuous-time modelling framework, realised kernels is consistent for quadratic variation, which incorporates the variation of both the continuous and jump component. In contrast, bipower variation is consistent for integrated variance (assuming no microstructure noise), which incorporates only the variation of the continuous component. *Theoretical* comparison of the BRM variance estimator to these two estimators is not meaningful, since the BRM uses a discrete-time modelling framework. However, an *empirical* comparison of their output provides some intuition of how the BRM responds when a single observation violates the vanishing variance assumption, that is, a jump, in continuous-time parlance.

Let $j = 1, ..., 20$ index the set of 20 New York Stock Exchange ticker codes listed in Footnote 45, and let $t = 1, ..., T$ index the set of trading days from January 2004 to December 2011, so $T \approx 2000$. For each asset and trading day I calculate the BRM variance estimator ($BRM_{j,t}$), realised kernels ($RK_{j,t}$), and 5-minute bipower variation ($BP_{j,t}$). A metric for the percentage deviation of one estimator from another is:

$$d_{j,t,X:Y} = 100 \left(\ln(X_{j,t}) - \ln(Y_{j,t}) \right), \tag{2.15}$$

where $X, Y \in \{BRM, RK, BP\}$. It is interesting to compare sample means of $d_{j,t,X:Y}$ on days with jumps versus days without jumps. Using the jump detection procedure proposed in Barndorff-Nielsen & Shephard (2006),[61] for each asset, I construct a set containing the indices of all days with jumps, and denote it $\mathcal{J}_j$. $\mathcal{K}_j$ contains the indices of days without jumps, i.e. the complement of $\mathcal{J}_j$ on $1, ..., T$. For each asset, consider the statistic constructed by differencing sample means as follows:

$$\bar{\delta}_{j,X:Y} = \frac{1}{|\mathcal{J}_j|} \sum_{t \in \mathcal{J}_j} d_{j,t,X:Y} - \frac{1}{|\mathcal{K}_j|} \sum_{t \in \mathcal{K}_j} d_{j,t,X:Y} \tag{2.16}$$

[58]Lahiri (2003) chapter 1.

[59]Barndorff-Nielsen et al. (2008*a*).

[60]Barndorff-Nielsen & Shephard (2004).

[61]I use a 5-minute sampling frequency for realised variance, bipower variation, and tripower quarticity.

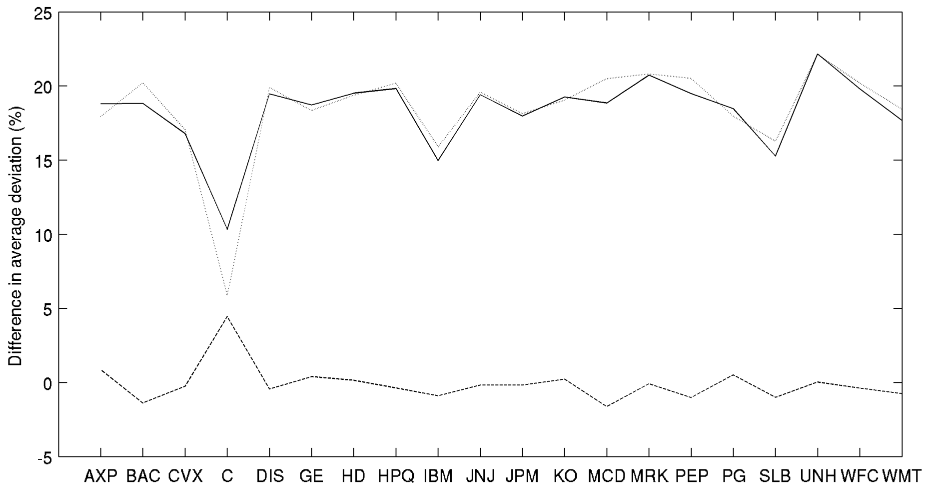


Figure 2.4: The three lines describe the difference in average percentage deviations of variance estimators on jump versus non-jump days, i.e. Equation (2.16), across 20 assets. The solid line corresponds to the BRM variance estimator versus bipower variation, the dashed line corresponds to the BRM variance estimator versus realised kernels, and the dotted line corresponds to realised kernels versus bipower variation.

The three lines in Figure 2.4 correspond to $\bar{\delta}_{j,BRM:BP}$ (solid line), $\bar{\delta}_{j,BRM:RK}$ (dashed line), and $\bar{\delta}_{j,RK:BP}$ (dotted line). As can be seen, the dashed line for $\bar{\delta}_{j,BRM:RK}$ hovers stochastically around zero, indicating that the difference in average percentage deviation of the BRM variance estimator from realised kernels is roughly the same, regardless of whether the day contains a jump. In contrast, the solid line for $\bar{\delta}_{j,BRM:BP}$ hovers stochastically around 20 percent. This indicates that on days containing jumps, the BRM variance estimator is approximately 20 percent larger than bipower variation (on average). Interestingly, almost identical performance is observed in $\bar{\delta}_{j,RK:BP}$ (dotted line). That is, on days containing jumps, realised kernels is also approximately 20 percent large than bipower variation (on average). In summary, the evidence suggests that empirically, the BRM variance estimator responds to the presence of jumps in much the same manner as realised kernels, incorporating the extra variation into the daily variance estimator.

Appendix 2.B A Comparison of Modelling Assumptions

In this appendix, the modelling assumptions of the present chapter are contrasted with the popular continuous-time modelling framework from the extant literature. The basic approach to continuous-time modelling of intraday return sequences models increments as discrete realizations of a continuous-time semi-martingale. This is the approach taken by Goncalves & Meddahi (2009); specifically, they employ a stochastic volatility model for the log-price.

It is well known that intraday return data violate this continuous-time model at the highest frequencies. In particular, microstructure effects such as bid-ask bounce induce a weak dependence structure on the data. One approach to duplicating this feature of the data is to hypothesize the existence of a latent "efficient" log-price process that is a continuous-time semi-martingale. The observed log-price process is set equal to the latent continuous-time semi-martingale plus an additive noise term that may be weakly dependent. This model can be found in, for example, Zhou (1996) or Zhang et al. (2005). I will refer to this model as the CSMN (Continuous Semi-Martingale plus Noise).

The CSMN differs from the modelling assumptions in the present chapter in three important respects. Specifically, in a CSMN framework:

1. it is the properties of a hypothesized, unobservable process that are of interest,

2. the variance of the intraday return sequence is stochastic, and can exhibit long-memory, and

3. the variance of an individual intraday return does not vanish as $N \to \infty$.

Regarding the first point above, there is some *economic* justification for the hypothesized, unobservable price process, although it is rarely stated in the *econometric* literature. For example, Jouini & Kallal (1995) demonstrate that the absence of arbitrage is equivalent to the existence of an equivalent probability measure that is able to transform a process between the best bid and best ask prices into a martingale. That is to say, a fundamental theorem of asset pricing can exist, even in the presence of microstructure effects. From a more practical perspective, for returns that span a long interval of time, the properties of a latent return and an observed return are likely to be similar.

However, it is worth emphasizing that the approach advocated in this chapter is strictly simpler in the sense that it does not depend on a no-arbitrage condition or any underlying economic theory. Further, the modelling assumptions are easier to verify in any particular case since they apply directly to the *observable* data. In short, the modelling approach in the present chapter can be recommended over the CSMN framework by appealing to Occam's razor: among competing models, the one with fewest assumptions should be selected.

The inability of the modelling assumptions in this chapter to address the second point above, are, in my opinion, the weakest feature of the BRM procedure. Interestingly, the BRM variance estimator was not negatively affected by the simulated $I(1)$ volatility process for intraday returns in Section 2.4. Nonetheless, a more thorough theoretical treatment of this issue in future work on this topic is desirable.

I have already discussed the third point above in some detail in Section 2.2, as well as Appendix 2.A but there are two additional points worth mentioning.

First, although the majority of papers in the literature have focused on the CSMN framework, i.e. non-vanishing variance of increments,[62] there is some precedent for vanishing variance.[63] Interestingly, the relevant asymptotic theory for a microstructure noise term with variance vanishing in N has been thoroughly treated in the continuous-time framework, and is frequently used in Physics.[64] An interesting possibility for future work would be an investigation of whether the discrete-time results in this chapter also apply in continuous time given an asymptotic framework where the variance of microstructure noise vanishes in N.

Second, it turns out the CSMN model with non-vanishing variance of increments is not meaningful in a re-sampling context. Intuitively, in a CSMN framework, the noise terms form a telescoping sum, such that all but the first and last term cancel when one constructs a daily return. However, if one samples with replacement from the intraday sequence using a dependent bootstrap, then a telescoping sum exists only in each re-sampled block. Thus the number of noise terms in the re-sampled daily return will increase with the number of blocks. In turn, this implies a seemingly paradoxical result: even though the variance of the observable daily return is modelled as finite for any N, the variance of a re-sampled daily return diverges with N. I conclude this discussion with Proposition 2.B.1, which proves this assertion for a simple specification of CSMN.

Proposition 2.B.1 *Let $\tilde{p}_s$ denote a latent log-price process generated by the continuous-time Ito process $dp_s = \sigma dW_s$, where W_s denotes a Wiener process and $0 < \sigma < \infty$ is a constant diffusion parameter. Let $p_s = \tilde{p}_s + \epsilon_s$, where ϵ_s denotes mean-zero* iid *microstructure noise with non-zero, finite variance. Consider $n = 0, 1, ..., N$ realizations of this process on the finite interval $[t-1, t]$, each spaced at the constant time increment $\Delta \equiv N^{-1}$. Define $q_n = p_{s_n} - p_{s_{n-1}}$ and $r_{t,N} = \sum_n q_n$, and let $r^*_{t,N}$ denote a SB resample of $r_{t,N}$. Then:*

$$\mathbb{V}r_{t,N} = \sigma^2 + 2\mathbb{V}\epsilon, \forall N, \text{ while } \mathbb{V}^* r^*_{t,N} \to \infty, \text{ as } N \to \infty, \text{ almost surely,} \tag{2.17}$$

where $\mathbb{V}^*$ denotes variance conditional on W_s, ϵ_s, and the resampling indices.

Proof The stated continuous-time model exhibits the discretization:

$$\tilde{p}_n - \tilde{p}_{n-1} = \sqrt{\Delta}\sigma Z_n, \tag{2.18}$$

where $Z_n \overset{iid}{\backsim} \mathcal{N}(0,1)$, see Merton (1980) Appendix A. Adding in the microstructure noise, we get:

$$p_n - p_{n-1} = \sqrt{\Delta}\sigma Z_n + \epsilon_n - \epsilon_{n-1}. \tag{2.19}$$

[62] See footnote 30.
[63] See Rosenbaum (2009) or Large (2011).
[64] Jacod & Protter (2011) Chapter 16.

Thus:

$$
\begin{aligned}
r_{t,N} =& \sum_{n=1}^{N} q_n && (2.20a)\\
=& \sum_{n=1}^{N} \sqrt{\Delta}\sigma Z_n + \sum_{n=1}^{N} \epsilon_n - \sum_{n=1}^{N} \epsilon_{n-1} && (2.20b)\\
=& \sigma N^{\frac{1}{2}} \bar{Z}_N + \epsilon_N - \epsilon_0, && (2.20c)
\end{aligned}
$$

which demonstrates that $\mathbb{V} r_{t,N} = \sigma^2 + 2\mathbb{V}\epsilon$.

Employing the notation from Appendix 2.C, assume for simplicity that $L_1 + ... + L_K = N$ and that the sequences $Z_1, ..., Z_N$ and $\epsilon_0, \epsilon_1, ..., \epsilon_N$ wrap in a circle so that $Z_{N+1} = Z_1$ and $\epsilon_{N+1} = \epsilon_0$. Thus:

$$
\begin{aligned}
\mathbb{V}^* r^*_{t,N} =& \sum_{n=1}^{N} q^*_n && (2.21a)\\
=& \sum_{k=1}^{K} \sum_{i=0}^{L_k-1} \sqrt{\Delta}\sigma Z_{I_k+i} + \sum_{k=1}^{K} \sum_{i=0}^{L_k-1} (\epsilon_{I_k+i} - \epsilon_{I_k+i-1}) && (2.21b)\\
=& \sqrt{\Delta}\sigma \sum_{k=1}^{K} \sum_{i=0}^{L_k-1} Z_{I_k+i} + \sum_{k=1}^{K} (\epsilon_{I_k+L_k-1} - \epsilon_{I_k-1}), && (2.21c)\\
=& \sqrt{\Delta}\sigma \sum_{k=1}^{K} \sum_{i=0}^{L_k-1} Z_{I_k+i} + \sum_{k=1}^{K} \epsilon_{I_k+L_k-1} - \sum_{k=1}^{K} \epsilon_{I_k-1}, && (2.21d)\\
=& S_{1,K} + S_{2,K} - S_{3,K}. && (2.21e)
\end{aligned}
$$

$S_{1,K}$ is by definition a stationary bootstrap re-sample of $\sigma Z_n, n = 1, ..., N$, which converges to σ^2 by Politis & Romano (1994*b*) Theorem 2.

Naively, one might guess there are $2K$ terms in $S_{2,K}$ and $S_{3,K}$ combined. However, this ignores the possibility that $I_k + L_k - 1 = I_j - 1$, $k = 1, ..., K$, $j = 1, ..., K$. If this event occurs, then obviously two of the error terms will cancel. This possibility is examined now. In the analysis that follows, I drop the -1 term from the subscript of $S_{2,K}$ and $S_{3,K}$, since it is irrelevant.

Independence of I_k and L_k implies that

$$
\mathbb{P}(I_k + L_k \neq I_1 \cap ... \cap I_k + L_k \neq I_K) = \prod_{j=1}^{K} \mathbb{P}(I_k + L_k \neq I_j). \tag{2.22}
$$

Independence also implies that for $j \neq k$:

$$
P(I_k + L_k \neq I_j) = \mathbb{P}(I_k \neq I_j) = \frac{N-1}{N}. \tag{2.23}
$$

For $j = k$, note that it is possible that $I_k + L_k = I_k$, since I assume the indices wrap in a circle for the stationary bootstrap. However, the definition of the geometric distribution guarantees that:

$$\mathbb{P}(I_k + L_k \neq I_k) > \frac{N-1}{N}. \tag{2.24}$$

Taken together, it follows that:

$$\mathbb{P}(I_k + L_k \neq I_1 \cap ... \cap I_k + L_k \neq I_K) > \left(\frac{N-1}{N}\right)^K. \tag{2.25}$$

Using a binomial expansion:

$$\left(\frac{N-1}{N}\right)^K = 1 + \sum_{k=1}^{K} (-1)^k \cdot \frac{1}{k!} \cdot \frac{K!}{(K-k)!} \cdot \frac{1}{N^k}. \tag{2.26}$$

Since:

$$\frac{K!}{(K-k)!} < K^k, \tag{2.27}$$

for any integer $k > 0$, it follows that:

$$\left|\frac{K!}{(K-k)!} \cdot \frac{1}{N^k}\right| < \left(\frac{K}{N}\right)^k \to 0, \tag{2.28}$$

if K diverges at a slower rate than N. This is a condition of the stationary bootstrap. Thus:

$$\left(\frac{N-1}{N}\right)^K \to 1, \text{ as } N \to \infty. \tag{2.29}$$

Therefore it has been shown that as $N \to \infty$, each noise term in $S_{2,K}$ will *not* cancel with a noise term in $S_{3,K}$, almost surely. It follows that the number of noise terms that remain is $2K$ almost surely. Some of the noise terms that remain may stack, that is, the same noise term may be encountered several times. However, this will only increase the overall variance of the two sums. Taken together, the analysis is sufficient to assert that:

$$\mathbb{V}^* r^*_{t,N} \geq \sigma^2 + 2K\mathbb{V}\epsilon, \text{ as } N \to \infty, \text{a.s.}. \tag{2.30}$$

Since $K \to \infty$ as $N \to \infty$, this proves the result. $\square$

Appendix 2.C The Stationary Bootstrap

In this appendix, the stationary bootstrap is reviewed for the benefit of any readers who may be unfamiliar with it. Consider the sequence of random variables: $\mathcal{X}_N = \{X_1, ..., X_N\}$. The stationary bootstrap provides a consistent and non-parametric method for estimating the distribution of a wide class of test statistics constructed from $\mathcal{X}_N$. In this chapter, the "test statistic" of interest is a daily return, so attention can be restricted to the sample mean, defined:

$$\bar{X}_N \equiv N^{-1} \sum_{n=1}^{N} X_n. \tag{2.31}$$

To describe the stationary bootstrap method requires some additional notation: Let $I_1, I_2, \ldots$ denote an *iid* sequence from the discrete uniform distribution over the integers $1, \ldots, N$. Let $L_1, L_2, \ldots$ denote an *iid* sequence from the geometric distribution characterized by the probability p_N. Let K denote the smallest integer such that $L_1 + \ldots + L_K \geq N$. Consider the countably infinite sequence Y_n that forms a circle over $\mathcal{X}_N$, so that $Y_1 \equiv X_1$, ..., $Y_N \equiv X_N$, $Y_{N+1} \equiv X_1$, $Y_{N+2} \equiv X_2, \ldots$.

A stationary bootstrap resample of $\mathcal{X}_N$ is obtained by sampling K blocks of sequential observations from $Y_1, Y_2, \ldots$. The k^{th} block ($k = 1, \ldots, K$) is defined:

$$\mathcal{B}_k = \{Y_{I_k}, Y_{I_k+1}, \ldots, Y_{I_k+L_k-1}\}, \tag{2.32}$$

and the full resample is defined:

$$\{X_1^*, \ldots, X_N^*\} \equiv \{\mathcal{B}_1, \ldots, \mathcal{B}_K\}. \tag{2.33}$$

Note, since the size of block $\mathcal{B}_K$ is random, it may happen that $L_1 + \ldots + L_K > N$. If this occurs, standard practice is to truncate the length of the final block so that there are exactly N re-sampled observations.

The only quantity that requires estimation in this process is p_N; the parameter of the geometric distribution that determines the average block length. The optimal choice for p_N will be determined by the type of weak dependence exhibited within $\mathcal{X}_N$. If data are close to *iid* then setting $p_N = 1$ is a sensible choice, and the SB simplifies to the *iid* bootstrap of Efron (1979). If the data exhibit strong serial dependence, then p_N should be chosen closer to 0. Several methods have been proposed for data-driven estimates of p_N.[65]

The re-sampled mean can now be defined:

$$\bar{X}_N^* = N^{-1} \sum_{n=1}^{N} X_n^*. \tag{2.34}$$

Given a weak dependence condition and an appropriate moment bound on X_n, and requiring that $p_N \to 0$ and $Np_N^2 \to \infty$ as $N \to \infty$, $\bar{X}_N^*$ can be characterized as an infinite sum over (asymptotically) independent blocks. Thus given appropriate regularity conditions, a CLT follows. Specifically, let $T_N = \sqrt{N}(\bar{X}_N - \mathbb{E}\bar{X}_N)$ and

[65] In particular, see Patton et al. (2009).

$T_N^* = \sqrt{N}(\bar{X}_N^* - \bar{X}_N)$. Politis & Romano (1994*b*) provide a set of sufficient conditions for:

$$\sup_{x \in R} \left| \mathbb{P}\left(T_N \leq x\right) - \mathbb{P}^*\left(T_N^* \leq x\right) \right| \xrightarrow{\mathbb{P}} 0, \text{ as } N \to \infty, \tag{2.35}$$

where $\mathbb{P}^*$ indicates the probability is conditional on $\mathcal{X}_N$. The conditions for this result are significantly weakened in Goncalves & White (2002) and Goncalves & de Jong (2003).

Chapter 3

An Empirical Analysis of Value-at-Risk Forecasting Models

3.1 Introduction

The number of value-at-risk (VaR) forecasting models proposed in the literature is daunting, particularly to market participants with limited research time. Choosing the most accurate forecasting model(s) is of measurable value to such participants. For example, accurate VaR forecasts save large financial institutions from setting aside too much working capital when satisfying Basel Committee requirements.

In this chapter I provide a set of recommendations pertaining to the most accurate VaR forecasting models. Since the modelling frameworks of different forecasting models are frequently not nested, choosing the best forecast models based on theoretical criteria is not possible. Instead, I analyse the empirical accuracy of the forecasting models *ex post*.

This is not the first paper to rank VaR forecast models based on empirical criteria. A brief survey of other results in the literature is provided in Section 3.4. The appeal of the present chapter is three-fold:

- the evaluation approach proposed in Chapter 2 is used and, as discussed in that chapter, is much more powerful than the procedures used in other comparable studies,
- a new class of VaR forecasting models that employ the intraday data-based VaR estimator from Chapter 2 is proposed and shown to yield accurate forecasts, and
- the analysis is of a large scale; 351 VaR forecast models are considered on equity data spanning two distinct exchanges and covering the global financial crisis period as well as the more stable period that immediately followed.

The methodology used in the present chapter is exactly that described in Section 2.5 of Chapter 2, so this chapter is essentially an expansion of the empirical VaR analysis from Chapter 2 to a larger dataset and incorporating a larger set of VaR forecast models. For this reason, I will not dwell on the exact details of the methodology since they are readily available in Chapter 2.

Using this method, I rank forecasting models across a wide range of equities from the New York Stock Exchange (NYSE) and my home market, the Australian Stock Exchange (ASX). I focus on forecasting VaR during the Global Financial Crisis and the period of (relatively) lower volatility that followed.

Briefly, the results indicate that forecast models that utilize intraday data strongly outperform those that use daily data. Of particular interest: a class of simple time series models defined over the Bootstrap Return Method (BRM) Quantile Estimator (proposed in Chapter 2) are the best performers. Among models that do not exploit intraday data, it is shown that, contrary to several other findings in the literature, unit-root conditional volatility models that employ a Gaussian assumption perform well. This set of models includes the oft-maligned RiskMetrics of JPMorgan. These findings are consistent both during and after the global financial crisis. It is also worth noting that, unsurprisingly, the ability to distinguish between forecast models is reduced during the global financial crisis.

The rest of the chapter proceeds as follows: in Section 3.2, the exact details of the methodology are briefly summarised, and the issue of parameter estimation error is discussed. In Section 3.3 I discuss the VaR forecast models that are to be compared. In Section 3.4 some empirical results and rankings from the prior literature are reviewed. Section 3.5 provides some additional detail on the data and computational choices, and then the empirical results are presented and discussed in Section 3.6. Section 3.7 concludes. In Appendix 3.A the Model Confidence Set procedure is discussed, and in Appendix 3.B a more comprehensive set of results tables are provided.

3.2 Method

Let $X_t, t = 1, ..., T$ denote T realizations of the unobservable random variable of interest; true VaR. Let $Y_{t,k}$ denote a forecast of X_t, constructed from time $t-1$ information, where there are K different forecasts indexed $k = 1, ..., K$, and where $Y_{t,k}$ possibly depends on a parameter vector. Let $\tilde{X}_t$ denote a proxy for X_t, where $\tilde{X}_t$ is constructed from information from the interval $[t-1, t]$. Where it is appropriate, I write $\tilde{X}_{t,N}$ to indicate a proxy that is constructed from an intraday sequence of length N from the interval $[t-1, t]$.

Let $L(X_t, Y_{t,j})$ denote a loss function (i.e. a distance measure). The best forecast model is defined as the one that minimises expected loss, i.e. $\mathbb{E}L(X_t, Y_{t,j})$. Given two forecast models, the simplest approach is to define the loss differential: $d_t = L(X_t, Y_{t,j}) - L(X_t, Y_{t,i})$ and then test the null hypothesis $H_0 : \mathbb{E}d_t = 0$ (see Diebold & Mariano (1995) and West (1996)).

Evaluating VaR forecast models is difficult for two reasons. First, VaR is unobservable. This problem is resolved by replacing X_t with a proxy, $\tilde{X}_t$, and defining the proxy loss differential: $\tilde{d}_t = L(\tilde{X}_t, Y_{t,j}) - L(\tilde{X}_t, Y_{t,i})$. Under conditions described in Patton & Li (2013) and Chapter 2, the rankings implied by analysis of $\mathbb{E}\tilde{d}_t$ and the rankings implied by analysis of $\mathbb{E}d_t$ are the same with unit probability in large sample . A detailed description of the framework is in Chapter 2. Second, there are typically more than two forecast models of interest. This problem is solved by employing the Model Confidence Set of Hansen et al. (2011). This procedure controls for the family-wise error rate implicit in analysing more than 2 models. The procedure is described in Appendix 3.A.

Employing the Patton & Li (2013) framework, in combination with the BRM Quantile estimator as a proxy for the true VaR, results in VaR forecast evaluations tests that are much more powerful than existing methods in the literature, such as the Conditional Coverage test of Christoffersen (1998), the Dynamic Quantile test of Engle & Manganelli (2004), and the regulatory loss-function approach found in, for example, Lopez (1998), Bao et al. (2006), and Sener et al. (2012). This is demonstrated in Chapter 2 and so the point will not be further expounded upon here.

This section is concluded with a brief discussion on the issue of parameter estimation. Some of the forecast models considered in this chapter have parameter vectors which must be estimated. This introduces an additional source of error into the forecast evaluation procedure, which some researchers are interested in explicitly accounting for.[66] For the purposes of the present chapter, it is assumed that the audience is not interested in explicitly accounting for this additional source of error. The stated goal of this chapter is to provide a set of rankings for practitioners. Since estimation error is of concern for practitioners, I believe it should inherently be a part of the forecast evaluation procedure. For example, consider two forecast models, j and i. Assume that if the population parameters of both models are known, then model j will exhibit slightly better forecast accuracy than model i. However, assume also that the parameters of model j are much more difficult to estimate accurately, and that inaccurate parameter estimation results in reduced forecast ability. In this hypothetical, it is likely that a practitioner will prefer to use model i over model j. The rankings provided in this chapter should reflect this, and so I do not explicitly attempt to remove the effects of estimation error from the forecast evaluation procedure. Nonetheless, it is also worth noting that several of the estimation choices in this chapter will possibly result in the issue of estimation error being moot. For example,

[66] See West (1996), Clark & McCracken (2001), and Giacomini & White (2006).

the estimation period is always longer than the forecast period which implies the estimation error may satisfy a negligibility criteria.[67] Further, a rolling estimation window is employed which is known in some cases to render the issue of estimation error moot, at least asymptotically.[68]

3.3 Models

In this chapter, the analysis is limited to the one-day ahead forecast horizon, since this is the most popular horizon in industry. The VaR forecast models considered can be divided into four categories, labelled:

1. historic VaR,
2. historic volatility,
3. conditional volatility, and
4. CAViaR (Conditional Autoregressive Value-at-Risk by Regression Quantiles).[69]

Further, within each category, I consider the sub-categories of those forecast models that utilize intraday data versus those that do not. Note that forecasts based on option-implied volatility are not included. This is because these forecasts are for the volatility of a return spanning from the current time to the expiration of the contract. Converting this forecast into a one-day ahead forecast is not possible without making additional strong assumptions. Additional assumptions are also necessary to choose the appropriate set of option strike prices to consider.

The models are now described, by category.

3.3.1 Historic Value-at-Risk

This category includes the naive historic method and basic time series models of a proxy for the true VaR (specifically, the BRM Quantile Estimator).

The naive historic method is the simplest to implement and easily the most popular among large financial institutions.[70] Let $R_{(s)}(t,\tau)$ denote the order statistics of the sequence $R_{t-\tau+1},...,R_t$. Then the empirical λ-quantile at time t is $R_{(\lfloor\lambda\tau\rfloor)}(t,\tau)$. In other words, at time t, sort all daily returns over the window $[t-\tau+1,t]$, and then choose the $\lfloor\lambda\tau\rfloor$ element of the sorted sequence. The forecast is then $Y_t = R_{(\lfloor\lambda\tau\rfloor)}(t-1,\tau)$. In this chapter, I consider $\tau = \{100, 250, 500\}$. Clearly, this method does *not* employ intraday data.

[67] West (1996) Theorem 4.1.
[68] Giacomini & White (2006).
[69] Engle & Manganelli (2004).
[70] See Berkowitz & O'Brien (2002) and Mehta, Neukirchen, Pfetsch & Poppensieker (2012).

In contrast, all other methods in this category employ intraday data, since they utilize the proxy for VaR proposed in Chapter 2; the BRM Quantile Estimator, denoted $\tilde{X}_t$. I consider several basic time series models over the proxy, with both fixed and optimized parameter choices. In the cases where parameters are chosen optimally, this is done at 3-month rolling windows by minimizing the loss over the most recent historical window of 1000 days. The time series models considered are:

1. random walk: $Y_t = \tilde{X}_{t-1}$,
2. moving average: $Y_t = \tau^{-1} \sum_{s=1}^{\tau} \tilde{X}_s$, for $\tau = \{5, 10, 20, 50, 100, 250\}$,
3. exponential smoothing: $Y_t = \beta Y_{t-1} + (1-\beta)\tilde{X}_{t-1}$, for $\beta = \{0.1, 0.25, 0.5, 0.75, 0.9\}$, as well as choosing β optimally as described above,
4. exponentially weighted moving average: $Y_t = (\sum_{s=1}^{\tau} \beta^s)^{-1} \sum_{s=1}^{\tau} \beta^s \tilde{X}_s$, for each combination of $\beta = \{0.1, 0.25, 0.5, 0.75, 0.9\}$ and $\tau = \{5, 10\}$, as well as choosing β optimally as described above, and
5. smooth transition exponential smoothing: $Y_t = \alpha_t \tilde{X}_{t-1} + (1 - \alpha_t) Y_{t-1}$, where $\alpha_t = 1/(1 + \exp(\beta + \gamma V_{t-1}))$, for $V_t = \left| Y_t - \tilde{X}_t \right|$ (MAE adjustment) and $V_t = (Y_t - \tilde{X}_t)^2$ (MSE adjustment), with β and γ chosen optimally as described above.

This is the first paper in the literature to examine time series models defined over the BRM Quantile Estimator.

3.3.2 Historic Volatility

This category consists entirely of the basic time series models described above, but applied to estimates of the variance. The VaR forecast is then obtained from a variance forecast using several different methods.

Specifically, I consider the following three estimates of variance:

1. squared return,
2. 5-minute realised variance,[71] and
3. realised kernels.[72]

I apply each of the basic time series models described above[73] to each of these estimates, which results in a variance forecast, denoted $\hat{\sigma}_t^2$. Each variance forecast is then transformed to a VaR forecast using each of the following two methods:

[71] See Andersen & Bollerslev (1998) or Barndorff-Nielsen & Shephard (2002*b*).
[72] Barndorff-Nielsen et al. (2008*b*).
[73] Random walk, moving average, exponential smoothing, exponentially weighted moving average, and smooth transition exponential smoothing.

1. Normality: $Y_t = \hat{\sigma}_t \Phi^{-1}(\lambda)$, where Φ^{-1} is the inverse Gaussian distribution function and λ is the quantile probability.

2. Filtering: let $Z_t = \tilde{\sigma}_t^{-1} R_t$ denote a standardized return. The VaR forecast is then $\hat{\sigma}_t Z_{(\lfloor 1000\lambda \rfloor)}(t-1, 1000)$, ie multiply the volatility forecast by the empirical quantile of the standardized returns over the most recent 1000 days. This method is used in, for example, Barone-Adesi, Giannopoulos & Vosper (2002), who set $\tilde{\sigma}_t^2 = \hat{\sigma}_t^2$ (the forecast). I also consider $\tilde{\sigma}_t^2$ equal to 5-minute realised variance and realised kernels, since given simple modelling assumptions it is straightforward to show that the accuracy of the empirical quantile increases as $\mathbb{V}(\tilde{\sigma}_t^{-1} - \sigma_t^{-1}) \to 0$, where σ_t denotes true volatility.

There is reason to believe that parametric assumptions other than Normality would be suitable.[74] However, in order to keep the total number of VaR forecast models tractable, I consider only the Normal assumption, and rely on the filtering method to capture situations in which the Normal assumption is not appropriate.

I consider every possible combination of estimator, time series model, and VaR transform method described above. Any combination that uses 5-minute realised variance or realised kernels is defined to be an intraday data-based method.

More complex volatility forecast models that incorporate intraday data do exist,[75] and have even been used in a VaR forecast context.[76] However, I avoid these more complex variants here in order to keep the overall number of forecast models tractable, and also because it is illustrative to draw direct comparison between the two intraday variance proxies (5-minute realised variance and realised kernels) and the intraday VaR proxy (BRM Quantile Estimator). This is facilitated by restricting attention to the same set of simple time series models across all three proxies.

3.3.3 Conditional Volatility

This category consists entirely of conditional volatility models from the GARCH family, transformed to VaR using several different methods.

The standard framework for conditional volatility models daily returns as $R_t = \sqrt{h_t} Z_t$, $h_t \perp\!\!\!\perp Z_t$, $Z_t \overset{iid}{\backsim} (0,1)$, with h_t known conditional on time $t-1$ information. The literature is then characterized by different models for h_t. Rather than cite a large number of individual papers, I refer the interested reader to the summaries in Hansen & Lunde (2005*a*) or Poon & Granger (2003).

Following the findings in Hansen & Lunde (2005*a*), I use only the first lag of relevant conditioning variables in all conditional volatility models. The models are:

---

[74] Giot & Laurent (2004).

[75] See Anderson & Vahid (2007), Koopman, Jungbacker & Hol (2005), Brownlees & Gallo (2010), Shao, Lian & Yin (2009), Patton & Sheppard (2009) and Shephard & Sheppard (2010) in a univariate framework, and Noureldin, Shephard & Sheppard (2012) in a multivariate framework.

[76] See Giot & Laurent (2004) and Clements, Galvao & Kim (2008).

- TARCH family: $h_t^\delta = \omega + \alpha \left| R_{t-1} \right|^\delta + \gamma \mathbb{I}(R_{t-1} < 0) \left| R_{t-1} \right|^\delta + \beta h_{t-1}^\delta$, $\delta \in \{1, 2\}$; this is RiskMetrics for $\{\omega = 0, \gamma = 0, \alpha = 0.06, \beta = 0.94, \delta = 2\}$, GARCH(1,1) for $\{\gamma = 0, \delta = 2\}$, GJR-GARCH(1,1,1) for $\delta = 2$, IGARCH for $\{\alpha + \beta = 1, \delta = 2\}$ and IAVARCH for $\{\alpha + \beta = 1, \delta = 1\}$ - I also consider the models defined by conditions $\{\delta = 1, \gamma = 0\}$ (TARCH11) and $\delta = 1$ (TARCH111),

- AGARCH: $h_t = \omega + \alpha(R_{t-1} - \gamma)^2 + \beta h_{t-1}$,

- NAGARCH: $h_t = \omega + \alpha(R_{t-1} - \gamma\sqrt{h_{t-1}})^2 + \beta h_{t-1}$, and

- APARCH: $h_t^\delta = \omega + \alpha(|R_{t-1}| - \gamma R_{t-1})^\delta + \beta h_{t-1}^\delta$.

Estimation of the parameters of these models requires specification of the distribution of Z_t. I consider every combination of the above models with the Normal, Student-T, and Skewed-T[77] distributions, except RiskMetrics which explicitly assumes Normality.

The VaR transform is obtained using the relevant parametric assumption (Normal, Student-T, or Skewed-T) or the filtering method described above. The only intraday-based models in this category are those that use 5-minute realised variance and realised kernels in the filtering method.

3.3.4 CAViaR

CAViaR models (Conditional Autoregressive Value-at-Risk by Regression Quantiles) are described in detail in Engle & Manganelli (2004). They are analogous to conditional volatility models, but where the quantity of interest is the quantile. I consider the following variations:

- Adaptive: $Y_t = Y_{t-1} + \beta((1 + \exp(G(R_{t-1} - Y_{t-1})))^{-1} - \lambda$, for positive, large G, where λ denotes the quantile probability.

- Symmetric absolute value: $Y_t = \beta_1 + \beta_2 Y_{t-1} + \beta_3 \left| R_{t-1} \right|$.

- Asymmetric slope: $Y_t = \beta_1 + \beta_2 Y_{t-1} + (\beta_3 \mathbb{I}(R_{t-1} > 0) R_{t-1} + \mathbb{I}(R_{t-1} < 0)) R_{t-1}$.

- Indirect GARCH(1,1): $Y_t = (\beta_1 + \beta_2 Y_{t-1}^2 + \beta_3 R_{t-1}^2)^{1/2}$.

None of the CAViaR models use intraday data.

This concludes the list of models. In total, the descriptions above result in 351 different forecast methods for VaR.

For many of the models, numerical optimization is required. I employed the *patternsearch* algorithm from Matlab's global optimization toolbox, with checks to ensure that termination of the routine was due to the mesh falling below a tolerance

[77] Hansen (1994).

of 1e-6. In testing, this approach proved more robust than gradient-based methods. The exception to this is the conditional volatility models, where all parameters are estimated using Kevin Sheppard's Matlab toolbox, discussed in Chapter 1.

3.4 Literature

There is a vast literature on VaR forecast models. In this section, I limit the discussion to the most prominent papers.

The primary question this paper attempts to address is whether there is an advantage to using intraday data in a forecast context. Results on this topic are mixed. Giot & Laurent (2004) find no appreciable difference between an APARCH conditional volatility model, transformed to VaR using a Skewed-T assumption, and a time series model of realised volatility. They evaluate the forecasts using the Likelihood Ratio test of Kupiec (1995) and the Dynamic Quantile test of Engle & Manganelli (2004). However, it is worth emphasizing that since both these tests analyse VaR breaks, it is possible that competing models exhibit correct coverage and are therefore considered equal, even though one model may be significantly more accurate (in terms of distance from true VaR) than another. That is to say, the results in Giot & Laurent (2004) are not inconsistent with intraday data-based forecasts providing greater accuracy. Similar results can also be found in Angelidis & Degiannakis (2008).

Brownlees & Gallo (2010) find that VaR forecasts based on several intraday variance estimators generally outperform standard conditional volatility models, although they do not significantly outperform VaR forecasts based on daily range volatility estimators. The analysis is done in a loss-based framework, employing the tick loss of Komunjer (2005), sometimes also called the linear regulatory loss function.[78] As discussed in Chapter 2, this approach lacks power relative to the approach advocated in this chapter - nonetheless, the result is notable and worth keeping in mind when considering the results of the present chapter.

Similar results to those of Brownlees & Gallo (2010) can also be found in Clements et al. (2008) and Shao et al. (2009). Of particular interest, the novel approach to forecast evaluation in Fuertes & Olmo (2013) provides evidence that intraday models do not encompass inter-day models, and vice-versa. This suggests that useful information exists in both intraday and inter-day forecasts, although does not provide a recommendation as to which is more accurate. In that paper, combined forecasts are suggested.

Using the Mincer-Zarnowitz regression, further support for realised volatility based forecasts of VaR can be found in Andersen, Bollerslev, Diebold & Labys (2003), although the primary focus of that paper is volatility forecasts. Similarly,

[78] Sener et al. (2012).

using the test for Superior Predictive Ability[79], Koopman et al. (2005) find evidence that realised volatility based forecasts (for volatility) are more accurate than standard stochastic volatility models, as well as GARCH.

For many risky assets intraday data may not be available. In Section 3.6 I also consider the best performing among the set of daily data-based models and so it is of interest to consider the literature on these models here.

Fully parametric or filtered historical methods that incorporate an asymmetric leverage term tend to perform well, e.g. GJRGARCH. More general conditional volatility models, such as GARCH, also have good performance if the source of randomness is assumed to follow a Skewed-T distribution. Support for these observations can be found in Christoffersen (1998), Giot & Laurent (2003), Bao et al. (2006), Kuester, Mittnik & Paolella (2006), and Sener et al. (2012). Sener et al. (2012) also find that CAViaR methods that incorporate asymmetry obtain good performance, although the results of Kuester et al. (2006) do not agree with this. However, Kuester et al. (2006) acknowledge that the poor performance of basic CAViaR methods is likely due to serial correlation early in their sample (which goes back to 1971). They propose an AR(1) adjustment in their paper for this early period which achieves much better performance. Since there is little evidence of serial correlation in daily returns in the data used for the present chapter, I do not consider this adjustment.

Kuester et al. (2006) also note that GARCH-Normal models, i.e. GARCH with a Normal assumption for Z_t, tend to perform better for the larger quantiles; specifically, 0.08 and above, while the GARCH-Skewed-T performs better for smaller quantiles. This is part of a broader set of results across almost the entire VaR forecast evaluation literature that suggests that different forecast models tend to perform better for different quantiles; see also for example Christoffersen, Hahn & Inoue (2001). Given the increased power of the methods used in the present chapter, it will be interesting to examine whether this trend continues.

There are also several papers that emphasize that the most popular VaR forecast method, the naive historical simulation, performs very poorly compared to the basic GARCH-Normal parametric model; see for example, Berkowitz & O'Brien (2002) or Gaglianone et al. (2011).

3.5 Data and Computational Choices

Before stating the results, a brief discussion of the data and computational choices beyond what is provided earlier is necessary.

The analysis in the present chapter is performed on 50 New York Stock Exchange and 50 Australian Stock Exchange publicly listed companies, spanning January 2004 to December 2013. As discussed in Chapter 1, all data are cleaned following the

[79] Hansen (2005).

methods discussed in Barndorff-Nielsen et al. (2009). Post cleaning, I visually investigated any remaining jumps of large magnitude. I am satisfied that the aforementioned cleaning methods are sufficient for the New York Stock Exchange data, but *not* sufficient for the Australian data. Therefore an additional clean was imposed on the Australian data. For each transaction, I examine the median of the next 6 and previous 6 transactions. If the present transaction deviates from both medians *in the same direction* by more than 10 times the most common non-zero absolute increment on that day, then it is removed.[80] Even after this additional measure, I chose to completely remove 19 from the (approximate) 125,000 asset/day combinations in the dataset due to large swathes of suspicious transactions.[81]

In this chapter, a "day" is defined to be the interval between the first and last transaction in a given day. A return over this span is referred to as a daily return, and similarly 5-minute realised variance, realised kernels, and the BRM Quantile Estimator are also calculated over this span.

I do not consider the overnight return as this would contradict the asymptotic assumptions of several of the estimators used in this chapter, and therefore the asymptotic conditions of the Patton & Li (2013) framework. This is because incorporating the overnight return prevents the mesh of the partition of intraday returns from approaching zero. An alternate possibility would have been to scale all intraday estimates up to a 24 hour span using a parameter estimated across all days.[82] Although feasible, the asymptotics are non-trivial since consistency of an intraday estimator on a given day now requires a dual asymptotic framework, where both the number of intraday returns, and number of days, grows. In particular, it is not clear how such a dual-asymptotic framework would affect the dual-asymptotic framework in Patton & Li (2013).

I do not use auction determined opening and closing prices for two reasons. First, such a price is theoretical, i.e. transactions do not necessarily occur at this price. Second, surprisingly, these prices are frequently some distance from the first and last (respectively) transaction of the day, suggesting that participants may be "gaming" the auction with unrealistic bids.

From the estimators and daily returns defined above, all VaR forecasts are computed. Any forecast model parameters are estimated using a 1000 day window, at rolling 3 month intervals. This means the first legitimate forecast is made on 2nd January, 2008, and so the two forecast intervals considered are January 2008 to December 2010, and January 2011 to December 2013. The former is defined to be the global financial crisis period. The interval from January 2004 to December 2007 is

[80] i.e. it is removed if above both medians, or below both medians, but is *not* removed if above one and below the other.

[81] An example of "suspicious" is transaction prices around \$10 for most of the day, except for a short interval where they plummet to around \$0.01.

[82] Hansen & Lunde (2005*b*).

used to “warm-up” the forecast models, as well as providing the initial parameter estimation period.

For quantiles $\lambda = \{0.01, 0.05\}$ and for both the Mean Absolute Error (MAE) and Mean Square Error (MSE) loss functions, I apply the Model Confidence Set algorithm to 3 sets of models. The first set consists of every model discussed in Section 3.3. For the second set, consider that I employ the BRM Quantile Estimator both as a proxy for true VaR and as a predictor in basic time series models. This is perfectly valid as long as the proxy error is a martingale difference sequence. Although this assumption is common in the literature,[83] for the sake of rigour, the second set of models omits all those that employ the BRM Quantile Estimator as a predictor. Finally, the third set omits all models that incorporate intraday data. That is, any model that uses 5-minute realised variance, realised kernels, or the BRM Quantile Estimator as a predictor are omitted.

The above procedure is performed on the 50 most liquid stocks on the New York Stock Exchange and Australian Stock Exchange over the sample period, since this should yield $N \gg T$. I am satisfied that N is sufficiently large for the NYSE assets, but for the ASX assets I also check the results against an analysis of only the 20 most liquid stocks due to liquidity concerns for the rest. These results were consistent with those obtained using all 50 stocks.

The statistic of interest is the number of times a given model appeared in the Model Confidence Set, expressed as a proportion, by summing across assets. In this way, I identify forecast models that work well for a large cross section of individual equities, in the hope that such models will exhibit good performance in other equity sequences not included in the dataset.

3.6 Results

For simplicity, in this section I only provide data on the very best performing models in each subset, and further, only provide the New York Stock Exchange results, since they are very similar to Australian Stock Exchange results. A more comprehensive set of tables can be found in Appendix 3.B.

Table 3.1 demonstrates that when all models are considered, the best performing forecasts are the time series models defined over the consistent proxy from Chapter 2: the BRM Quantile Estimator. The result is consistent across both quantiles, across both forecast intervals, and across both exchanges.

This result is indicative of a dominant theme across the extended set of results. Forecast models that incorporate intraday data outperform those that don’t. Further, the forecast models need not be sophisticated. In fact, a simple exponential smoothing rule with a fixed parameter choice is the overall best performing model. The

[83] See Patton (2011*a*) or Andersen et al. (2003).

Table 3.1: Members of Model Confidence Set: All Models

0.01 quantile	2008-2010	2011-2013	Average
ES($\beta = 0.5$) on BRMQE	0.78	0.78	0.78
STES:MSE(Opt) on BRMQE	0.66	0.60	0.63
ES(Opt) on BRMQE	0.46	0.64	0.55
0.05 quantile	2008-2010	2011-2013	Average
ES($\beta = 0.5$) on BRMQE	0.84	0.90	0.87
STES:MSE(Opt) on BRMQE	0.72	0.64	0.68
ES(Opt) on BRMQE	0.54	0.66	0.60

BRMQE = Bootstrap Return Method Quantile Estimator, ES = Exponential Smoothing, STES:MSE = Smooth Transition Exponential Smoothing with Mean Square Error Adjustment, Opt = model parameters estimated optimally

success of exponential smoothing is not overly surprising. Exponential smoothing has an ARIMA(0,1,1) representation, and $I(1)$ processes have long been considered reasonable models for volatility. Given the relationship between volatility and quantiles, it is not unreasonable to conjecture that the time series of VaR should exhibit similar qualities.

The success of the fixed parameter rule over an optimally chosen parameter is more surprising. This suggests that the appropriate smoothing parameter is similar across all assets and close to 0.5 (although it is worth adding that 0.75 performs well for Australian stocks). In particular, it would appear that attempting to estimate the parameter optimally for each individual asset can prove sub-optimal, due to the additional source of parameter estimation error in the forecast. It would seem that better performance could be achieved by estimating a single parameter using data from all assets, such as is explicitly done by RiskMetrics in the daily data case.

In Table 3.2 I remove all BRM Quantile Estimator based forecast models from the feasible set. Interestingly, the best model is still an exponential smoothing with fixed parameter of 0.5, albeit this time defined over realised kernels, with the inverse Gaussian distribution used to transform the volatility forecast to VaR.

This result strengthens the argument that VaR forecasts based on intraday data outperform those based on daily data, since there is no reason to believe that VaR forecasts based on time series models of realised kernels would be unfairly advantaged by the use of the BRM Quantile estimator as a proxy for true VaR. As discussed, it is also not surprising that an $I(1)$ model is a reasonable choice for realised kernels, and, given the results in Table 3.1, the dominance of the fixed parameter rule seems reasonable.

An interesting result worth emphasizing is the dominance of forecasts based on

Table 3.2: Members of Model Confidence Set: No Bootstrap Return Method Quantile Estimator

0.01 quantile	2008-2010	2011-2013	Average
ES($\beta = 0.5$) on RK, Normal	0.76	0.56	0.66
ES($\beta = 0.75$) on RK, Normal	0.30	0.72	0.51
STES:MAE(Opt) on RK, Normal	0.58	0.36	0.47
0.05 quantile	2008-2010	2011-2013	Average
ES($\beta = 0.5$) on RK, Normal	0.72	0.48	0.60
ES($\beta = 0.75$) on RK, Normal	0.34	0.68	0.51
STES:MAE(Opt) on RK, Normal	0.68	0.32	0.50

RK = realised kernels, ES = Exponential Smoothing, STES:MAE = Smooth Transition Exponential Smoothing with Mean Absolute Error Adjustment, Opt = model parameters estimated optimally, Normal = volatility forecast transformed to VaR with inverse Gaussian cdf

realised kernels over forecasts based on 5-minute realised variance. Of course, this is not a direct test of the efficacy of these estimators, but their relative performance in forecasting applications is nonetheless a question of interest. I note that this result, for the most part, accords with the findings in Liu et al. (2013), who perform direct tests of these estimators. This is not to say 5-minute realised variance is not useful. Analysis of Table 3.5 in Appendix 3.B indicates that a fixed parameter exponential smoothing model for 5-minute realised variance, with inverse Gaussian transform to obtain VaR, performed well, especially in the 2011-2013 period. In contrast, daily data-based methods do not even make an appearance in this table.

A final point worth mentioning from Table 3.2 is the dominance of the Normal assumption to obtain VaR from a volatility forecast. There has previously been evidence in the literature that daily returns, standardized by intraday volatility estimators, appear close to Normal.[84] However, the results in Table 3.2 suggest (indirectly) that the Normal assumption is also good for returns standardized by volatility forecasts based on intraday measures. In particular, it is interesting that this assumption outperforms empirical quantiles estimated from filtered data. This could be because of the additional estimation error inherent in estimating empirical quantiles, or it could be due to other forms of nonstationarity in daily returns aside from that caused by heteroskedasticity.

Although the primary purpose of this chapter is to compare the performance of VaR forecasts based on intraday data versus those based on daily data, it is also of interest to examine the dominant models from the subset of daily data-based models. As discussed in Section 3.4, there has been some suggestion in the literature that for equity data, conditional volatility models that incorporate asymmetry and utilize

[84]Andersen et al. (2001).

Table 3.3: Members of Model Confidence Set: No Intraday Models

0.01 quantile	2008-2010	2011-2013	Average
RiskMetrics	0.76	0.82	0.79
IAVARCH:Normal, Parametric	0.90	0.68	0.79
IGARCH:Normal, Parametric	0.88	0.52	0.70
0.05 quantile	2008-2010	2011-2013	Average
RiskMetrics	0.72	0.60	0.66
IAVARCH:Skewed-T, Parametric	0.84	0.38	0.61
IAVARCH:Normal, Parametric	0.86	0.34	0.60

Parametric = volatility forecast transformed to VaR using parametric assumption from model estimation. Note, RiskMetrics uses inverse Gaussian cdf

a Skewed-T distribution will outperform. In contrast, I find the best VaR forecasts result from $I(1)$ conditional volatility models combined with a Gaussian assumption. Of particular note: the oft-maligned RiskMetrics of JPMorgan is the overall best performer.

It is possible that these results are simply a feature of the forecast interval under analysis, or the forecast evaluation method employed. However, given the results in Table 3.1 and Table 3.2, I suggest that estimation error might be the determining factor. RiskMetrics is a fixed parameter model and so exhibits no parameter estimation error. While parameter estimation is part of IGARCH and IAVARCH, these models are notable among conditional volatility models in that both of them impose a parameter constraint such that the resulting process is $I(1)$. If the true process is $I(1)$, this constraint will significantly reduce the estimation error. Previous studies that find in favour of asymmetric models with a Skewed-T assumption generally estimate parameters over a much longer estimation window than that used in the present chapter, hence estimation error is less likely to be a determining factor in these studies. In the present chapter, I was constrained by the need for good quality intraday transactions data over the entire estimation and forecast interval - which is difficult to obtain for the Australian Stock Exchange prior to 2004.

A further result that can be garnered from all three tables, as well as the additional tables in Appendix 3.B, is that in all the tests performed, the best performing forecast models are fairly constant, across equity exchanges, across forecast intervals, and even across both quantiles. This is a reassuring result for any practitioners who adopt a one-size-fits-all rule in their risk modelling, although it is worth emphasizing that, in this chapter, only equity data are considered. Results could differ dramatically for other asset types.

One final point worth mentioning is the poor performance of empirical quantiles based on daily data. In most cases, this approach is not included in a Model Confidence Set for any of the assets analysed. This result accords with the literature, as discussed in Section 3.4. Nonetheless, it is something of a concern, given the popularity of this model in industry. On the basis of the present study, application of this model appears entirely irrational[85] given the excellent performance of RiskMetrics, which is just as easy to implement.

3.7 Conclusion

In this chapter, I perform the first large-scale evaluation of VaR forecast models using the BRM Quantile estimator as a proxy for true VaR in combination with tests from the loss-based forecast evaluation literature. As discussed in Chapter 2, this approach exhibits greater power to distinguish between competing VaR forecasts than other comparable methods. The approach is used to address the question of whether there is any advantage to using VaR forecast models that utilize intraday data.

I find strong evidence in favour of simple time series models for intraday data-based estimators. In particular, time series models for the BRM Quantile Estimator yielded particularly accurate forecasts, while time series models for realised kernels, combined with an inverse Gaussian transform to obtain VaR, were also good performers. In particular, these methods appear to strongly dominate daily data-based forecasts across all forecast intervals, both quantiles, and for both the New York Stock Exchange and the Australian Stock Exchange data.

It is worth emphasizing that this chapter contains the first application of time series models to the BRM Quantile Estimator. For this initial, exploratory study, the analysis was deliberately constrained to simple time series models. However, the strong results for this approach suggest that future work could look at the development of more sophisticated time series models for this estimator. A paper devoted entirely to the time series properties of the BRM Quantile Estimator would likely prove a valuable addition to the literature.

Appendix 3.A The Model Confidence Set

In this appendix the Model Confidence Set (MCS) of Hansen et al. (2011) is described. I begin the discussion by noting that the use of the MCS within the framework described in Patton & Li (2013) is theoretically justified.[86] The procedure itself works as follows.

Let $\mathcal{M}^0$ denote the set of all forecast models, $k = 1, ..., K$, and let $\mathcal{M} \subseteq \mathcal{M}^0$. Let $\mathcal{M}^*$ denote the MCS, that is, the set of all forecast models with minimum expected

[85] Of course, forecast accuracy may not be the primary incentive in industry.
[86] See Section 4.3 of Patton & Li (2013).

forecast error. This set may only contain one element. The MCS algorithm is a step-wise procedure that consists of a hypothesis test, followed by an elimination rule. Initially, set $\mathcal{M} = \mathcal{M}^0$. At each step, test the null hypothesis that all models in $\mathcal{M}$ are equivalent. If the null hypothesis is rejected, then apply the elimination rule to remove the worst performing models from $\mathcal{M}$. This procedure is repeated until the null hypothesis of equivalence is accepted, at which point denote $\mathcal{M}$ as $\hat{\mathcal{M}}^*$, i.e. an estimator of the MCS. For a given confidence level α, Hansen et al. (2011) show that asymptotically, $\mathbb{P}(\mathcal{M}^* \subset \hat{\mathcal{M}}^*) \geq 1 - \alpha$, and further, that asymptotically, $\mathbb{P}(k \in \hat{M}^*) = 0$, $\forall k \notin \mathcal{M}^*$, where k denotes a forecast model. In words, we can reasonably expect the MCS to contain the true set of best models with a given confidence level and further, we can reasonably expect it not to contain inferior models. It is worth adding that in the ideal case where the MCS contains only one model, Hansen et al. (2011) show that asymptotically $\mathbb{P}(\mathcal{M}^* = \hat{\mathcal{M}}^*) = 1$.

A strength of the MCS algorithm is that if the equivalence test lacks power, then the algorithm will terminate quite early, and the resulting MCS will contain many models. If, on the other hand, the data contain more information, then the equivalence test will have greater power, and the resulting MCS will contain a smaller number of models. In practice, this means that the power of the testing procedure may be evaluated by counting the number of models in the MCS. In the present analysis, this is used to demonstrate that greater power to distinguish between competing models obtains when one employs the Mean Absolute Error loss function in place of the Mean Square Error, and also when one examines data from outside the global financial crisis period.

A second strength of the MCS algorithm is that it controls for the pairwise error rate associated with testing more than two forecast models. In other words, it implicitly controls for K, i.e. the size of $\mathcal{M}^0$. For example, if $\mathcal{M}^0$ contains a large number of purely random processes, none of them are likely to be singled out by the MCS algorithm as the best model. In contrast, if one performs a large number of bivariate comparisons in a Diebold & Mariano (1995) type framework, then one is likely to find an independent random process with strong apparent forecasting ability, purely by chance.

A practical description of the equivalence test and elimination rule can be found in Hansen et al. (2011). In particular, in the present analysis, I employ the equivalence test and elimination rule constructed from t-statistics,[87] with relevant confidence bounds estimated using the stationary bootstrap of Politis & Romano (1994*b*). As discussed in Hansen et al. (2011), sufficient conditions justifying this procedure are that $\tilde{d}_t$ is a strictly stationary strong mixing process of size $-r/(r-2)$, with non-zero variance and $\mathbb{E}|\tilde{d}_t|^{r+\gamma} < \infty$, $r > 2$, $\gamma > 0$, where $\tilde{d}_t$ denotes the proxy loss differen-

[87]See Section 3.2.1 of Hansen et al. (2011).

tial,[88] although it would not be unreasonable to conjecture that these conditions can be weakened to those discussed in Goncalves & de Jong (2003)[89]

Appendix 3.B Additional Results

In this appendix I provide more comprehensive versions of the tables in Section 3.6. For each table, I display the smaller of:

- the top 14 forecast models, and
- all models that appear in the Model Confidence Set with a proportion greater than 0.2 for at least one of the time periods .

These tables contain results for the MAE loss function. The MSE loss function results were very similar, albeit with slightly less power, and so are omitted.

[88] It is defined precisely in Section 3.2.

[89] A heterogeneous process exhibiting L_2 near epoch dependence on a strong mixing base and a moment bound of $r + \delta$, $r > 2$, $\delta > 0$.

Table 3.4: Members of Model Confidence Set: All Models - New York Stock Exchange

0.01 quantile	2008-2010	2011-2013	Average
ES($\beta = 0.5$) on BRMQE	0.78	0.78	0.78
STES:MSE(Opt) on BRMQE	0.66	0.60	0.63
ES(Opt) on BRMQE	0.46	0.64	0.55
EWMA(Opt)($\tau = 10$) on BRMQE	0.46	0.60	0.53
STES:MAE(Opt) on BRMQE	0.48	0.58	0.53
ES($\beta = 0.75$) on BRMQE	0.34	0.54	0.44
EWMA($\beta = 0.75$)($\tau = 10$) on BRMQE	0.34	0.30	0.32
EWMA($\beta = 0.5$)($\tau = 10$) on BRMQE	0.56	0	0.28
EWMA($\beta = 0.5$)($\tau = 5$) on BRMQE	0.46	0	0.23
ES($\beta = 0.25$) on BRMQE	0.38	0	0.19
EWMA(Opt)($\tau = 5$) on BRMQE	0.38	0	0.19
ES($\beta = 0.5$) on RK, Normal	0.32	0	0.16
STES:MAE(Opt) on RK, Normal	0.30	0	0.15
EWMA($\beta = 0.75$)($\tau = 5$) on BRMQE	0.24	0	0.12
0.05 quantile	**2008-2010**	**2011-2013**	**Average**
ES($\beta = 0.5$) on BRMQE	0.84	0.90	0.87
STES:MSE(Opt) on BRMQE	0.72	0.64	0.68
STES:MAE(Opt) on BRMQE	0.66	0.60	0.63
ES(Opt) on BRMQE	0.54	0.66	0.60
EWMA(Opt)($\tau = 10$) on BRMQE	0.60	0.60	0.60
ES($\beta = 0.75$) on BRMQE	0.32	0.50	0.41
EWMA($\beta = 0.5$)($\tau = 10$) on BRMQE	0.72	0	0.36
EWMA($\beta = 0.75$)($\tau = 10$) on BRMQE	0.28	0.40	0.34
EWMA($\beta = 0.5$)($\tau = 5$) on BRMQE	0.60	0	0.30
ES($\beta = 0.25$) on BRMQE	0.44	0	0.22
EWMA(Opt)($\tau = 5$) on BRMQE	0.36	0	0.18
STES:MAE(Opt) on RK, Normal	0.30	0	0.15
ES($\beta = 0.5$) on RK, Normal	0.28	0	0.14
ES(Opt) on RK, Normal	0.22	0	0.11

BRMQE = Bootstrap Return Method Quantile Estimator, RK = realised kernels, ES = Exponential Smoothing, EWMA = Exponentially Weighted Moving Average, STES:MSE[MAE] = Smooth Transition Exponential Smoothing with Mean Square Error Adjustment [Mean Absolute Error Adjustment], Opt = model parameters estimated optimally, Normal = volatility forecast transformed to VaR with inverse Gaussian cdf

Table 3.5: Members of Model Confidence Set: No Bootstrap Return Method Quantile Estimator - New York Stock Exchange

0.01 quantile	2008-2010	2011-2013	Average
ES($\beta = 0.5$) on RK, Normal	0.76	0.56	0.66
ES($\beta = 0.75$) on RK, Normal	0.30	0.72	0.51
STES:MAE(Opt) on RK, Normal	0.58	0.36	0.47
STES:MSE(Opt) on RK, Normal	0.46	0.38	0.42
ES(Opt) on RK, Normal	0.42	0.38	0.40
EWMA(Opt)($\tau = 10$) on RK, Normal	0.42	0.36	0.39
ES($\beta = 0.75$) on RV5Min, Normal	0.24	0.42	0.33
EWMA($\beta = 0.75$)($\tau = 10$) on RK, Normal	0.22	0.36	0.29
STES:MAE(Opt) on RV5Min, Normal	0.30	0.22	0.26
EWMA($\beta = 0.5$)($\tau = 10$) on RK, Normal	0.26	0	0.13
STES:MSE(Opt) on RV5Min, Normal	0.24	0	0.12
ES(Opt) on RV5Min, Normal	0.22	0	0.11
EWMA($\beta = 0.75$)($\tau = 10$) on RV5Min, Normal	0.22	0	0.11
ES($\beta = 0.9$) on RK, Normal	0	0.22	0.11
0.05 quantile	**2008-2010**	**2011-2013**	**Average**
ES($\beta = 0.5$) on RK, Normal	0.72	0.48	0.60
ES($\beta = 0.75$) on RK, Normal	0.34	0.68	0.51
STES:MAE(Opt) on RK, Normal	0.68	0.32	0.50
EWMA(Opt)($\tau = 10$) on RK, Normal	0.52	0.34	0.43
STES:MSE(Opt) on RK, Normal	0.52	0.34	0.43
ES(Opt) on RK, Normal	0.52	0.30	0.41
EWMA($\beta = 0.75$)($\tau = 10$) on RK, Normal	0.30	0.40	0.35
ES($\beta = 0.75$) on RV5Min, Normal	0	0.42	0.21
EWMA($\beta = 0.5$)($\tau = 10$) on RK, Normal	0.34	0	0.17
STES:MAE(Opt) on RV5Min, Normal	0.30	0	0.15
ES($\beta = 0.5$) on RV5Min, Normal	0.26	0	0.13
ES($\beta = 0.25$) on RK, Normal	0.26	0	0.13
EWMA($\beta = 0.5$)($\tau = 5$) on RK, Normal	0.26	0	0.13
EWMA(Opt)($\tau = 5$) on RK, Normal	0.26	0	0.13

RK = realised kernels, RV5Min = 5-minute realised variance, ES = Exponential Smoothing, EWMA = Exponentially Weighted Moving Average, STES:MSE[MAE] = Smooth Transition Exponential Smoothing with Mean Square Error Adjustment [Mean Absolute Error Adjustment], Opt = model parameters estimated optimally, Normal = volatility forecast transformed to VaR with inverse Gaussian cdf

Table 3.6: Members of Model Confidence Set: No Intraday Models - New York Stock Exchange

0.01 quantile	2008-2010	2011-2013	Average
RiskMetrics	0.76	0.82	0.79
IAVARCH:Normal, Parametric	0.90	0.68	0.79
IGARCH:Normal, Parametric	0.88	0.52	0.70
GJRGARCH:Normal, Parametric	0.68	0.70	0.69
GARCH:Normal, Parametric	0.68	0.68	0.68
NAGARCH:Normal, Parametric	0.64	0.68	0.66
TARCH11:Normal, Parametric	0.66	0.60	0.63
AGARCH:Normal, Parametric	0.64	0.58	0.61
APARCH:Normal, Parametric	0.58	0.64	0.61
ES($\beta = 0.9$) on SqRet, Normal	0.76	0.42	0.59
TARCH111:Normal, Parametric	0.54	0.60	0.57
MA($\tau = 50$) on SqRet, Normal	0.22	0.76	0.49
IAVARCH:Skewed-T, Parametric	0.74	0	0.37
NAGARCH:Skewed-T, Parametric	0.68	0	0.34
0.05 quantile	**2008-2010**	**2011-2013**	**Average**
RiskMetrics	0.72	0.60	0.66
IAVARCH:Skewed-T, Parametric	0.84	0.38	0.61
IAVARCH:Normal, Parametric	0.86	0.34	0.60
IGARCH:Skewed-T, Parametric	0.74	0.40	0.57
GJRGARCH:Normal, Parametric	0.62	0.44	0.53
AGARCH:Skewed-T, Parametric	0.62	0.44	0.53
NAGARCH:Normal, Parametric	0.64	0.42	0.53
IGARCH:Normal, Parametric	0.78	0.28	0.53
GJRGARCH:Skewed-T, Parametric	0.54	0.50	0.52
TARCH11:Student-T, Filtered	0.46	0.58	0.52
AGARCH:Normal, Parametric	0.64	0.40	0.52
AGARCH:Student-T, Filtered	0.42	0.62	0.52
IAVARCH:Skewed-T, Filtered	0.50	0.54	0.52
ES($\beta = 0.9$) on SqRet, Normal	0.70	0.30	0.50

SqRet = Squared Returns, ES = Exponential Smoothing, MA = Moving Average, Normal = volatility forecast transformed to VaR with inverse Gaussian cdf, Parametric = volatility forecast transformed to VaR using parametric assumption from model estimation. Note, RiskMetrics uses inverse Gaussian cdf, Filtered = volatility forecast transformed to VaR using filtering method

Table 3.7: Members of Model Confidence Set: All Models - Australian Stock Exchange

0.01 quantile	2008-2010	2011-2013	Average
STES:MSE(Opt) on BRMQE	0.88	0.84	0.86
ES($\beta = 0.75$) on BRMQE	0.90	0.80	0.85
ES(Opt) on BRMQE	0.78	0.82	0.80
STES:MAE(Opt) on BRMQE	0.74	0.86	0.80
ES($\beta = 0.75$) on RK, Normal	0.72	0.66	0.69
EWMA(Opt)($\tau = 10$) on BRMQE	0.70	0.56	0.63
EWMA($\beta = 0.75$)($\tau = 10$) on BRMQE	0.70	0.54	0.62
ES(Opt) on RK, Normal	0.44	0.68	0.56
STES:MAE(Opt) on RK, Normal	0.34	0.76	0.55
STES:MSE(Opt) on RK, Normal	0.28	0.70	0.49
ES($\beta = 0.9$) on RK, Normal	0.24	0.56	0.40
EWMA($\beta = 0.5$)($\tau = 10$) on RK, Normal	0.42	0.38	0.40
ES($\beta = 0.9$) on BRMQE	0.28	0.48	0.38
EWMA(Opt)($\tau = 10$) on RK, Normal	0.36	0.36	0.36
0.05 quantile	2008-2010	2011-2013	Average
STES:MAE(Opt) on BRMQE	0.86	0.96	0.91
STES:MSE(Opt) on BRMQE	0.88	0.90	0.89
ES($\beta = 0.75$) on BRMQE	0.92	0.82	0.87
ES(Opt) on BRMQE	0.82	0.84	0.83
EWMA($\beta = 0.75$)($\tau = 10$) on BRMQE	0.76	0.52	0.64
ES($\beta = 0.75$) on RK, Normal	0.64	0.60	0.62
EWMA(Opt)($\tau = 10$) on BRMQE	0.70	0.48	0.59
ES(Opt) on RK, Normal	0.38	0.54	0.46
STES:MAE(Opt) on RK, Normal	0.28	0.60	0.44
STES:MSE(Opt) on RK, Normal	0.24	0.56	0.40
EWMA($\beta = 0.75$)($\tau = 10$) on RK, Normal	0.34	0.34	0.34
ES($\beta = 0.9$) on BRMQE	0.22	0.44	0.33
EWMA(Opt)($\tau = 10$) on RK, Normal	0.32	0.28	0.30
ES($\beta = 0.9$) on RK, Normal	0.22	0.36	0.29

BRMQE = Bootstrap Return Method Quantile Estimator, RK = realised kernels, ES = Exponential Smoothing, EWMA = Exponentially Weighted Moving Average, STES:MSE[MAE] = Smooth Transition Exponential Smoothing with Mean Square Error Adjustment [Mean Absolute Error Adjustment], Opt = model parameters estimated optimally, Normal = volatility forecast transformed to VaR with inverse Gaussian cdf

Table 3.8: Members of Model Confidence Set: No Bootstrap Return Method Quantile Estimator - Australian Stock Exchange

0.01 quantile	2008-2010	2011-2013	Average
ES($\beta = 0.75$) on RK, Normal	0.88	0.86	0.87
ES(Opt) on RK, Normal	0.60	0.78	0.69
STES:MAE(Opt) on RK, Normal	0.44	0.82	0.63
STES:MSE(Opt) on RK, Normal	0.36	0.76	0.56
EWMA($\beta = 0.75$)($\tau = 10$) on RK, Normal	0.54	0.54	0.54
ES($\beta = 0.9$) on RK, Normal	0.36	0.60	0.48
EWMA(Opt)($\tau = 10$) on RK, Normal	0.48	0.48	0.48
ES($\beta = 0.5$) on RK, Normal	0.30	0.32	0.31
EWMA($\beta = 0.9$)($\tau = 10$) on RK, Normal	0.22	0.36	0.29
ES($\beta = 0.75$) on RV5Min, Normal	0.30	0	0.15
STES:MAE(Opt) on RV5Min, Normal	0	0.26	0.13
EWMA($\beta = 0.75$)($\tau = 10$) on RV5Min, Normal	0.24	0	0.12
STES:MSE(Opt) on RV5Min, Normal	0	0.22	0.11
0.05 quantile	**2008-2010**	**2011-2013**	**Average**
ES($\beta = 0.75$) on RK, Normal	0.92	0.84	0.88
ES(Opt) on RK, Normal	0.62	0.80	0.71
STES:MAE(Opt) on RK, Normal	0.48	0.84	0.66
STES:MSE(Opt) on RK, Normal	0.34	0.78	0.56
EWMA($\beta = 0.75$)($\tau = 10$) on RK, Normal	0.52	0.44	0.48
EWMA(Opt)($\tau = 10$) on RK, Normal	0.48	0.48	0.48
ES($\beta = 0.9$) on RK, Normal	0.32	0.56	0.44
ES($\beta = 0.5$) on RK, Normal	0.24	0.26	0.25
EWMA($\beta = 0.9$)($\tau = 10$) on RK, Normal	0.24	0.24	0.24
ES($\beta = 0.75$) on RV5Min, Normal	0.34	0	0.17
EWMA($\beta = 0.75$)($\tau = 10$) on RV5Min, Normal	0.32	0	0.16
EWMA($\beta = 0.75$)($\tau = 5$) on RK, Normal	0.28	0	0.14
STES:MAE(Opt) on RV5Min, Normal	0	0.24	0.12

RK = realised kernels, RV5Min = 5-minute realised variance, ES = Exponential Smoothing, EWMA = Exponentially Weighted Moving Average, STES:MSE[MAE] = Smooth Transition Exponential Smoothing with Mean Square Error Adjustment [Mean Absolute Error Adjustment], Opt = model parameters estimated optimally, Normal = volatility forecast transformed to VaR with inverse Gaussian cdf

Table 3.9: Members of Model Confidence Set: No Intraday Models - Australian Stock Exchange

0.01 quantile	2008-2010	2011-2013	Average
IAVARCH:Normal, Parametric	0.66	0.90	0.78
IAVARCH:Skewed-T, Parametric	0.72	0.80	0.76
RiskMetrics	0.82	0.58	0.70
IGARCH:Normal, Parametric	0.64	0.72	0.68
TARCH11:Normal, Parametric	0.38	0.90	0.64
TARCH11:Skewed-T, Parametric	0.46	0.82	0.64
IGARCH:Skewed-T, Parametric	0.62	0.66	0.64
GARCH:Normal, Parametric	0.40	0.82	0.61
TARCH111:Skewed-T, Parametric	0.44	0.78	0.61
GJRGARCH:Normal, Parametric	0.44	0.76	0.60
TARCH111:Normal, Parametric	0.38	0.76	0.57
GARCH:Skewed-T, Parametric	0.38	0.74	0.56
AGARCH:Normal, Parametric	0.36	0.74	0.55
AGARCH:Skewed-T, Parametric	0.34	0.74	0.54
0.05 quantile	2008-2010	2011-2013	Average
IAVARCH:Normal, Parametric	0.66	0.84	0.75
RiskMetrics	0.82	0.54	0.68
IGARCH:Normal, Parametric	0.74	0.58	0.66
TARCH11:Normal, Parametric	0.44	0.86	0.65
IGARCH:Skewed-T, Parametric	0.64	0.64	0.64
IAVARCH:Skewed-T, Parametric	0.56	0.72	0.64
TARCH111:Normal, Parametric	0.36	0.80	0.58
IAVARCH:Student-T, Parametric	0.68	0.48	0.58
IAVARCH:Student-T, Filtered	0.60	0.54	0.57
IAVARCH:Skewed-T, Filtered	0.60	0.52	0.56
CAViaR(Indirect GARCH)	0.52	0.60	0.56
GARCH:Normal, Parametric	0.38	0.72	0.55
CAViaR(Symmetric Absolute Value)	0.58	0.50	0.54
GARCH:Skewed-T, Parametric	0.34	0.72	0.53

Parametric = volatility forecast transformed to VaR using parametric assumption from model estimation. Note, RiskMetrics uses inverse Gaussian cdf, Filtered = volatility forecast transformed to VaR using filtering method

Chapter 4

Ranking Intraday Volatility Estimators Using Empirical Criteria

4.1 Introduction

This chapter is concerned with the problem of ranking intraday volatility estimators purely via empirical criteria. Volatility estimators based on intraday data have proven very useful in the literature in recent times. For example, Andersen & Bollerslev (1998) famously use the realised variance estimator[90] to improve statistical methods for evaluating volatility forecast models, while Andersen et al. (2003) incorporate realised variance into the forecast itself. Since then, the literature has seen a proliferation of methods for estimating daily variance, or in a continuous time framework, quadratic variation, from intraday data.[91]

Despite the fact that choosing the most accurate from among these estimators would be useful in both a forecasting and forecast evaluation context, the topic has received surprisingly little attention over the past decade. Typically, attention has focused on theoretical optimality[92], rather than empirical optimality, but this is of limited use when modelling assumptions are open to debate. Further, most of the analysis that does exist is limited to the theoretically optimal sampling frequency of realised variance.

A notable exception is Patton (2011*a*) (with an extension in Patton & Sheppard (2009)), who states a set of sufficient conditions under which several tests from the

[90]First proposed in Merton (1980).

[91] A short list of candidates includes realised variance, see Merton (1980) and Barndorff-Nielsen & Shephard (2002*a*); the estimator of Zhou (1996); the two-scale realised variance of Zhang et al. (2005); the multi-scale realised variance of Zhang (2006); the realised range estimator of Christensen & Podolskij (2007); the realised kernels of Barndorff-Nielsen et al. (2008*a*); and the pre-averaging approach of Jacod et al. (2009).

[92]See Bandi & Russell (2008) and Barndorff-Nielsen et al. (2008*a*).

extensive literature on volatility *forecast* evaluation can be used directly to evaluate volatility *estimators*, with one potential application being a purely data-driven approach to selecting the optimal sampling frequency for realised variance. In order to accomplish this, Patton assumes a specific parametric model for the true dynamics of the volatility process, and describes it as an "initial approximation". Specifically, Patton (2011*a*) considers two models for volatility: a random walk, and a stationary AR(p) process for which the order of p is assumed to be known. He notes that empirical evidence in support of these parametric models is mixed. Thus, an approach that allows for a wider variety of volatility processes is desirable.

In essence, the contribution of this chapter is to propose a trade-off in modelling assumptions. Patton (2011*a*) makes strict assumptions regarding the true volatility process, but broad assumptions regarding the choice of loss function. In this chapter, the opposite is true: the choice of loss function is restricted, but the range of allowable volatility processes is broad. This trade-off is desirable as the true volatility process is unobservable, and so verifying strict assumptions on this process is difficult. In contrast, the choice of a loss function is transparent, and the restriction in this chapter is a popular one.

Specifically, I assume that the volatility process belongs to a particular class of near epoch dependent processes. It is worth emphasising that the class of near epoch dependent processes includes the stationary AR(p) model assumed by Patton (2011*a*). The class also includes a wide range of weakly dependent processes, including many popular conditional volatility and stochastic volatility models, e.g. autoregressive moving average, GARCH variants, log-normal stochastic volatility, and autoregressive conditional duration, given appropriate parameter restrictions.[93]

The cost of this generality is the restriction of attention to the popular Mean Squared Error (MSE) loss function. In many cases, this cost is worth bearing. The MSE is a popular choice of loss function in practical work. Furthermore, it is a member of the set of robust loss functions described in Patton (2011*b*), which is a necessary condition for the use of loss-based evaluation tests when the proxy error has non-zero variance. A particular moment restriction is also assumed. However, it is shown below that, given a reasonable true data generating process, this restriction is satisfied by any estimator conforming to a particular quadratic form. Specific examples include realised variance, two-scales realised variance, multi-scale realised variance and realised kernels.

Regarding performance, it is shown via simulation that the approach advocated in the present chapter typically outperforms that of Patton (2011*a*) with a stationary AR(1) assumption and MSE loss function. In cases where the true variance strongly violates a random walk assumption, the approach advocated in this chapter also

[93]See Hansen (1991), Engle & Russell (1998), Carrasco & Chen (2002), Davidson (2002), and Davidson (2004).

outperforms the approach of Patton with a random walk assumption and MSE loss function. Interestingly, the simulations also suggest that the size and power of the approach advocated in this chapter is approximately invariant to the variance of the proxy error. This is important, because all methods discussed in this paper, including those of Patton (2011*a*), require an unbiased proxy, and the most reliable candidate in this regard is daily squared returns, which are particularly noisy. Importantly, the size and power of the methods proposed in Patton (2011*a*) deteriorate as the variance of the proxy error increases.

Empirically, the restriction to the MSE loss function can be problematic for time-intervals that include large spikes in volatility, since a small number of days with large volatility will dominate the analysis.[94] Despite this, the approach advocated in the present chapter generates results that match *a priori* expectations.

A final point worth mentioning is that the proofs in Appendix 4.A include a technical lemma on the product of near epoch dependent processes which, to the best of my knowledge, is new in the literature and of general application.

The remainder of this chapter is arranged as follows: Section 4.2 defines a set of common notation, while Section 4.3 examines the prior literature. Section 4.4 provides the main results. The theory is analysed in Section 4.5 using several simulations. Section 4.6 contains a short empirical example, then Section 4.7 concludes. All proofs are contained in Appendix 4.A.

4.2 Notation

The notation in this chapter follows Patton (2011*a*) and Patton & Sheppard (2009), and differs from the rest of the thesis. Let θ_t denote the latent parameter of interest; in the present application, this is the daily variance of a risky asset. A set of estimators for θ_t are denoted $x_{j,t}$, $j = 1, ..., J$. An estimator $x_{j,t}$ can be decomposed into latent variance and estimation error, i.e. $x_{j,t} = \theta_t + u_{j,t}$. A proxy for θ_t is denoted $\tilde{\theta}_t$, and it can similarly be decomposed into latent variance and proxy error, i.e. $\tilde{\theta}_t = \theta_t + \tilde{u}_t$. All random variables are transformations from the underlying space of events Ω, and I use $\mathcal{F}_{t-1}^{t}$ to denote the σ-algebra generated by random variables from the span $[t-1, t]$.

The accuracy of $x_{j,t}$ for θ_t is analysed using a distance measure known as a loss function, denoted $L(\cdot, \cdot)$. The main problem investigated in this chapter is the estimation of the expected loss differential, defined:

$$\mathbb{E}\Delta L(\theta_t, \mathbf{x}_t) \equiv \mathbb{E}L(\theta_t, x_{i,t}) - \mathbb{E}L(\theta_t, x_{j,t}), \forall i \neq j, \tag{4.1}$$

and the subsequent statistical testing of null hypotheses based on the expected loss

[94] QLIKE generally works better in these types of situations.

differential. Since the assumption set allows $\mathbb{E}\Delta L(\theta_t, \mathbf{x}_t)$ to be dependent on t, the actual parameter of interest is:

$$\gamma \equiv T^{-1} \sum_{t=1}^{T} \mathbb{E}\Delta L(\theta_t, \mathbf{x}_t). \tag{4.2}$$

For ease of notation, the sample mean of any arbitrary random variable a_t is denoted $\overline{a}$ while the sample variance is denoted $\hat{\mathbb{V}}a$. The sample covariance of any two arbitrary random variables a_t and b_t is denoted $\widehat{\text{cov}}\,(a, b)$.

4.3 Literature

Empirical evaluation of volatility *forecast* models is a popular topic in financial econometrics. It is a more difficult problem than standard forecast evaluation, as the variable of interest, θ_t, is latent. The typical approach is to replace the latent variance with a conditionally unbiased proxy, denoted $\tilde{\theta}_t$. Popular choices include squared returns or low sampling frequency realised variance.[95]

Replacing θ_t with $\tilde{\theta}_t$ is a reasonable approach if the proxy error, $\tilde{u}_t$, is mean-zero and independent of the forecast error, as the effect of $\tilde{u}_t$ will vanish given a suitable large number law. Specifically, in a conditional volatility framework, Hansen & Lunde (2006*b*) and Patton (2011*b*) show that model rankings obtained using $\tilde{\theta}_t$ will be asymptotically consistent with the model rankings obtained using θ_t, assuming the loss function belongs to a class of "robust and homogeneous" loss functions.[96] Intuitively, this works since $\tilde{\theta}_t \in \mathcal{F}_{t-1}^{t}$, while the forecasts belong to $\mathcal{F}_{t-2}^{t-1}$, so the proxy error and appropriately transformed forecast errors can be treated as uncorrelated (given certain regularity conditions).

However, in a volatility estimation framework, $\tilde{\theta}_t \in \mathcal{F}_{t-1}^{t}$ and $x_{j,t} \in \mathcal{F}_{t-1}^{t}$, so the proxy error and estimation errors are likely to be contemporaneously dependent, particularly for intraday variance estimators (of which there are many).[97] The implication of this is that model (or estimator) rankings obtained using the proxy may be inconsistent with model rankings obtained using the true variance.[98]

Patton (2011*a*) addresses this issue by stating a set of sufficient conditions under which the model rankings will be preserved if $\tilde{\theta}_t \in \mathcal{F}_{t}^{\infty}$. The simplest way to accomplish this is to set $\tilde{\theta}_t = y_{t+1}$, where y_t denotes a proxy such as squared returns or low sampling frequency realised variance.[99] Assuming $y_{t+1} - \theta_{t+1}$ is a martingale difference sequence and θ_t is a random walk, this approach ensures the proxy error

[95] See Poon & Granger (2003) or Hansen & Lunde (2005*a*).
[96] Patton (2011*b*).
[97] See footnote 91.
[98] Patton (2011*a*) Proposition 1b.
[99] Patton (2011*a*) Assumption P2.

and an appropriate transformation of the estimation errors will be uncorrelated. This in turn justifies the use of the numerous popular statistical tests from the volatility forecasting literature.[100] Alternatively, Patton also allows for θ_t to follow an AR(p) process, although this introduces new parameters which must be estimated.

In my opinion, there are two drawbacks to Patton's approach. First, the assumption of a particular model for true variance is, at best, an approximation. If true variance follows a unit root process, e.g. RiskMetrics, then the random walk approximation could be quite good. Interestingly, Hansen & Lunde (2010) fail to reject the null hypothesis of a unit root for many of the Dow Jones 30 assets. On the other hand, Wright (1999) strongly rejects non-stationarity in the volatility of stock returns and exchange rates, implying that the empiricist may be forced to use the AR(p) approximation.

Second, even if the random walk assumption is correct, by setting $\tilde{\theta}_t = y_{t+1}$, an additional source of error is introduced into the proxy, i.e. $y_{t+1} - y_t$. If y_t is itself quite volatile, this additional error will be large. In finite samples, this will impact on the efficiency of any statistical procedure.

In the present chapter I set $\tilde{\theta}_t = y_t$, which eliminates the additional source of error, and avoids the need to assume a specific time series model for θ_t. Of course, this also introduces the problem of contemporaneous dependence between the proxy error and the estimation errors. However, by restricting attention to the MSE loss function, I show that it is possible to estimate the inevitable contemporaneous dependence between $\tilde{u}_t$ and $u_{j,t}$ using two additional covariance parameters and a moment restriction, and so retrieve a consistent and asymptotically Normal estimator for the true expected loss differential. Furthermore, this approach does not require $\tilde{u}_t$ to be a martingale difference sequence. The process is described in detail in the next section.

4.4 Modelling Assumptions and Theory

In this section I derive an expression for the true expected loss differential, provide an estimator for this expression, and then show that it obeys a central limit theorem and a bootstrap central limit theorem. To accomplish this, I provide a technical lemma on the product of near epoch dependent processes (contained in Appendix 4.A) which, to the best of my knowledge, is new in the literature and of general application.

The first proposition provides an expression for the true expected loss differential in terms of population moments of observable random variables, given arbitrary dependence between the proxy error and estimation error. The proposition employs the following assumptions:

[100] See, for example, Diebold & Mariano (1995), West (1996), White (2000), Hansen (2005), Romano & Wolf (2005) and Hansen et al. (2011).

Assumptions 4.1

1. *$L(x, y) = (x - y)^2$, i.e. $L(\cdot, \cdot)$ is the MSE loss function,*

2. *$\mathbb{E}\tilde{u}_t = 0$, i.e. the proxy is unconditionally unbiased, and*

3. *$cov(\theta_t, u_{j,t}) - cov(\theta_t, u_{i,t}) = 0, \forall i, j$.*

Remark 4.1 Assumption 4.1.1 states that $L(\cdot, \cdot)$ denotes the MSE loss function, which is one of the most popular distance measures in use. It is also a member of the set of homogeneous, robust loss functions discussed in Patton (2011*b*), and further, is the only symmetric loss function that is a member of this set. Proposition 4.4.1 will demonstrate that restricting attention to this loss function allows the dependence between the proxy error and estimation error to be captured by two covariance terms.

Remark 4.2 Assumptions 4.1.2 and 4.1.3, respectively, allow the proxy error to be averaged out via a law of large numbers, and eliminate some nuisance parameters. These assumptions are somewhat analogous to Assumption P1 in Patton (2011*a*), which requires $\mathbb{E}[\tilde{u}_t | \theta_t, \mathcal{F}_{t-1}] = 0$. Assumption 4.1.2 is strictly weaker than P1, both because conditional unbiasedness is a stronger assumption than unconditional unbiasedness, and because I do not assume the proxy error is a martingale difference sequence. Assumption 4.1.3 is weaker than P1 in some respects, but stronger in others. Using the law of iterated expectations it is simple to demonstrate that P1 necessarily implies that $\text{cov}(\tilde{u}_t, \theta_t) = 0$. In contrast, Assumption 4.1.3 places no restriction on the covariance between the proxy error and latent variance, but does restrict the covariance between the estimation error and latent variance.

Remark 4.3 As is made clear in the proofs provided in Appendix 4.A,[101] Assumption 4.1.3 is necessary for identification in the present framework, so a more detailed consideration of this assumption is useful.

Let $q_{n,t} = \sqrt{\nu_{n,t}} z_{n,t}$, $n = 1, ..., N_t$, define a sequence of intraperiod returns, where $\sum_n \nu_{n,t} = \theta_t$ and $z_{n,t} \overset{iid}{\backsim} (0, 1)$, with $\mathbb{E}[z_{n,t} | \nu_{m,t}] = 0$, $\forall n, m$, and $\mathbb{E}[z_{n,t}^2 | \nu_{m,t}] = 1$, $\forall n, m$. Let $x_{j,t}$ denote an estimator that can be expressed as a quadratic form of intraperiod returns, i.e.

$$x_{j,t} = \sum_{n=1}^{N_t} \sum_{m=1}^{N_t} a_{n,m} q_{n,t} q_{m,t}, \tag{4.3}$$

where $a_{n,m}$ is the (n, m) element of a non-stochastic weighting matrix. Sun (2006) and Andersen, Bollerslev & Meddahi (2011) demonstrate that a large number of estimators can be expressed in this form, including realised variance (of any sampling frequency), two-scales realised variance, multi-scale realised variance and realised

[101] Equation (4.26).

kernels. By definition, $u_{j,t} = x_{j,t} - \theta_t$, so (4.3) implies that:

$$\operatorname{cov}(\theta_t, u_{j,t}) = \sum_{n=1}^{N_t} \sum_{m=1}^{N_t} \operatorname{cov}\left(\theta_t, a_{m,n} \sqrt{\nu_{n,t} \nu_{m,t}} z_{n,t} z_{m,t}\right) - \operatorname{cov}(\theta_t, \theta_t). \tag{4.4}$$

The stated assumptions are sufficient for the double sum covariance term to vanish whenever $n \neq m$, and to reduce to $\operatorname{cov}(\theta_t, \sum_n a_{n,n} \nu_n)$ whenever $n = m$. Thus (4.4) will reduce to zero if $a_{n,n} = 1$, $\forall n$. It turns out that this condition is satisfied for almost all quadratic form intraday variance estimators, including realised variance (of any sampling frequency), Subsampled realised variance, two-scales realised variance, and realised kernels.[102] Thus we might reasonably expect Assumption 4.1.3 to be satisfied for most realised variance-type estimators.

Next, I derive an expression for the true expected loss differential in terms of observable random variables. Note, to keep the notation simple, $\mathbf{x}_t$ is assumed to store two estimators: $x_{j,t}$ and $x_{i,t}$. This does not imply any loss in generality.

Proposition 4.4.1 *Assume 4.1.1 to 4.1.3. Then:*

$$\mathbb{E}\Delta L(\theta_t, \mathbf{x}_t) = \mathbb{E}\Delta L(\tilde{\theta}_t, \mathbf{x}_t) + 2\left(\operatorname{cov}(x_{j,t}, \tilde{\theta}_t) - \operatorname{cov}(x_{i,t}, \tilde{\theta}_t)\right). \tag{4.5}$$

Proof See Appendix 4.A.

The next step is to construct an estimator based on Proposition 4.4.1 and demonstrate a CLT. Consider the additional assumptions:

Assumptions 4.2

$\exists \delta > 0$, $r > 2$ such that:

1. *$u_{j,t}$ and $\tilde{u}_t$ are $L_{\frac{(2+\delta)(2+2\delta)}{\delta}}$-NED (near epoch dependent) of size -1 on a strong mixing process of size $-(2+\delta)(r+\delta)/(r-2)$, and*

2. *θ_t is $L_{2+2\delta}$-NED of size -1 on a strong mixing process of size $-(2+\delta)(r+\delta)/(r-2)$.*

Remark 4.4 Assumptions 4.2.1 and 4.2.2 provide the properties of the estimation errors, proxy error and latent variance. The assumptions allow for serial dependence and heterogeneity in these random variables, as well as arbitrary contemporaneous dependence between them. They also imply that all the random variables of interest will obey a central limit theorem, and so motivate the construction of an estimator for the expected loss differential with an associated asymptotic distribution.

[102] Andersen et al. (2011).

Remark 4.5 I am referring to the same δ and r in both assumptions 4.2.1 and 4.2.2. Thus there is a trade-off in the moment bound of $u_{j,t}$ and the moment bound of θ_t. Lemma 4.A.1 demonstrates why this trade-off is required. In practice, I envisage δ as small, so that only a low moment bound is required for θ_t. I do not see this as overly restrictive for $u_{j,t}$ as most estimators are asymptotically Normal around θ_t, so higher moments on the estimation error might reasonably be expected to exist.

Remark 4.6 The mixing size and moment bound in assumptions 4.2.1 and 4.2.2 are required for Proposition 4.4.3, but are *slightly* stronger than is required for Proposition 4.4.2. For simplicity, I only write the stronger form.

Remark 4.7 Regarding Assumption 4.2.2, note that determining that a particular process is NED typically needs to be done on a case-by-case basis. It is not my aim in this thesis to extend the literature in this regard. Instead, I refer the interested reader to Davidson (2002), where a wide variety of interesting time series models are shown to satisfy the NED condition, including Auto-regressive Moving Average processes (thus Patton's AR(p) assumption is nested by the present modelling assumptions), GARCH processes, and switching and threshold auto-regressions.[103] Also, Carrasco & Chen (2002) show that a wide variety of conditional and stochastic volatility models are stationary, geometric ergodic and β-mixing with exponential decay. These properties are all a strict subset of Assumption 4.2.2 and so all the results of that paper apply, although I emphasize that Assumption 4.2.2 is *much* more general than the conditions of that paper.

Remark 4.8 The weak dependence of the true variance process implied by 4.2.2 may not be necessary in some cases. As in Patton (2011*a*), it is only necessary that the sample mean loss differential obeys a central limit theorem, and additionally in the present paper, the difference of two covariance estimators, see Equation 4.6. These central limit theorems may still obtain, even if θ_t is $I(1)$, as long as a suitable co-integrating relationship exists. A more thorough theoretical treatment of this point is deferred to future work.

Next, consider the estimator:

$$\hat{\gamma} = \overline{\Delta L(\tilde{\theta}_t, \mathbf{x}_t)} + 2\left(\widehat{\text{cov}}(x_1, \tilde{\theta}) - \widehat{\text{cov}}(x_2, \tilde{\theta})\right). \tag{4.6}$$

Proposition 4.4.2 *Assume each sum in $\hat{\gamma}$ has non-zero variance.*[104] *Then given assumptions 4.1.1 to 4.2.2:*

$$\sqrt{T}\left(\hat{\gamma} - \gamma\right) \xrightarrow{d} \mathcal{N}(0, V), \tag{4.7}$$

[103]See Davidson (2004) for some interesting extensions, as well as Hansen (1991) for the GARCH(1,1) case.
[104]This rules out degenerate scenarios such as telescoping sum representations.

where $V < \infty$.

Proof See Appendix 4.A.

Proposition 4.4.2 provides a set of sufficient conditions under which $\hat{\gamma}$ obeys a CLT, which allows the use of $\hat{\gamma}$ in the Diebold & Mariano (1995) and West (1996) tests for bivariate comparison of two estimators. Estimating the variance of the test statistic for these two tests via analytical methods may prove problematic, so instead, the next proposition justifies the use of the stationary bootstrap of Politis & Romano (1994*b*) to estimate the standard error of $\hat{\gamma}$. A stationary bootstrap also allows the use of the Reality Check test of White (2000), the Superior Predictive Ability test of Hansen (2005), the stepwise multiple testing method of Romano & Wolf (2005), the test of Hsu, Hsu & Kuan (2010), the Model Confidence Set of Hansen et al. (2011), and the hybrid test of Song (2012).

Two additional assumptions are needed to justify the use of a stationary bootstrap:

Assumptions 4.3

1. *let X_t represent any of the random variables in the set $\{u_{j,t}, u_{j,t}^2, \tilde{u}_t, \theta_t, \theta_t u_{j,t}\}$, $\forall j$, and let $\mu_X = T^{-1} \sum_t \mathbb{E} X_t$, then $T^{-1} \sum_t (\mathbb{E} X_t - \mu_X)^2 = o(T^{-\frac{1}{2}})$, and*

2. *$p_T \in (0,1)$, $p_T \to 0$ and $T p_T^2 \to \infty$, where p_T is the parameter of the geometric distribution used to determine each block length in the stationary bootstrap of Politis & Romano (1994*b*).*

Remark 4.9 Assumption 4.3.1 is Assumption 2.2 from Goncalves & White (2002) and constrains the degree of heterogeneity of the random variables under analysis. It allows for non-stationarity, but rules out some behaviour such as first moments trending to infinity. A more detailed discussion can be found in the aforementioned paper.

Remark 4.10 Assumption 4.3.2 is a standard assumption given application of the stationary bootstrap of Politis & Romano (1994*b*), who also show that $p_T = O(T^{-\frac{1}{3}})$ is an optimal rate of convergence.

In what follows, $*$ is used to indicate a stationary bootstrap re-sample of a random variable. Then $\mathbb{P}(\sqrt{T}(\hat{\gamma}_T^* - \hat{\gamma}_T) \leq c)$ denotes the empirical stationary bootstrap cdf of $\hat{\gamma}$, conditional on the observable data, such that it is an estimate of the true cdf $\mathbb{P}(\sqrt{T}(\hat{\gamma}_T - \gamma) \leq c)$. Note, for clarity, the dependence of $\hat{\gamma}$ on the number of observations T is made explicit:

Proposition 4.4.3 *Let d denote any metric suitable for weak convergence. Given assumptions 4.1.1 to 4.3.2:*

$$d\left(\mathbb{P}^*\left(\sqrt{T}(\hat{\gamma}_T^* - \hat{\gamma}_T) \leq c\right), \mathbb{P}\left(\sqrt{T}(\hat{\gamma}_T - \gamma) \leq c\right)\right) \xrightarrow{\mathbb{P}} 0, c \in \mathbb{R}. \tag{4.8}$$

Proof See Appendix 4.A.

In summary, it has been shown that $\hat{\gamma}$, defined in (4.6), is a consistent and asymptotically Normal estimator for the true expected loss differential. Further, it is also shown that the density of $\hat{\gamma}$ can be estimated using the stationary bootstrap. This motivates the use of $\hat{\gamma}$ in the large battery of tests from the volatility forecasting literature, including those whose construction requires the use of a dependent bootstrap.

4.5 Simulation

In this section, I perform a similar simulation exercise to that conducted in Patton (2011*a*). First, I show that for four popular continuous-time models from the literature, the approach advocated in this chapter typically outperforms that of Patton (2011*a*) with the AR(1) assumption, and has comparable performance to the random walk assumption. Second, I show that for three different auto-regressive specifications, the approach advocated in this chapter outperforms that of Patton for both the AR(1) and random walk assumption. This is particularly interesting since for two of these specifications, the AR(1) assumption is correctly specified. Third, I consider the case where Patton (2011*a*) Assumption P1 is violated by a proxy error with an MA(1) structure. Finally, I provide simulation-based evidence that suggests the approach proposed in this paper is invariant to the variance of the proxy error. In contrast, the same simulations demonstrate that the power of both approaches described in Patton (2011*a*) deteriorates as the variance of the proxy error increases. This result is important since the best way to ensure realised variance is an unbiased proxy is to lower the sampling frequency, thus increasing the variance of the proxy error.

All simulations are performed over 5000 iterations. The first four simulations model log-price increments via a standard Euler discretization of the zero-drift model:

$$d \ln p_t^{\#} = \frac{1}{100}\nu_t \left(\kappa_1 dW_{1,t} + \kappa_2 dW_{2,t} + \sqrt{1 - \kappa_1^2 - \kappa_2^2} dW_{3,t}\right). \tag{4.9}$$

The diffusion parameter ν_t is scaled by $\frac{1}{100}$ to convert it from percentage units to fractional units, as log-price increments are, strictly speaking, expressed in fractional units. Volatility dynamics - expressed in percentage units - are simulated using each of the following four models:

1. The GARCH diffusion from Patton (2011*a*), Goncalves & Meddahi (2009) and Andersen & Bollerslev (1998): $d\nu_t^2 = 0.035(0.636 - \nu_t^2)dt + 0.144\nu_t^2 dW_{1,t}$.

2. The log-normal diffusion from Patton (2011*a*) and Andersen et al. (2005): $d \ln \nu_t^2 = -0.0136(0.8382 + \ln \nu_t^2)dt + 0.1148 dW_{1,t}$.

3. The two factor diffusion from Patton (2011*a*), Goncalves & Meddahi (2009) and Huang & Tauchen (2005): $\nu_t^2 = \exp(-1.2 + 0.04\nu_{1,t}^2 + 1.5\nu_{2,t}^2)$, where $d\nu_{1,t}^2 = -0.00137\nu_{1,t}^2 dt + dW_{1,t}$ and $d\nu_{2,t}^2 = -1.386\nu_{2,t}^2 dt + (1 + 0.25\nu_{2,t}^2)dW_{2,t}$.

4. The two factor affine diffusion from Andersen et al. (2005) and Bollerslev & Zhou (2002): $\nu_t^2 = \nu_{1,t}^2 + \nu_{2,t}^2$, where $d\nu_{1,t}^2 = 0.5708(0.3257 - \nu_{1,t}^2)dt + 0.2286\nu_{1,t}dW_{1,t}$ and $d\nu_{2,t}^2 = 0.0757(0.1786 - \nu_{2,t}^2)dt + 0.1096\nu_{2,t}dW_{2,t}$.

For GARCH and Log-normal diffusion, $\kappa_1 = -0.576$ and $\kappa_2 = 0$;[105] for the two factor diffusion, $\kappa_1 = \kappa_2 = -0.3$;[106] and for the two factor affine diffusion, $\kappa_1 = 0.9$ and $\kappa_2 = -0.4$.[107]

$T = 500$ periods are simulated, where each period contains $N = 720$ steps indexed $n = 1, ..., N$. Log prices are corrupted with additive microstructure noise via the model $\ln p_{n,t} = \ln p_{n,t}^{\#} + \xi_{n,t}$, where $\xi_{n,t} \overset{iid}{\backsim} \mathcal{N}(0, \sigma_{\xi,t}^2)$, and $p_{n,t}$ is the observable price. Following Patton (2011*a*), Ait-Sahalia et al. (2005) and Huang & Tauchen (2005), $\sigma_{\xi,t}^2$ is the solution to:

$$\frac{2\sigma_{\xi,t}^2}{\frac{5}{390}\mathbb{V}r_t + 2\sigma_{\xi,t}^2} = 0.2, \tag{4.10}$$

where r_t denotes a single period return. In other words, the proportion of variance of a 5 minute log-price increment that is attributable to microstructure noise is set to 20%. Note, a closed form solution to (4.10) is $\sigma_{\xi,t}^2 = \frac{1}{624}\theta_t$.

Diebold & Mariano (1995) tests are used to compare a base case estimator to five other estimators. For each of the five estimators, I test the null hypothesis $H_0 : \mathbb{E}L\Delta(\theta_t, \mathbf{x}_t) = 0$ against $H_A : \mathbb{E}L\Delta(\theta_t, \mathbf{x}_t) \neq 0$. The present chapter suggests testing H_0 by estimating $\mathbb{E}L\Delta(\theta_t, \mathbf{x}_t)$ using the estimator from Proposition 4.4.2. For the remainder of the paper, this estimator will be denoted $\hat{\gamma}_{[BH]}$. This approach is contrasted with the method suggested in Patton (2011*a*) Proposition 2, i.e. the random walk case, and Patton (2011*a*) Proposition 3, i.e. the AR(p) case. Regarding Patton (2011*a*) Assumption P2, following Patton & Sheppard (2009), unit weight is placed on the proxy at time $t + 1$, and in the AR(p) case, I set $p = 1$. I denote these specifications of Patton's estimator $\hat{\gamma}_{[P:RW]}$ and $\hat{\gamma}_{[P:AR]}$, respectively. The MSE loss function is used for both $\hat{\gamma}_{[P:RW]}$ and $\hat{\gamma}_{[P:AR]}$. This is because it provides the most transparent comparison between the different methods, as $\hat{\gamma}_{[BH]}$ must also use

[105] See Patton (2011*a*), Andersen et al. (2005), or Andersen, Benzoni & Lund (2002).
[106] See Patton (2011*a*) or Goncalves & Meddahi (2009).
[107] See Andersen et al. (2005) or Chernov et al. (2003).

the MSE loss function. However, Patton notes that the simulation results in Patton (2011*a*) Table 1 are similar for MSE and QLIKE. For both methods, the variance of the test statistic is estimated using 500 draws from the stationary bootstrap of Politis & Romano (1994*b*) with optimal block length selected following Politis & White (2004) and Patton et al. (2009).

Following Patton (2011*a*), I focus on 30-minute realised variance as the unbiased proxy of choice for all simulations, although I investigate other proxies at the end of this section.

All estimators are simulated as in Patton (2011*a*). That is:

$$x_{j,t} = \theta_t + \zeta_{j,t}, \tag{4.11}$$

where $j = \{b, 1, 2, 3, 4, 5\}$, where b stands for the base case. The estimation error is:

$$\zeta_{j,t} = \omega \tilde{u}_t + (1 - \omega)\sigma_u Z_{1,j,t} + \sqrt{\sigma^2_{\zeta,j} - \sigma^2_{\zeta,0}} Z_{2,j,t}, \tag{4.12}$$

where $(Z_{1,j,t}, Z_{2,j,t})^\intercal \overset{iid}{\backsim} \mathcal{N}(0, \mathbf{I})$. The first term on the right hand side of Equation (4.12) ensures the proxy and estimation error are correlated. The correlation coefficient is set to $\rho = 0.5$. This is ensured by setting ω and σ^2_u using:

$$\omega = \frac{\rho \sigma_{\zeta,0}}{\hat{\mathbb{V}}\tilde{u}}, \text{ and} \tag{4.13}$$

$$\sigma^2_u = \frac{(1 - \rho^2)\hat{\mathbb{V}}\tilde{u}\sigma^2_{\zeta,0}}{(\hat{\mathbb{V}}\tilde{u} - 0.5\sigma_{\zeta,0})^2}. \tag{4.14}$$

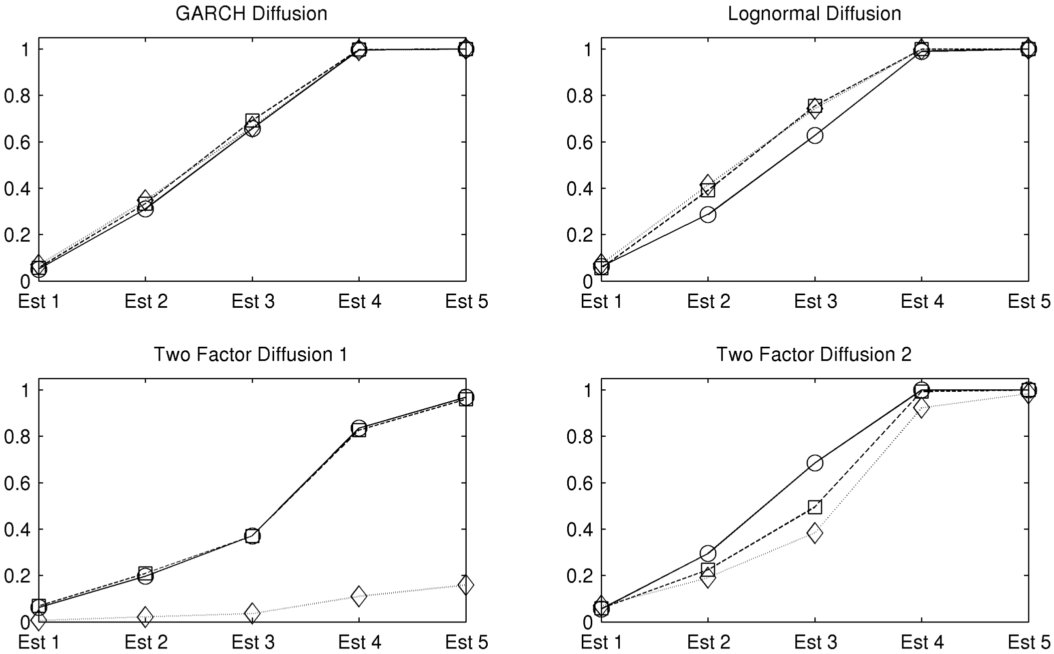
Set $\sigma^2_{\zeta,0} = 0.1\hat{\mathbb{V}}\theta$,and then the magnitude of the noise for each estimator can be controlled by $\sigma^2_{\zeta,j} = \lambda_j \hat{\mathbb{V}}\theta$, where λ_j controls the size of the noise for estimators 1 to 5. Note that the source of randomness for the estimation error and the true volatility models is independent, so Assumption 4.1.3 is satisfied.

The size of the statistical method is tested by setting $\lambda_j = 0.1, j = \{b, 1\}$, which causes the last term in (4.12) to vanish. Thus the base case and estimator 1 have identical expected loss, ie a true null. $\lambda_j, j = \{2, 3, 4, 5\}$ is set to $\{0.15, 0.2, 0.5, 1\}$ respectively, so that as j increases, the expected loss increases and rejection of $H_0 : \mathbb{E}\Delta L(\theta_t, \mathbf{x}_t) = 0$ grows progressively easier. This facilitates a study of the power of the statistical method.

Table 4.1 demonstrates that $\hat{\gamma}_{[BH]}$ and $\hat{\gamma}_{[P:RW]}$ have appropriate size across all four volatility models, as the rejection frequency is generally close to 0.05. In contrast, $\hat{\gamma}_{[P:AR]}$ slightly over-rejects a true null hypothesis, except for the two factor diffusion from Patton (2011*a*) where $\hat{\gamma}_{[P:AR]}$ shows significant size distortion in the form of under-rejection.

Next I consider the power of the statistical tests. When volatility follows the

Figure 4.1: Power curves for the four diffusion processes



The rejection frequencies are on the Y-axis. The filled line with a circle marker denotes the estimator from Equation 4.6. The dashed line with a square marker denotes the estimator from Patton (2011*a*) with the random walk assumption. The dotted line with a diamond marker denotes the estimator from Patton (2011*a*) with the AR(1) assumption. $T = 500$ in all four plots.

Test sizes corresponding to Figure 4.1

	GD	LD	TFD1	TFD2
$\hat{\gamma}_{[BH]}$	0.050	0.061	0.062	0.055
$\hat{\gamma}_{[P:RW]}$	0.057	0.055	0.069	0.058
$\hat{\gamma}_{[P:AR]}$	0.070	0.072	0.006	0.069

Table 4.1: $\hat{\gamma}_{[BH]}$ is the estimator from Equation (4.6), while $\hat{\gamma}_{[P:RW]}$ and $\hat{\gamma}_{[P:AR]}$ are the estimators from Patton (2011*a*) with the random walk and the AR(1) assumption respectively. GD = GARCH diffusion, LD = lognormal diffusion, TFD1 = two factor diffusion from Patton (2011*a*), and TFD2 = two factor affine diffusion from Andersen et al. (2005).

GARCH diffusion, all three methods have near-identical performance, while for the lognormal diffusion, $\hat{\gamma}_{[P:RW]}$ and $\hat{\gamma}_{[P:AR]}$ slightly outperform $\hat{\gamma}_{[BH]}$. This is probably because the log-diffusion volatility dynamics are quite mild, so the loss of efficiency that $\hat{\gamma}_{[P:RW]}$ and $\hat{\gamma}_{[P:AR]}$ suffer due to using a proxy from time $t+1$ is minor. $\hat{\gamma}_{[BH]}$, on the other hand, always suffers some loss of efficiency due to the burden of estimating two additional covariance parameters.

For the more challenging two factor diffusion from Patton (2011*a*), $\hat{\gamma}_{[BH]}$ and $\hat{\gamma}_{[P:RW]}$ exhibit near-identical performance. In contrast, $\hat{\gamma}_{[P:AR]}$ exhibits very poor performance; it is mostly unable to reject a false null across all estimators. Interestingly, Patton (2011*a*) also observed poor performance when true volatility follows this model.

The most challenging volatility dynamics are provided by the two factor affine diffusion which exhibits larger movements in latent volatility, as well as a strong correlation between the volatility equations and the log-price increment equation. For this model, Figure 4.1 demonstrates that $\hat{\gamma}_{[BH]}$ has superior power to both $\hat{\gamma}_{[P:RW]}$ and $\hat{\gamma}_{[P:AR]}$.

Next, I propose four additional simulations to further demonstrate the robustness of $\hat{\gamma}_{[BH]}$. In the first three simulations, daily variance is modelled via the AR(1) process:

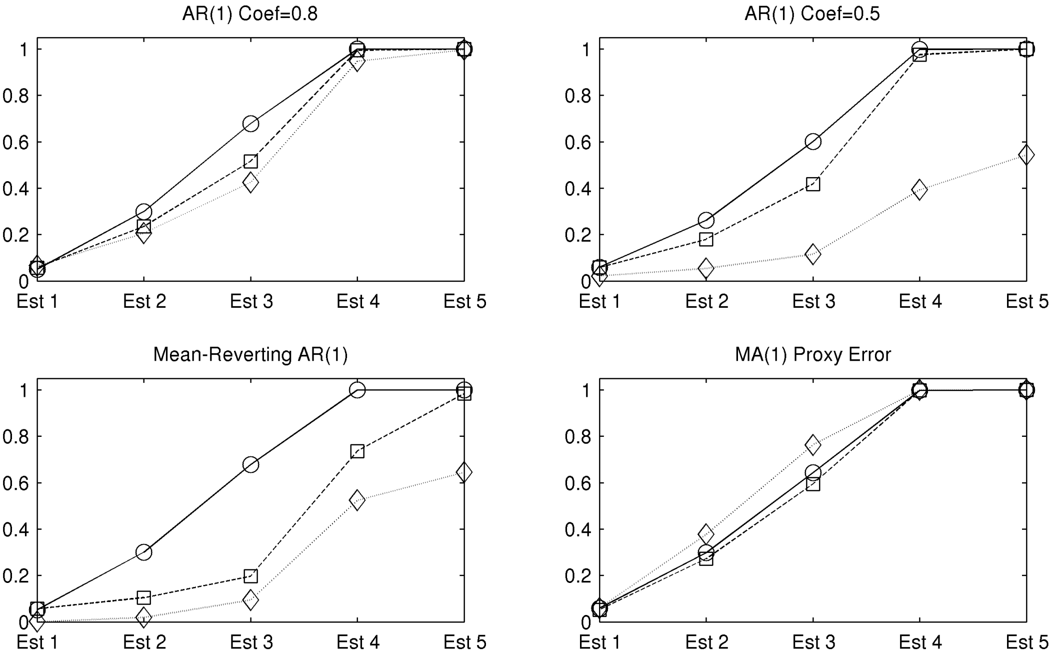
$$\theta_t = \mu + \phi\theta_{t-1} + \eta_t, \tag{4.15}$$

for the specifications $\{\mu = 0.15, \phi = 0.8\}$, $\{\mu = 0.2, \phi = 0.5\}$, and $\{\mu = 1.75, \phi = -0.75\}$. Thus the first simulation is the closest to a random walk, while the third is a strong violation of the random walk model. The second lies between the two. For all the auto-regressive models, $\eta_t \overset{iid}{\backsim} \mathcal{N}(0, 0.01)$, with constant intraday dynamics. Log-price increments are modelled using *iid* standard Normals multiplied by the corresponding intraday volatility.

The fourth simulation employs the GARCH diffusion for which all three approaches had near identical performance (see Figure 4.1). However, this time the proxy error is modelled as a zero-mean MA(1) process, scaled such that it has sample variance equal to the sample variance of the estimation error of 30 minute realised variance. This ensures that the only real difference between the simulation results for the GARCH diffusion in Figure 4.1 and the present case is the MA(1) component of the proxy error. This is designed to violate the martingale difference sequence assumption for the proxy error in Patton (2011*a*). The comparative power of the three approaches across these four models is illustrated in Figure 4.2.

Table 4.2 demonstrates that, as before, $\hat{\gamma}_{[BH]}$ and $\hat{\gamma}_{[P:RW]}$ have appropriate rejection rates of a true null hypothesis across all four volatility models. $\hat{\gamma}_{[P:AR]}$ slightly under-rejects a true null hypothesis for the AR(1) with coefficient of 0.5, and exhibits significant size distortion for AR(1) with coefficient of -0.75.

Figure 4.2: Power curves for AR(1) processes and MA(1) proxy error



The rejection frequencies are on the Y-axis. The filled line with a circle marker denotes the estimator from Equation 4.6. The dashed line with a square marker denotes the estimator from Patton (2011*a*) with the random walk assumption. The dotted line with a diamond marker denotes the estimator from Patton (2011*a*) with the AR(1) assumption. $T = 500$ in all plots.

Test sizes corresponding to Figure 4.2

	AR(1)0.8	AR(1)0.5	AR(1)-0.75	MA(1)PE
$\hat{\gamma}_{[BH]}$	0.049	0.059	0.051	0.058
$\hat{\gamma}_{[P:RW]}$	0.057	0.059	0.056	0.052
$\hat{\gamma}_{[P:AR]}$	0.065	0.021	0.001	0.059

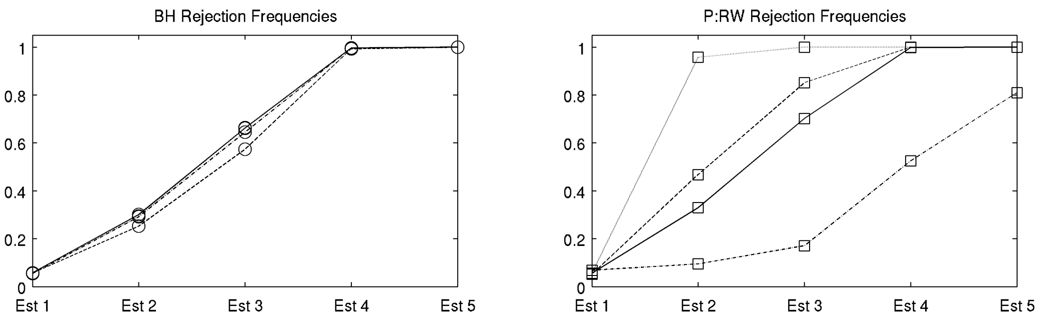
Table 4.2: $\hat{\gamma}_{[BH]}$ is the estimator from Equation (4.6), while $\hat{\gamma}_{[P:RW]}$ and $\hat{\gamma}_{[P:AR]}$ are the estimators from Patton (2011*a*) with the random walk and the AR(1) assumption respectively. AR(1)0.8, AR(1)0.5, and AR(1)-0.75 represent the AR(1) model for true variance with coefficients of 0.8, 0.5, and -0.75 respectively. MA(1)PE represents the GARCH diffusion with MA(1) proxy error.

Figure 4.2 shows that across all three auto-regressive models, $\hat{\gamma}_{[BH]}$ outperforms both of Patton's approaches. This is particularly interesting, as for the first two auto-regressive models, $\hat{\gamma}_{[P:AR]}$ is correctly specified. The results suggest that the additional estimation error inherent in using the AR(1) correction term in Patton (2011*a*) Proposition 3 is not justified by the overall reduction in bias. Thus if the empiricist suspects true variance follows a weakly dependent process that is not "close" to a random walk, Figure 4.2 provides strong evidence that the approach advocated in the present chapter should be used.

Figure 4.2 also demonstrates that, surprisingly, $\hat{\gamma}_{[P:AR]}$ slightly outperforms when an MA(1) proxy error structure is introduced to the GARCH diffusion. Further, $\hat{\gamma}_{[P:RW]}$ only exhibits a slight reduction in power relative to the pure GARCH diffusion case. This is good news for the approach of Patton (2011*a*) as it suggests that violations of the martingale difference property of Patton (2011*a*) assumption P1 do not appear to translate to overall loss of power or size distortion.

I conclude this section with some simulation-based evidence that suggests that the method proposed in this paper is approximately invariant to the variance of the proxy error, while the power of the approaches proposed in Patton (2011*a*) deteriorates as the variance of the proxy error increases. To demonstrate this result, I use the same GARCH diffusion simulation that generated the results in Table 4.1, but this time I repeat the simulation for four proxies: true variance, 5-minute realised variance, 30-minute realised variance, and squared daily returns.[108]

Figure 4.3: Power curves for different proxies



The rejection frequencies are on the Y-axis. The left-side plot depicts power curves for the estimator advocated in this paper. The right-side plot depicts power curves for the estimator from Patton (2011*a*) with the random walk assumption. In both plots, the line-style indicates which proxy is being used. A dotted line indicates true variance, a dashed line indicates 5-minute realised variance, a solid line indicates 30-minute realised variance, and a dot-dash line indicates squared daily returns. $T = 500$ in all plots.

Figure 4.3 contains results for $\hat{\gamma}_{[BH]}$ and $\hat{\gamma}_{[P:RW]}$. The results for $\hat{\gamma}_{[P:AR]}$ are omitted as they are very similar to those for $\hat{\gamma}_{[P:RW]}$. Figure 4.3 demonstrates that the size

[108]The variance of the proxy error is zero when true variance is used as a proxy, but it increases as the sampling frequency of realised variance is lowered (squared daily returns are the limit of this process).

and power of $\hat{\gamma}_{[BH]}$ is unaffected by the choice of proxy. In contrast, $\hat{\gamma}_{[P:RW]}$ has more power when the proxy is true variance, but less power, and some small size distortion, when the proxy is a squared daily return. This is an important finding, since all approaches discussed require an unbiased proxy and currently, a squared daily return is the most reliable proxy if the main priority is zero bias.

In summary, Figure 4.3 suggests that the methodology advocated in this chapter should be preferred when the variance of the proxy error is large. Further, Figure 4.2 suggests that the methodology advocated in this chapter should be used in situations where true variance is suspected to deviate significantly from the random walk assumption. In cases where the random walk assumption provides a good approximation, and the variance of the proxy error is not large, the approach advocated in this chapter still provides a useful complement to the approach of Patton (2011*a*) with a random walk assumption. Given its sporadic reliability, the approach of Patton with an AR(1) assumption should probably be avoided.

4.6 Empirical

In this section I examine estimates of the expected loss differential generated by the approach proposed in the present chapter and the approaches of Patton (2011*a*), and I compare these estimates to those of a naive approach that does not account for possible dependence between the forecast error and proxy error. The analysis is performed on IBM - a publicly listed asset on the New York Stock Exchange - over the interval January 2004 to December 2007.

I estimate the expected loss differential between different sampling frequency realised variance estimators. All frequencies are specified in calendar time. I set 5-minute realised variance as the base-case sampling frequency and the alternative cases are a range of sampling frequencies from 1 second up to 90 minutes. I use 30-minute realised variance as the conditionally unbiased proxy.[109] Note that this rules out a comparison of 5 and 30 minute realised variance, since the proxy must be different to the estimators under consideration to apply the method proposed in this chapter.

Figure 4.4 contains the estimated expected loss differentials obtained using the approach proposed in the present chapter, as well as the random walk and AR(1) approaches proposed in Patton (2011*a*). It also contains a sequence of naive expected loss differentials that are obtained simply by applying the MSE loss function without any lag methods or adjustments. A Diebold & Mariano (1995) test is performed for

[109]Ideally, one would choose squared daily returns as a proxy since by construction it must be unbiased (assuming daily returns are mean zero). However, squared returns are usually too noisy for statistical significance to obtain in the results. In contrast, the higher the chosen sampling frequency, the more likely that the proxy will not be unbiased. The 30-minute frequency balances these various concerns.

all three sequences with confidence intervals generated via a stationary bootstrap. Rejections of the null are indicated with a mark on each plot-line.

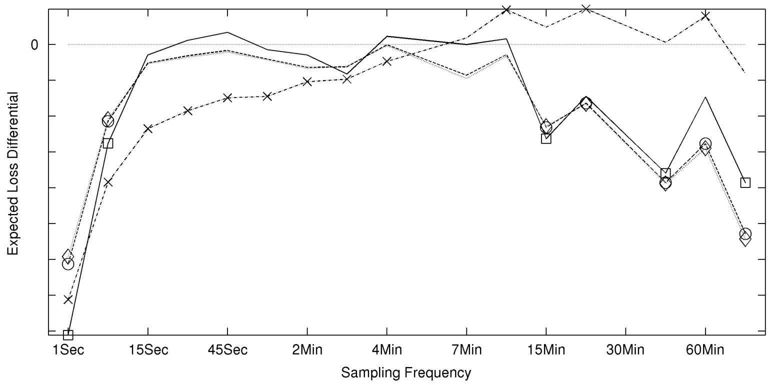


Figure 4.4: Four different estimates of the expected loss differential across a range of realised variance sampling frequencies. Rejections of the null hypothesis of the expected loss differential equal to zero are marked. The methods are: 1) the estimator from equation 4.6 (full line with square marker), 2) the approach of Patton (2011*a*) with the random walk assumption (dashed line with circle marker), 3) the approach of Patton (2011*a*) with the AR(1) assumption (dotted line with diamond marker), and 4) the naive MSE estimate (dot-dash line with cross marker). Note, no tests are performed at the 5 and 30 minute frequencies as they correspond to the base case frequency and the proxy frequency.

The most important point to take away from Figure 4.4 is that the approach advocated in the present chapter and the approaches of Patton (2011*a*) generate very similar estimates. This is a reassuring result, as the assumption sets employed by these methods are quite different. It is worth emphasizing that results for the approach recommended in this chapter match the *a priori* expected pattern: the very low and very high sampling frequencies perform poorly relative to 5 minute realised variance, and I am able to reject the null hypothesis of equal expected loss in several of these extreme cases.

In contrast, the naive estimate of the expected loss differential is markedly different. It generates an unlikely sequence of estimates of the expected loss differential; one that is close to monotonically increasing as the sampling frequency decreases. Further, the naive estimate finds that the 60 minute sampling frequency significantly outperforms the 5 minute sampling frequency - a very unlikely result for an asset that is as liquid as IBM.

Taken together, these points suggest that a correction such as the one proposed in this chapter, is almost certainly a good idea for any empiricist attempting to rank variance estimators via empirical criteria. Taking the simulation results into account, the empiricist would use the approach advocated in this chapter if the true variance process appears to be weakly dependent, and the approach of Patton (2011*a*) with the random walk assumption if the true variance process is suspected to be close to a random walk.

A second point worth noting is that the methods show little ability to reject the null hypothesis of equal expected loss across the mid-range sampling frequencies. This is unfortunate, since statistical significance in this region would allow the methods discussed in this paper to serve as a model-free approach to determining the optimal sampling frequency for realised variance. One might imagine that a stronger statistical significance could be obtained by increasing the number of observations. In practice, this approach often fails, as the poor ability to reject the null hypothesis is usually the result of a small number of days dominating the sample. For example, if the interval of analysis was increased to include the global financial crisis period (2007 to 2009) of spiking volatility, then statistical significance would almost certainly *decrease*.

Currently, the only solution to this problem is to employ a loss function that places less weight on outliers, such as the QLIKE loss function advocated in Patton (2011*b*). Unfortunately, QLIKE exhibits asymmetric loss, which may be undesirable given the utility functions of economic agents. Further, QLIKE is incompatible with the methods proposed in this paper, and as demonstrated in Patton (2011*a*), for some popular models of true volatility it will lead to significant size distortions. Thus it must be regarded as an imperfect solution.

One other possible solution to this problem is to employ a robust estimator, such as the trimmed mean, with the MSE loss function. However, preliminary work undertaken by the present author on this topic suggests that the unconditional distribution of the loss differential sequence is frequently asymmetric, so ad hoc rules would be needed to adjust a robust location estimator. For this reason, fat-tails in the loss differential sequence must still be regarded as an open problem deserving of future work.

4.7 Conclusion

This chapter presents a new methodology for estimating the expected loss differential between two estimators (or forecasts) when the latent quantity of interest is replaced by a proxy. The method is robust to arbitrary contemporaneous dependence between the proxy error and estimation errors. This is particularly important if one is attempting to evaluate the empirical performance of intraday variance estimators, as opposed to forecasts. The chapter also contains, in Appendix 4.A, a new result on the product of near epoch dependent processes that applies generally to the time-series literature.

The approach advocated in this paper builds on the initial work on this problem by Patton (2011*a*) by proposing a trade-off: specific modelling assumptions for the true volatility process are generalized, at the cost of restricting attention to the MSE loss function. The present approach explicitly nests that of Patton (2011*a*) with a stationary AR(p) assumption, since the near epoch dependence assumption explicitly nests autoregressive models with coefficients inside the unit circle. Further, the present approach outperforms the stationary AR(p) assumption in most of the simulations considered in Section 4.5. In contrast, the present approach has performance comparable to that of Patton (2011a) with a random walk assumption, as long as the true variance process is "close" to a random walk. In cases where the random walk assumption is strongly violated, the evidence presented suggests that the present approach should be preferred.

Although popular, the MSE loss function restriction can prove problematic for certain time-intervals. Since this loss function emphasizes large losses, periods that include large spikes in volatility can result in loss of statistical power, since a small number of observations dominate the analysis. For example, analysis using the MSE loss function will not perform well during the global financial crisis period.

Ultimately, I recommend that in any analysis an empiricist uses the method suggested in this chapter alongside the method of Patton (2011a) with the random walk assumption and both MSE and QLIKE loss functions. Sensible conclusions can then be obtained given careful consideration of both the data and the differing assumption sets of each approach.

Appendix 4.A Proofs

This appendix contains proofs of Proposition 4.4.1 to Proposition 4.4.3. However, before providing these proofs, we require a technical lemma on the product of near epoch dependent processes which, to the best of my knowledge, is new in the literature.

Lemma 4.A.1 *Let X_t be L_r-NED of size $-\phi_X$ on any process V_t and let Y_t be L_s-NED of size $-\phi_Y$ on V_t, where $r, s \geq 1$ and $\phi_X, \phi_Y > 0$. Then if $\ell p \leq r$ and $\ell q \leq s$, where $(1/p) + (1/q) = 1$, then $X_t Y_t$ is L_ℓ-NED of size $-\min\{\phi_X, \phi_Y\}$ on V_t.*

Proof This proof is a straightforward extension of the proof of Theorem 17.9 in Davidson (1994).

In what follows, I use the shorthand $\mathbb{E}_m \bullet \equiv \mathbb{E}[\bullet | \mathcal{F}_{t-m}^{t+m}]$. Also, let d_t^X and d_t^Y denote a sequence of positive (finite) constants associated with X_t and Y_t respectively, and let $\nu_m^X = O(m^{-\phi_X})$ and $\nu_m^Y = O(m^{-\phi_Y})$ denote the sequence of mixing coefficients associated with X_t and Y_t respectively, so that by definition, $||X_t - \mathbb{E}_m X_t||_r \leq d_t^X \nu_m^X$ and $||Y_t - \mathbb{E}_m Y_t||_s \leq d_t^Y \nu_m^Y$, where $||\bullet||_r \equiv (\mathbb{E}|\bullet|^r)^{1/r}$. Consider:

$$||X_t Y_t - \mathbb{E}_m X_t Y_t||_\ell \tag{4.16a}$$

$$= ||X_t(Y_t - \mathbb{E}_m Y_t) + (X_t - \mathbb{E}_m X_t)\mathbb{E}_m Y_t - \mathbb{E}_m (X_t - \mathbb{E}_m X_t)(Y_t - \mathbb{E}_m Y_t)||_\ell \tag{4.16b}$$

$$\leq ||X_t(Y_t - \mathbb{E}_m Y_t)||_\ell + ||(X_t - \mathbb{E}_m X_t)\mathbb{E}_m Y_t||_\ell + ||\mathbb{E}_m (X_t - \mathbb{E}_m X_t)(Y_t - \mathbb{E}_m Y_t)||_\ell , \tag{4.16c}$$

where (4.16c) follows from Minkowski's Inequality. I derive an appropriate upper bound for each of these three norms separately. For the first two norms, note that Holder's Inequality implies that for any random variables A and B:

$$||AB||_{\ell} \leq \left|\left|A^{\ell}\right|\right|_{p}^{1/\ell} \left|\left|B^{\ell}\right|\right|_{q}^{1/\ell} = ||A||_{\ell p} \, ||B||_{\ell q} \, , \tag{4.17}$$

where $(1/p) + (1/q) = 1$. So for the first norm, note that:

$$\begin{aligned} ||X_t(Y_t - \mathbb{E}_m Y_t)||_{\ell} &\leq ||X_t||_{\ell p} \, ||Y_t - \mathbb{E}_m Y_t||_{\ell q} && (4.18a) \\ &\leq ||X_t||_{\ell p} \, d_t^Y \nu_m^Y, && (4.18b) \end{aligned}$$

where the final term follows if $\ell q \leq s$. For the second norm, note that:

$$\begin{aligned} ||(X_t - \mathbb{E}_m X_t)\mathbb{E}_m Y_t||_{\ell} &\leq ||X_t - \mathbb{E}_m X_t||_{\ell p} \, ||\mathbb{E}_m Y_t||_{\ell q} && (4.19a) \\ &\leq ||Y_t||_{\ell q} \, d_t^X \nu_m^X, && (4.19b) \end{aligned}$$

where the final term follows if $\ell p \leq r$. For the third norm, note that:

$$\begin{aligned} & ||\mathbb{E}_m (X_t - \mathbb{E}_m X_t)(Y_t - \mathbb{E}_m Y_t)||_{\ell} && (4.20a) \\ =& (\mathbb{E} \left|\mathbb{E}_m (X_t - \mathbb{E}_m X_t)(Y_t - \mathbb{E}_m Y_t)\right|^{\ell})^{1/\ell} && (4.20b) \\ \leq& (\mathbb{E}\mathbb{E}_m \left|(X_t - \mathbb{E}_m X_t)(Y_t - \mathbb{E}_m Y_t)\right|^{\ell})^{1/\ell} && (4.20c) \\ =& (\mathbb{E} \left|(X_t - \mathbb{E}_m X_t)(Y_t - \mathbb{E}_m Y_t)\right|^{\ell})^{1/\ell} && (4.20d) \\ =& \, ||(X_t - \mathbb{E}_m X_t)(Y_t - \mathbb{E}_m Y_t)||_{\ell} && (4.20e) \\ \leq& \, ||X_t - \mathbb{E}_m X_t||_{\ell p} \, ||Y_t - \mathbb{E}_m Y_t||_{\ell q} && (4.20f) \\ \leq& d_t^X \nu_m^X d_t^Y \nu_m^Y, && (4.20g) \end{aligned}$$

where (4.20c) follows from Jensen's Inequality (for conditional expectations), (4.20d) follows from the Law of Iterated Expectations, (4.20f) follows from Holder's Inequality (as discussed above), and (4.20g) follows if $\ell p \leq r$ and $\ell q \leq s$.

Combining the three upper bounds demonstrates that:

$$||X_t Y_t - \mathbb{E}_m X_t Y_t||_{\ell} \leq d_t \nu_m, \tag{4.21}$$

where:

$$d_t = \max\{||X_t||_{\ell p} \, d_t^Y, ||Y_t||_{\ell q} \, d_t^X, d_t^X d_t^Y\}, \tag{4.22}$$

and:

$$\nu_m = \nu_m^X + \nu_m^Y + \nu_m^X \nu_m^Y = O(m^{-\min\{\phi_X, \phi_Y\}}). \tag{4.23}$$

As discussed, sufficient conditions for this are $\ell p \leq r$ and $\ell q \leq s$. These are also sufficient conditions for $||X_t||_{\ell p} < \infty$ and $||Y_t||_{\ell q} < \infty$, thus the result follows. □

Remark 4.11 Let $r = s = p = q = 2$. Then the conditions of Lemma 4.A.1 are only satisfied for $\ell \leq 1$, so $X_t Y_t$ is L_1-NED. This special case is precisely Theorem 17.9 of Davidson (1994). More generally, the proof of Lemma 4.A.1 reveals that ℓ is essentially determined by Holder's inequality, and so if $\min\{r, s\}$ is small, ℓ can be kept close to $\min\{r, s\}$ if $\max\{r, s\}$ is allowed to be correspondingly large. This result is particularly useful in circumstances such as the present chapter, where I wish to keep the bound on the moments of latent variance as low as possible, but am happy for the bound on the moments of the estimation error to be large.

Remark 4.12 A straightforward application of Lemma 4.A.1 implies that both $u_{j,t}\theta_t$ and $u_{j,t}^2$ are $L_{2+\delta}$-NED of size -1 on a strong mixing process of size $-(2+\delta)(r+\delta)/(r-2)$, $r > 2$, and δ the same as that referred to in assumptions 4.2.1 and 4.2.2.[110] This result will be used extensively in the proofs that follow.

Proof of Proposition 4.4.1 Using Assumption 4.1.1, a Taylor expansion of $L(\tilde{\theta}_t, x_{j,t})$ around $(\theta_t, x_{j,t})$ yields:

$$L(\tilde{\theta}_t, x_{j,t}) = L(\theta_t, x_{j,t}) + 2(\theta_t - x_{j,t})(\tilde{\theta}_t - \theta_t) + (\tilde{\theta}_t - \theta_t)^2. \tag{4.24}$$

Taking the difference of (4.24) for two estimators, j and i, yields:

$$\Delta L(\theta_t, \mathbf{x}_t) = \Delta L(\tilde{\theta}_t, \mathbf{x}_t) + 2(x_{j,t}\tilde{\theta}_t - x_{i,t}\tilde{\theta}_t - x_{j,t}\theta_t + x_{i,t}\theta_t), \tag{4.25}$$

as the last term in (4.24) does not depend on j or i and so vanishes. Taking expectations of both sides of (4.25), and adding and subtracting $(\mathbb{E}x_{j,t})(\mathbb{E}\theta_t) + (\mathbb{E}x_{i,t})(\mathbb{E}\theta_t)$ yields:

$$\begin{aligned}\mathbb{E}\Delta L(\theta_t, \mathbf{x}_t) =& \mathbb{E}\Delta L(\tilde{\theta}_t, \mathbf{x}_t) + 2\Big(\mathrm{cov}(x_{j,t}, \tilde{\theta}_t) - \mathrm{cov}(x_{i,t}, \tilde{\theta}_t)\Big) \\ &+ 2\Big(\mathrm{cov}(u_{i,t}, \theta_t) - \mathrm{cov}(u_{j,t}, \theta_t)\Big),\end{aligned} \tag{4.26}$$

where Assumption 4.1.2 was used to convert $\mathbb{E}\theta_t$ to $\mathbb{E}\tilde{\theta}_t$. The final term in (4.26) vanishes by Assumption 4.1.3. Proposition 4.4.1 immediately follows. □

Proof of Proposition 4.4.2 The proof proceeds by demonstrating that $\hat{\gamma}$ can be decomposed into a linear function of terms that, when scaled by $\sqrt{T}$, will each satisfy a CLT.

Exploiting Assumption 4.1.1, $\hat{\gamma}$ admits the decomposition:

$$\hat{\gamma} = \overline{u_j^2 - u_i^2} + 2(\overline{(u_j - u_i)\theta} - \overline{u_j - u_i}\,\overline{\theta} - \overline{u_j - u_i}\,\overline{\tilde{u}}). \tag{4.27}$$

[110]Obviously, the bound on the moments of $u_{j,t}^2$ could actually be made much tighter, but the one stated here is sufficient for the purposes of this chapter.

Assumptions 4.2.1 and 4.2.2, along with Lemma 4.A.1, guarantee that each of $\{u_{j,t}, u_{j,t}^2, u_{j,t}\theta_t, \theta_t, \tilde{u}_t\}$ are $L_{2+\delta}$-NED of size -1 on a strong mixing process of size $-(2+\delta)(r+\delta)/(r-2)$, $\delta > 0$, $r > 2$. Using Davidson (1994) Equation 24.29, it is straightforward to verify that each process will, when centred and scaled appropriately, obey the conditions of the central limit theorem in Davidson (1994) Theorem 24.6 and Corollary 24.7 (the above conditions are actually slightly stronger than is required by the theorem and corollary). This immediately takes care of the first two RHS terms of (4.27). For the third term, note that the CLT is sufficient for $\overline{\theta} \xrightarrow{\text{m.s.}} \mathbb{E}\overline{\theta} \equiv \mu_\theta$, so an application of Cramér's theorem[111] demonstrates that:

$$\overline{\theta}\left(\sqrt{T}\,\overline{u_j - u_i - (\mu_{u_j} - \mu_{u_i})}\right) \xrightarrow{\text{d}} \mu_\theta Z, \tag{4.28}$$

where Z is a zero-mean Normal random variable. The argument for the fourth term is identical to the third: simply replace θ with $\tilde{u}$. Since a central limit theorem holds for each right-hand side term in (4.27), it immediately follows from the continuous mapping theorem[112] that $\hat{\gamma}$ obeys a CLT. Note that Assumption 4.1.3 guarantees the asymptotic distribution of $\hat{\gamma}$ is centred on γ. $\square$

Proof of Proposition 4.4.3 As in Proposition 4.4.2, $\hat{\gamma}_T^*$ admits the following decomposition:

$$\hat{\gamma}_T^* = \overline{u_j^{2*} - u_i^{2*}} + 2(\overline{(u_j^* - u_i^*)\theta^*} - \overline{u_j^* - u_i^*}\,\overline{\theta^*} - \overline{u_j^* - u_i^*}\,\overline{\tilde{u}^*}). \tag{4.29}$$

Assumptions 4.2.1 and 4.2.2, along with Lemma 4.A.1, guarantee that each of $\{u_{j,t}, u_{j,t}^2, u_{j,t}\theta_t, \theta_t, \tilde{u}_t\}$ are $L_{2+\delta}$-NED of size -1 on a strong mixing process of size $-(2+\delta)(r+\delta)/(r-2)$, $\delta > 0$, $r > 2$. In combination with assumptions 4.3.1 and 4.3.2, all assumptions of Goncalves & de Jong (2003) Theorem 2 are satisfied, so the first two RHS terms of (4.29), scaled by $\sqrt{T}$, obey a stationary bootstrap CLT. For the third term, since $\sqrt{T}\,\overline{u_j^* - u_i^*}$ obeys a stationary bootstrap CLT, it remains to show that $\overline{\theta^*} \xrightarrow{\text{m.s.}} \mu_\theta$, and then to apply Cramér's theorem,[113] much as was done in the proof of Proposition 4.4.2. By construction of the stationary bootstrap, θ_t^* is a stationary sequence (even if θ_t is not), and $\mathbb{E}[\theta_t^* | \theta_1, ..., \theta_T] = \overline{\theta}$. Let $\mu_\theta = T^{-1}\sum_t \mathbb{E}\theta_t$, then assumption 4.2.2 is sufficient for $\overline{\theta} \xrightarrow{\text{m.s.}} \mu_\theta$ and for $\overline{\theta^*} \xrightarrow{\text{m.s.}} \overline{\theta}$, and so $\overline{\theta^*} \xrightarrow{\text{m.s.}} \mu_\theta$.

The argument for the fourth term is identical, simply substitute $\tilde{u}$ for θ.

Having demonstrated a stationary bootstrap CLT for the RHS terms of (4.29), a bootstrap CLT for $\hat{\gamma}^*$ now follows immediately upon application of the continuous mapping theorem. This argument is used in, for example, Politis & Romano (1994*b*) Theorem 3. $\square$

[111] Davidson (1994) Theorem 22.14.
[112] Davidson (1994) Theorem 22.11.
[113] Davidson (1994) Theorem 22.14.

Chapter 5

Conclusion

This thesis is concerned with estimation and forecast of daily risk parameters using high frequency financial data. The material is drawn from three distinct papers that are available on the *Social Science Research Network*.[114]

Accurate forecasting of daily risk parameters is important, since these parameters are used in the every-day decisions of financial institutions. For example, value-at-risk is used to determine the magnitude of the capital cushions held by financial institutions. Poor forecasts will lead to over or under-allocation of capital to the cushion. The former implies inefficient investment of capital, while the latter implies that in the event of a crisis, sufficient capital may not be available to prevent foreclosure.

The role of accurate estimates is a little more subtle. Under certain modelling conditions, accurate estimates can be used to proxy the true target of interest in forecast evaluation procedures. Moreover, the estimates themselves may prove useful as predictive variables in forecast models.

Chapter 2 contributes to this topic by providing a methodology for consistently estimating a wide variety of risk parameters associated with the distribution of a daily asset return. The only requirement is an appropriate sequence of intraday data. The procedure is referred to as the Bootstrap Return Method. Two specific applications were considered. First, it was shown that the Bootstrap Return Method variance estimator has performance comparable to realised kernels.[115] Second, a single unified framework that combines the Bootstrap Return Method with the dual-asymptotic approach to forecast evaluation of Patton & Li (2013) was described, and for the specific case of value-at-risk, this framework was shown to exhibit much greater power at distinguishing between competing forecast models than other methods in the literature.

There are many interesting theoretical extensions to the Bootstrap Return Method. For example, extending it to a multivariate framework, deriving an asymptotic distribution theory, or investigating the possibility of other re-sampling methods.

[114] See Bowers & Heaton (2014*a*), Bowers & Heaton (2014*b*), and Bowers & Heaton (2014*c*).

[115] Arguably the best performing intraday data-based variance estimator.

In Chapter 3, the empirical analysis of value-at-risk in Chapter 2 is expanded to include a large suite of value-at-risk forecast models, and to utilize data from two exchanges, spanning two distinct time periods.[116] The results provide strong evidence that forecast models that employ intraday data outperform those that utilize only daily data. Further, several simple time series models of the Bootstrap Return Method value-at-risk estimator are shown to outperform all other forecast methods considered.

Possible extensions of this chapter include the analysis of more sophisticated time series models, as well as the investigation of forecast models of other parameters that are estimable via the Bootstrap Return Method, such as Expected Shortfall. It may also be illustrative to investigate possible co-integrating relationships between the Bootstrap Return Method variance estimator, Quantile estimator, and Expected Shortfall estimator.

Chapter 4 contributes to the topic by providing a method for determining which intraday data-based estimates of daily variance are the most accurate. The method generalizes modelling assumptions in the initial work on this topic[117] at the cost of restricting the distance metric used in evaluation to the Mean Square Error loss function. As described in Chapter 4 this restriction limits the usefulness of the methods in periods when volatility spikes, since the effect of a small number of large volatility measurements are over-weighted by the choice of loss function. For this reason, I regard the method proposed in Chapter 4 as a step towards an ideal solution, rather than an ideal solution itself. Empiricists are encouraged to use the method in Chapter 4 to complement that in Patton (2011*a*), rather than to replace it.

Future work on this topic could seek other loss functions besides the Mean Square Error where the approach advocated in Chapter 4 can be applied. Another alternative is to design indirect tests of the variance estimators. For example, one could analyse the properties of the standardized return series (daily returns divided by volatility estimate) and see if they accord with rational modelling assumptions.

I would like to conclude this thesis by emphasizing a common theme: that high frequency data can be used to greatly improve estimation and forecasts of classical daily financial parameters. In Chapter 2, it was shown estimators derived from an intraday sequence can be used to greatly improve the power of forecast evaluation procedures. In Chapter 3, it was shown that forecasts based on intraday data outperform those based on only daily data. In Chapter 4, the proposed method was used to demonstrate that higher frequency realised variance estimators are more accurate than those with very low frequencies (especially the limiting case of squared daily returns).

[116] Specifically, the global financial crisis period (2008 to 2010) and the period of relatively lower volatility that followed (2011 to 2013).

[117] Patton (2011*a*).

I think that there is a lot more work to be done before we can declare a thorough understanding of the data-generating-process behind high frequency financial data. The literature on complete modelling of the order book is still in its formative stages, while other issues such as the fixed exchange regulated minimum tick size are largely ignored.[118] In this thesis, as with most scientific work, I provide only incremental steps on the path to understanding, in the hope that one day in the future a fuller picture can be obtained.

[118] Rosenbaum (2009) does look into this issue, although is forced to assume a minimum tick size that vanishes as the number of intraday observations grows. A more general discussion of some of the issues can be found in Jacod & Protter (2011) Chapter 16.

Bibliography

Acerbi, C. & Tasche, D. (2002), 'Expected shortfall: A natural coherent alternative to value at risk', *Economic Notes* 31, pp. 379–388.

Ait-Sahalia, Y., Mykland, P. A. & Zhang, L. (2005), 'How often to sample a continuous-time process in the presence of market microstructure noise', *The Review of Financial Studies* 18 (2), pp. 351–416.

Andersen, T. G., Benzoni, L. & Lund, J. (2002), 'An empirical investigation of continuous-time equity return models', *The Journal of Finance* 57 (3), pp. 1239–1284.

Andersen, T. G. & Bollerslev, T. (1998), 'Answering the skeptics: Yes, standard volatility models do provide accurate forecasts', *International Economic Review* 39 (4)(4), pp. 885–905.

Andersen, T. G., Bollerslev, T., Diebold, F. X. & Ebens, H. (2001), 'The distribution of realized stock return volatility', *Journal of Financial Economics* 61 (1), pp. 43–76.

Andersen, T. G., Bollerslev, T., Diebold, F. X. & Labys, P. (2003), 'Modeling and forecasting realized volatility', *Econometrica* 71 (2)(2), pp. 579–625.

Andersen, T. G., Bollerslev, T. & Meddahi, N. (2005), 'Correcting the errors: Volatility forecast evaluation using high-frequency data and realized volatilities', *Econometrica* 73 (1), pp. 279–296.

Andersen, T. G., Bollerslev, T. & Meddahi, N. (2011), 'Realized volatility forecasting and market microstructure noise', *Journal of Econometrics* 160 (1), pp. 220–234.

Anderson, H. M. & Vahid, F. (2007), 'Forecasting the volatility of Australian stock returns', *Journal of Business and Economic Statistics* 25 (1), pp. 76–90.

Angelidis, T. & Degiannakis, S. (2008), 'Volatility forecasting: Intra-day versus inter-day models', *Journal of International Financial Markets, Institutions & Money* 18, pp. 449–465.

Bajgrowicz, P., Scaillet, O. & Treccani, A. (2013), 'Jumps in high-frequency data: Spurious detections, dynamics, and news', *SSRN* .

Bandi, F. M. & Russell, J. R. (2008), 'Microstructure noise, realized variance, and optimal sampling', *Review of Economic Studies* 75, pp. 339–369.

Bao, Y., Lee, T.-H. & Saltoglu, B. (2006), 'Evaluating predictive performance of value-at-risk models in emerging markets: A reality check', *Journal of Forecasting* 25, pp. 101–128.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A. & Shephard, N. (2008*a*), 'Designing realized kernels to measure the ex post variation of equity prices in the presence of noise', *Econometrica* 76 (6), pp.1481–1536.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A. & Shephard, N. (2008*b*), 'Multivariate realized kernels: Consistent positive semidefinite estimators of the covariation of equity prices with noise and nonsynchronous trading', *Unpublished Manuscript* .

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A. & Shephard, N. (2009), 'Realized kernels in practice: Trades and quotes', *Econometrics Journal* 12, pp. C1–C32.

Barndorff-Nielsen, O. E. & Shephard, N. (2002*a*), 'Econometric analysis of realized volatility and its use in estimating stochastic volatility models', *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* 64 (2)(2), pp. 253–280.

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

Barndorff-Nielsen, O. E. & Shephard, N. (2004), 'Power and bipower variation with stochastic volatility and jumps', *Journal of Financial Econometrics* 2 (1), pp. 1–37.

Barndorff-Nielsen, O. E. & Shephard, N. (2006), 'Econometrics of testing for jumps in financial economics using bipower variation', *Journal of Financial Econometrics* 4 (1), pp. 1–30.

Barone-Adesi, G., Giannopoulos, K. & Vosper, L. (2002), 'Backtesting derivative portfolios with filtered historical simulations (fhs)', *European Financial Management* 8 (1), pp. 31–58.

Basel Committee On Banking Supervision (2012), Fundamental review of the trading book, Technical report, Bank for International Settlements.

Berkowitz, J. & O'Brien, J. (2002), 'How accurate are value-at-risk models at commercial banks?', *The Journal of Finance* 57 (3), pp. 1093–1111.

Bollerslev, T. (1986), 'Generalized autoregressive conditional heteroskedasticity', *Journal of Econometrics* 31, pp. 307–327.

Bollerslev, T. & Zhou, H. (2002), 'Estimating stochastic volatility diffusion using conditional moments of integrated volatility', *Journal of Econometrics* 109, pp. 33–65.

Bowers, C. T. & Heaton, C. (2014*a*), 'Bootstrapping daily returns', *SSRN* .

Bowers, C. T. & Heaton, C. (2014*b*), 'An empirical analysis of value-at-risk forecasting models', *SSRN* .

Bowers, C. T. & Heaton, C. (2014*c*), 'Ranking intraday volatility estimators using empirical criteria', *SSRN* .

Brownlees, C. T. & Gallo, G. M. (2010), 'Comparison of volatility measures: a risk management perspective', *Journal of Financial Econometrics* 8 (1), pp. 29–56.

Carrasco, M. & Chen, X. (2002), 'Mixing and moment properties of various garch and stochastic volatility models', *Econometric Theory* 18 (1), pp. 17–39.

Chernov, M., Gallant, A. R., Ghysels, E. & Tauchen, G. (2003), 'Alternative models for stock price dynamics', *Journal of Econometrics* 116, pp. 225–257.

Christensen, K. & Podolskij, M. (2007), 'Realized range-based estimation of integrated variance', *Journal of Econometrics* 141, pp. 323–349.

Christoffersen, P. F. (1998), 'Evaluating interval forecasts', *International Economic Review* 39 (4), pp. 841–862.

Christoffersen, P., Hahn, J. & Inoue, A. (2001), 'Testing and comparing value-at-risk measures', *Journal of Empirical Finance* 8, pp. 325–342.

Clark, T. E. & McCracken, M. W. (2001), 'Tests of equal forecast accuracy and encompassing for nested models', *Journal of Econometrics* 105, pp. 85–110.

Clements, M. P., Galvao, A. B. & Kim, J. H. (2008), 'Quantile forecasts of daily exchange rate returns from forecasts of realized volatility', *Journal of Empirical Finance* 15, pp. 729–750.

Davidson, J. (1994), *Stochastic Limit Theory*, Oxford University Press.

Davidson, J. (2002), 'Establishing conditions for the functional central limit theorem in nonlinear and semiparametric time series processes', *Journal of Econometrics* 106, pp. 243–269.

Davidson, J. (2004), 'Moment and memory properties of linear conditional heteroscedasticity models, and a new model', *Journal of Business and Economic Statistics* 22 (1), pp. 16–29.

Diebold, F. X. & Mariano, R. S. (1995), 'Comparing predictive accuracy', *Journal of Business and Economic Statistics* 13 (3), pp. 253–263.

Dionne, G., Duchesne, P. & Pacurar, M. (2009), 'Intraday value at risk (IVaR) using tick-by-tick data with application to the Toronto stock exchange', *Journal of Empirical Finance* 16, pp. 777–792.

Dovonon, P., Goncalves, S. & Meddahi, N. (2013), 'Bootstrapping realized multivariate volatility measures', *Journal of Econometrics* 172 (1), pp. 49–65.

Efron, B. (1979), 'Bootstrap methods: Another look at the jackknife', *The Annals of Statistics* 7 (1), pp. 1–26.

Engle, R. F. & Manganelli, S. (2004), 'CAViaR: Conditional autoregressive value at risk by regression quantiles', *Journal of Business and Economic Statistics* 22 (4), pp. 367–381.

Engle, R. F. & Russell, J. R. (1998), 'Autoregressive conditional duration: A new model for irregularly spaced transaction data', *Econometrica* 66 (5), pp. 1127–1162.

Fleming, J., Kirby, C. & Ostdiek, B. (2001), 'The economic value of volatility timing', *The Journal of Finance* 56 (1).

Fleming, J., Kirby, C. & Ostdiek, B. (2003), 'The economic value of volatility timing using "realised" volatility', *Journal of Financial Economics* 67.

Fuertes, A.-M. & Olmo, J. (2013), 'Optimally harnessing inter-day and intra-day information for daily value-at-risk prediction', *International Journal of Forecasting* 29, pp. 28–42.

Gaglianone, W. P., Lima, L. R., Linton, O. & Smith, D. R. (2011), 'Evaluating value-at-risk models via quantile regression', *Journal of Business and Economic Statistics* 29 (1), pp. 150–160.

Giacomini, R. & White, H. (2006), 'Tests of conditional predictive ability', *Econometrica* 74 (6), pp. 1545–1578.

Giot, P. (2005), 'Market risk models for intraday data', *The European Journal of Finance* 11 (4), pp. 309–324.

Giot, P. & Grammig, J. (2006), ‘How large is liquidity risk in an automated auction market?’, *Empirical Economics* 30, pp. 867–887.

Giot, P. & Laurent, S. (2003), ‘Value-at-risk for long and short trading positions’, *Journal of Applied Econometrics* .

Giot, P. & Laurent, S. (2004), ‘Modelling daily value-at-risk using realized volatility and arch type models’, *Journal of Empirical Finance* 11, pp.379–398.

Glosten, L. R., Jagannathan, R. & Runkle, D. E. (1993), ‘On the relation between the expected value and the volatility of the nominal excess return on stocks’, *The Journal of Finance* 48 (5), pp. 1779–1801.

Goncalves, S. & de Jong, R. (2003), ‘Consistency of the stationary bootstrap under weak moment conditions’, *Economics Letters* 81, pp. 273–278.

Goncalves, S., Hounyo, U. & Meddahi, N. (Forthcoming), ‘Bootstrap inference for pre-averaged realized volatility based on non-overlapping returns’, *Journal of Financial Econometrics* .

Goncalves, S. & Meddahi, N. (2009), ‘Bootstrapping realized volatility’, *Econometrica* 77 (1), pp. 283–306.

Goncalves, S. & White, H. (2002), ‘The bootstrap of the mean for dependent heterogeneous arrays’, *Econometric Theory* 18 (6), pp. 1367–1384.

Hansen, B. E. (1991), ‘Garch(1,1) processes are near epoch dependent’, *Economics Letters* 36, pp. 181–186.

Hansen, B. E. (1994), ‘Autoregressive conditional density estimation’, *International Economic Review* 35 (3), pp. 705–730.

Hansen, P. R. (2005), ‘A test for superior predictive ability’, *Journal of Business and Economic Statistics* 23 (4), pp. 365–380.

Hansen, P. R. & Lunde, A. (2005*a*), ‘A forecast comparison of volatility models: Does anything beat a garch(1,1)?’, *Journal of Applied Econometrics* 20 (7), pp. 873–889.

Hansen, P. R. & Lunde, A. (2005*b*), ‘A realized variance for the whole day based on intermittent high-frequency data’, *Journal of Financial Econometrics* 3 (4), pp. 525–554.

Hansen, P. R. & Lunde, A. (2006*a*), ‘Consistent ranking of volatility models’, *Journal of Econometrics* 131, pp. 97–121.

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

Hansen, P. R. & Lunde, A. (2010), 'Estimating the persistence and the autocorrelation function of a time series that is measured with error', *Working paper* .

Hansen, P. R., Lunde, A. & Nason, J. M. (2011), 'The model confidence set', *Econometrica* 79 (2), pp. 453–497.

Hogg, R. V. (1974), 'Adaptive robust procedures: A partial review and some suggestions for future applications', *Journal of the American Statistical Association* 69 (348), pp. 909–923.

Hsu, P.-H., Hsu, Y.-C. & Kuan, C.-M. (2010), 'Testing the predictive ability of technical analysis using a new stepwise test without data snooping bias', *Journal of Empirical Finance* 17 (3), pp. 471–484.

Huang, X. & Tauchen, G. (2005), 'The relative contribution of jumps to total price variance', *Journal of Financial Econometrics* 3 (4), pp. 456–499.

Huber, P. J. (1981), *Robust Statistics*, John Wiley & Sons.

Hwang, E. & Shin, D. W. (2013), 'Stationary bootstrapping realized volatility under market microstructure noise', *Electronic Journal of Statistics* 7, pp. 2032–2053.

Jacod, J., Li, Y., Mykland, P. A., Podolskij, M. & Vetter, M. (2009), 'Microstructure noise in the continuous case: The pre-averaging approach', *Stochastic Processes and their Applications* 119, pp. 2249–2276.

Jacod, J. & Protter, P. (2011), *Discretization of Processes*, Springer.

Jouini, E. & Kallal, H. (1995), 'Martingales and arbitrage in securities markets with transaction costs', *Journal of Economic Theory* 66, pp. 178–197.

Komunjer, I. (2005), 'Quasi-maximum likelihood estimation for conditional quantiles', *Journal of Econometrics* 128, pp. 137–164.

Koopman, S. J., Jungbacker, B. & Hol, E. (2005), 'Forecasting daily variability of the s&p 100 stock index using historical, realised and implied volatility measurements', *Journal of Empirical Finance* 12, pp. 445–475.

Kuester, K., Mittnik, S. & Paolella, M. S. (2006), 'Value-at-risk prediction: A comparison of alternative strategies', *Journal of Financial Econometrics* 4 (1), pp. 53–89.

Kupiec, P. H. (1995), 'Techniques for verifying the accuracy of risk measurement models', *The Journal of Derivatives* 3 (2), pp. 73–84.

Lahiri, S. N. (1999), 'Theoretical comparisons of block bootstrap methods', *The Annals of Statistics* 27 (1), pp. 386–404.

Lahiri, S. N. (2003), *Resampling Methods for Dependent Data*, Springer-Verlag New York.

Large, J. (2011), 'Estimating quadratic variation when quoted prices change by a constant increment', *Journal of Econometrics* 160, pp. 2–11.

Laurent, S., Rombouts, J. V. K. & Violante, F. (2013), 'On loss functions and ranking forecasting performances of multivariate volatility models', *Journal of Econometrics* 173, pp. 1–10.

Liu, L., Patton, A. J. & Sheppard, K. (2013), 'Does anything beat 5-minute rv? a comparison of realized measures across multiple asset classes', *Working paper* .

Lopez, J. A. (1998), 'Methods for evaluating value-at-risk estimates', *FRBNY Economic Policy Review* 4 (3), pp. 119–124.

Markowitz, H. M. (1991), *Portfolio Selection: Efficient Diversification of Investments*, John Wiley & Sons.

Mehta, A., Neukirchen, M., Pfetsch, S. & Poppensieker, T. (2012), 'Managing market risk: Today and tomorrow', *McKinsey Working Papers on Risk* 32, pp. 1–16.

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

Newey, W. A. & West, K. D. (1987), 'A simple, positive semi-definite, heteroskdasticity and autocorrelation consistent covariance matrix', *Econometrica* 55 (3), pp. 703–708.

Nordman, D. J. (2009), 'A note on the stationary bootstrap's variance', *The Annals of Statistics* 37 (1), pp. 359–370.

Noureldin, D., Shephard, N. & Sheppard, K. (2012), 'Multivariate high-frequency-based volatility (heavy) models', *Journal of Applied Econometrics* 27 (6), pp. 907–933.

Paparoditis, E. & Politis, D. N. (2001), 'Tapered block bootstrap', *Biometrika* 88 (4), pp. 1105–1119.

Paparoditis, E. & Politis, D. N. (2002), 'The tapered block bootstrap for general statistics from stationary sequences', *The Econometrics Journal* 5 (1), pp. 131–148.

Patton, A. J. (2011*a*), 'Data-based ranking of realised volatility estimators', *Journal of Econometrics* 161, pp. 284–303.

Patton, A. J. (2011*b*), 'Volatility forecast comparison using imperfect volatility proxies', *Journal of Econometrics* 160, pp. 246–256.

Patton, A. J. & Li, J. (2013), 'Asymptotic inference about predictive accuracy using high frequency data', *SSRN* .

Patton, A. J. & Sheppard, K. (2009), 'Optimal combinations of realised volatility estimators', *International Journal of Forecasting* 25, pp. 218–238.

Patton, A., Politis, D. N. & White, H. (2009), 'Correction to: Automatic block-length selection for the dependent bootstrap, by D. Politis and H. White', *Econometric Reviews* 28 (4), pp. 372–375.

Politis, D. N. & Romano, J. P. (1994*a*), 'Large sample confidence regions based on subsamples under minimal assumptions', *The Annals of Statistics* 22 (4), pp. 2031–2050.

Politis, D. N. & Romano, J. P. (1994*b*), 'The stationary bootstrap', *Journal of the American Statistical Association* 89 (428), pp. 1303–1313.

Politis, D. N., Romano, J. P. & Wolf, M. (1999), *Subsampling*, Springer-Verlag New York.

Politis, D. N. & White, H. (2004), 'Automatic block-length selection for the dependent bootstrap', *Econometric Reviews* 23 (1), pp. 53–70.

Poon, S. & Granger, C. W. J. (2003), 'Forecasting volatility in financial markets: A review', *Journal of Economic LIterature* 41 (2), pp. 478–539.

R Core Team (2014), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.

Romano, J. P. & Wolf, M. (2005), 'Stepwise multiple testing as formalized data snooping', *Econometrica* 73 (4), pp. 1237–1282.

Rosenbaum, M. (2009), 'Integrated volatility and round-off error', *Bernoulli* 15 (3), pp. 687–720.

Rowe, D. (2012), 'The false promise of expected shortfall', *http://www.risk.net/risk-magazine/opinion/2216891/the-false-promise-of-expected-shortfall* .

Sener, E., Baronyan, S. & Menguturk, L. A. (2012), 'Ranking the predictive performances of value-at-risk estimation methods', *International Journal of Forecasting* 28, pp. 849–873.

Shao, X. (2010), 'Extended tapered block bootstrap', *Statistica Sinica* 20, pp. 807–821.

Shao, X.-D., Lian, Y.-J. & Yin, L.-Q. (2009), 'Forecasting value-at-risk using high frequency data: The realised range model', *Global Finance Journal* 20, pp. 128–136.

Shephard, N. & Sheppard, K. (2010), 'Realising the future: Forecasting with high-frequency-based volatility (heavy) models', *Journal of Applied Econometrics* 25 (2), pp. 197–231.

Song, K. (2012), 'Testing predictive ability and power robustification', *Journal of Business and Economic Statistics* 30 (2), pp. 288–296.

Sun, Y. (2006), 'Best quadratic unbiased estimators of integrated variance in the presence of market microstructure noise', *Working paper* .

West, K. D. (1996), 'Asymptotic inference about predictive ability', *Econometrica* 64 (5), pp. 1067–1084.

White, H. (2000), 'A reality check for data snooping', *Econometrica* 68 (5), pp. 1097–1126.

Wright, J. H. (1999), 'Testing for a unit root in the volatility of asset returns', *Journal of Applied Econometrics* 14 (3), pp. 309–318.

Wu, C. F. J. (1986), 'Jackknife, bootstrap and other resampling methods in regression analysis', *The Annals of Statistics* 14 (4), pp. 1261–1295.

Zhang, L. (2006), 'Efficient estimation of stochastic volatility using noisy observations: A multi-scale approach', *Bernoulli* 12 (6), pp. 1019–1043.

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

Zhou, B. (1996), 'High-frequency data and volatility in foreign-exchange rates', *Journal of Business & Economic Statistics* 14 (1), pp. 45–52.