Improving Probabilistic Forecasts by Using Intra-Day Data: An Application to Financial and Temperature Data

Xiaochun Meng

Merton College

University of Oxford

Supervisor: Professor James W. Taylor

A thesis submitted for the degree of *Doctor of Philosophy* in Management Studies

1st June 2018

Abstract

The thesis consists of three studies. The first two contribute to financial market risk modelling and the third contributes to the modelling of temperature extremes.

Value at risk (VaR) is a popular measure of market risk. The first study proposes new approximate long-memory VaR models that incorporate intra-day price ranges. These models use lagged intra-day range with the feature of considering different range components calculated over different time horizons. We also investigate the impact of the market overnight return on the VaR forecasts, which has not yet been considered with the range in VaR estimation. Model estimation is performed using linear quantile regression. An empirical analysis is conducted on 18 market indices. In spite of the simplicity of the proposed methods, the empirical results show that they successfully capture the main features of the financial returns and are competitive with established benchmark methods. The empirical results also show that several of the proposed range-based VaR models, utilising both the intra-day range and the overnight returns, are able to outperform generalised autoregressive conditional heteroskedasticity (GARCH) based methods and conditional autoregressive value at risk (CAViaR) models.

The second study contributes to the literature of VaR and expected shortfall (ES) forecasting. To convey information regarding potential exceedances beyond the VaR, ES has been advocated for future regulatory frameworks. However, the estimation of VaR and ES is challenging, as it requires the estimation of the tail behaviour of daily returns. Furthermore, ES is not elicitable, which means it is difficult to estimate and evaluate. In this paper, we take advantage of recent research that shows that VaR and ES are jointly elicitable, and that provides scoring functions for the joint estimation and evaluation of these two risk measures. We consider the use of intra-day data in this context. We focus on the intra-day range, which is the absolute difference between the highest and lowest intra-day log prices. In contrast to intra-day observations, the intra-day low and high are

widely available for many financial assets. To alleviate the challenge of modelling extreme risk measures, we propose the use of the intra-day low series. We draw on a theoretical result for Brownian motion to show that a quantile of the daily returns can be estimated as the product of a constant term and a less extreme quantile of the intra-day low returns, which we define as the difference between the lowest log price of the day and the log closing price of the previous day. In view of this, we use estimates of the VaR and ES of the intra-day low returns to estimate the VaR and ES of the daily returns. We provide empirical support for the new proposals using daily stock index data.

The third study contributes to extreme temperature forecasting, which is related to environmental risk. Understanding changes in the frequency, severity and seasonality of daily temperature extremes is important for public policy decisions regarding heat waves and cold snaps. Temperature forecasts are needed to trigger warnings, and to enable adequate preparation of public services. The uncertainty in the weather implies that forecasting daily temperature is inherently a probabilistic forecasting problem. Autoregressive moving average and GARCH (ARMA-GARCH) models have been applied to daily temperature time series to produce density forecasts. However, a heat wave is sometimes defined in terms of both the daily minimum and maximum temperature, which necessitates the generation of forecasts of the joint distribution of these two variables. We consider the modelling of the daily minimum and maximum temperature using a bivariate vector ARMA and multivariate GARCH (VARMA-MGARCH) model, with conditional dependency modelled using a dynamic copula. A useful by-product is the implicit modelling of the daily average and the diurnal temperature range, which has been used as an index of climate change. Using Spanish data recorded over a 65-year period, we find that a bivariate VARMA-MGARCH model is able to outperform univariate models in terms of the forecast accuracy of both the marginal and joint distributions.

Acknowledgements

First and foremost, I would like to express my sincerest gratitude to my supervisor, Prof. James W. Taylor for his continuous guidance and generous support for my DPhil research. I would like to thank Dr. Jooyoung Jeon, Dr. Ho-Yin Mak, Dr. Dolores Morales and Dr. Athanasios Tsanas for their useful comments and suggestions for improving my thesis. I would also like to thank my parents Yuwei and Zhaowen for their loving support for me.

Contents

List of Figures

List of Tables

Table 1: Table of Abbreviations.

AL score	Scores based on asymmetric Laplace distribution
AR	Autoregressive model
ARMA	Autoregressive moving average model
CAViaR	Conditional autoregressive value at risk model
CAViaR-AS	CAViaR asymmetric slope model
CAViaR-FZ	CAViaR model estimated using FZ score
CAViaR-FZ-AS	CAViaR-AS model estimated using FZ score
CAViaR-FZ-Range	CAViaR-Range model estimated using FZ score
CAViaR-FZ-SAV	CAViaR-SAV model estimated using FZ score
CAViaR-IndG	CAViaR indirect GARCH model
CAViaR-RV	CAViaR model based on realised volatility
CAViaR-SAV	CAViaR symmetric absolute value model
CAViaR-RV-N	CAViaR model based on realised volatility and the overnight return
CAViaR-Range	CAViaR model based on the intra-day range
CAViaR-Range-C	CAViaR model based on the close-to-close range
CAViaR-Range-N	CAViaR model based on the intra-day range and the overnight return
CDF	Cumulative distribution function
CRPS	Continuous ranked probability score
DTR	Diurnal temperature range
ES	Expected shortfall
EVT	Extreme value theory
FIGARCH	Fractionally integrated GARCH model
FZ score	Scores proposed by Fissler and Ziegel (2016)
GARCH	Generalised autoregressive conditional heteroskedasticity model
GARCH-t	Generalised autoregressive conditional heteroskedasticity model with Student's t distribution
Gen asym skew-t	Generalised asymmetric skew-t distribution
GJR-GARCH	The GARCH model proposed by Glosten, Jagannathan and Runkle
GJR-GARCH-t	The GARCH model proposed by Glosten, Jagannathan and Runkle with Student's t distribution
MGARCH	Multivariate GARCH model
NZ score	Scores proposed by Nolde et al. (2017)
HAR	Heterogeneous autoregressive model
HAR-Range	HAR model for the intra-day range
HAR-Range-t	HAR model for the intra-day range with Student's t distribution

Table 2: Table of Abbreviations (cont.).

HAR-RV	HAR model for realised volatility
HAR-RV-t	HAR model for realised volatility with Student's t distribution
PDF	Probability density function
PIT	Probability integral transform
QRHAR	Quantile regression HAR model
QRHAR-Range	QRHAR model for intra-day range
QRHAR-Range-C	QRHAR model for the intra-day volatility measure based on the close-to-close intra-day range
QRHAR-Range-N	QRHAR model for the intra-day volatility measure based on the intra-day range and the overnight return
QRHAR-RV	QRHAR model for realised volatility
QRHAR-RV-N	QRHAR model for the intra-day volatility measure based on realised volatility and the overnight return
Realised-GARCH	Realised GARCH model
Realised-GARCH-Range	Realised GARCH model for intra-day range
Realised-GARCH-Range-t	Realised GARCH model for intra-day range with Student's t distribution
RV	Realised volatility
Realised-GARCH-RV	Realised GARCH model for realised volatility
Realised-GARCH-RV-t	Realised GARCH model for realised volatility with Student's t distribution
SBC	Schwarz bayesian criterion
TN	Daily minimum temperature
TX	Daily maximum temperature
TG	Daily average temperature
VaR	Value at risk
VARMA	Vector ARMA model

Table 3: Table of Mathematical Expressions.

$a(\cdot)$	One function in FZ score
$\boldsymbol{A}_T$	One matrix obtained for daily returns from in-sample period of length T, involved in Appendix 4.5
$\tilde{\boldsymbol{A}}_T$	Matrix obtained for intra-day low returns from in-sample period of length T, involved in Appendix 4.5
$\lvert\cdot\rvert$	Absolute value
α_i	Scalar parameter
$\boldsymbol{\beta}$	Vector of parameters
β_i	Scalar parameter
$\boldsymbol{\beta}^0$	True vector of parameters for daily returns, involved in Appendix 4.5
$\tilde{\boldsymbol{\beta}}^0$	True vector of parameters for intra-day low returns, involved in Appendix 4.5
$\hat{\boldsymbol{\beta}}_T$	Parameter estimator using in-sample observations of length T, involved in Appendix 4.5
$\boldsymbol{B}_{j,i}$	Matrix of parameters
$close_t$	Daily closing log price
$\boldsymbol{C}$	Matrix
$C(\cdot,...,\cdot)$	Copula CDF
$C_{\boldsymbol{R}_t,\nu}(\cdot,...,\cdot)$	Student's t or Gaussian copula CDF, where ν is the degrees of freedom for the Student's t copula (ν is ∞ for the Gaussian copula), and $\boldsymbol{R}_t$ is the correlation parameter.
d	Fractionally integrated parameter for FIGARCH model
$d(\cdot)$	Repeating step function that numbers the days of the year from 1 to 365 within each year
$\boldsymbol{D}_T$	Matrix obtained for daily returns from in-sample period of length T, involved in Appendix 4.5
$\tilde{\boldsymbol{D}}_T$	Matrix obtained for intra-day low returns from in-sample period of length T, involved in Appendix 4.5
dim	Dimension of a vector
Δ	Divide the market interval span of market opening hours S equally into M intervals
$\boldsymbol{E}_i$	Matrix of parameters
$ES(\theta)$	θ expected shortfall of a distribution
ES_t	Conditional expected shortfall of the corresponding time series
$\lVert\cdot\rVert$	Euclidean norm
$\mathbb{E}(\cdot)$	Expectation of a function of random variables
$\mathbb{E}_{F_t}(\cdot)$	Expectation of a random variable with CDF F_t
$\boldsymbol{\varepsilon}_t$	Residual of ARMA model
$\boldsymbol{\epsilon}_t$	Residual of VARMA model
η_t	Standardised error with mean 0 and variance 1
$\tilde{\eta}_t$	Standardised error with mean 0 and variance 1
$\boldsymbol{\eta}_t$	Standardised error with mean $\mathbf{0}$ and identity covariance matrix
es_t	Conditional expected shortfall of the corresponding time series
$\widetilde{es}_t$	Conditional expected shortfall of the intra-day low returns, involved in Appendix 4.5

Table 4: Table of Mathematical Expressions (cont.).

$f(\cdot)$	PDF
$f_t(\cdot)$	Conditional PDF of the corresponding time series
$F(\cdot)$	CDF of the corresponding distribution
$F_t(\cdot)$	Conditional CDF of the corresponding time series
$F_{\varepsilon_{i,t}}(\cdot)$	CDF of $\varepsilon_{i,t}$
$F_{\boldsymbol{\epsilon}_t}(\cdot)$	Bivariate CDF of $\boldsymbol{\epsilon}_t$
$F^{-1}(\cdot)$	Inverse CDF function (or quantile function)
γ_i	Scalar parameter
g_T	Function for daily returns from in-sample period of length T, involved in Appendix 4.5
$\tilde{g}_T$	Function for intra-day low returns from in-sample period of length T, involved in Appendix 4.5
$G_1(\cdot)$	Function in FZ score
$G_2(\cdot)$	Function in FZ score
$\Gamma(\cdot)$	Gamma function
$\nabla(\cdot)$	Gradient of a function
$High_t$	Intra-day highest log price
$\boldsymbol{H}_t$	Covariance matrix, or a positive definite matrix used in VARMA-MGARCH model with copula
$h_{j,t}$	Variance of $\varepsilon_{j,t}$
$\mathscr{F}_t$	Information set up to day t
$\mathbb{1}(\cdot)$	Indicator function
k	Number of parameters
l_t	log likelihood
L	Lag operator
Low_t	Intra-day lowest log price
$\boldsymbol{\lambda}_i$	Vector of parameters
m	AR order in ARMA or VARMA model
M	Number of intervals the interval span of market opening hours is divided into
μ_t	Conditional mean of a univariate time series
$\boldsymbol{\mu}_t$	Conditional mean of a multivariate time series
n	MA order in ARMA or VARMA model
ν	Degrees of freedom parameter for Student's t distribution or Student's t copula
ν_i	Degrees of freedom for generalised asymmetric skew-t distribution
$open_t$	Daily opening log price
o	GJR order in GARCH or MGARCH model
$\upsilon_i(x)$	Polynomial
ω	Scalar parameter
$\boldsymbol{\Omega}$	Matrix of parameters
$\mathscr{G}_2(x)$	One function in FZ score
p	ARCH order in GARCH or MGARCH model
$P_{\cdot,\cdot}$	Intra-day log price of a financial return

Table 5: Table of Mathematical Expressions (cont.).

$\Pr(\cdot)$	Probability of an event
ϕ_i	Vector of parameters
ψ_i	Vector of parameters
$\mathbf{\Psi}_t$	Vector of explanatory variables
q	GARCH order in GARCH or MGARCH model
q_t	Conditional quantile of the corresponding time series
$\tilde{q}_t$	Conditional quantile of the intra-day low returns, involved in Appendix 4.5
$Range_t$	Intra-day Range
$Range_{N,t}$	Intra-day range with overnight return
$Range_{NC,t}$	Close-to-close range
RV_t	Realised volatility
$RV_{N,t}$	Realised volatility with overnight return
$\boldsymbol{R}$	Unconditional correlation matrix of Gaussian or Student's t copula
R_t	Conditional correlation parameter
$\boldsymbol{R}_t$	Conditional correlation matrix of Gaussian or Student's t copula
r_t	Stochastic part of the conditional mean of a univariate time series when it contains a deterministic part
$\boldsymbol{r}_t$	Stochastic part of the conditional mean of a multivariate time series when it contains a deterministic part
$\boldsymbol{\rho}_i$	Vector of parameters
s	Scale parameter of a generalised Pareto distribution
S	Interval span of market opening hours
$SE(\cdot,\cdot)$	Squared error
σ_t	Conditional volatility of a time series
T	Number of observations
T_t	Univariate temperature series
$\boldsymbol{T}_t$	Bivariate temperature series
$(\cdot)'$	Derivative of a function
$(\cdot)^T$	Transpose of a matrix
θ	Probability level
$\hat{\theta}$	Probability level
$\text{vech}(\cdot)$	Operator that stacks the lower triangular portion of a matrix as a column vector
ξ	Skewness parameter for generalised asymmetric skew-t distribution
u_i	Variable between 0 and 1
v	Number of lags involved in the DQ test
$\{Y_1, Y_2, ...\}$	Univariate time series
y_t	Realisation of $\{Y_1, Y_2, ...\}$ for day t
$y_{N,t}$	Overnight return
$\{\boldsymbol{Y}_1, \boldsymbol{Y}_2, ...\}$	Multivariate time series
$\boldsymbol{y}_t$	Realisation of $\{\boldsymbol{Y}_1, \boldsymbol{Y}_2, ...\}$ for day t
ζ	Shape parameter of a generalised Pareto distribution
Z_t^i	Univariate random variable
$\boldsymbol{Z}_t^i$	Multivariate random variable

Table 6: Glossary.

Asset	Economic resource, typical examples equities, commodities and high-yield bonds
Diurnal Range	The range between the daily maximum and minimum temperature
Daily Mean Temperature	Average of the daily maximum and minimum temperature
Expected Shortfall	The expected value of exceedances beyond the VaR
Forecast	It is the estimate of a future value of a variable
Intra-day Low	The lowest log price within a day
Intra-day Low Return	The difference between the lowest log price of a day and the log closing price of the previous day
Intra-day High	The highest log price within a day
Intra-day Range	The absolute difference between the intra-day high and the intra-day low
Overnight Return	The difference between the log opening price of a day and the log closing price of the previous day
Quantile	The θ quantile of a distribution is defined to be $\mathrm{F}^{-1}(\theta) = \inf\{x : \mathrm{F}(x) \geq \theta\}$ where F is the CDF
Daily Return	The difference between the log closing price of a day and the log closing price of the previous day
Risk	'(exposure to) the possibility of loss, injury, or other adverse or unwelcome circumstance'. Major types of risks are market risk, liquidity risk, operational risk and credit risk.
VaR	Value at risk, which is defined to be the quantile of a distribution corresponding to some probability level
Volatility	The square root of the second moment of a financial return

Chapter 1

Introduction

1.1 Motivation and Background Information

We are often exposed to different types of risk. Risk, according to the Oxford English dictionary, is defined as '*(exposure to) the possibility of loss, injury, or other adverse or unwelcome circumstance*'. The thesis contributes to probabilistic forecasting in risk management. The thesis consists of three studies. The first two (Chapters 3 and 4) focus on market risk, and the third (Chapter 5) focuses on environmental risk.

In financial markets, there are various types of risk, such as market risk, liquidity risk, operational risk and credit risk. Market risk is the risk to a financial portfolio from movements in market prices, liquidity risk is the risk that an asset might have to be sold below its fundamental value, operational risk is the risk of loss due to errors in the operation, and credit risk is the risk that a counterparty may be unable to fulfill its obligation on the agreed date (Ross et al. 2011). These risks can cause serious loss to the investors if not dealt with properly. We focus exclusively on financial market risk. Other types of risk require special treatment that lie beyond the scope of this thesis.

In particular, market risk contains uncertainty about the future price of an *asset*. Asset means economic resource, typical examples of assets are equities, commodities and high-yield bonds. Two important market risk measures are discussed in this thesis: *value at risk* (VaR) and *expected shortfall* (ES). Mathematically, VaR is defined to be the quantile

of a probabilistic distribution, with a certain probability level. In this document, we will sometimes use 'VaR' interchangeably with 'quantile'. The θ quantile of a distribution is defined to be

$$F^{-1}(\theta) = \inf\{x : F(x) \geq \theta\}$$

where F is the *cumulative distribution function* (CDF) of the distribution; and $F^{-1}(\cdot)$ is referred to as the inverse CDF or the quantile function. The ES associated with probability θ of this distribution is defined to be the expectation of the variable with this distribution beyond the θ quantile, which can be expressed as follows:

$$ES(\theta) = \frac{1}{\theta} \int_{-\infty}^{F^{-1}(\theta)} x f(x)\, dx$$

where f is the *probability density function* (PDF) of the distribution. VaR is the most commonly reported market risk measure (Manganelli and Engle 2004). The first appearance of VaR was in the 1980s when it was used by major financial firms to measure the risk of a portfolio (Linsmeier and Pearson 1997). Since then, there has been a massive growth in the use of VaR. It has been widely adopted as a risk measure by many financial institutions and nonfinancial firms. The success and prevalence of VaR drew the attention of regulators. Since 1996, the Basel Committee on Banking Supervision has imposed capital requirements to banks and investment firms based on VaR estimation. Therefore, accurate forecasting for VaR is of great importance to both investors and regulators. Intuitively, it measures the amount a certain portfolio can lose over some period of time, given a probability level, which is usually chosen as 1% or 5%. On the other hand, VaR has been criticised for two major weaknesses. Firstly, VaR does not take into account the severity of the loss beyond it. Secondly, VaR is not a sub-additive risk measure, which means that the measure for a portfolio can be less than the sum of the measures of the components of the portfolio. This could potentially be a problem in practice. In view of these, ES has been proposed as a risk measure for future regulatory frameworks (Embrechts et al. 2014).

We are interested in the distribution of a time series conditional on its past information. For simplicity, we use the word 'conditional' and omit 'past information'. Time series of the type considered in this thesis typically have a time-varying conditional distribution, which cause the quantile and ES to be time-varying. Figure 1.1 plots an example of the estimated 1 and 5% % VaR and ES for S&P500. In the first two studies (Chapters 3 and 4) in this thesis, we focus on financial data. For financial data, as a convention, the log return of the price, which is defined as the log of the ratio of the price on one day to the price on the previous day, is studied. We refer to the log return simply as the 'return' in this thesis. Financial returns have the notable feature that the conditional mean has little variation around zero but there is strong autocorrelation in the squared returns (Bollerslev 1986). The small correlation together with the little variation, shown in Figure 1.1, in the returns lead to the convention of assuming the conditional mean of a financial return to be zero (see, for example, Žikeš and Baruník 2015); and the strong autocorrelation in the squared returns, shown in Figure 1.2, prompts the introduction of autoregressive models such as the *generalised autoregressive conditional heteroskedasticity* (GARCH) models for the volatility, which is the conditional standard deviation of a financial return, and the *conditional autoregressive value at risk* (CAViaR) models for direct VaR estimation. These models have been shown to perform competitively in VaR and ES estimation (see, for example, Chen et al. 2012; Hansen et al. 2012; Gerlach and Wang 2016). Therefore, we use autoregressive methods in the first two studies (Chapters 3 and 4) in this thesis to estimate the VaR and ES of the daily returns. We consider 1 day-ahead VaR and ES forecasting as the Basel Committee on Banking Supervision has imposed capital requirements to banks and investment firms on a daily basis.

VaR and ES are important for assessing financial risk, however, the full density estimates are required for some other applications. In our third study (Chapter 5), we use density estimation in environmental risk assessment. Although environmental risk does not seem to be directly related to market risk, these studies do have a unified focus, which is to utilise intra-day data to improve forecasting accuracy. In practice, extreme weather

Figure 1.1: Illustration of conditional 1% and 5% VaR and ES for S&P500.

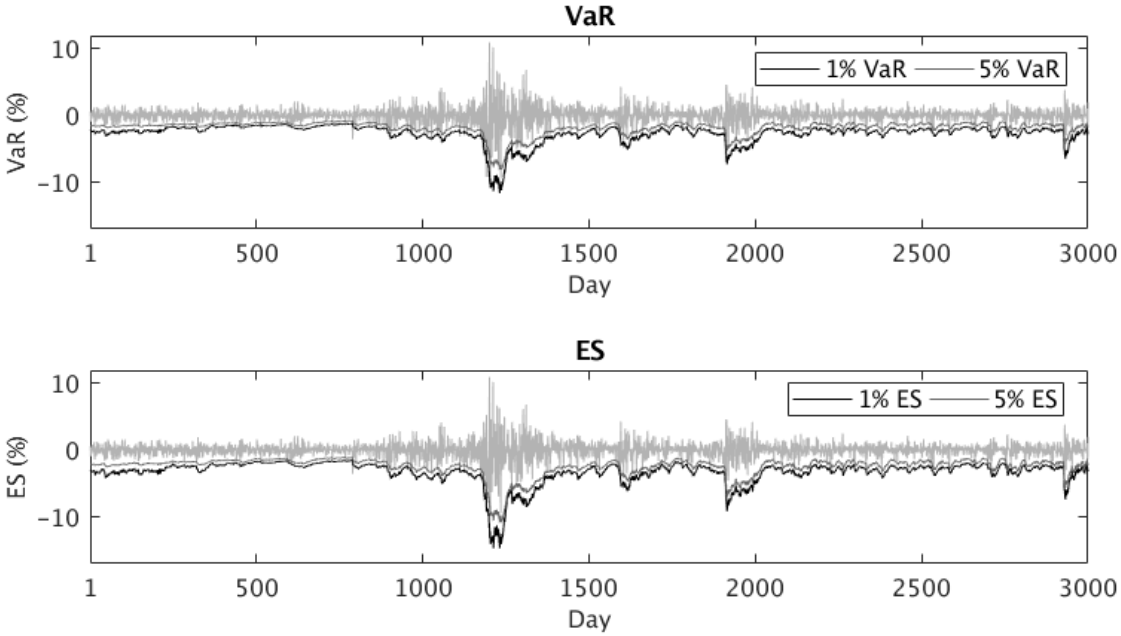
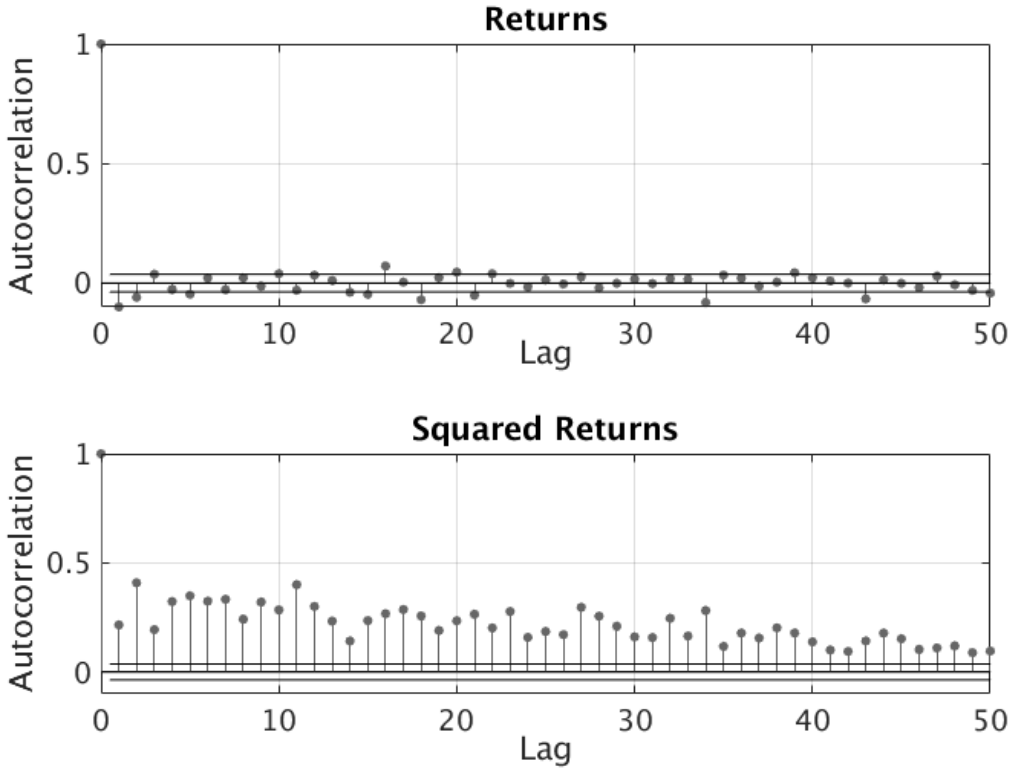


Figure 1.2: Illustration of the autocorrelation for the returns and the squared returns of S&P500.

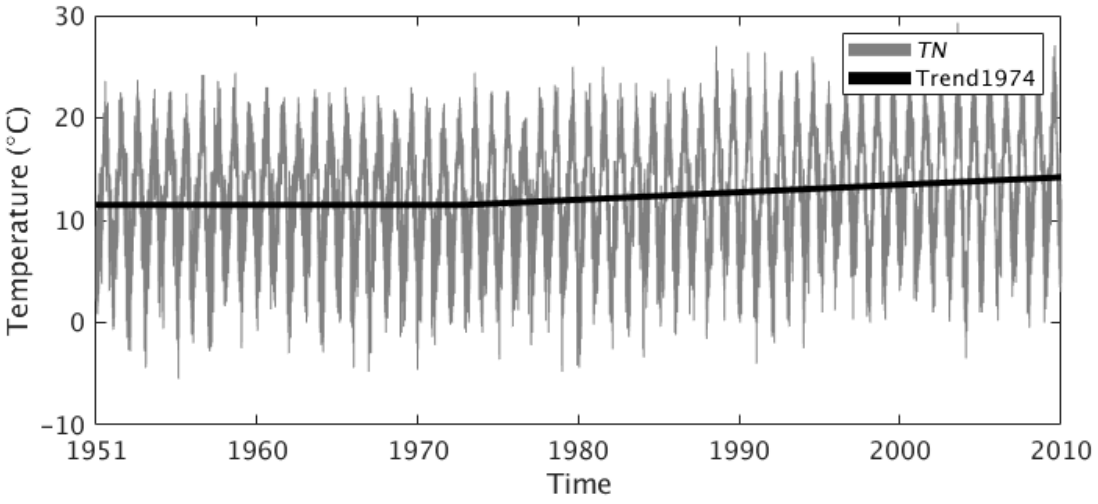


events can have huge impact on our society and ecosystem. For example, heat waves and cold snaps can cause serious health problems and even increased morbidity. The definition of a heat wave involves the exceedances of the daily minimum and maximum temperature over some thresholds over a period of days. Hence, the essence of this study is to estimate the joint distribution of the daily minimum and maximum temperature. An alternative approach to time series modelling is the numerical weather prediction (NWP) system used by meteorological offices. Through an NWP model with multiple different initial conditions, an ensemble of predictions can be obtained before they can be used as the basis of probabilistic prediction. Time series temperature models have three main advantages when compared to the NWP system. Firstly, the NWP system requires various types of weather data, which are not necessarily available for a less populated area, while time series model only requires historical temperature data, which is easy to obtain for many locations. Secondly, weather ensemble prediction is numerically challenging, as a procedure is required to convert the ensemble into a probabilistic forecast (see, for example, Taylor and Buizza 2004), while time series models are fast to estimate. Thirdly, weather ensemble prediction does not give insight into temperature evolution, while statistical inference of a time series model can deliver intuitive insight. On the other hand, temperature data exhibits very different features when compared to a financial return series. The most noticeable difference is that there is a need to model rich structure in the dynamics of the mean. Figure 1.3 plots the daily minimum temperature in Seville and the fitted trend using the proposed model that will be discussed in Chapter 5. In this figure, we can also see an annual cycle in the temperature series.

1.2 Contribution of this Document

The main contributions in the thesis are as follows. Chapter 3 investigates the benefits of including intra-day data, in VaR forecasting for financial returns. Since the first appearance of VaR in the 1980s, it has become the most prevalent risk measure and is currently a standard tool for risk management in financial and insurance institutions

Figure 1.3: Daily minimum temperature in Seville.



(Berkowitz et al. 2011). Traditional VaR methods only utilise the historical returns in model specification. With intra-day data becoming increasingly available, efforts have been made to use them in the forecasting of VaR for daily returns (see, for example, Corsi 2009; Shephard and Sheppard 2010; Chen et al. 2012). Chapter 3 considers the use of realised volatility and the intra-day range in model specification for various methods. The first part of Chapter 3 contains an up-to-date review of the established VaR methods and the methods for evaluating VaR forecasting accuracy. In the second part of Chapter 3, a comprehensive empirical study is performed for a large number of asset returns.

Chapter 4 proposes a novel use of intra-day data to estimate the quantile and ES of the daily returns. Chapter 4 is motivated by the fact that the accuracy of VaR and ES forecasting is influenced by the number of observations in the tail of the distribution of the returns. When the data is scarce in the tail, the performance of VaR and ES estimation methods may be unreliable. We find an interesting relationship between the daily return and the intra-day low series, namely, the dynamics of the VaR and ES of the daily returns can be well approximated by that of a pair of less extreme VaR and ES of the intra-day low series. Such an approach essentially provides us with more observations in the tail, and this can improve the parameter estimation efficiency for VaR and ES models. The empirical results are promising.

Chapter 5 focuses on probabilistic forecasting of heat waves. Understanding changes in the frequency, severity, and seasonality of daily temperature extremes is important for public policy decisions regarding heat waves and cold snaps. Extreme weather events can have negative impacts on our society and economy and lead to serious health problems and increased morbidity (Dupuis 2012; 2014). Understanding the evolution of the heat waves and the cold snaps requires an accurate density forecast for the daily minimum and maximum temperature. Previously proposed methods model the densities of the daily minimum and maximum temperature separately, which neglects the dependence structure between them. Our work proposes a bivariate time series model building on the work of Bollerslev et al. (1988) and Campbell and Diebold (2005), to improve the estimation of both the marginal and joint distributions of the daily minimum and maximum temperature. Empirical study shows that the proposed bivariate method outperforms the benchmark methods. We focus on daily temperature data in this paper, as heat waves are defined based on daily temperature data. However, there is scope for improvement by considering temperature data of higher frequencies, such as hourly temperature data. For example, Dupuis (2017) studies the impact of hourly temperature on daily extreme temperature. We will explore hourly temperature data in future.

Chapter 4 links very closely to Chapter 3, as they both consider the use of the intra-day data for financial data. Although Chapter 5 does not deal with financial data, it has, in common with Chapters 3 and 4, a focus on intra-day data. The daily minimum and maximum temperature are extreme observations that are conceptually similar to the intra-day lowest/highest log prices for financial data. Chapter 5 models the joint distribution of daily minimum and maximum temperature using bivariate time series models.

1.3 Structure of the Document

Each of the three substantive chapters, Chapters 3-5, is self-contained, and has been submitted, as a research paper, to an academic journal. Chapter 3 has been accepted for publication in the International Journal of Forecasting. The structure of this document is

as follows. Chapter 2 contains a brief literature review and some background information for this thesis. Chapter 3-5 present three studies. Chapter 6 summarises the key results and findings of this work, and proposes some future research areas.

Chapter 2

Background Information

2.1 Introduction

We provide some essential background information in this Chapter. The thesis focuses on quantile, ES, and density forecasting for a time series. We first review briefly methods generating density forecasts. Existing methods can be classified into three categories, following the classification of Manganelli and Engle (2004): *nonparametric*, *parametric* and *semi-parametric*. In particular, we review parametric methods for multivariate time series, which will be discussed further in Chapter 5.

Next, we review the methods for quantile and ES estimation. As the density forecasting methods provide full density estimates, we can easily derive the quantile and ES from the estimated density. In this context, the nonparametric, parametric, and semiparametric density forecasting methods are termed nonparametric, parametric, and semiparametric methods for quantile and ES estimation, respectively. In addition, there is another class of semiparametric methods that can directly estimate the quantile and ES based on linear or autoregressive regression of some chosen score (Engle and Manganelli 2004; Taylor 2017a).

Finally, we review evaluation measures for evaluating model performance in density, quantile, and ES forecasting. Evaluating model performance is of vital importance in the forecasting literature. We also review evaluation methods that can be used to select the

best performing models given some confidence level.

The rest of the chapter is structured as follows. Section 2.2 reviews density estimation methods, Section 2.3 reviews forecasting methods for quantile and ES, and Section 2.4 reviews evaluation measures for evaluating density, quantile and ES.

2.2 Density Forecasting

2.2.1 An Overview of Density Forecasting Methods

Parametric models involve a specific parameterisation of the behaviour of a time series. Typical examples are the *autoregressive moving average* (ARMA) and GARCH class of models. ARMA models are used to model the conditional mean of a time series, and the GARCH models are used to model the conditional variance of a time series. Among the GARCH models, there are the standard GARCH model (Bollerslev 1987), the *GJR-GARCH* model (Glosten et al. 1993) and the long memory *fractionally integrated generalised autoregressive conditional heteroskedasticity* (FIGARCH) model (Baillie et al. 1996). There are other types of variance models that directly estimate a certain intra-day variance measure, and use this variance measure as a proxy for unobservable daily variance. Typical examples are the *heterogeneous autoregressive* (HAR) and *realised GARCH* models based on realised volatility and intra-day range. Parametric models also require a distributional assumption, typical examples of which are Gaussian and Student's t distributions. Parametric models are always subject to some degree of error in model specification or distributional assumption.

Nonparametric methods make no parametric form for the distribution of a time series. A typical example that lies in this class is historical simulation. Historical simulation estimates the CDF of a time series as the empirical CDF over a certain window length. Kernel density estimation assigns some kernel function to each observation within a certain window length to get an estimated density of a time series. Two major advantages of nonparametric approaches are that they are easy to implement, and they have no parametric

form. However, Manganelli and Engle (2004) point out that, in the historical simulation and kernel density estimation approaches, an implicit assumption lies behind this method, namely, the distribution of the time series remains at least roughly the same within the window of observations, which may not be appropriate.

With regard to parametric models, there are also multivariate parametric models, which are applied to temperature data in Chapter 5. These methods enable parametric modelling of the co-movement for a multivariate time series. One example is the *multivariate generalised autoregressive conditional heteroskedasticity* (MGARCH) models, which are natural extensions of the GARCH models. Another approach is to apply dynamic copula approach to capture the dependence structure for a multivariate time series.

Semiparametric methods involve the *extreme value theory* (EVT) and *filtered historical simulation* methods, which are generally based on a chosen parametric model. The EVT approach fits a *generalised Pareto distribution* to the tail of a time series beyond a certain threshold. The filtered historical simulation method replaces the distributional assumption in a parametric method by the empirical distribution of the in-sample standardised residuals obtained from this parametric method.

2.2.2 Nonparametric Methods

One of the most commonly used methods for density forecasting is historical simulation. Mathematically, the estimated CDF can be expressed as:

$$F_t(y) = \sum_{i=1}^{N} \frac{1}{N} \mathbb{1}(y \geq y_{t-i})$$

where y_t is a realisation of the time series $\{Y_{1,}Y_{2,\ldots}\}$ for day t; $F_t(y)$ denotes the estimated CDF of Y_t; and N is the chosen window-length. A review of historical simulation methods can be found in Sheather et al. (2004). The choice of window length N is a nontrivial question, for which no consensus has been reached. In each study, we experiment with

different window-lengths.

2.2.3 Parametric Methods

Parametric methods involve a specific parameterisation of the behaviour of a time series. The parametric methods usually decompose the distribution of a time series as the sum of a conditional mean and the product of the volatility, which is the square root of the variance of a time series, and an i.i.d. random variable with mean zero and variance one, which is independent of the past information. Mathematically, the representation is as follows,

$$y_t = \mu_t + \varepsilon_t \tag{2.1}$$

$$\varepsilon_t = \sigma_t \eta_t \tag{2.2}$$

μ_t denotes the mean of Y_t; ε_t denotes the error term; σ_t denotes the volatility of Y_t; and η_t denotes a realisation of an independent and identically distributed (i.i.d.) random variable with mean 0 and variance 1.

2.2.3.1 ARMA-GARCH Models

For financial data, the conditional mean μ_t generally has little variation and is usually small. We use a widely adopted convention in Chapters 3 and 4, which is to assume the conditional mean μ_t to be 0. In Chapter 5, the temperature series exhibits strong seasonality and also perhaps an increasing trend, as well as short term autocorrelation. Therefore, we consider an ARMA model with explanatory variables to capture these features. The ARMA model has the following expression:

$$\mu_t = \Omega^T \Psi_t + r_t \tag{2.3}$$

$$r_t = \sum_{i=1}^{m} \phi_i r_{t-i} + \sum_{i=1}^{n} \delta_i \varepsilon_{t-i} + \varepsilon_t \tag{2.4}$$

where $\mathbf{\Psi}_t$ is a vector of explanatory variables, such as a deterministic trend or a seasonal term; $\mathbf{\Omega}$ is a vector of parameters; and ϕ_i and δ_i are scalar parameters; r_t is the residual of y_t after removing $\mathbf{\Omega}^T\mathbf{\Psi}_t$. $\mathbf{\Omega}^T\mathbf{\Psi}_t$ and r_t can be considered as the deterministic and stochastic parts of y_t, respectively. In particular, the ARMA models with δ_i being zero in expression (2.4) are referred to as the *autoregressive* (AR) models.

Next, we discuss the models for the conditional volatility σ_t. A series of financial returns typically exhibits volatility clustering, that is, the return series is more volatile following a volatile period and is less volatile following a tranquil period. This has led to the development of a variety of univariate time series methods for modelling the conditional volatility since the seminal paper of Engle (1982). Among the proposed GARCH models, the GARCH model proposed by Bollerslev (1987), the GJR-GARCH model proposed by Glosten et al. (1993), and the FIGARCH model proposed by Baillie et al. (1996) are the most popular ones.

A GARCH(p,q) process is defined as:

$$\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{q} \beta_i \sigma_{t-i}^2$$

where ω, α_i and β_i are scalar parameters; σ_t is the conditional volatility; and ε_t is defined as in expressions (2.1) and (2.2).

A GJR-GARCH(p,l,q) process is defined as:

$$\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{l} \gamma_i \varepsilon_{t-i}^2 \mathbb{1}(\varepsilon_{t-i} < 0) + \sum_{i=1}^{q} \beta_i \sigma_{t-i}^2$$

where $\mathbb{1}(\cdot)$ is the indicator function. The GJR-GARCH model captures the leverage effect, namely, the common occurrence that the volatility tends to be greater following a negative return than following a positive return of the same magnitude. Before we introduce the FIGARCH model, we first introduce the lag operator L. Given a series of realisation

$\{y_1, y_2, ...\}$ of a time series $\{Y_1, Y_2, ...\}$, the lag operator L is defined as:

$$L(y_t) = y_{t-1} \tag{2.5}$$

For a polynomial $\upsilon_1(x) = \sum_{i=0}^{n} \alpha_i x^i$, we can define the following lag operator $\upsilon_1(L)$ as:

$$\upsilon_1(L)(y_t) = \sum_{i=0}^{n} \alpha_i y_{t-i}$$

A FIGARCH(p,d,q) model has the following form:

$$\sigma_t^2 = \alpha_0 \left[1 - \upsilon_1(1)\right]^{-1} + \left\{1 - \left[1 - \beta(L)\right]^{-1} \upsilon_2(L)(1-L)^d\right\} \varepsilon_t^2$$

where L represents the lag operator; $\upsilon_1(\cdot)$ and $\upsilon_2(\cdot)$ are polynomials in the lag operator of orders p and q respectively; α_0 is a scalar parameter; d is the fractional parameter satisfying $0 < d < 1$. The FIGARCH model is proposed to capture the long memory property of the daily volatility, that is, the financial returns typically have autocorrelation that is slow to decay. Practical business implications can be drawn from the fractional parameter d: if $d = 0$, the FIGARCH models degenerate into the standard GARCH models, where any price shock goes to zero at a fast exponential rate of decay; if $0 < d < 1$, any price shock still converges zero, but by a much slower hyperbolic rate of decay, which means price shocks persist for a much longer period; if $d = 1$, price shocks persist indefinitely; when $d > 1$, the variance process is explosive, which is unrealistic in practice (Baillie et al. 1996). This long memory property is the notable feature of several of the models considered in Chapter 3. In practice, the number of lags included in the GARCH models is often chosen as 1, which works well in empirical studies. The GARCH(1,1), GJR-GARCH(1,1,1), and FIGARCH(1,d,1) are widely accepted as benchmark models for volatility and VaR estimation (see, for example, Chen et al. 2012). As is the case for the conditional mean in expression (2.3), the GARCH models can be augmented by incorporating explanatory variables to capture additional features such as seasonality and trend for non-financial data.

The parameters of a GARCH model are estimated by maximum likelihood estimation, that is, we choose the parameters which maximise the log likelihood in the in-sample period. The likelihood function can be obtained based on the distributional assumption. Widely-used distributions are Gaussian and Student's t distributions. In addition, we also consider a third distribution in Chapter 5, which is called the *generalised asymmetric skew-t distribution* with the following PDF:

$$f(x;\nu_1,\nu_2,\xi) = \begin{cases} \frac{\xi}{\xi^*}K(\nu_1)\left[1+\frac{1}{\nu_1}\left(\frac{x}{2\xi^*}\right)^2\right] & \text{if } x \le 0 \\ \frac{1-\xi}{1-\xi^*}K(\nu_2)\left[1+\frac{1}{\nu_2}\left(\frac{x}{2(1-\xi^*)}\right)^2\right] & \text{if } x > 0 \end{cases}$$

$$K(x) = \frac{\Gamma(\frac{x+1}{2})}{\sqrt{\pi x}\Gamma(\frac{x}{2})}$$

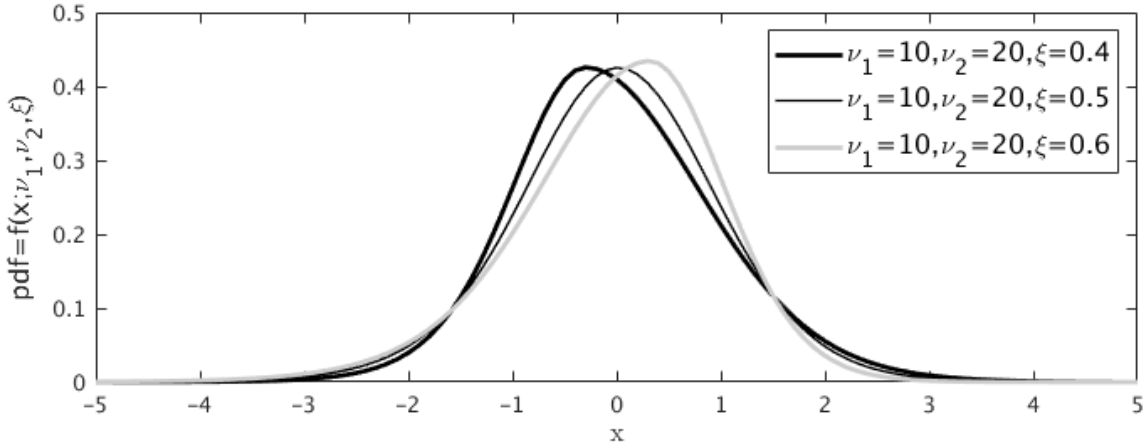
$$\xi* = \frac{\xi K(\nu_1)}{\xi K(\nu_1)+(1-\xi)K(\nu_2)}$$

where $\xi \in (0,1)$ is a skewness parameter; ν_1 and ν_2 are degrees of freedom parameters; $\Gamma(\cdot)$ is the Gamma function. The generalised asymmetric skew-t distribution allows positive and negative parts to have different degrees of freedom, as well as a parameter of skewness to give additional flexibility. This distribution encompasses the Gaussian and Student's t distributions. Figure 2.1 shows the PDF functions for the generalised asymmetric skew-t distribution for three sets of parameters.

2.2.3.2 HAR Models Based on Intra-day Volatility Measures

An alternative approach to GARCH modelling for financial returns is to model some intra-day volatility measure directly. A fundamental difficulty in volatility forecasting is that the daily volatility is non-observable. With intra-day high-frequency data gradually becoming available, various intra-day volatility measures have been studied to uncover the true daily volatility. Some popular intra-day volatility measures that have been considered are realised volatility, intra-day range, and close-to-close range. The *realised*

Figure 2.1: PDF plot of the generalised asymmetric skew-t distribution for three sets of parameters.



volatility of a certain stock is defined as follows:

$$\begin{aligned} RV_t &= \sqrt{\sum_{i=2}^{M} (P_{t,i\cdot\Delta} - P_{t,(i-1)\cdot\Delta})^2} \\ \Delta &= \frac{S}{M} \end{aligned} \tag{2.6}$$

where RV_t denotes the realised volatility within day t; S denotes the interval span of market opening hours; Δ divides S equally into M intervals; and $P_{t,i\cdot\Delta}$ denotes the log price at time $i\cdot\Delta$ of day t. It has been theoretically shown that, if the prices $P_{t,i\cdot\Delta}$ are observed without noise, then expression (2.6) is a consistent estimate of the daily volatility as M tends to infinity (Andersen et al. 2001b; Barndorff-Nielsen 2002). An alternative intra-day volatility measure, the *intra-day range*, is defined as follows:

$$Range_t = H_t - L_t \tag{2.7}$$

where H_t and L_t denote respectively the highest log price and the lowest log price of the day.

The realised volatility and the intra-day range do not take into account the market overnight information, as they are calculated only using the information when the market is open. One way to solve this problem is to incorporate, in the calculation of the realised

volatility and the intra-day range, the overnight return, which is defined as follows:

$$y_{N,t} = open_t - close_{t-1} \tag{2.8}$$

where $open_t$ denotes the log opening price, and $close_{t-1}$ denotes the log closing price on the previous day. Blair et al. (2001) and Hua and Manzan (2013) consider the following overnight return in the realised volatility:

$$RV_{N,t} = \sqrt{(RV_t)^2 + (y_{N,t})^2} \tag{2.9}$$

We incorporate the overnight return in the intra-day range as in expression (2.10):

$$Range_{N,t} = \sqrt{Range_t^2 + y_{N,t}^2} \tag{2.10}$$

To our knowledge, this consideration is novel, and we will discuss this further in Chapter 3. Instead of using the overnight return, Gerlach and Wang (2016) consider the *daily close-to-close return* as follows:

$$Range_{NC,t} = \max(H_t, close_{t-1}) - \min(L_t, close_{t-1}) \tag{2.11}$$

An intra-day volatility measure can be used as a proxy for daily volatility, thus, in this way, we can simply model the time series of the chosen intra-day volatility measure directly. Corsi (2009) proposes the heterogeneous autoregressive model of realised volatility (HAR-RV), which has the following expression:

$$RV_t = \beta_0 + \beta_1 RV_{t-1} + \beta_2 RV_{t-1}^w + \beta_3 RV_{t-1}^m + \varepsilon_t \tag{2.12}$$

$$RV_{t-1}^w = \frac{1}{5}\sum_{i=1}^{5} RV_{t-i} \tag{2.13}$$

$$RV_{t-1}^m = \frac{1}{22}\sum_{i=1}^{22} RV_{t-i} \tag{2.14}$$

where RV_t is the realised volatility; RV_t^w and RV_t^m denote the average realised volatility over a week and a month respectively. Another popular volatility model based on realised volatility is the realised GARCH model proposed by Hansen et al. (2012):

$$\ln(\sigma_t) = \omega + \sum_{i=1}^{p} \alpha_i \ln(\varepsilon_{t-i}) + \sum_{i=1}^{q} \beta_i \ln(RV_{t-1})$$

$$\ln(RV_t) = \tau_0 + \tau_1 \ln(\sigma_t) + \tau_3 \tilde{\eta}_t + \tau_4(\tilde{\eta}_t^2 - 1) + \eta_t$$

where $\tilde{\eta}_t = \frac{y_t}{\sigma_t}$ is an i.i.d. Student's t distribution with mean 0 and variance 1; η_t is a realisation of a random variable that is i.i.d. Gaussian distributed with mean 0 and variance 1; ω, α_i, β_i, τ_i are scalar parameters; and p and q are positive integers. The HAR model and the realised GARCH model we present here are based on realised volatility, which can be replaced by other intra-day volatility measures.

2.2.4 Multivariate Parametric Models

Understanding the co-movement between related time series is of great importance for many applications. For example, MGARCH models have been considered for financial data and *vector autoregressive moving average* and MGARCH (VARMA-MGARCH) models have been considered for wave height forecasting (see, for example, Bauwens et al. 2006; Taylor and Jeon 2018). In this thesis, we consider the VARMA-MGARCH-VEC model and the VARMA-MGARCH model based on copula. The VARMA-MGARCH-VEC model proposed by Bollerslev et al. (1988) estimates the conditional mean and covariance matrix of a multivariate time series. It has the following expression:

$$\boldsymbol{y}_t = \boldsymbol{E}_1 \boldsymbol{\Psi}_t + \boldsymbol{r}_t \tag{2.15}$$

$$\boldsymbol{r}_t = \sum_{i=1}^{m} \boldsymbol{B}_{1,i} \boldsymbol{r}_{t-i} + \sum_{i=1}^{n} \boldsymbol{B}_{2,i} \boldsymbol{\epsilon}_{t-i} + \boldsymbol{\epsilon}_t \tag{2.16}$$

$$\boldsymbol{\epsilon}_t = \boldsymbol{H}_t^{1/2} \boldsymbol{\eta}_t \tag{2.17}$$

$$\begin{aligned}\text{vech}(\boldsymbol{H}_t) &= \boldsymbol{E}_2\boldsymbol{\Psi}_t + \sum_{i=1}^{q} \boldsymbol{B}_{3,i}\text{vech}(\boldsymbol{\epsilon}_{t-i}\boldsymbol{\epsilon}_{t-i}^T) \\ &\quad + \sum_{i=1}^{o} \boldsymbol{B}_{4,i}\text{vech}(\boldsymbol{\epsilon}_{t-i}\boldsymbol{\epsilon}_{t-i}^T)\mathbb{1}\left(\text{vech}(\boldsymbol{\epsilon}_{t-i}\boldsymbol{\epsilon}_{t-i}^T) < \mathbf{0}\right) \\ &\quad + \sum_{i=1}^{p} \boldsymbol{B}_{5,i}\text{vech}(\boldsymbol{H}_{t-i}) \end{aligned} \tag{2.18}$$

where $\boldsymbol{y}_t$ is a realisation of the time series $\{\boldsymbol{Y}_1, \boldsymbol{Y}_{2,...}\}$ for day t; $\boldsymbol{\Psi}_t$ is a vector of explanatory variables; $\boldsymbol{E}_1$ and $\boldsymbol{E}_2$ are matrices of parameters; vech denotes the operator that stacks the lower triangular portion of a matrix as a column vector; $\boldsymbol{B}_{k,i}$ are parameter matrices for $k = 1,2,...5$; $\boldsymbol{r}_t$ contains the stochastic parts of $\boldsymbol{y}_t$; $\boldsymbol{\epsilon}_t$ is the error term of $\boldsymbol{r}_t$; $\boldsymbol{H}_t$ is the covariance matrix of $\boldsymbol{\epsilon}_t$; and m, n, q, o and p are the VARMA and MGARCH orders. The parameters can be estimated using the maximum likelihood approach. A commonly used distribution is the multivariate Gaussian distribution (Bauwens et al. 2006).

An alternative to covariance matrix estimation is to model the dependence structure of multivariate time series via the copula approach. Sklar's theorem enables the decomposition of a joint distribution into its marginal distributions and a dependence function, which is called the copula function (Kolev et al. 2006; Patton 2012). A copula function is a CDF defined on $[0,1]^{dim}$, where dim is the dimension of the random variable. Sklar's theorem can be expressed as follows:

$$\begin{aligned}\Pr(X_1 < x_1, ..., X_{dim} < x_{dim}) &= C(u_1, ..., u_{dim}) \\ u_i &= \Pr(X_i < x_i)\end{aligned}$$

where $(X_1, X_2, ..., X_{dim})$ is a multivariate random variable; x_i are real numbers; u_i are the probabilities $\Pr(X_i < x_i)$; and C is a copula function. It can be seen that the copula function and marginal distributions completely determine a multivariate random variable. In our third study (Chapter 5), we consider the Gaussian copula and the Student's t copula, which are standard in the literature (Patton 2006, 2012). The Gaussian copula has the following

expression:

$$C(u_1,...,u_{dim}) = F_{\boldsymbol{R}}(F^{-1}(u_1),...,F^{-1}(u_{dim}))$$

where $F_{\boldsymbol{R}}$ is the CDF of the multivariate Gaussian distribution with correlation matrix $\boldsymbol{R}$, and F^{-1} is the quantile function of a standard univariate Gaussian distribution. The Student's t copula can be expressed as follows:

$$C(u_1,...,u_{dim}) = F_{\boldsymbol{R},\nu}(F_{\nu}^{-1}(u_1),...,F_{\nu}^{-1}(u_{dim}))$$

where $F_{\boldsymbol{R},\nu}$ is a multivariate Student's t distribution with correlation matrix $\boldsymbol{R}$ and degrees of freedom ν, and F_{ν}^{-1} is the quantile function of a Student's t distribution with degrees of freedom ν. The Gaussian copula has one parameter, which is a correlation parameter, and the Student's t copula has two parameters, which are a correlation parameter and a parameter for the degrees of freedom. A VARMA-MGARCH model with copula can be written as follows:

$$\begin{aligned}
\boldsymbol{y}_t &= \boldsymbol{E}_1\boldsymbol{\Psi}_t + \boldsymbol{r}_t && (2.19)\\
\boldsymbol{r}_t &= \sum_{i=1}^{m}\boldsymbol{B}_{1,i}\boldsymbol{r}_{t-i} + \sum_{i=1}^{n}\boldsymbol{B}_{2,i}\boldsymbol{\epsilon}_{t-i} + \boldsymbol{\epsilon}_t && (2.20)\\
\boldsymbol{\epsilon}_t &= (\varepsilon_{1,t},\varepsilon_{2,t})^T && (2.21)\\
F_{\boldsymbol{\epsilon}_t}(z_1,z_2) &= C_{\boldsymbol{R}_t,\nu}(F_{\varepsilon_{1,t}}(z_1),F_{\varepsilon_{2,t}}(z_2)) && (2.22)\\
\text{var}(\varepsilon_{j,t}) &= h_{j,t} && (2.23)\\
\boldsymbol{H}_t &= \begin{pmatrix} h_{1,t} & R_t\sqrt{(h_{1,t}h_{2,t})} \\ R_t\sqrt{(h_{1,t}h_{2,t})} & h_{2,t} \end{pmatrix} && (2.24)\\
\text{vech}(\boldsymbol{H}_t) &= \boldsymbol{E}_2\boldsymbol{\Psi}_t + \sum_{i=1}^{q}\boldsymbol{B}_{3,i}\text{vech}(\boldsymbol{\epsilon}_{t-i}\boldsymbol{\epsilon}_{t-i}^T) \\
&\quad + \sum_{i=1}^{o}\boldsymbol{B}_{4,i}\text{vech}(\boldsymbol{\epsilon}_{t-i}\boldsymbol{\epsilon}_{t-i}^T)\mathbb{1}\left(\text{vech}(\boldsymbol{\epsilon}_{t-i}\boldsymbol{\epsilon}_{t-i}^T) < \boldsymbol{0}\right) \\
&\quad + \sum_{i=1}^{p}\boldsymbol{B}_{5,i}\text{vech}(\boldsymbol{H}_{t-i}) && (2.25)
\end{aligned}$$

where $\boldsymbol{y}_t$ is a realisation of the time series $\{\boldsymbol{Y}_1,\boldsymbol{Y}_{2,...}\}$ for day t; $h_{j,t}$ is the variance of

the error terms $\varepsilon_{j,t}$; $C_{\boldsymbol{H}_t,\nu}(\cdot,\cdot)$ is the Gaussian or Student's t copula; R_t is the correlation parameter in the copula; $\boldsymbol{R}_t$ is the correlation matrix corresponding to $\boldsymbol{H}_t$; ν is the degrees of freedom for the Student's t copula and ν is ∞ for the Gaussian copula; $F_{\boldsymbol{\epsilon}_t}(\cdot,\cdot)$ is the distribution function of $\boldsymbol{\epsilon}_t$; $F_{\varepsilon_{i,t}}(\cdot)$ is the CDF of $\varepsilon_{i,t}$; and the other notations are the same as in expressions (2.15)-(2.18). Note that R_t and $\boldsymbol{R}_t$ are time varying, which makes the copula also time-varying.

2.2.5 Semiparametric Methods

We review filtered historical simulation and the EVT approach in this section. The filtered historical simulation approach simply replaces the distributional assumption for the standardised residuals in a parametric model by the empirical distribution of the standardised residuals in the in-sample period. Such an approach has been shown to perform competitively in quantile and ES forecasting (Taylor 2017a).

EVT involves the *block maxima* and the *peaks over threshold* approaches. Peaks over threshold is a modern type of EVT, models of which are for all observations beyond a certain threshold (e.g. 10% quantile). This type of model is generally considered to be more useful in practice than block maxima EVT, because of its efficient use of data (McNeil 1999), so block maxima EVT will not be discussed in this document. In this thesis, we will simply refer to the peaks over threshold approach as the EVT approach when no confusion is incurred. Within the peaks over threshold class, there are two different styles of analysis. The first one is the semi-parametric model built around the Hill estimator, and the second one is the parametric model based on generalised Pareto distribution. McNeil and Frey (2000) show that the second approach tends to give more stable estimates with respect to the choice of threshold. In this document, we focus on the second approach and do not discuss the first approach.

Suppose we estimate a time series using a parametric model, the peaks over threshold approach fits a generalised Pareto distribution to the exceedances of the standardised residuals beyond the 10% unconditional quantile of the standardised residuals, where a

generalised Pareto distribution has the following CDF:

$$F(x) = \begin{cases} 1-(1-\zeta\frac{x}{s})^{1/\xi} & \zeta \neq 0 \\ 1-\exp(-\frac{x}{s}) & \zeta = 0 \end{cases} \tag{2.26}$$

where s is the scale parameter; and ζ is the shape parameter.

2.3 Quantile and Expected Shortfall Regression

We have reviewed density forecasting methods. Quantile and ES forecasts can be obtained easily from the estimated density. The nonparametric, parametric, and semiparametric methods for density forecasting are called the nonparametric, parametric, and semiparametric methods for quantile and ES forecasting. In this section, we review an alternative class of semiparametric methods, which model the desired quantile and ES directly, rather than obtaining them from an estimated density. We discuss solely such semiparametric quantile and ES models for financial returns since quantile and ES estimation for other types of data are not considered in this thesis. In this thesis, for simplicity, we omit the probability level θ for quantile and ES, when no confusion is incurred.

2.3.1 Quantile Regression

A generic quantile regression model can be expressed in the following expression:

$$q_t(\boldsymbol{\beta}) = q(\boldsymbol{\Psi}_{t-1}, q_{t-1}(\boldsymbol{\beta}), \ldots, \boldsymbol{\Psi}_1, q_1(\boldsymbol{\beta}); \boldsymbol{\beta}, \theta) \tag{2.27}$$

where θ is the probability level of interest; q_t is the θ quantile of the daily return; $\boldsymbol{\beta}$ is a vector of parameters; and $\boldsymbol{\Psi}_t$ is a vector of explanatory variables. The parameter vector $\boldsymbol{\beta}$

is chosen to minimise the following quantile score:

$$\sum_{t=1}^{T}[y_t - q_t(\boldsymbol{\beta})][\theta - \mathbb{1}(y_t < q_t(\boldsymbol{\beta}))] \tag{2.28}$$

where y_t is a daily return; T is the number of observations in the in-sample period; and other expressions are as in expression (2.27). Different quantile regression models involve different selections of $\boldsymbol{\Psi}_t$ and formulations of the $q_t(\cdot)$ in expression (2.27). Note that expression (2.27) degenerates into a simple linear regression model, when there is no lagged quantile estimate $q_{t-i}(\boldsymbol{\beta})$ in the model specification.

Next we review several benchmark quantile regression models. Engle and Manganelli (2004) propose four CAViaR models, which use lagged returns and quantile estimates as the explanatory variables. We present the following four representations proposed by Engle and Manganelli (2004):

CAViaR Symmetric Absolute Value (CAViaR-SAV):

$$q_t(\boldsymbol{\beta}) = \beta_1 + \beta_2 q_{t-1}(\boldsymbol{\beta}) + \beta_3 |y_{t-1}|$$

CAViaR Asymmetric Slope (CAViaR-AS):

$$q_t(\boldsymbol{\beta}) = \beta_1 + \beta_2 q_{t-1}(\boldsymbol{\beta}) + \beta_3 (y_{t-1})^+ + \beta_4 (y_{t-1})^-$$

CAViaR Indirect GARCH (CAViaR-IndG):

$$q_t(\boldsymbol{\beta}) = \text{sgn}(\theta - 0.5)(\beta_1 + \beta_2 q_{t-1}(\boldsymbol{\beta}) + \beta_3 y_{t-1}^2)^{\frac{1}{2}}$$

CAViaR Adaptive (CAViaR-Adaptive):

$$q_t(\boldsymbol{\beta}) = q_{t-1}(\boldsymbol{\beta}) + \beta_1(\theta - \mathbb{1}(y_{t-1} < q_{t-1}(\boldsymbol{\beta})))$$

where $(x)^+ = \max(x, 0)$, and $(x)^- = -\min(x, 0)$. The CAViaR-Adaptive model has been

shown to perform not as competitively as the other CAViaR models due to its inability to capture the underlying dynamics of the quantiles (Manganelli and Engle 2004). Therefore, we do not discuss the CAViaR-Adaptive model further in this thesis. The other CAViaR models have been shown to perform competitively in quantile estimation (see, for example, Chen et al. 2012; Jeon and Taylor 2013).

2.3.2 Expected Shortfall Regression

A fundamental problem for ES estimation is its lack of *elicitability*, meaning that the ES can not be estimated or evaluated directly. Recently, Fissler and Ziegel (2016) have shown that VaR and ES are jointly elicitable, and present the following set of joint scoring functions for quantile and ES:

$$h_{FZ}(y_t, q_t, es_t; \theta, G_1, G_2, a) = (\mathbb{1}\{y_t \leq q_t\} - \theta)\left(G_1(q_t) - G_1(y_t) + \frac{1}{\theta}G_2(es_t)q_t\right) - G_2(es_t)\left(\frac{1}{\theta}\mathbb{1}\{y_t \leq q_t\}y_t - es_t\right) - \mathscr{G}_2(es_t) + a(y_t) \quad (2.29)$$

where y_t is the daily return; θ is the probability level; $q_t(\cdot)$ is the estimated θ quantile of the daily return y_t; es_t is the ES; G_1, G_2, $\mathscr{G}_2$ and a are real valued functions; G_1 is weakly increasing; G_2 is strictly increasing and positive; and G_2 is the derivative of $\mathscr{G}$: $\mathscr{G}_2' = G_2$. We refer to scores of this type as *FZ scores*. Taylor (2017a) studies one such FZ score with $G_1(x) = 0$, $G_2(x) = -\frac{1}{x}$, $\mathscr{G}_2(x) = -\ln(-x)$ and $a(x) = 1 - \ln(1-\theta)$ in expression (2.29) to give:

$$h_{FZ}(y_t, q_t, es_t; \theta) = -\ln\left(\frac{\theta - 1}{es_t}\right) - \frac{(y_t - q_t)(\theta - \mathbb{1}\{y_t \leq q_t\})}{\theta es_t} + \frac{y_t}{es_t} \quad (2.30)$$

We refer to this as the *AL score*. Taylor (2017a) considers modelling the quantile using the CAViaR models, and modelling the ES as a constant or dynamic multiple of the quantile estimates. We refer to such models, estimated by minimising an FZ score, as *CAViaR-FZ* models. The empirical study of Taylor (2017a) shows that the CAViaR-FZ models

with the ES modelled as a constant multiple of its corresponding quantile perform very competitively. Therefore, for simplicity, we only discuss such models and do not discuss the CAViaR-FZ models with the ES modelled as a dynamic multiple of its corresponding quantile in this thesis. We present two such models in the following expressions:
CAViaR Symmetric Absolute Value estimated using an FZ score (CAViaR-FZ-SAV):

$$\begin{aligned} q_t(\boldsymbol{\beta}) &= \beta_1 + \beta_2 q_{t-1}(\boldsymbol{\beta}) + \beta_3 |y_{t-1}| \\ es_t(\boldsymbol{\beta}) &= \beta_4 q_t(\boldsymbol{\beta}) \end{aligned} \tag{2.31}$$

CAViaR Asymmetric Slope estimated using an FZ score (CAViaR-FZ-AS):

$$\begin{aligned} q_t(\boldsymbol{\beta}) &= \beta_1 + \beta_2 q_{t-1}(\boldsymbol{\beta}) + \beta_3 (y_{t-1})^+ + \beta_4 (y_{t-1})^- \\ es_t(\boldsymbol{\beta}) &= \beta_5 q_t(\boldsymbol{\beta}) \end{aligned} \tag{2.32}$$

where, in each model, $\boldsymbol{\beta}$ is a vector of parameters. The parameter β_2 captures the degree of autoregression existed in the VaR process; β_3 and β_4 represent the magnitude of the impact of the price shocks on the conditional VaR; and the parameter β_5 is always greater than 1, as the magnitude of the ES is always larger than that of VaR by definition.

2.4 Forecast Evaluation

2.4.1 Density Forecast Evaluation

To evaluate model performance, we utilise the scoring functions. A scoring function assigns a value for a forecast when a realisation is observed. A scoring function is said to be *proper* if the expected value of the scoring function is minimised by the perfect forecasts. In other words, proper scoring functions encourage honest assessments of model performance (Garthwaite et al. 2005).

We first use *root mean squared error* (RMSE) to evaluate forecasts of the mean. The

RMSE is based on the *squared error* with the following expression:

$$SE(\mu_t, y_t) = (\mu_t - y_t)^2$$

where y_t is the actual observation; μ_t is the estimated mean. In practice, the RMSE is calculated as the square root of the average squared error in the post-sample period. The RMSE is proper as the RMSE is minimised by the true mean of y_t (Gneiting and Raftery 2007).

As for density forecasts, we use *Schwarz bayesian criterion* (SBC) and *probability integral transformation* (PIT) for in-sample model analysis in Chapter 5. The SBC is based on the in-sample likelihood function, and penalises the excessive number of parameters. The SBC has the following expression:

$$SBC = \ln(T)k - 2\sum_{t=1}^{T} l_t$$

where T is the number of observations in the in-sample period; k is the number of parameters; and l_t is the in-sample log likelihood values. SBC is not proper but it is asymptotically proper as the sample size goes to ∞ (Gneiting and Raftery 2007). SBS requires the likelihood function, which is often not known in practice. This is considered as a major weakness as an evaluation measure (Gneiting and Raftery 2007).

The PIT is the value of the estimated CDF function evaluated at the actual observation. The PIT of the perfect forecasts of a time series should be uniformly distributed between 0 and 1 (Diebold et al. 1998). However, the PIT is only a necessary condition, which means a model with perfect PIT may not be the best performing model (Diebold et al. 1998).

For post-sample density evaluation, we use the *continuous ranked probability score* (CRPS) and *energy score* to evaluate univariate and multivariate density forecasts, respectively. The CRPS is widely used to evaluate univariate density forecasts (Gneiting

and Raftery 2007). The CRPS can be expressed as follows:

$$CRPS(F_t, y_t) = \int_{-\infty}^{\infty} (F_t(x) - \mathbb{1}(y_t \geq x))^2 dx \tag{2.33}$$

$$= -\frac{1}{2}\mathbb{E}_{F_t}|Z_t^1 - Z_t^2| + \mathbb{E}_{F_t}|Z_t^1 - y_t| \tag{2.34}$$

where y_t is the actual observation; $F_t(\cdot)$ is the predicted CDF; Z_t^1 and Z_t^2 are two independent copies of random variable with CDF F_t; and $\mathbb{E}_{F_t}$ is the expectation with respect to the CDF F_t.

For multivariate distributions, Gneiting and Raftery (2007) introduce the energy score, as a generalisation of the CRPS. The energy score is expressed as follows:

$$Energyscore(F_t, \boldsymbol{y}_t) = -\frac{1}{2}\mathbb{E}_{F_t}\|\boldsymbol{Z}_t^1 - \boldsymbol{Z}_t^2\| + \mathbb{E}_{F_t}\|\boldsymbol{Z}_t^1 - \boldsymbol{y}_t\| \tag{2.35}$$

where F_t is the estimated multivariate CDF; $\boldsymbol{y}_t$ is the actual observation; $\|\cdot\|$ is the Euclidean norm; $\boldsymbol{Z}_t^1$ and $\boldsymbol{Z}_t^2$ are two independent copies of a random vector drawn from the distribution with CDF F_t. Note that $\boldsymbol{y}_t$, $\boldsymbol{Z}_t^1$, and $\boldsymbol{Z}_t^2$ are multivariate.

The CRPS and energy score can easily be obtained, as expressions (2.34) and (2.35) can be calculated using numerical simulation. As is the case with the RMSE, the CRPS and energy score are averaged over the post-sample period to produce an overall measure of accuracy. The CRPS and energy score are proper because the expectation of the scores are minimised by the correct density forecast (Gneiting and Raftery 2007).

2.4.2 Evaluating Event Probability Forecasts

In Chapter 5, we will discuss the heat wave estimation, where a heat wave is defined as the exceedance of the daily maximum temperature beyond a threshold for a period days, or the exceedances of the daily minimum and maximum temperature simultaneously beyond some thresholds over a period of days. To evaluate probability forecasts of heat waves, we need to assess the accuracy of forecasts of the exceedance probability. We evaluate the

exceedance probability forecasts using the widely adopted *Brier score*, which is computed as:

$$BS(p_t, o_t) = (p_t - o_t)^2$$

where p_t is the predicted probability; and o_t is the actual outcome of the event, which equals 0 if the event occurs, and 1 otherwise. The Brier score is a proper score for evaluating event probability forecasts (Gneiting and Raftery 2007)

2.4.3 Quantile Forecast Evaluation

We first review the *unconditional coverage test* that assesses the hit percentage defined as the proportion of the returns below the estimated quantile corresponding to probability level θ. Significant difference from the ideal value of θ is tested using a binomial distribution: it evaluates whether the sequence of the variable defined as $\mathbb{1}(y_t < q_t(\boldsymbol{\beta}))$, is distributed i.i.d. Bernoulli with probability θ. The p-values for the for the null hypothesis $H_0 : \Pr(\mathbb{1}(y_t < q_t(\boldsymbol{\beta})) = \theta$ under the unconditional coverage test can be established through the following asymptotic expressions based on a likelihood ratio test:

$$l(\theta) = \log\left(\theta^{T_0}(1-\theta)^{T-T_0}\right)$$
$$-2(l(\theta) - l(\hat{\theta})) \overset{a}{\sim} \chi^2(1)$$

where $\hat{\theta}$ is the expectation of $\mathbb{1}(y_t < q_t(\boldsymbol{\beta}))$ with respect to t; T is the length of the test period; and $T_0 = T\hat{\theta}$; and $\chi^2(1)$ is the Chi-squared distribution with 1 degree of freedom. Engle and Manganelli (2004) introduce a new test for conditional coverage called the *dynamic quantile* (DQ) test, which uses a regression framework as in the following expression :

$$Hit_t = \delta_0 + \Sigma_{i=1}^{v} \delta_i Hit_{t-i} + \delta_{p+1} q_{t-1} + \varepsilon_t \tag{2.36}$$

where $Hit_t = \mathbb{1}(y_t < q_t) - \theta$; δ_i are parameters; and ε_t is the error term. If the model is correctly specified, the coefficients δ_i should all be zero. In practice, the coefficients δ_i

being not significantly different from zero usually means the corresponding model performs well. Note expression (2.36) is a standard linear regression, therefore the p-values under the null hypothesis $H_0 : \delta_i = 0 \; for \; all \; i$ can be established through the following asymptotic expressions:

$$\hat{\boldsymbol{\delta}}_{OLS} = (\boldsymbol{C}^T\boldsymbol{C})^{-1}\boldsymbol{C}^T Hit \quad \overset{a}{\sim} \quad N(0, \theta(1-\theta)(\boldsymbol{C}^T\boldsymbol{C})^{-1}) \tag{2.37}$$

$$\frac{\hat{\boldsymbol{\delta}}_{OLS}^T\boldsymbol{C}^T\boldsymbol{C}\hat{\boldsymbol{\delta}}_{OLS}}{\theta(1-\theta)} \overset{a}{\sim} \chi^2(v) \tag{2.38}$$

where $\boldsymbol{C}$ is a matrix where the ith row defined to be the row vector $(1, Hit_{t-1-i}, \ldots Hit_{t-p-i}, q_{t-1-i})$ and the number of rows equals the number of the in-sample observations; v is the number of lags; and $\chi^2(v)$ is the Chi-squared distribution with v degrees of freedom. The number of lags v is subjective, and we consider $v = 4$, which is used in Engle and Manganelli (2004).

We also evaluate the quantile forecasts using the *quantile score*, which is expression (3.8) summed over the post-sample period. The quantile score is a proper scoring function for evaluating quantile forecasts, implying that the expectation of the score is minimised by the perfect set of quantile forecasts (Gneiting and Raftery 2007).

2.4.4 Expected Shortfall Forecast Evaluation

To evaluate ES forecasts, we first use the bootstrap test of McNeil and Frey (2000), which focuses on the discrepancies between the VaR exceedances and the ES forecasts. Mathematically, the random variable $\frac{y_t - ES_t}{\sigma_t}$ should be i.i.d. and have mean zero conditional on $y_t < q_t$, where σ_t is the volatility estimate, q_t and ES_t are a pair of quantile and ES estimates associated with a certain probability level. We replace σ_t by ES_t when σ_t is not available (Acerbi and Szekely 2014). In view of the relatively small number of VaR exceedances, conditional coverage is typically not considered for ES forecasts. The distribution of $\frac{y_t - ES_t}{\sigma_t}$ is dependent on the distribution of y_t that is generally unknown. McNeil and Frey (2000) propose to use bootstrap to obtain the test statistics and p-values.

Next, we use the FZ scores described in expression (2.29) to evaluate the model performance. The FZ scores are proper, as is shown by Nolde and Ziegel (2017). This class consists of an infinite number of scores, and there is no consensus on which one has the best test power. We consider three different FZ scores in this thesis. The first FZ score we consider is the AL score in expression (2.30). Secondly, we consider an FZ score proposed by Nolde and Ziegel (2017) resulting from setting $G_1 = 0$, $G_2 = \frac{1}{2}(-x)^{-\frac{1}{2}}$, $\mathscr{G}_2(x) = -(-x)^{\frac{1}{2}}$ and $a(x) = 0$ in expression (2.29). We refer to this as the *NZ score*. The third FZ score we consider is the one considered by Fissler et al. (2016), which is produced by setting $G_1(x) = x$, $G_2 = \frac{\exp(x)}{1+\exp(x)}$, $\mathscr{G}_2(x) = \ln(1+\exp(x))$ and $a(x) = 0$. To make the values more easily comparable, we set $a(x) = \ln(2)$, which leads to positive values for the score.

2.4.5 Significance Testing

The scoring functions, such as the CRPS, energy score and the quantile score are appealing, as they enable an honest and intuitive comparison between different methods. However, in practice, there are two major weaknesses in interpreting the results. Firstly, the post-sample period usually only contains a finite number of observations. Hence, two methods having different values of a scoring function might not be significant statistically. Secondly, relative model performance is usually not transitive. Therefore, it is very difficult to select the best performing method, especially among a large set of competing methods. To solve these problems, Hansen et al. (2011) introduce the *model confidence set* that can be used to select a set of best performing methods, which contains the best performing model given some significance level.

The model confidence set procedure is as follows. Given a set of models that are to be evaluated and a pre-chosen significance level, apply an *equivalence test* to test the null hypothesis that all the methods are not statistically significantly different. If the null hypothesis is not rejected, the model confidence set consists of all the methods. If the null hypothesis is rejected, use an *elimination rule* to eliminate the worst performing method,

and revert back to the first step. Under some suitable regularity conditions and assumptions, Hansen et al. (2011) show that the model confidence set outcome contains the best performing method with the chosen confidence level.

The choice of the equivalence test or the elimination rule is not unique, and it does not affect the asymptotic results if the equivalence test and the elimination rule satisfy the conditions in Hansen et al. (2011). In this thesis, we use the equivalence test based on the Diebold-Mariano test (Diebold and Mariano 2002) and use the one-sided elimination rule (this is described as $T_{max,M}$ in Hansen et al. (2011)), as is discussed in Hansen et al. (2011). The Diebold-Mariano test and the one-sided elimination rule are described as follows. Given two methods and a scoring function, such as the CRPS or energy score, the values of the scoring functions for each method for each day of the post-sample period can be viewed as the values of the scoring functions as two time series. The Diebold-Mariano test simply tests whether the unconditional sample mean of the first time series is less than or equal to that of the second time series. We use the bootstrap method to obtain the test statistics as described in Hansen et al. (2011). The equivalence test now tests whether any method is not better than all the other methods under the Diebold-Mariano test given some confidence level. As for the elimination rule, each method is compared with all the other methods, and it is assigned the maximum resulting test statistics. The method with the highest value will be eliminated.

2.5 Summary

In this chapter, we briefly reviewed forecasting and evaluation methods that we use in the remainder of the thesis. Next, we present three self-contained studies in Chapters 3-5. Chapters 3 proposes several quantile regression models based on intra-day data in model specification to improve quantile forecasting accuracy. Chapters 4 considers the use of the intra-day low series in parameter estimation for ES regression model estimation. Chapters 5 considers the joint modelling of the daily minimum and maximum temperature to improve both the marginal and joint distributions of the daily minimum

and maximum temperature.

Chapter 3

An Approximate Long-Memory Range-Based Approach for Value at Risk Estimation

3.1 Introduction

Since the first appearance of VaR in the 1980s, it has become the most prevalent risk measure, and is currently a standard tool for risk management in financial and insurance institutions (Berkowitz et al. 2011; Nieto and Ruiz 2016). The VaR is the quantile of the conditional distribution of the return on a portfolio. Accurate forecasting of VaR is of great importance for internal risk control and financial regulation. Although the concept of VaR is not complex, its measurement has proved to be challenging. There has been a variety of approaches proposed for the forecasting of VaR, and yet no consensus has been reached as to the best method.

Classical approaches to VaR forecasting, such as the use of GARCH models, historical simulation and CAViaR models (Engle and Manganelli 2004), use only the historical returns. Intra-day data is becoming increasingly available, and it has been found to provide useful information for the estimation of the distribution of the daily returns

(Corsi et al. 2008). Therefore, efforts have been made to use intra-day data in the forecasting of VaR for daily returns. Realised volatility, which is a nonparametric measure of unobservable volatility, calculated using intra-day data, has been widely used as a basis for forecasting daily volatility (see, for example, Andersen and Bollerslev 1998; Andersen et al. 2001b; Barndorff-Nielsen 2002). However, intra-day data tends to be expensive, and often a long time series of observations is not available. Moreover, the effort and resources required to process the high-frequency data may prove excessive (Rogers and Zhou 2008).

In contrast, the daily opening, daily closing, intra-day low and intra-day high series for the last thirty years are readily available for most tradable assets. Instead of using intra-day data to produce the realised volatility for VaR estimation, we consider an alternative use of intra-day data, which is much easier to implement and yet highly efficient. We base VaR estimation on the intra-day range, which is the absolute difference between the daily high and the daily low log prices. Despite the fact that the intra-day range has been widely studied in volatility forecasting (Parkinson 1980; Andersen and Bollerslev 1998; Brandt and Jones 2006), little attention has been devoted to utilizing the intra-day range in VaR estimation. The only such literature, that the authors are aware of, are the studies of Brownlees and Gallo (2010), Chen et al. (2012) and Fuertes and Olmo (2013). Brownlees and Gallo (2010) use the intra-day range in a parametric framework. Chen et al. (2012) consider the use of the intra-day range in CAViaR models. Fuertes and Olmo (2013) includes the intra-day range in a GARCH model. Another variable that we consider in this study is the market overnight return. It has been pointed out that the overnight return is useful for volatility forecasting, because, while the market is closed, a great amount of highly relevant information arrives from markets abroad, and public announcements might be made after the previous day's closing time (Tsiakas 2008). The work of Brownlees and Gallo (2010) is the only study that has compared the performance of the intra-day range and realised volatility for VaR estimation, with only parametric methods being considered. The authors are not aware of any study that evaluates the performance of the overnight

return for VaR estimation.

This study has two contributions. First, we propose a number of new quantile regression models based on realised volatility, the intra-day range and the overnight return. Second, we carry out an empirical comparison between the proposed methods and a large number of benchmark methods. Moreover, this study is the first study that compares VaR estimation performance of a large set of methods based on the intra-day range and methods based on realised volatility for VaR estimation.

The remainder of the study is structured as follows. Section 3.2 reviews intra-day volatility measures. Section 3.3 gives a brief review of the established VaR methods, with particular focus on those that are closely related to our new proposals. Section 3.4 introduces our new VaR models. Section 3.5 uses 18 series of stock returns to evaluate the performance of the new models, and to compare their VaR estimation accuracy to established methods. Section 3.6 provides a summary and some concluding remarks.

3.2 Intra-day Volatility Measures

In this section, we introduce the intra-day volatility measures that are used in this study. A very popular intra-day volatility measure is realised volatility. The realised volatility of a certain stock is defined as follows:

$$\begin{aligned} RV_t &= \sqrt{\sum_{i=2}^{M} (P_{t,i\cdot\Delta} - P_{t,(i-1)\cdot\Delta})^2} \\ \Delta &= \frac{S}{M} \end{aligned} \tag{3.1}$$

where RV_t denotes the realised volatility within day t; S denotes the interval span of market opening hours; Δ divides S equally into M intervals; and $P_{t,i\cdot\Delta}$ denotes the log price at time $i\cdot\Delta$ of day t. It has been theoretically shown that, if the prices $P_{t,i\cdot\Delta}$ are observed without noise, then expression (3.1) is a consistent estimate of the daily volatility as M tends to infinity (Andersen et al. 2001b; Barndorff-Nielsen 2002). It has been found that models

incorporating the realised volatility can significantly improve daily volatility forecasts in comparison to the conventional GARCH models, which are applied to daily returns data (see, for example, Corsi et al. 2008; Hansen et al. 2012; Martens et al. 2009; Shephard and Sheppard 2010).

The calculation of realised volatility clearly requires access to high-frequency data. Although such data is gradually becoming available, it is still expensive and is not available for a long time series of observations. The calculation of expression (3.1) requires significant computational power. Moreover, microstructure noise, which , might contaminate the data (see, for example, Hounyo et al. 2017), making expression (3.1) an inconsistent estimate for daily volatility (Andersen et al. 2011). Typical sources of microstructure noise are bid-ask bounce, discreteness and rounding issues of price changes, recording errors of data measurement, and impact of a block trade on the prices (Aït-Sahalia et al. 2011). This prompts consideration of, as an alternative intra-day volatility measure, the intra-day range, which is readily available for most tradable assets and requires little computing resource. The intra-day range is defined as follows:

$$Range_t = H_t - L_t \tag{3.2}$$

where H_t and L_t denote respectively the highest log price and the lowest log price of the day. The intra-day range has been widely studied in volatility estimation. Parkinson (1980) shows that the properly scaled intra-day range is an unbiased estimator of daily volatility, and is more efficient than the squared daily return. Brandt and Jones (2006) show that the efficiency of the intra-day range is even comparable to that of realised variance calculated using 3-hour to 6-hour returns. Alizadeh et al. (2002) show that the intra-day range is more robust to market microstructure noise, in comparison with realised volatility.

It should be noted that both the realised volatility and the intra-day range ignore the market overnight return, which is defined as follows:

$$y_{N,t} = open_t - close_{t-1} \tag{3.3}$$

where $open_t$ denotes the log opening price; and $close_{t-1}$ denotes the log closing price on the previous day. The overnight return has raised the interest of volatility forecasters. Hansen and Lunde (2006) find that an intra-day volatility measure that ignores the overnight return might not be a good proxy for the true daily volatility. Ahoniemi and Lanne (2013) find that incorporating the overnight return can lead to a more accurate realised volatility measure and can influence the relative performance of different volatility forecasting models. Wang et al. (2015) find that including the overnight return along with other explanatory variables improves volatility forecasts from the HAR-RV of Corsi (2009). In this study, we consider the approach of Blair et al. (2001) and Hua and Manzan (2013), which incorporates the overnight return in the realised volatility as follows:

$$RV_{N,t} = \sqrt{(RV_t)^2 + (y_{N,t})^2} \tag{3.4}$$

We can also incorporate the overnight return in the intra-day range, as in expression (3.5). To our knowledge, this has not previously been considered for VaR and volatility estimation:

$$Range_{N,t} = \sqrt{Range_t^2 + y_{N,t}^2} \tag{3.5}$$

Another way of incorporating the overnight information in the intra-day range is to replace the intra-day low and high log prices in expression (3.2) by the lowest and highest log prices in the daily close-to-close period (Gerlach and Wang 2016):

$$Range_{NC,t} = max(High_t, close_{t-1}) - min(Low_t, close_{t-1}) \tag{3.6}$$

In this study, we focus on the use of the intra-day range of expression (3.2), and the measure based on the intra-day range and overnight return in expressions (3.5)-(3.6).

3.3 Review of Established VaR Methods

In this section, we review the established VaR methods. Following the classification of Manganelli and Engle (2004), we divide the VaR literature into three categories: parametric, nonparametric and semi-parametric. Figure 3.1 is a chart of the classification of the major established VaR methods. The methods in bold are our proposed new models, which will be discussed in Section 3.4. A recent comprehensive review of the VaR literature can be found in Nieto and Ruiz (2016). We now discuss the established methods in more detail.

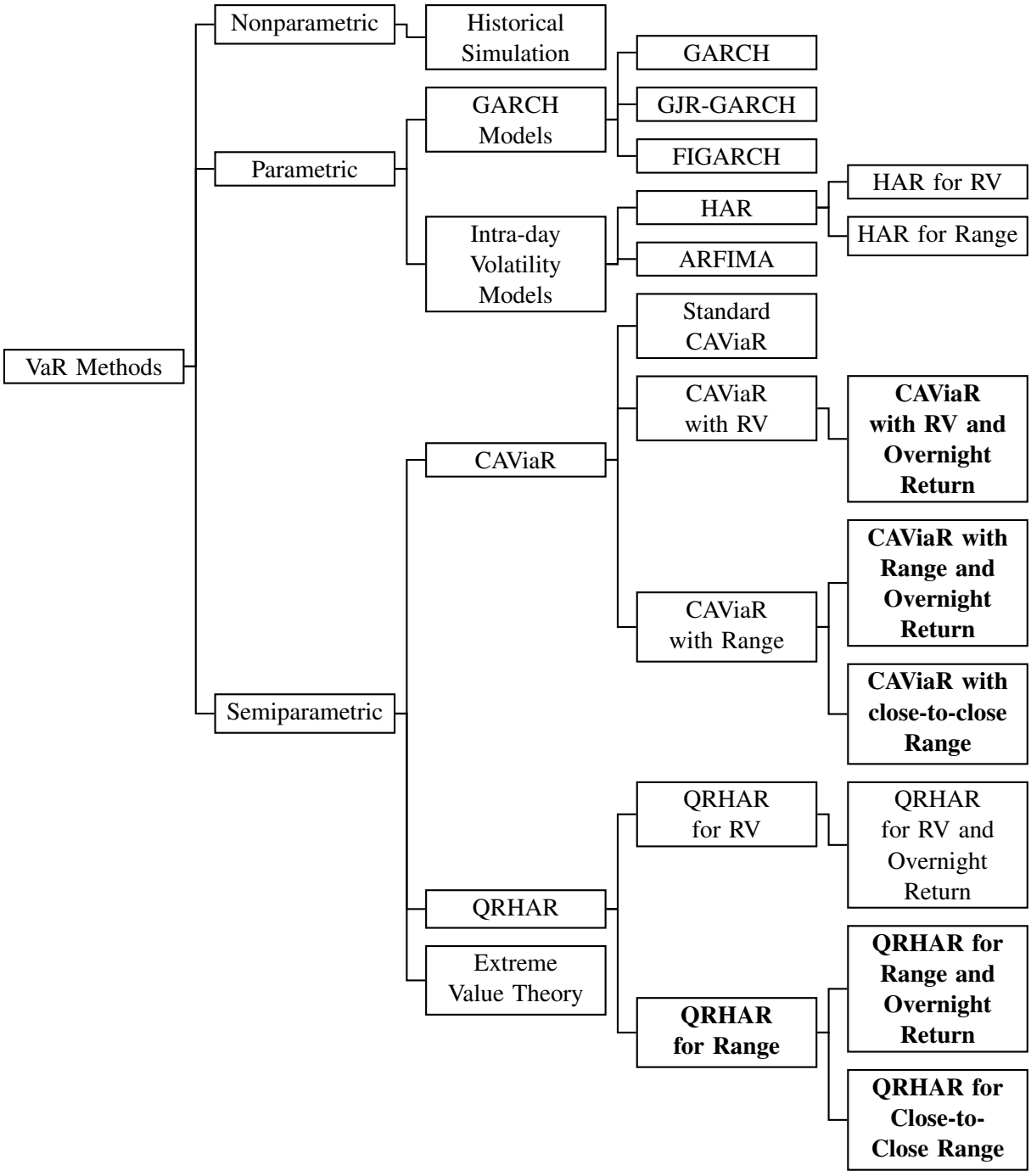
3.3.1 Nonparametric Methods

Nonparametric methods make no assumption about the distribution of the returns. The most common example in this class is historical simulation, which estimates the VaR as the sample quantile of the returns over a certain window length (Pritsker 2006). Nonparametric methods have the advantages of being model free and easy to implement (Manganelli and Engle 2004). However, historical simulation requires an implicit assumption that the distribution of the returns remains at least roughly the same within the window of observations, which may not be true (Manganelli and Engle 2004). Moreover, it is not easy to determine the window length: a short window might incur large sampling error while a long window may cause the method to be less responsive to the changes in the true distribution. In empirical comparisons, nonparametric methods have not performed well (Manganelli and Engle 2004).

3.3.2 Parametric Methods

In a parametric method, a quantile estimate is constructed by assuming a certain dynamic for the conditional volatility, $\sigma_t = \sqrt{var(y_t|\mathscr{F}_{t-1})}$, of the log return y_t, at time t conditional upon $\mathscr{F}_{t-1}$, the information set of all past information up to time $t-1$, and a specific conditional distribution for the log return. This can be viewed as the volatility of

Figure 3.1: Classification of established VaR methods with proposed new methods. The methods in bold are our proposed new models.



an error term $\varepsilon_t = y_t - \mathbb{E}(y_t|\mathscr{F}_{t-1})$, where $\mathbb{E}(y_t|\mathscr{F}_{t-1})$ is the conditional mean, which is often assumed to be zero or a small constant. The error term ε_t is usually referred to as the price 'shock'. The distribution of y_t is not observable, and so various distributional assumptions have been proposed. A Gaussian distribution is computationally convenient, but the fact that the distribution of the return is often fat-tailed has prompted the use of other distributions, such as the Student's t distribution.

For the conditional volatility σ_t, there are two widely-used types of models. Firstly, there are models based on the daily return, in which the daily volatility appears as a latent variable. Typical examples are the GARCH models. For example, the GARCH and GJR-GARCH models (Glosten et al. 1993) have been widely studied in VaR estimation (Gerlach et al. 2012). A slightly different type of GARCH model, the long-memory FIGARCH model, is proposed by Baillie et al. (1996) to capture the high persistence found in daily volatility.

The second type of volatility model induces an estimate of a certain intra-day volatility measure. Realised volatility RV_t, defined in expression (3.1), is undoubtedly the most popular intra-day volatility measure. Among the established realised volatility models, the HAR-RV model proposed by Corsi (2009), is a simple and pragmatic approach, whereby the forecast is constructed from the realised volatility over different time horizons. The HAR-RV model has been shown to perform well for volatility estimation, and its success and simplicity have made it popular among researchers (Corsi 2009; Bollerslev et al. 2016; Maheu and McCurdy 2011). These models have also been considered for VaR forecasting. Brownlees and Gallo (2010) show that the HAR-RV model improves VaR forecast accuracy in comparison with various GARCH models. Brownlees and Gallo (2010) also consider the HAR model, but with realised volatility replaced by the intra-day range, which is defined in expression (3.2). They find that the HAR model for the intra-day range is not outperformed by the HAR model for the realised volatility, and that both models improve VaR estimation accuracy compared to various GARCH models. We refer to the HAR model for the intra-day range as

HAR-Range. Another class of models based on realised volatility is the realised GARCH model of Hansen et al. (2012). These model the realised volatility and the volatility simultaneously, where the realised volatility is assumed to be dependent on the contemporaneous volatility, and the volatility is assumed to be modeled using its own lagged values and the lagged value of the estimated realised volatility. Hansen et al. (2012) find that the realised GARCH model outperforms a standard GARCH model. Gerlach and Wang (2016) consider the use of the intra-day range instead of the realised volatility in the realised volatility model. The model is shown to be competitive compared to the realised GARCH model. We refer to the realised GARCH model with realised volatility as the Realised-GARCH-RV model, and we refer to the realised GARCH model with the intra-day range as the Realised-GARCH-Range model.

3.3.3 Semiparametric Methods

The semiparametric methods include applications of extreme value theory and applications based on the use of quantile regression (Manganelli and Engle 2004). We focus on the semiparametric methods based on quantile regression, which estimates the desired quantile directly. A generic quantile regression model can be expressed in the following expression:

$$q_t(\boldsymbol{\beta}) = q(\boldsymbol{\Psi}_{t-1}, q_{t-1}(\boldsymbol{\beta}), \ldots, \boldsymbol{\Psi}_1, q_1(\boldsymbol{\beta}); \boldsymbol{\beta}, \theta) \tag{3.7}$$

where θ is the probability level of interest; $q_t(\boldsymbol{\beta})$ is the θ quantile of some daily return y_t; $\boldsymbol{\beta}$ is a vector of parameters; and $\boldsymbol{\Psi}_t$ is a vector of explanatory variables. The parameter vector $\boldsymbol{\beta}$ is chosen to minimise the following quantile regression objective function:

$$\sum_{t=1}^{T} [y_t - q_t(\boldsymbol{\beta})][\theta - \mathbb{1}(y_t < q_t(\boldsymbol{\beta}))] \tag{3.8}$$

where T is the number of observations in the in-sample period. Different quantile regression models involve different selections of $\boldsymbol{\Psi}_t$ and formulations of the $q_t(\cdot)$ in expression (3.7). Next we review several benchmark quantile regression models.

3.3.3.1 CAViaR Models and Quantile Regression

Engle and Manganelli (2004) propose four CAViaR models that use an autoregressive framework to model the evolution of the desired quantile directly. CAViaR models have several advantages. They are free from distributional misspecification since no distributional assumption is needed; they allow the conditional distribution to be time-varying; and the models can have different parameters for the quantiles associated with different probability levels, so that different quantiles are able to have distinct dynamics. Due to space constraints, we present only three CAViaR models here:

CAViaR-SAV:

$$q_t(\boldsymbol{\beta}) = \beta_1 + \beta_2 q_{t-1}(\boldsymbol{\beta}) + \beta_3 |y_{t-1}| \tag{3.9}$$

CAViaR-AS:

$$q_t(\boldsymbol{\beta}) = \beta_1 + \beta_2 q_{t-1}(\boldsymbol{\beta}) + \beta_3 (y_{t-1})^+ + \beta_4 (y_{t-1})^- \tag{3.10}$$

CAViaR-IndG:

$$q_t(\boldsymbol{\beta}) = \text{sgn}(\theta - 0.5)(\beta_1 + \beta_2 q_{t-1}(\boldsymbol{\beta})^2 + \beta_3 y_{t-1}^2)^{\frac{1}{2}} \tag{3.11}$$

where $(x)^+ = \max(x,0)$; and $(x)^- = -\min(x,0)$ in expression (3.10).

The lagged values of the returns in the CAViaR models serve as the innovation terms, which can be replaced or augmented by other explanatory variables, such as the intra-day volatility measures. Žikeš and Baruník (2015) use the realised volatility in the CAViaR models, and show that the proposed models improve VaR forecasting accuracy. Similar in spirit to this, Chen et al. (2012) consider several CAViaR models with the intra-day range, and show that the proposed models perform well when compared to a variety of VaR models. For illustrative purposes, we present here the CAViaR-SAV model with realised volatility and with the intra-day range:

CAViaR with Realised Volatility (CAViaR-RV):

$$q_t(\boldsymbol{\beta}) = \beta_1 + \beta_2 q_{t-1}(\boldsymbol{\beta}) + \beta_3 RV_t \tag{3.12}$$

CAViaR Range Value (CAViaR-Range):

$$q_t(\beta) = \beta_1 + \beta_2 q_{t-1}(\beta) + \beta_3 Range_{t-1} \tag{3.13}$$

3.3.3.2 Quantile Regression HAR Models Based on Realised Volatility

Quantile models with HAR-type specifications have also been considered. To distinguish these semiparametric HAR models from the parametric HAR models in Section 3.3.2, we refer to these approaches as *quantile regression HAR* (QRHAR) models. Hua and Manzan (2013) consider four QRHAR models based on realised volatility and the overnight return to estimate the quantiles of the returns. They find that the QRHAR models with intra-day volatility measures outperform benchmark GARCH models. However, they do not find that the inclusion of the overnight return brings any improvement. We present two QRHAR models here:

QRHAR with Realised Volatility (QRHAR-RV):

$$q_t(\beta) = \beta_0 + \beta_1 RV_{t-1} + \beta_2 RV^w_{t-1} + \beta_3 RV^m_{t-1} \tag{3.14}$$

$$RV^w_{t-1} = \frac{1}{5}\sum_{i=1}^{5} RV_{t-i} \tag{3.15}$$

$$RV^m_{t-1} = \frac{1}{22}\sum_{i=1}^{22} RV_{t-i} \tag{3.16}$$

QRHAR with Realised Volatility and overNight return (QRHAR-RV-N):

$$q_t(\beta) = \beta_0 + \beta_1 RV_{N,t-1} + \beta_2 RV^w_{N,t-1} + \beta_3 RV^m_{N,t-1} \tag{3.17}$$

$$RV^w_{N,t-1} = \frac{1}{5}\sum_{i=1}^{5} RV_{N,t-i} \tag{3.18}$$

$$RV^m_{N,t-1} = \frac{1}{22}\sum_{i=1}^{22} RV_{N,t-i} \tag{3.19}$$

where RV_t is the realised volatility; RV^w_t and RV^m_t in expressions (3.14)-(3.16) denote the average realised volatility over a week and a month respectively; and $RV_{N,t}$ in

expressions (3.17)-(3.19) is defined as in expression (3.4).

3.4 Proposed New Models for VaR Forecasting

In this section, we propose four new semiparametric models. Each is a quantile model estimated using quantile regression. Quantile regression modelling is appealing because it avoids the need for a distributional assumption, it allows the quantiles associated with different probability levels to have different dynamics, and such models have performed well in empirical studies (Manganelli and Engle 2004; Taylor 2008; Chen et al. 2012). The new quantile regression models that we consider are CAViaR and QRHAR formulations. Our new models incorporate the overnight return and intra-day range in these models.

3.4.1 Two New CAViaR Models Incorporating the Overnight Return

We first consider the use of the overnight return. As we explained in Section 3.2, information regarding public announcements and relevant information arriving from markets abroad, after the previous day's closing time, can be useful for forecasting the distribution of daily returns. Intra-day volatility measures, such as the intra-day range and the realised volatility, contain only the information during the market opening hours of the day. In general, the use of the overnight return has been neglected in the VaR literature. To the authors' knowledge, the only study that has used the overnight return for VaR forecasting methods is the study of Hua and Manzan (2013), who incorporate the overnight return in QRHAR models based on realised volatility. In that study, the overnight return did not lead to improved VaR forecast accuracy. In this study, we revisit the use of the overnight return, but instead consider its use within CAViaR models. As the overnight return is, in essence, a measure based on intra-day data, we opt to use the versions of CAViaR models that are based on intra-day data. The first is based on the CAViaR-RV model of expression (3.12) and the other two are based on the CAViaR-Range model of expression (3.13):

CAViaR with Realised Volatility and overNight return (CAViaR-RV-N):

$$q_t(\boldsymbol{\beta}) = \beta_1 + \beta_2 q_{t-1}(\boldsymbol{\beta}) + \beta_3 RV_{t-1} + \beta_4 |y_{N,t-1}| \tag{3.20}$$

CAViaR with intra-day Range and overNight return (CAViaR-Range-N):

$$q_t(\boldsymbol{\beta}) = \beta_1 + \beta_2 q_{t-1}(\boldsymbol{\beta}) + \beta_3 Range_{t-1} + \beta_4 |y_{N,t-1}| \tag{3.21}$$

CAViaR with intra-day close-to-close Range (CAViaR-Range-C):

$$q_t(\boldsymbol{\beta}) = \beta_1 + \beta_2 q_{t-1}(\boldsymbol{\beta}) + \beta_3 Range_{NC,t-1} \tag{3.22}$$

where $y_{N,t}$ is the overnight return, as defined in expression (3.3). Note that, we could also add the overnight return to the other CAViaR models of Section 3.3.3.1, but for simplicity we only consider the above two models in this study.

3.4.2 Two New QRHAR Models Incorporating the Intra-day Range

Next, we consider the use of the intra-day range within a QRHAR model. The HAR model has been shown to perform well for volatility estimation, and its success and simplicity have made it popular among researchers (Corsi 2009; Brownlees and Gallo 2010; Maheu and McCurdy 2011; Bollerslev et al. 2016). The HAR model considers the use of intra-day volatility aggregated over different horizons to capture the asymmetry and long-memory property in volatility dynamics. The motivation for this is that it has been observed that the volatility over a long period has a strong impact on the volatility over a short period, and there is significant evidence of long-memory in daily volatility (Baillie et al. 1996; Andersen et al. 2001a). The HAR model is able to capture these features in a simple formulation, which contrasts with standard GARCH models (Corsi et al. 2008). Given the success of the HAR models in volatility estimation, it seems somewhat surprising that the QRHAR models, which use the HAR structure within a quantile regression framework,

have not attracted more attention in VaR forecasting. The only such study, of which the authors are aware, is the work of Hua and Manzan (2013), which, as we described in the previous section, is based on realised volatility. In this study, we introduce the idea of incorporating the intra-day range within a QRHAR model. Our motivation for this is threefold. Firstly, the intra-day range is an efficient estimator of the daily volatility, and so it is a suitable candidate explanatory variable in a VaR model. Secondly, various studies have shown the intra-day range has competitive predictive power compared to the realised volatility in terms of volatility and VaR forecasting (Brownlees and Gallo 2010; Chen et al. 2012; Gerlach and Wang 2016). Thirdly, a practical advantage of the intra-day range is that it is free and readily available for most tradable assets over long historical periods, which contrasts with realised volatility, which has the added inconvenience of needing to be processed prior to use (Rogers and Zhou 2008). We propose the following two new QRHAR models based on the intra-day range:

QRHAR with intra-day Range (QRHAR-Range):

$$q_t(\boldsymbol{\beta}) = \beta_1 + \beta_2 Range_{t-1} + \beta_3 Range^w_{t-1} + \beta_4 Range^m_{t-1} \tag{3.23}$$

$$Range^w_{t-1} = \frac{1}{5}\sum_{i=1}^{5} Range_{t-i} \tag{3.24}$$

$$Range^m_{t-1} = \frac{1}{22}\sum_{i=1}^{22} Range_{t-i} \tag{3.25}$$

QRHAR with intra-day Range and overNight return (QRHAR-Range-N):

$$q_t(\boldsymbol{\beta}) = \beta_1 + \beta_2 Range_{N,t-1} + \beta_3 Range^w_{N,t-1} + \beta_4 Range^m_{N,t-1} \tag{3.26}$$

$$Range^w_{N,t-1} = \frac{1}{5}\sum_{i=1}^{5} Range_{N,t-i} \tag{3.27}$$

$$Range^m_{N,t-1} = \frac{1}{22}\sum_{i=1}^{22} Range_{N,t-i} \tag{3.28}$$

QRHAR with close-to-close Range (QRHAR-Range-C):

$$q_t(\boldsymbol{\beta}) = \beta_1 + \beta_2 Range_{NC,t-1} + \beta_3 Range^w_{NC,t-1} + \beta_4 Range^m_{NC,t-1} \tag{3.29}$$

$$Range^w_{NC,t-1} = \frac{1}{5}\sum_{i=1}^{5} Range_{NC,t-i} \tag{3.30}$$

$$Range^m_{NC,t-1} = \frac{1}{22}\sum_{i=1}^{22} Range_{NC,t-i} \tag{3.31}$$

where $Range_{N,t}$ in expressions (3.26)–(3.28) is defined as in expression (3.5); $Range_{Nc,t}$ in expressions (3.29)–(3.31) is defined as in expression (3.6); $Range^w_t$, $Range^w_{N,t}$ and $Range^w_{NC,t}$ denote the average of $Range_t$, $Range_{N,t}$ and $Range_{NC,t}$ over a week; and $Range^m_t$, $Range^m_{N,t}$ and $Range^m_{NC,t}$ denote the average of $Range_t$, $Range_{N,t}$ and $Range_{NC,t}$ over a month. As we stated in Section 3.2, we are not aware of the previous use of $Range_{N,t}$ in VaR forecasting. $Range_{NC,t}$ has been considered in parametric models in Gerlach and Wang (2016), but has not been considered in a semiparametric quantile regression model. Expressions (3.23)-(3.25), expressions (3.26)-(3.28) and expressions (3.29)-(3.31) are essentially the QRHAR-RV model of expressions (3.14)-(3.16), and the QRHAR-RV-N model of expressions (3.17)-(3.19), with the realised volatility terms replaced by the intra-day range. We will show in the empirical study that the use of the intra-day range leads to better forecasting accuracy, in spite of the similarity between model structures.

In order to illustrate our contribution, consider again Figure 3.1. It can be seen from Figure 3.1 that we focus on the use of the intra-day range and the overnight return for semiparametric quantile regression models. We are proposing two new CAViaR models that include the overnight return in two existing CAViaR models, and two new QRHAR models that differ from the existing QRHAR models in that they are based on the intra-day range, rather than the realised volatility.

3.5 Empirical Results

3.5.1 Data

To explore the forecasting performance of our four proposed new VaR models, and a set of benchmark methods, we used the daily opening, closing, high and low prices of the following 18 global stocks: S&P500, FTSE100, Nikkei225, DAX30, Russel2000, All Ordinaries, DJIA, Nasdaq100, CAC40, KOSPI Composite Index, AEX Index, Swiss Market Index, IBEX35, IPC Mexico, Bovespa Index, Euro STOXX 50, FT Straits Times Index, and FTSE Mib Index. We used the realised volatility of all the global stocks sampled at a 5-min sampling frequency. We obtained the data from the Oxford-Man Institute's realised library Version 0.2 (Heber et al. 2009). We excluded only two indices from the Oxford-Man Institute's realised library, the Hang Seng and S&P/TSX Composite Index, due to their insufficient number of observations. The sample period used in our study was from 3 January 2000 to 20 May 2014. We chose a post-sample period consisting of 1500 days to evaluate day-ahead quantile estimates. We employed a rolling window approach, where each method was re-estimated for each day in the post-sample period using data from the previous 1800 days. Note that different markets tend to have different public holidays, and so they can have different numbers of observations within a certain calendar period. We chose our post-sample period in such a way that the observations from all the markets shared the same end date, which was 20 May 2014. Following common practice, we did not estimate the conditional mean of the returns, and assumed it to be zero (see, for example, Žikeš and Baruník 2015).

3.5.2 VaR Forecasting Methods

For each method considered, one-day-ahead forecasts of VaR were generated for each day in the post sample period. We considered the 1% and the 5% quantiles, as is common in the VaR literature. We implemented the historical simulation nonparametric method with a variety of different window lengths. However, as the results were poor, for simplicity,

we do not discuss the results further.

As parametric benchmarks, we implemented the GARCH(1,1) model, the GJR-GARCH(1,1) model, the FIGARCH(1,d,1) model, where d is a parameter between 0 and 1, the HAR-RV model, the HAR-Range model, the Realised-GARCH-RV model, and the Realised-GARCH-Range model. For each model, we used a Student's t distribution for parameter estimation and for VaR forecasting.

In terms of semiparametric methods, we included a variety of quantile regression benchmark methods in our empirical study. These are models from the literature, rather than new models proposed by us. We first implemented the standard CAViaR models based on the daily return: the CAViaR-SAV model of expression (3.9), the CAViaR-AS model of expression (3.10), and the CAViaR-IndG model of expression (3.11). For the CAViaR models based on realised volatility, we implemented all the models proposed by Žikeš and Baruník (2015). We found that these models performed similarly to each other in our study. Therefore, we only report the results of the CAViaR-RV model of expression (3.12), which uses just the realised volatility alongside the lagged quantile. For the CAViaR models based on the intra-day range, we also implemented all the models proposed by Chen et al. (2012). We once again found that these models performed similarly to each other in our empirical study. In view of this, to save space, we only present the results for the CAViaR-Range model of expression (3.13). Moving on to the QRHAR models, we implemented the QRHAR-RV model of expressions (3.14)-(3.16), and the QRHAR-RV-N model of expressions (3.17)-(3.19).

With regard to our proposed new methods, we implemented the six models in Section 3.4: the CAViaR-RV-N model based on realised volatility and the overnight return of expression (3.20); the CAViaR-Range-N model based on the intra-day range and the overnight return of expression (3.21); the CAViaR-Range-C model based on the close-to-close range of expression (3.22); the QRHAR-Range model based on the intra-day range of expressions (3.23)-(3.25); the QRHAR-Range-N model based on the intra-day range and the overnight return of expressions (3.26)-(3.28); and the

QRHAR-Range-N model based on the intra-day range and the overnight return of expressions (3.29)-(3.31).

3.5.3 In-Sample Parameters

Table 3.1 presents parameters for our new CAViaR-Range-N model of expression (3.21) and our new QRHAR-Range-N model of expressions (3.26)-(3.28) estimated using the FTSE100 data. Note that we estimated 1500 different parameter vectors, since the parameters are re-estimated as the rolling window passes through the post-sample period. In Table 3.1, we present the estimated parameters derived using the first moving window, together with their p-values. For the CAViaR-Range-N model, the parameters of $Range_{t-1}$ and $|y_{N,t-1}|$ are negative, indicating that larger values of these variables will result in a lower estimated quantile, which is intuitive. The parameters of q_{t-1} are positive, which is consistent with the well-known volatility clustering effect. For the QRHAR-Range-N model, all the parameters except β_4 for the 5% quantiles are negative, indicating that larger values of $Range_{t-1}$, $Range^w_{t-1}$, and $Range^m_{t-1}$ will result in a lower value of the estimated quantile. The exception of β_4 might stem from multicollinearity between $Range_{t-1}$, $Range^w_{t-1}$, and $Range^m_{t-1}$. We will discuss this issue and the p-values in more detail later.

In Table 3.2, we report the number of times, out of 1500, that each parameter was statistically significantly different from zero at the 5% significance level for the CAViaR-Range-N model and QRHAR-Range-N models. For the CAViaR-Range-N model, it can be seen from Table 3.2 that the coefficient β_2 of the lagged quantile is almost always statistically significant; the coefficient β_3 of the intra-day range was statistically significant for more than half the periods; and the coefficient β_4 for the overnight return was significant reasonably often.

The results for the QRHAR-Range-N model in Table 3.2 are somewhat more complex. The coefficient β_3 of $Range^w_{N,t-1}$ was always significant for both quantiles; and the coefficient β_4 of $Range^m_{N,t-1}$ was significant for a reasonable proportion of the periods,

Table 3.1: Estimated parameters of the CAViaR-Range-N model and the QRHAR-Range-N model for the FTSE100, estimated using the first moving window of 1800 days.

	CAViaR-Range-N		QRHAR-Range-N	
		1% Quantiles		
	Parameters	p-values	Parameters	p-values
β_1	-0.116	0.113	-0.182	0.078
β_2	0.588	0.003	-0.229	0.014
β_3	-0.622	0.145	-2.045	0.000
β_4	-0.460	0.067	0.414	0.009
		5% Quantiles		
β_1	-0.035	0.276	-0.227	0.014
β_2	0.792	0.000	-0.089	0.305
β_3	-0.217	0.151	-0.967	0.000
β_4	-0.133	0.241	-0.053	0.662

Table 3.2: Number of times, out of 1500, that each estimated parameter, was significantly different from zero at the 5% significance level, for the CAViaR-Range-N model and the QRHAR-Range-N model applied to the FTSE100.

	CAViaR-Range-N		QRHAR-Range-N	
	1%	5%	1%	5%
β_1	0	0	102	778
β_2	1498	1500	1500	0
β_3	1043	823	1500	1500
β_4	202	107	226	608

especially for the 5% quantile, which suggests the presence of long-memory in VaR. The coefficient β_2 of $Range_{N,t-1}$ was always significant for the 1% quantiles, but it was never significant for the 5% quantiles. Our explanation for this is that $Range_{N,t-1}$ and $Range^{w}_{N,t-1}$ are highly correlated. In fact, the correlation between these two variables for our FTSE100 data is 0.85. However, multicollinearity is not necessarily a problem for forecasting (see, for example, Brooks 2014). In fact, we estimated the QRHAR-Range-N model without the $Range_{N,t}$ term, but this worsened the post-sample VaR forecasting performance. Thus, we conclude that $Range_{N,t-1}$, $Range^{w}_{N,t-1}$ and $Range^{m}_{N,t-1}$ should all be included in the model.

3.5.4 Post-Sample Evaluation

3.5.4.1 Evaluation Measures

We evaluated unconditional and conditional coverage in the post-sample VaR forecasts. The unconditional coverage test assesses the hit percentage defined as the proportion of the returns below the estimated quantile corresponding to probability level θ. Significant difference from the ideal value of θ is tested using a binomial distribution. To evaluate conditional coverage, we used the dynamic quantile test proposed by Engle and Manganelli (2004), which is a joint test of correct unconditional coverage and independence of violations. It evaluates whether the sequence of the hit variable, defined as $Hit_t = \mathbb{1}(y_t < q_t(\beta)) - \beta$, is distributed i.i.d. Bernoulli with probability θ, and is independent of lags of the Hit_t variable and the conditional quantile forecast. Following Engle and Manganelli (2004), we used four lags of the Hit_t variable in the test's regression.

We also evaluated the VaR forecasts using the quantile score, which is the quantile regression loss function used in the summation of expression (3.8). The quantile score is a proper scoring rule for evaluating quantile forecasts, implying that the expectation of the score is minimised by the perfect set of quantile forecasts (Gneiting and Raftery 2007). For each method, and each of the 18 stock indices, we computed the average quantile score. As the resulting values have no intuitive interpretation, for each method and each series, we calculated the ratio of the score to that of a benchmark method, then subtracted this ratio from one, and multiplied the result by 100. We term this the quantile skill score. For the calculation of the skill score, we used the GARCH-t model as the benchmark. To summarize performance across the 18 indices, we calculated the geometric mean of the ratios of the score for each method to the score for the GARCH-t benchmark method, then subtracted this from one, and multiplied the result by 100. Higher values are preferable for the resulting geometric mean values, and indeed the skill scores, reflecting superiority over the benchmark method.

Finally, we use the model confidence set evaluation method to select the

best-performing models (Hansen et al. 2011). In practice, limited sample size can decrease the power of statistical tests. Two models might not be distinguishable when there is not a sufficiently large number of post-sample observations. The model confidence set approach does not assume that the confidence sets contain the best method, instead, the approach selects a set of best-performing models that will contain the best model with a given level of confidence. We use the quantile score as the loss function, and select methods for the confidence set based on the Diebold-Mariano test for statistical significance between the mean of the quantile scores of pairs of methods (Diebold and Mariano 2002). Following Hansen et al. (2011), we used 75% and 90% as the confidence levels for the model confidence sets.[1]

3.5.4.2 Evaluation Results

We now consider the post-sample evaluation results. For the FTSE100, Table 3.3 presents the hit percentage, the p-value for the dynamic quantile test, and the quantile skill score for each method. Larger p-values are preferred for the dynamic quantile test. The results for our six new proposed models are presented in the bottom six rows of the table. We present this table merely for illustrative purposes, as we must consider the results for all 18 stock indices in order to compare the methods. This is done in Table 3.4, which presents the number of rejections for the unconditional coverage test and the dynamic quantile test at the 5% significance level for each model applied to the 18 indices, as well as the skill score averaged across all 18 indices. Our discussion focuses on this summary table.

Let us first consider the results for unconditional coverage in Table 3.4. We find that all the parametric methods and the four semiparametric methods based on realised volatility have large numbers of rejections. All the other models perform reasonably well. The poor performance of the parametric methods suggests misspecification of the volatility model or inappropriate choice of distribution. The poor performance of the

[1]To implement the model confidence set approach to evaluation, we used the MATLAB code provided by Kevin Sheppard in his MFE Toolbox, which is freely available on https://www.kevinsheppard.com/MFE_Toolbox.

semiparametric methods utilizing the realised volatility is somewhat surprising, since the other semiparametric methods all perform well. One explanation is that the realised volatility is contaminated by microstructure noise. We next consider the dynamic quantile test results. The proposed new models have fewer rejections than the benchmark methods, with the CAViaR-Range-N model, the CAViaR-Range-C model, the QRHAR-Range-N model and the QRHAR-Range-C model performing the best. Now we move on to the skill scores using the GARCH-t model as the benchmark. A positive skill score indicates the model has outperformed the benchmark method. Naturally, the skill scores for the benchmark method are zero. The six proposed new methods have positive skill scores for both the 1% and 5% quantiles. One general observation is that the models including the realised volatility or the intra-day range perform better than the methods based only on the daily returns. Despite their relatively poor coverage test results, the realised GARCH models have the best skill scores for the 1% quantile. The proposed QRHAR-Range-C and QRHAR-Range-N models perform the best for the 5% quantiles. Overall, the proposed QRHAR-Range-N and the CAViaR-Range-N model perform very competitively.

To further illustrate the benefit of including the intra-day range and the overnight return in quantile regression models, we compare the numbers of rejections between the models based on realised volatility and the corresponding models based on the intra-day range, and the numbers of rejections between the models without the overnight return and the corresponding models with the overnight return. We provide this comparison for the unconditional coverage and dynamic quantile tests in Table 3.5. For each pair of models, the numbers are obtained by subtracting the numbers of test rejections of the latter models from those of the former models. A positive number indicates that the former model is outperformed by the latter. In Table 3.5, the upper block of results compares the performance between the quantile regression models using realised volatility and the intra-day range. The lower block of results compares the performance between the quantile regression models with and without the overnight return. The results suggest that

the models based on the intra-day range outperform the corresponding models based on realised volatility, and the models using the overnight return outperform the corresponding models without the overnight return.

Table 3.3: Evaluation of 1% and 5% VaR forecast accuracy for the FTSE100. The hit percentage, the *p*-values for the dynamic quantile test, and the quantile skill scores using the GARCH-t model as the benchmark.

	1% Quantiles			5% Quantiles		
	Hit%	Dynamic Quantile	Skill Score%	Hit%	Dynamic Quantile	Skill Score%
Parametric Benchmarks						
GARCH-t	1.5	0.003**	0.0	6.7**	0.029*	0.0
GJR-GARCH-t	1.5	0.200	1.7	6.1	0.266	1.3
FIGARCH-t	1.1	0.169	0.4	6.4*	0.035*	0.4
HAR-RV-t	1.7*	0.101	-1.5	6.7**	0.043*	1.3
HAR-Range-t	1.5	0.003**	-1.9	6.6**	0.031*	-0.9
Realised-GARCH-RV-t	1.6*	0.078	0.6	6.2*	0.132	1.8
Realised-GARCH-Range-t	1.5	0.002**	-0.7	6.6**	0.04**	0.1
Semiparametric Benchmarks						
CAViaR-SAV	0.9	0.061	-5.9	5.6	0.131	-0.4
CAViaR-AS	1.2	0.536	2.6	5.3	0.597	1.0
CAViaR-IndG	0.7	0.154	-1.7	5.3	0.170	0.0
CAViaR-RV	1.0	0.991	0.2	5.1	0.636	1.9
CAViaR-Range	1.0	0.471	-1.3	5.2	0.937	1.1
QRHAR-RV	1.1	0.986	-4.1	5.3	0.505	2.1
QRHAR-RV-N	1.0	0.992	-2.6	5.2	0.884	0.1
New Semiparametric Models						
CAViaR-RV-N	1.1	0.975	2.4	5.1	0.594	2.1
CAViaR-Range-N	1.1	0.532	1.1	4.9	0.982	1.0
CAViaR-Range-C	1.1	0.470	2.0	5.3	0.953	1.1
QRHAR-Range	0.9	0.426	0.4	5.6	0.727	1.5
QRHAR-Range-N	1.1	0.986	1.3	5.3	0.924	1.8
QRHAR-Range-C	1.4	0.557	-0.5	5.4	0.512	1.2

Note: Larger *p*-values are better. Significance at 5% and 1% levels is indicated by * and **, respectively. Larger skill scores are better.

Table 3.4: Summary of the VaR forecast accuracy results. Number of test rejections at the 5% significance level for the unconditional coverage and dynamic quantile tests, and the skill scores using the GARCH-t model as the benchmark.

	Unconditional Coverage			Dynamic Quantile			Skill Score	
	1%	5%	Total	1%	5%	Total	1%	5%
Parametric Benchmarks								
GARCH-t	6	7	13	9	9	18	0.0	0.0
GJR-GARCH-t	8	6	14	8	3	11	2.5	1.7
FIGARCH-t	2	4	6	8	10	18	1.0	0.6
HAR-RV-t	10	8	18	9	8	17	3.3	2.2
HAR-Range-t	9	8	17	6	9	15	2.7	1.1
Realised-GARCH-RV-t	10	9	19	10	7	17	3.7	2.1
Realised-GARCH-Range-t	6	7	13	7	9	16	3.7	1.5
Semiparametric Benchmarks								
CAViaR-SAV	1	0	1	8	3	11	-3.1	-0.9
CAViaR-AS	3	0	3	7	3	10	0.2	1.1
CAViaR-IndG	1	0	1	9	4	13	-1.1	-0.5
CAViaR-RV	4	4	8	6	6	12	3.0	2.3
CAViaR-Range	0	0	0	6	3	9	2.5	2.1
QRHAR-RV	6	4	10	9	5	14	1.2	2.0
QRHAR-RV-N	5	2	7	6	3	9	0.9	1.8
New Semiparametric Models								
CAViaR-RV-N	6	1	7	5	3	8	2.6	2.0
CAViaR-Range-N	1	0	1	1	1	2	2.5	2.0
CAViaR-Range-C	2	0	2	3	2	5	2.6	2.4
QRHAR-Range	3	1	4	6	1	7	1.4	2.2
QRHAR-Range-N	2	0	2	4	0	4	1.7	2.4
QRHAR-Range-C	2	0	2	3	1	4	1.5	2.2

Note: Smaller numbers of rejections and larger skill scores are better.

Table 3.5: Comparison of the number of rejections for the unconditional coverage and dynamic quantile tests at the 5% significance level.

	Unconditional Coverage		Dynamic Quantile	
	1%	5%	1%	5%
Realised volatility and intra-day range				
CAViaR-RV - CAViaR-Range	4	4	0	3
CAViaR-RV-N - CAViaR-Range-N	5	1	4	2
QRHAR-RV - QRHAR-Range	3	3	3	4
QRHAR-RV-N - QRHAR-Range-N	3	2	2	3
Without and with the overnight return				
CAViaR-RV - CAViaR-RV-N	-2	3	1	3
CAViaR-Range - CAViaR-Range-N	-1	0	5	2
CAViaR-Range - CAViaR-Range-C	-2	0	3	1
QRHAR-RV - QRHAR-RV-N	1	2	3	2
QRHAR-Range - QRHAR-Range-N	1	1	2	1
QRHAR-Range - QRHAR-Range-C	1	1	3	0

Note: The numbers are obtained by subtracting the number of test rejections in Table 3.4 of the latter model from that of the former model. A positive number indicates that the former model is outperformed by the latter.

Finally, we report the number of series for which each model lies inside the model confidence set for the 75% and 90% confidence levels in Table 3.6. A large number is preferred as it indicates the corresponding model is not frequently eliminated in the selection process. A value of 18 is the best possible performance, and this indicates that the method was not eliminated from the model confidence set for any of the 18 indices. Although the results are not particularly helpful in differentiating between the accuracy of the methods, we can still draw some conclusions from the test results. Firstly, the proposed models are the best-performing models, with very few times being excluded from the model confidence set. Secondly, in general, the models including the intra-day data perform well, while the models solely depending on the historical returns are more often excluded from the model confidence set, with the GJR-GARCH-t model being the only exception.

To summarize, the results of the proposed new methods are promising. In terms of the unconditional coverage and dynamic quantile tests, the new CAViaR-Range-N and

QRHAR-Range-N models perform the best. They have the lowest numbers of rejections for the dynamic quantile test and have very low numbers of rejections for the unconditional coverage test. The results were also good for our other two newly proposed methods, as well as the QRHAR-RV-N model proposed by Hua and Manzan (2013), and the CAViaR-Range model proposed by Chen et al. (2012). In terms of the quantile scores and the model confidence set evaluation method, the methods based on the intra-day data are better than the methods solely depending on the historical returns, and the methods using the intra-day range were generally more accurate than the methods using the realised volatility.

Table 3.6: Summary of the model confidence set results for the 1% and 5% quantiles using the Diebold-Mariano test on based on quantile scores. Number of series each model is inside the model confidence set for 75% and 90% confidence level.

	1% Quantiles		5% Quantiles	
Confidence Level	75%	90%	75%	90%
Parametric Benchmarks				
GARCH-t	16	18	12	16
GJR-GARCH-t	17	18	16	17
FIGARCH-t	17	18	13	17
HAR-RV-t	18	18	17	18
HAR-Range-t	18	18	17	18
Realised-GARCH-RV-t	18	18	17	18
Realised-GARCH-Range-t	18	18	18	18
Semiparametric Benchmarks				
CAViaR-SAV	12	16	9	11
CAViaR-AS	16	18	14	16
CAViaR-IndG	15	15	10	14
CAViaR-RV	18	18	17	18
CAViaR-Range	18	18	18	18
QRHAR-RV	18	18	18	18
QRHAR-RV-N	16	18	17	18
New Semiparametric Models				
CAViaR-RV-N	18	18	17	18
CAViaR-Range-N	18	18	17	18
CAViaR-Range-C	18	17	18	18
QRHAR-Range	17	18	18	18
QRHAR-Range-N	17	18	18	18
QRHAR-Range-C	16	17	18	18

Note: Large numbers are better.

3.6 Concluding Remarks

In this study, we have considered the use of realised volatility, the intra-day range, and the overnight return in quantile regression models to improve VaR forecasting accuracy. Our proposed new methods delivered promising results in our empirical study. The proposed QRHAR-Range-N model, QRHAR-Range-C model, and the CAViaR-Range-N model, and the and the CAViaR-Range-C model, which use the intra-day range and the overnight return, performed particularly well. The empirical study also included previously proposed methods, and enabled a comparison of methods based on the intra-day range with methods based on realised volatility, and a comparison between methods including and excluding the overnight return. We found that the intra-day range is more useful than realised volatility for VaR forecasting, and that it is beneficial to include the overnight return. A further contribution of the study is that we provide empirical support for the previous work of Chen et al. (2012) and Hua and Manzan (2013).

In the next chapter, we will continue focusing on the use of intra-day data, but we will be focusing on ES as well as VaR. In addition, we will consider the use of the intra-day data in model estimation, which is not discussed in this chapter.

Chapter 4

Using the Intra-day Low and Range to Forecast Value-at-Risk and Expected Shortfall

4.1 Introduction

VaR has become a standard tool for risk management in financial and insurance institutions. It involves measuring the amount a certain portfolio can lose for a given probability level. Formally, VaR is defined as a tail quantile of the distribution of a financial return. The accurate forecasting of VaR is fundamental for internal risk control and financial regulation. Despite its importance and popularity, VaR has some undesirable features. Firstly, it is not a sub-additive risk measure. In practice, lack of sub-additivity violates the risk diversification principle, which means the risk of a portfolio should be no more than that of the aggregate of its individual components (Righi and Ceretta 2015). Secondly and more importantly, it does not take into account the magnitude of the potential loss beyond the VaR. In view of these concerns, ES has been proposed as a risk measure for future regulatory frameworks (Embrechts et al. 2014). ES is defined as the expected value of exceedances beyond the VaR. However, ES is not elicitable, which

means there does not exist a scoring function that is minimised by the true ES. This has posed difficulties for both the estimation and evaluation of ES in practice. In recent work, Fissler and Ziegel (2016) find that ES and VaR are jointly elicitable, and provide a family of scoring functions that are minimised by the true VaR and ES. Such scoring functions enable the joint estimation and evaluation of VaR and ES. For example, Patton et al. (2017) and Taylor (2017a) estimate autoregressive models for VaR and ES based on joint scoring functions of this type.

Classical approaches to forecasting daily VaR and ES use only historical daily returns, such as GARCH models (Bollerslev 1987), historical simulation, and CAViaR models (Engle and Manganelli 2004). The studies of Patton et al. (2017) and Taylor (2017a) are also solely based on historical daily returns. With intra-day data becoming increasingly available, efforts have been made to use it in the forecasting of VaR and ES for daily returns (see, for example, Clements et al. 2008; Brownlees and Gallo 2010). However, the use of intra-day data in VaR and ES estimation has tended to involve high-frequency data, which is generally expensive, with only limited availability (Kumar and Maheswaran 2015). In this study, we consider the use of the intra-day low and high series to improve VaR and ES forecast accuracy. These data are readily available for most tradable assets for the past 30 years.

Our aim in this study is to improve CAViaR modelling in order to generate more accurate VaR and ES forecasts. CAViaR models are appealing because they estimate the quantiles directly, and have been shown to perform competitively in comparison with other VaR models (Manganelli and Engle 2004; Chen et al. 2012; Gerlach et al. 2012). Using a joint scoring function, Taylor (2017a) estimates the VaR and ES using a CAViaR model for the VaR, and ES modelled as a constant multiple of the VaR, or assumed to be equal to the VaR plus a dynamic model for the exceedances. In this study, we make two contributions. First, to enrich the models proposed by Taylor (2017a), we incorporate the intra-day range, which is defined as the absolute difference between the lowest and highest log prices of a day. Secondly, and more significantly, we show that the dynamics

of the VaR and ES of the daily returns can be well approximated by the dynamics of the VaR and ES of the *intra-day low returns*, which we define as the difference between the lowest log price of the day and the log closing price of the previous day. Intuitively, the intra-day low returns should contain valuable information about the lower tails of the daily return. We propose a method, in which the VaR and ES of the intra-day low returns are used as the basis for estimating the VaR and ES of the daily returns.

In Section 4.2, we review established VaR and ES methods, and introduce a new model, which is a simple extension of existing approaches. Section 4.3 describes how VaR and ES estimated for the intra-day low returns can be used to estimate VaR and ES for the daily returns. Section 4.4 uses five series of stock index returns to evaluate the performance of the proposed new methods, and to compare their estimation accuracy with the established benchmark methods. Section 4.5 provides a summary and some concluding remarks.

4.2 VaR and ES Methods

In this section, we review time series methods for VaR and ES, and the recently proposed joint scoring functions of Fissler and Ziegel (2016), which enable joint modelling. We then describe VaR models that incorporate intra-day information. We end the section by proposing a new model that uses the intra-day range, and can be estimated using a joint scoring function to enable both VaR and ES prediction.

4.2.1 Time Series Methods for VaR and ES

The methods available for VaR and ES forecasting can be classified as parametric, nonparametric or semiparametric. Nonparametric methods involve no parameterisation for the conditional volatility and no assumption for the returns distribution. Examples are historical simulation and kernel density estimation, which are applied to moving windows of observations. Although these methods are appealingly easy to implement, they implicitly assume that the returns distribution remains reasonably constant within the

specified window length, which may be inappropriate (Manganelli and Engle 2004). Moreover, the choice of the window length is not clear.

Parametric models involve a parameterisation of the conditional volatility and an assumption for the distribution of the return. The VaR and ES estimates can then be obtained from the estimated distribution. A common example is a GARCH model with a Gaussian or Student t distribution. Although parametric models provide an estimate of the complete conditional distribution of the return series, they usually suffer from some degree of misspecification, either from the model structure or from the distributional assumption.

Semiparametric methods involve either the use of EVT, filtered historical simulation, or a direct modelling of VaR and ES in a regression framework. A popular EVT approach fits a generalised Pareto distribution to exceedances over some threshold of the standardised residuals from a parametric model (McNeil and Frey 2000). Filtered historical simulation avoids a distributional assumption by using the empirical distribution of standardised residuals. In a semiparametric regression approach, a model is estimated by minimising a score function summed over the in-sample period. For example, the CAViaR models of Engle and Manganelli (2004) are estimated using quantile regression, which amounts to minimising the quantile score. By directly modelling individual quantiles, CAViaR models have the advantage of making no distributional assumption, and allowing different quantiles to have distinct dynamics. An apparent limitation of these models is that they do not deliver ES predictions. However, this concern has been alleviated by the recent development of joint scoring functions for VaR and ES.

4.2.2 Joint Scoring Functions for VaR and ES

A fundamental problem for ES estimation has been that it is not elicitable, meaning that scoring functions do not exist for its estimation and evaluation. This has recently been addressed by Fissler and Ziegel (2016), who show that VaR and ES are jointly elicitable,

and present the following set of proper joint scoring functions for these two risk measures:

$$h_{FZ}(y_t, q_t, es_t; \theta, G_1, G_2, a) = (\mathbb{1}\{y_t \leq q_t\} - \theta)\left(G_1(q_t) - G_1(y_t) + \frac{1}{\theta} G_2(es_t) q_t\right) \quad (4.1)$$
$$- G_2(es_t)\left(\frac{1}{\theta}\mathbb{1}\{y_t \leq q_t\} y_t - es_t\right) - \mathscr{G}_2(es_t) + a(y_t)$$

where y_t is the daily return; θ is the probability level; q_t is the VaR; es_t is the ES; G_1, G_2, $\mathscr{G}_2$ and a are real valued functions; G_1 is weakly increasing; G_2 is strictly increasing and positive; and G_2 is the derivative of $\mathscr{G}$: $\mathscr{G}_2^{'} = G_2$. We refer to scores of this type as FZ scores.

Taylor (2017a) studies the joint scoring function obtained by selecting $G_1(x) = 0$, $G_2(x) = -\frac{1}{x}$, $\mathscr{G}_2(x) = -\ln(-x)$ and $a(x) = 1 - \ln(1-\theta)$ in expression (4.1) to give:

$$h_{FZ}(y_t, q_t, es_t; \theta) = -\ln\left(\frac{\theta - 1}{es_t}\right) - \frac{(y_t - q_t)(\theta - \mathbb{1}\{y_t \leq q_t\})}{\theta es_t} + \frac{y_t}{es_t} \quad (4.2)$$

Taylor (2017a) notes that minimising this score is equivalent to maximising a likelihood function based on the asymmetric Laplace distribution, which links the optimisation to quantile regression. We refer to this as the AL score. In independent work, Patton et al. (2017) uses a score that differs only by a constant term from expression (4.2), and so their analysis applies equally to the AL score. Their work shows that the AL score is homogeneous of degree zero. Scores of this type have been shown to have nice properties, such as they lead to better test power (Diebold and Mariano 2002; Patton and Sheppard 2009). Patton et al. (2017) use the score to estimate dynamic models for VaR and ES, and to establish the asymptotic consistency and normality of the resulting parameter estimators.

Other FZ scores have also been considered. In their numerical analysis, Nolde and Ziegel (2017) consider the score resulting from setting $G_1 = 0$, $G_2 = \frac{1}{2}(-x)^{-\frac{1}{2}}$, $\mathscr{G}_2(x) = -(-x)^{\frac{1}{2}}$ and $a(x) = 0$ in expression (4.1). We refer to this as the *NZ score*. Fissler et al. (2016) consider the score produced by setting $G_1(x) = x$, $G_2 = \frac{\exp(x)}{1+\exp(x)}$, $\mathscr{G}_2(x) = \ln(1 + \exp(x))$ and $a(x) = 0$. In our use of this score, we found that the first three significant figures did not differ between different forecasting methods. To make the values more

easily comparable, we set $a(x) = \ln(2)$, which led to positive values for the score. We refer to the resulting score as the FZG score. We summarize the AL, NZ and FZG scores in Table 4.1.

Table 4.1: A selection of FZ scoring functions for the joint estimation and evaluation of VaR and ES.

	$G_1(x)$	$G_2(x)$	$\mathscr{G}_2(x)$	$a(x)$
AL	0	$-\frac{1}{x}$	$-\ln(-x)$	$1-\ln(1-\theta)$
NZ	0	$\frac{1}{2}(-x)^{-\frac{1}{2}}$	$-(-x)^{\frac{1}{2}}$	0
FZG	x	$\frac{\exp(x)}{1+\exp(x)}$	$\ln(1+\exp(x))$	$\ln(2)$

Taylor (2017a) finds that estimation based on minimising the AL score is particularly competitive when used for models composed of a CAViaR expression for the VaR, and ES expressed as a constant multiple of the VaR. We refer to such models, estimated by minimising an FZ score, as CAViaR-FZ models, and present two in the following expressions:

CAViaR-FZ-SAV:

$$\begin{aligned} q_t(\boldsymbol{\beta}) &= \beta_1 + \beta_2 q_{t-1}(\boldsymbol{\beta}) + \beta_3 |y_{t-1}| \\ es_t(\boldsymbol{\beta}) &= \beta_4 q_t(\boldsymbol{\beta}) \end{aligned} \tag{4.3}$$

CAViaR-FZ-AS:

$$\begin{aligned} q_t(\boldsymbol{\beta}) &= \beta_1 + \beta_2 q_{t-1}(\boldsymbol{\beta}) + \beta_3 (y_{t-1})^+ + \beta_4 (y_{t-1})^- \\ es_t(\boldsymbol{\beta}) &= \beta_5 q_t(\boldsymbol{\beta}) \end{aligned} \tag{4.4}$$

where, in each model, $\boldsymbol{\beta}$ is a vector of parameters.

4.2.3 Incorporating Intra-day Information in VaR and ES modelling

Intra-day data provides information regarding the distribution of the price process within a day, and this can be useful in modelling the daily volatility (see, for example, Corsi 2009; Shephard and Sheppard 2010; Hansen et al. 2012). However, intra-day data, recorded

at a relatively high frequency, such as minute-by-minute, is generally expensive and not available for a long period. In contrast, the daily opening, closing, intra-day high and intra-day low prices are readily available for most tradable assets for the past 30 years. Although an abundance of literature can be found on volatility estimation based on the intra-day range (see, for example, Alizadeh et al. 2002; Brandt and Jones 2006; Corsi 2009), very few studies have used the intra-day range in the daily VaR and ES context. Chen et al. (2012) find that including range information can be beneficial for some daily CAViaR models. Gerlach and Wang (2016) show that the intra-day range is competitive in terms of VaR and ES forecasting when compared to other volatility measures based on high-frequency data for the realised GARCH models proposed by Hansen et al. (2012).

4.2.4 A New Model Using the Intra-day Range and an FZ Score

The CAViaR models of Chen et al. (2012) provide a pragmatic synthesis of autoregressive quantile modelling and the intra-day range. However, as they are simply quantile models, they do not provide ES predictions. To rectify this, we propose the model of expression (4.5), which incorporates the intra-day range, and is estimated by minimising an FZ score. Quite apart from the appeal of capturing intra-day information, the intra-day range has the advantage of being a more efficient and less noisy estimator of the daily volatility than the magnitude of the return, which is used in the more standard CAViaR formulations of expressions (4.3) and (4.4).

CAViaR with intra-day range estimated using an FZ score (CAViaR-FZ-Range):

$$\begin{aligned} q_t(\boldsymbol{\beta}) &= \beta_1 + \beta_2 q_{t-1}(\boldsymbol{\beta}) + \beta_3 Range_{t-1} \\ es_t(\boldsymbol{\beta}) &= \beta_4 q_t(\boldsymbol{\beta}) \end{aligned} \tag{4.5}$$

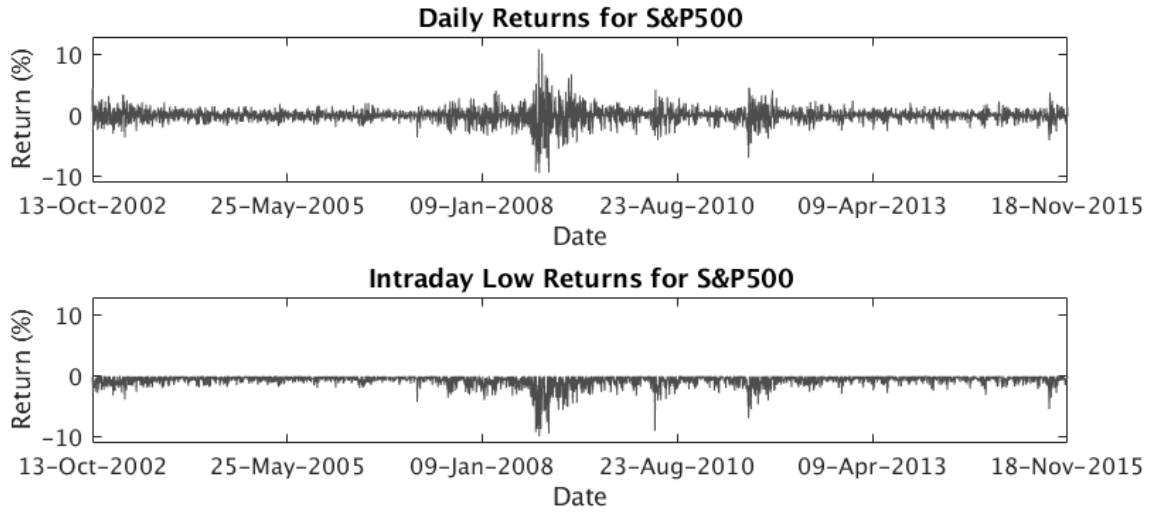
4.3 Model Estimation based on the Intra-day Low

The accuracy of VaR and ES forecasting is influenced by the number of observations in the tail of the returns distribution. When the data is scarce in the tail, estimation can be

challenging. The essence of the new approach that we propose is that extreme VaR and ES of the daily returns can be approximated by less extreme VaR and ES of the intra-day low returns, which, as we stated in Section 4.1, we define as the difference between the lowest log price of the day and the log closing price of the previous day. (We define the daily return, in the usual way, as the difference between the log closing price on successive days.) Figure 4.1 plots a sample of daily returns and intra-day low returns for the S&P 500 stock index.

In Section 4.3.1, we describe how, for standard Brownian motion, there is an equivalence of a quantile of the daily return to a less extreme quantile of the intra-day low return. Building on this, we propose that the intra-day low return can be used as the basis for estimating VaR and ES for the daily return. We theoretically establish the estimation efficiency of this proposal. In Section 4.3.2, we describe an approach for implementing our proposal in practice.

Figure 4.1: Daily returns and intra-day low returns for the S&P 500.



4.3.1 Brownian Motion and the Intra-day Low

Brownian motion has been widely used to model the log price in the context of stochastic volatility and option price modelling (Kou 2002; Andersen and Piterbarg 2007). It is well-established that the distribution of the infimum Low_t of standard Brownian motion

y_s, starting at 0 over a closed interval $[0,t]$, and the distribution of y_t are related via the following expression (see, for example, Mörters and Peres 2010):

$$\Pr(Low_t < x) = 2\Pr(y_t < x) \qquad if\ x < 0 \tag{4.6}$$

For Brownian motion with nonzero starting point, we can simply subtract the starting value from the Brownian motion. Suppose the intra-day log price follows a Brownian motion starting from the log closing price of the previous day. In expression (4.6), y_t can be considered the daily return, and Low_t can be considered the intra-day low return. Expression (4.6) implies that the VaR and ES of the daily return, corresponding to probability level θ, are equal to the VaR and ES of the intra-day low return, corresponding to probability level 2θ.

In the VaR and ES context, this is a potentially interesting result, because it implies that, instead of estimating the VaR and ES of the daily return, we can estimate a less extreme VaR and ES of the intra-day low return, which is appealing in terms of estimation accuracy. Indeed, there will be twice as many observations beyond the VaR if we use the intra-day low return than if we use the daily return. Moreover, suppose we use some semiparametric model based on an FZ score to estimate the VaR and ES of the daily return for probability level θ, and use the same approach to estimate the VaR and ES of the intra-day low return for probability level 2θ, it is more efficient to estimate the model parameters for the intra-day low return, as will be shown in Theorem 1 below. In fact, we can loosen the assumption (in expression (4.6)) that the ratio $\frac{\Pr(Low_t<x)}{\Pr(y_t<x)}$ is equal to 2. As we show in Theorem 1, for any stochastic process y_t, which is not necessarily Brownian motion, provided the ratio $\frac{\Pr(Low_t<x)}{\Pr(y_t<x)}$ is equal to a constant (greater than 1) within the estimation window, improved efficiency will be achieved by estimating the quantile for the intra-day low return. In the rest of the study, we use the notation $\tilde{}$ to distinguish all the quantities related to the intra-day low return from those of the daily return. For example, we use $\tilde{q}_t$ and $\tilde{e}_t$ to denote the VaR and ES of the intra-day low return.

Assumption 4.1. *Let y_t and Low_t be two time series. Assume the following:*

1. There exist a scalar x_0 and a constant $\lambda > 1$, such that

$$\Pr(Low_t < x) = \lambda \Pr(y_t < x) \qquad \text{for all } x < x_0. \tag{4.7}$$

2. The dynamics of VaR and ES of y_t can be captured by the same parametric form as the dynamics of VaR and ES of Low_t, and the VaR of y_t or Low_t of interest is less than the threshold x_0 in 1.

3. y_t, Low_t and their corresponding VaR and ES satisfy the regularity conditions in Assumptions 1 and 2 in Patton et al. (2017).

Theorem 4.2. *Under Assumption 4.1, estimating the VaR and ES of the intra-day low return Low_t for probability level $\lambda\theta$ is more efficient than estimating the VaR and ES of the daily return y_t for probability level θ.*

A proof of this theorem is presented in the appendix. The theorem provides support for using the intra-day low return to estimate VaR and ES of the daily return. However, we should emphasis that the theorem requires Assumption 4.1. The second point in Assumption 4.1 means that the dynamics of VaR and ES of y_t contain the same explanatory variables and model structure as the dynamics of VaR and ES of Low_t, and that they differ only in their parameter values. An important aspect of Theorem 4.2 is the practical issue of obtaining the value of λ, or, more specifically, the value of the probability level $\lambda\theta$ to use with the intra-day low returns. This is the focus of Section 4.3.2.

4.3.2 Estimating a Model for VaR and ES using the Intra-day Low Returns

Our ultimate goal is to use the VaR and ES of the intra-day low returns to improve the forecasting accuracy of the VaR and ES of the daily returns. For this, we propose the following two alternative approaches:

1. As forecasts of the VaR and ES of the daily returns with probability level θ, use forecasts of the VaR and ES of the intra-day low returns with probability level 2θ. To enable this, model the VaR and ES of the intra-day low returns using one of the CAViaR-FZ models presented in expressions (4.3)-(4.5).

2. When modelling the VaR and ES of the intra-day low returns, instead of using a probability level of 2θ, estimate the appropriate probability level as $\tilde{\theta} = \frac{1}{T}\sum_{t=1}^{T} \mathbb{1}(Low_t < Q_\theta)$, where Q_θ is the unconditional θ quantile of the daily returns, and T is the size of the estimation sample.

4.4 Empirical Study of VaR and ES Forecasting

4.4.1 Data

To evaluate VaR and ES forecast accuracy, we used the daily opening, daily closing, intra-day low and intra-day high price series of the following five stocks: CAC 40, DAX 30, FTSE 100, Nikkei 225 and S&P 500. For each index, our sample period consisted of the series of 3300 observations ending on 19 November 2015. We subtracted the log closing price of the previous day from the log closing price, and from the lowest log price of the current day, to obtain the daily return y_t and the intra-day low return Low_t, respectively. The intra-day range $Range_t$ is the absolute difference between the highest and lowest log price. We considered the 0.5%, 1% and 5% probability levels for the VaR and ES of the daily returns. We used a rolling window of 1800 observations for repeated re-estimation of the parameters of each method. This enabled us to produce post-sample forecasts from each method for the final 1500 days in each series.

4.4.2 VaR and ES Forecasting Methods

In this section, we describe the methods that we included in the empirical study. We implemented historical simulation with window lengths ranging from 25 to 300 days, but the methods performed poorly, and so we omit the results from this study. We also

implemented the standard GARCH(1,1) and GJR-GARCH(1,1) models, as well as the log-linear realised GARCH model based on the intra-day range (see Hansen et al. 2012). For each GARCH model, we estimated the parameters using a Student t distribution. To estimate VaR and ES, we used the following three different approaches: the Student t distribution; filtered historical simulation applied to the 1800 in-sample observations; and the EVT approach of McNeil and Frey (2000), with threshold chosen as the 10% quantile.

We implemented Taylor's (2017a) CAViaR-FZ-SAV and CAViaR-FZ-AS models of expressions (4.3) and (4.4), and our CAViaR-FZ-Range model of expression (4.5). These models are based on daily returns. All of the CAViaR-FZ models in this study were estimated using the AL score in expression (4.2), as the asymptotic theories have been established in Patton et al. (2017). We use a numerical procedure similar to that described by Engle and Manganelli (2004) for CAViaR models. For each model, 10^d initial trial vectors of parameters were randomly generated, where d is the number of parameters. The first $d-1$ entries of each vector were sampled from a uniform distribution between 0 and 1, and correspond to the VaR models. The final entry in each vector corresponds to the ES; it was sampled from a uniform distribution between 1 and 10. For each initial trial vector, we calculated the particular AL score. The six vectors producing the lowest values of the AL score were passed to the Nelder-Mead algorithm as starting vectors. The resulting vector that produced the lowest score was selected as the optimal parameter vector.

We implemented the two new approaches of Section 4.3.2. Each involves fitting one of the CAViaR-FZ models of expressions (4.3)-(4.5) to the intra-day low returns. In the first approach, we used the model corresponding to probability level 2θ to estimate the VaR and ES of the daily returns for probability level θ. Table 4.2 presents the estimated parameters for three CAViaR-FZ-Range models derived using the first estimation window of 1800 days of the S&P 500 series of intra-day low returns. The three models correspond to probability levels of 2θ, where $\theta = 0.5\%$, 1% and 5%.

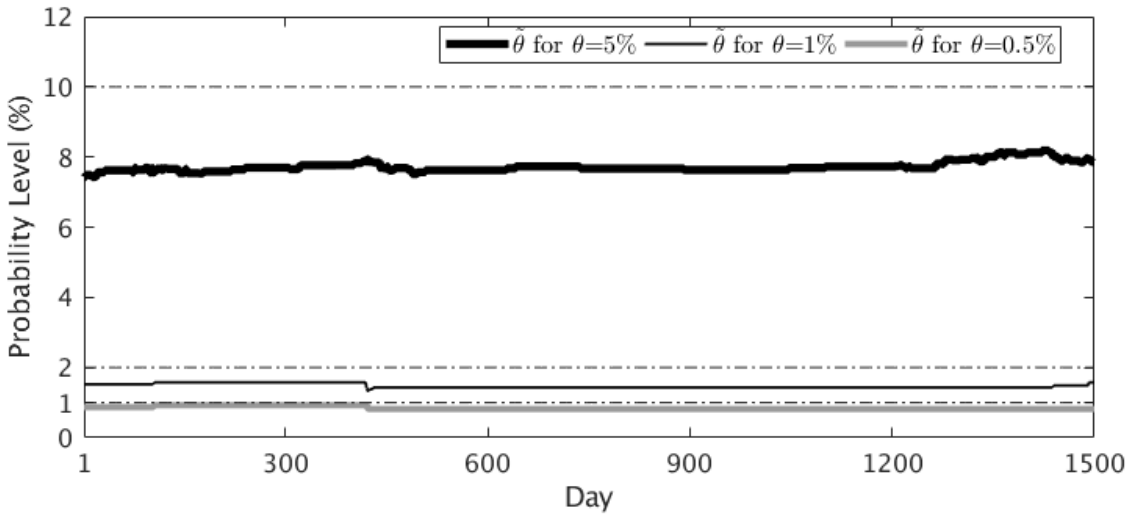
The second approach in Section 4.3.2 estimates the probability level $\tilde{\theta}$ to use with the

intra-day low returns. Figure 4.2 plots the estimated values of $\tilde{\theta}$ that we obtained for the 1500 moving estimation windows for the S&P 500 returns. In the figure, the $\tilde{\theta}$ values deviate quite clearly from 2θ, which would be appropriate for Brownian motion. Indeed, it would seem that to use a probability level of 2θ with the S&P 500 intra-day low returns would deliver VaR and ES forecasts that are not sufficiently extreme.

Table 4.2: Parameter estimates and their standard errors for the CAViaR-FZ-Range model with probability level 2θ, fitted to the intra-day low returns Low_t, derived using the first moving window of 1800 days of the S&P 500.

$100 \times 2\theta$	1	2	10
β_1	0.068	0.136	0.060
	(0.034)	(0.048)	(0.137)
β_2	0.699	0.657	0.735
	(0.025)	(0.035)	(0.091)
β_3	0.601	0.557	0.279
	(0.047)	(0.066)	(0.165)
β_4	1.202	1.209	1.367
	(0.151)	(0.105)	(0.055)

Figure 4.2: Appropriate probability levels for the S&P 500 intra-day low returns estimated as $\tilde{\theta} = \frac{1}{T}\sum_{t=1}^{T} \mathbb{1}(Low_t < Q_\theta)$. Dashed horizontal lines correspond to $\tilde{\theta}$=2θ, which is appropriate for Brownian motion.



4.4.3 Post-Sample Results

We first evaluate the post-sample VaR and ES forecasts using coverage tests. The unconditional VaR coverage test focuses on the hit percentage, which is the proportion of the observations that fall below the estimated quantiles. For the quantile with probability level θ, the ideal value of the proportion is θ, and we used this as the null hypothesis in a standard test for a sample proportion based on the binomial distribution. To evaluate VaR conditional coverage, we used the dynamic quantile test of Engle and Manganelli (2004) with four lags in the test's regression. To evaluate the unconditional coverage of the ES forecasts, we used the bootstrap test of McNeil and Frey (2000), which focuses on the discrepancies between the VaR exceedances and the ES forecasts. The test examines whether the standardised discrepancies have zero unconditional expectation. As volatility estimates are not available for semiparametric methods, we adopt the approach of Acerbi and Szekely (2014), which standardises the discrepancies by dividing each by the corresponding ES estimate. In view of the relatively small number of VaR exceedancies, conditional coverage is typically not considered for ES forecasts.

Table 4.3 summarizes the coverage test results for the five stock indices. For each of the three CAViaR-FZ models, the results of three methods are presented. The first method fits the model to the daily returns y_t using probability level θ, while the other two fit the model to the intra-day low returns Low_t using the two different approaches described in Section 4.3.2. Table 4.3 reports the numbers of test rejections at the 5% significance level for each method across the five indices. Smaller numbers of rejections are preferred. All the methods perform well in terms of unconditional coverage. For the dynamic quantile test, it is interesting to see that the methods tend to perform better for the 5% probability level than the other probability levels, which reflects the tendency for the more extreme quantiles to be more difficult to estimate. The best results correspond to GJR-GARCH, and the CAViaR-FZ-AS models, with one of the CAViaR-FZ-Range models also performing well. For the ES bootstrap test, little insight can be observed in terms of model performance. This is partly due to the fact that there are only few observations beyond the VaR. This

motivates us to evaluate the model performance using FZ scores.

In terms of scoring functions, we used the quantile score and the three FZ scores described in Section 4.2.2 to evaluate the model performance. For each series and method, we averaged each score over the post-sample period. We then calculated skill scores as the ratio of each method's score to the score of the GARCH model with Student-t distribution, then subtracted this ratio from 1, and multiplied the result by 100. For all skill scores, higher values are preferable. To summarize performance across the five stock indices, for each method, we calculated the geometric mean of the ratios of the method's score to the score of the GARCH-t method, then subtracted this from one, and multiplied the result by 100. We present the results in Tables 4.4 and 4.5, where we have indicated in bold the best two performing methods for each probability level in each table. Recall that the AL, NZ and FZG scores are all types of FZ score, which enable the joint evaluation of the VaR and ES. In Table 4.4, for the 0.5% probability level, the best quantile and AL skill score results correspond to the CAViaR-FZ-Range model fitted to the intra-day low returns using either of the approaches described in Section 4.3.2. This is also the case when using the AL skill score to evaluate the results for the 1% probability level. For the 5% probability level, the best quantile and AL skill score results again correspond to the models fitted to the intra-day low returns, but for this less extreme probability level, the CAViaR-FZ-AS model is the best. The ranking of the models, in terms of the NZ and FG skill scores in Table 4.5, are similar to those for the AL score in Table4.4. Overall, the proposed CAViaR-FZ-Range models and the CAViaR-FZ-AS models based in the intra-day low returns perform well. These results are supportive of the new proposals put forward in this study, namely the use of the CAViaR-FZ-Range model of expression (4.5), and the use of the intra-day low returns to estimate CAViaR-FZ models.

Table 4.3: Summary of the coverage test results for the five stock indices.

	VaR Hit			VaR DQ			ES Bootstrap		
$\theta \times 100$	0.5	1	5	0.5	1	5	0.5	1	5
GARCH									
Student-t	0	1	0	3	4	3	0	0	0
Filtered historical simulation	0	0	0	3	4	3	0	0	0
EVT	0	0	0	3	2	3	0	0	0
GJR-GARCH									
Student-t	0	0	2	2	1	1	1	0	0
Filtered historical simulation	0	0	0	2	1	0	1	1	0
EVT	0	0	0	1	0	0	1	0	0
Realised GARCH									
Student-t	0	1	0	3	3	3	1	0	2
Filtered historical simulation	0	0	0	3	2	0	1	0	0
EVT	0	0	0	3	3	0	0	0	0
CAViaR-FZ-SAV									
θ VaR and ES of y_t	0	0	0	3	2	2	0	0	1
2θ VaR and ES of Low_t	0	1	1	3	3	2	0	0	0
$\tilde{\theta}$ VaR and ES of Low_t	0	0	0	3	2	2	0	1	0
CAViaR-FZ-AS									
θ VaR and ES of y_t	0	0	0	2	0	0	0	0	0
2θ VaR and ES of Low_t	0	1	0	2	1	2	0	0	0
$\tilde{\theta}$ VaR and ES of Low_t	0	0	0	2	0	0	0	0	1
CAViaR-FZ-Range									
θ VaR and ES of y_t	0	0	0	3	2	1	1	0	0
2θ VaR and ES of Low_t	0	0	1	3	3	1	1	1	0
$\tilde{\theta}$ VaR and ES of Low_t	0	0	0	3	0	0	1	1	0

Note: Smaller rejections are preferred. y_t is daily return and Low_t is intra-day low return.

Table 4.4: Summary of the quantile and AL skill score results for the five stock indices.

	Quantile Score			AL Score		
$\theta \times 100$	0.5	1	5	0.5	1	5
GARCH						
Student-t	0.0	0.0	0.0	0.0	0.0	0.0
Filtered historical simulation	-0.4	0.1	-0.1	-0.2	0.2	0.2
EVT	-0.7	0.2	0.0	0.2	0.4	0.2
GJR-GARCH						
Student-t	4.4	**4.6**	2.5	1.0	1.7	1.6
Filtered historical simulation	3.8	4.2	2.5	1.0	1.8	1.9
EVT	3.2	3.9	2.6	1.2	1.9	1.9
Realised GARCH						
Student-t	3.6	3.8	1.9	1.7	2.0	1.4
Filtered historical simulation	2.5	4.0	2.0	1.2	2.3	1.7
EVT	3.1	4.1	2.1	1.8	2.6	1.8
CAViaR-FZ-SAV						
θ VaR and ES of y_t	2.2	0.6	-0.1	1.0	0.6	0.3
2θ VaR and ES of Low_t	0.4	0.9	0.5	0.5	0.7	0.5
$\tilde{\theta}$ VaR and ES of Low_t	1.0	0.8	0.2	0.6	0.8	0.5
CAViaR-FZ-AS						
θ VaR and ES of y_t	5.1	**4.8**	2.6	2.1	2.3	1.8
2θ VaR and ES of Low_t	4.1	**4.6**	**3.0**	1.5	2.0	**2.0**
$\tilde{\theta}$ VaR and ES of Low_t	3.7	4.1	**2.9**	1.1	1.9	**2.0**
CAViaR-FZ-Range						
θ VaR and ES of y_t	4.7	3.8	2.1	3.1	**2.7**	1.7
2θ VaR and ES of Low_t	**5.9**	4.2	1.9	**3.3**	**2.7**	1.5
$\tilde{\theta}$ VaR and ES of Low_t	**6.8**	4.5	2.1	**3.6**	**3.0**	1.9

Note: Large skill scores are better. The best two methods in each column are indicated in bold. y_t is daily return and Low_t is intra-day low return.

Table 4.5: Summary of the NZ and FZG skill score results for the five stock indices.

	NZ Score			FZG Score		
$\theta \times 100$	0.5	1	5	0.5	1	5
GARCH						
Student-t	0.0	0.0	0.0	0.0	0.0	0.0
Filtered historical simulation	-0.2	0.1	0.1	-0.1	0.2	0.2
EVT	-0.1	0.3	0.1	0.3	0.4	0.2
GJR-GARCH						
Student-t	1.8	2.2	1.5	-0.6	0.3	1.4
Filtered historical simulation	1.6	2.1	1.6	-0.2	0.7	1.6
EVT	1.5	2.1	1.6	0.1	0.9	1.6
Realised GARCH						
Student-t	1.9	2.0	1.2	0.9	1.3	1.3
Filtered historical simulation	1.3	2.3	1.3	0.6	1.5	1.6
EVT	1.8	2.4	1.4	1.1	1.8	1.6
CAViaR-FZ-SAV						
θ VaR and ES of y_t	1.2	0.5	0.1	0.5	0.5	0.3
2θ VaR and ES of Low_t	0.4	0.6	0.3	0.3	0.5	0.4
$\tilde{\theta}$ VaR and ES of Low_t	0.6	0.7	0.3	0.2	0.6	0.5
CAViaR-FZ-AS						
θ VaR and ES of y_t	2.6	2.5	1.6	0.6	1.1	1.6
2θ VaR and ES of Low_t	2.0	2.3	**1.8**	0.2	0.8	**1.8**
$\tilde{\theta}$ VaR and ES of Low_t	1.7	2.1	**1.7**	-0.1	0.8	**1.8**
CAViaR-FZ-Range						
θ VaR and ES of y_t	3.0	2.4	1.4	**1.8**	**1.9**	1.6
2θ VaR and ES of Low_t	**3.4**	**2.6**	1.2	1.6	**1.9**	1.4
$\tilde{\theta}$ VaR and ES of Low_t	**3.8**	**2.8**	1.4	**1.7**	**2.1**	1.7

Note: Large skill scores are better. The best two methods in each column are indicated in bold. y_t is daily return and Low_t is intra-day low return.

Table 4.6: Summary of the model confidence set results for the quantile and AL scores.

Confidence Level	75%						90%					
	Quantile Score			AL Score			Quantile Score			AL Score		
$\theta \times 100$	0.5	1	5	0.5	1	5	0.5	1	5	0.5	1	5
GARCH												
Student-t	3	2	4	4	3	4	4	5	5	5	5	4
Filtered historical simulation	3	1	2	3	1	1	5	4	4	5	4	4
EVT	2	1	3	3	3	2	3	4	4	4	3	4
GJR-GARCH												
Student-t	4	5	5	4	5	5	4	5	5	5	5	5
Filtered historical simulation	4	4	5	4	4	5	4	5	5	5	5	5
EVT	4	4	5	4	4	5	4	4	5	4	5	5
Realised GARCH												
Student-t	5	5	5	5	5	5	5	5	5	5	5	5
Filtered historical simulation	5	5	5	5	5	5	5	5	5	5	5	5
EVT	5	5	5	5	5	5	5	5	5	5	5	5
CAViaR-FZ-SAV												
θ VaR and ES of y_t	5	3	1	5	5	3	5	5	4	5	5	4
2θ VaR and ES of Low_t	3	3	5	3	3	4	4	4	5	5	5	5
$\tilde{\theta}$ VaR and ES of Low_t	3	4	2	3	3	4	4	4	4	5	5	5
CAViaR-FZ-AS												
θ VaR and ES of y_t	4	5	5	5	5	5	5	5	5	5	5	5
2θ VaR and ES of Low_t	3	5	5	4	5	5	5	5	5	5	5	5
$\tilde{\theta}$ VaR and ES of Low_t	4	5	5	4	5	5	5	5	5	5	5	5
CAViaR-FZ-Range												
θ VaR and ES of y_t	5	5	5	5	5	5	5	5	5	5	5	5
2θ VaR and ES of Low_t	5	5	5	5	5	5	5	5	5	5	5	5
$\tilde{\theta}$ VaR and ES of Low_t	5	5	5	5	5	5	5	5	5	5	5	5

Note: Values shown are the number of indices for which the method remained in the model confidence set. Large numbers are better. y_t is daily return and Low_t is intra-day low return.

Finally, we tested the statistical significance of the differences between the scores of competing methods. We used the model confidence set evaluation approach to select the best-performing models (Hansen et al. 2011). The approach selects a set of best-performing models that will contain the best model with a given level of confidence based on some chosen statistical test and loss function. We used each of the four scores as the loss function, and selected the best performing methods based on the

Diebold-Mariano test. Following Hansen et al. (2011), we used 75% and 90% as the confidence levels for the model confidence sets. For simplicity, in reporting the results, we focus on just the quantile and AL scores. Table 4.6 reports the number of series for which each model lies inside the model confidence set, where the confidence levels are chosen to be the 75% and 90%. Large counts in this table are preferable. The value for the best possible performance is 5, indicating that the method was not eliminated from the model confidence set for any of the five indices. The results are not particularly informative, as several methods perform equally well. Nevertheless, it is encouraging to see competitive results for the new CAViaR-FZ-Range model of expression (4.5), and the CAViaR-FZ models estimated using the intra-day low returns.

4.5 Conclusion

In this study, we have introduced a new approach to estimating conditional VaR and ES. In this approach, we not only consider the use of intra-day information in model specification, but also use it in parameter estimation. Based on theory for Brownian motion, we argue that the dynamics of VaR and ES of the intra-day low return contains useful information for forecasting the dynamics of VaR and ES of the daily returns. For a chosen probability level, we estimate the VaR and ES of the daily returns using the VaR and ES of the intra-day low returns corresponding to a less extreme probability level. Our empirical results suggest that the approach is particularly useful for more extreme probability levels. A further contribution of the study is that we support the previous finding that the intra-day range is a useful explanatory variable for VaR and ES modelling, and that minimising a joint VaR and ES scoring function is a useful way to estimate models with CAViaR structure for VaR and ES prediction.

We have studied financial data in Chapters 3 and 4, in which we focus on the use of the intra-day data, especially the intra-day range and the intra-day low/high log price, to improve the forecasting accuracy of VaR and ES estimates. In the next chapter (Chapter 5), we will study daily extreme temperature of intra-day observations: the daily

minimum and maximum temperature. Even though temperature data has very different features when compared to financial data, our focus of studying the daily minimum and maximum temperature has commonalities with Chapters 3 and 4. The daily minimum and maximum temperature are analogous to the intra-day low/high log prices for financial data, and the absolute difference between the daily minimum and maximum temperature, which is called *diurnal range* (DTR) in the climate literature, is analogous to the intra-day range for financial data. We have reviewed and implemented parametric methods for density forecasting in the context of VaR and ES forecasting. In Chapter 5, we will use the parametric ARMA-GARCH models and their multivariate extensions to obtain the marginal and joint distributions of the daily minimum and maximum temperature.

Appendix

In this appendix, we provide a proof for Theorem 4.2 under Assumption 4.1. Before we prove Theorem 4.2, we state the results in Theorem 2 of Patton et al. (2017): Under suitable regularity conditions, which are given in Assumption 2 of Patton et al. (2017), the parameter vector estimator $\hat{\boldsymbol{\beta}}$ for y_t satisfies the following asymptotic normality:

$$\sqrt{T}\boldsymbol{A}_T^{-\frac{1}{2}}\boldsymbol{D}_T(\hat{\boldsymbol{\beta}}_T-\boldsymbol{\beta}_0)\xrightarrow{d}N(0,I)\ as\ T\to\infty$$

where

$$\begin{aligned}
\boldsymbol{D}_T &= \mathbb{E}\left[\frac{1}{T}\sum_{t=1}^{T}\frac{f_t(q_t(\boldsymbol{\beta}_0)|\mathscr{F}_{t-1})}{-es_t(\boldsymbol{\beta}_0)\theta}\nabla q_t(\boldsymbol{\beta}_0)^T\nabla q_t(\boldsymbol{\beta}_0)+\frac{1}{es_t(\boldsymbol{\beta}_0)^2}\nabla es_t(\boldsymbol{\beta}_0)^T\nabla es_t(\boldsymbol{\beta}_0)\right]\\
\boldsymbol{A}_T &= \mathbb{E}\left[\frac{1}{T}\sum_{t=1}^{T}g_t(\boldsymbol{\beta}_0)g_t(\boldsymbol{\beta}_0)^T\right]\\
g_t(\boldsymbol{\beta}_0) &= \nabla q_t(\boldsymbol{\beta}_0)^T\frac{1}{-es_t(\boldsymbol{\beta}_0)}\left(\frac{1}{\theta}\mathbb{1}\{y_t\le q_t(\boldsymbol{\beta}_0)\}-1\right)\\
&\quad+\nabla es_t(\boldsymbol{\beta}_0)^T\frac{1}{es_t(\boldsymbol{\beta}_0)^2}\left(\frac{1}{\theta}\mathbb{1}\{y_t\le q_t(\boldsymbol{\beta}_0)\}(q_t(\boldsymbol{\beta}_0)-y_t)-q_t(\boldsymbol{\beta}_0)+es_t(\boldsymbol{\beta}_0)\right)
\end{aligned}$$

$f_t(q_t(\boldsymbol{\beta}_0)|\mathscr{F}_{t-1})$ denotes the density of y_t conditional on past information $\mathscr{F}_{t-1}$ evaluated at the θ conditional quantile $q_t(\boldsymbol{\beta}_0)$; $\boldsymbol{\beta}_0$ is the true parameter; and $\hat{\boldsymbol{\beta}}_T$ is the parameter estimator using the in-sample observations of length T.

Efficiency is shown by proving positive semi-definiteness of the difference between two asymptotic parameter covariance matrices (see, for example, Komunjer and Vuong 2010). To prove our Theorem 4.2 in Section 4.3.1, we need to prove the positive semi-definiteness of $(\boldsymbol{D}_T)^{-1}\boldsymbol{A}_T(\boldsymbol{D}_T)^{-1} - (\tilde{\boldsymbol{D}}_T)^{-1}\tilde{\boldsymbol{A}}_T(\tilde{\boldsymbol{D}}_T)^{-1}$, which is the difference between the asymptotic parameter covariance matrices of the θ VaR and ES for the daily returns and the $\lambda\theta$ VaR and ES for the intra-day low returns. Now we begin our proof:

Proof. Under Assumption 4.1, we have that the $\lambda\theta$ VaR and ES of Low_t are equal to the θ VaR and ES of y_t, that is, $\tilde{q}_t(\tilde{\boldsymbol{\beta}}^0) = q_t(\boldsymbol{\beta}^0)$ and $\widetilde{es}_t(\tilde{\boldsymbol{\beta}}^0) = es_t(\boldsymbol{\beta}^0)$. Therefore, $\tilde{\boldsymbol{\beta}}^0 = \boldsymbol{\beta}^0$ by Assumption 1 (B) in Patton et al. (2017). As a consequence , $\nabla q_t(\tilde{\boldsymbol{\beta}}^0) = \nabla q_t(\boldsymbol{\beta}^0)$. For convenience, we use q_t to denote $\tilde{q}_t(\tilde{\boldsymbol{\beta}}^0)$ and $q_t(\boldsymbol{\beta}^0)$, and we use es_t to denote $\widetilde{es}_t(\tilde{\boldsymbol{\beta}}^0)$ and $es_t(\boldsymbol{\beta}^0)$, in the rest of the proof. First observe that:

$$\begin{aligned}
\tilde{\boldsymbol{D}}_T &= \mathbb{E}\left[\frac{1}{T}\sum_{t=1}^{T}\frac{\tilde{f}_t(q_t|\mathscr{F}_{t-1})}{-es_t\lambda\theta}\nabla q_t^T\nabla q_t + \frac{1}{es_t^2}\nabla es_t^T\nabla es_t\right] \\
&= \mathbb{E}\left[\frac{1}{T}\sum_{t=1}^{T}\frac{f_t(q_t|\mathscr{F}_{t-1})}{-es_t\theta}\nabla q_t^T\nabla q_t + \frac{1}{es_t^2}\nabla es_t^T\nabla es_t\right] \\
&= \boldsymbol{D}_T
\end{aligned}$$

where the second equality uses the fact that $\tilde{f}_t(q_t|\mathscr{F}_{t-1}) = \lambda f_t(q_t|\mathscr{F}_{t-1})$, which is a direct consequence of Assumption 4.1 (1). Next we look at $\mathbb{E}\left[\tilde{g}_t(\tilde{g}_t)^T\right]$:

$$\begin{aligned}
\mathbb{E}\left[\tilde{g}_t(\tilde{g}_t)^T\right] = \mathbb{E}\Bigg[&\left(\nabla q_t^T\frac{1}{-es_t}\left(\frac{1}{\lambda\theta}\mathbb{1}\{Low_t \le q_t\} - 1\right)\right. \\
&\left.+\nabla es_t^T\frac{1}{es_t^2}\left(\frac{1}{\lambda\theta}\mathbb{1}\{Low_t \le q_t\}(q_t - Low_t) - q_t + es_t\right)\right) \\
&\left(\nabla q_t^T\frac{1}{-es_t}\left(\frac{1}{\lambda\theta}\mathbb{1}\{Low_t \le q_t\} - 1\right)\right. \\
&\left.+\nabla es_t^T\frac{1}{es_t^2}\left(\frac{1}{\lambda\theta}\mathbb{1}\{Low_t \le q_t\}(q_t - Low_t) - q_t + es_t\right)\right)^T\Bigg]
\end{aligned}$$

$$
\begin{aligned}
&= \mathbb{E}\Bigg[\frac{1}{es_t^2}\Bigg(\frac{\mathbb{1}\{Low_t \le q_t\}}{\lambda\theta}\left(-\nabla q_t + \frac{\nabla es_t}{es_t}q_t - \frac{\nabla es_t}{es_t}Low_t\right) \\
&\quad + \left(\frac{\nabla es_t}{es_t}(es_t - q_t) + \nabla q_t\right)\Bigg)^T \\
&\quad \left(\frac{\mathbb{1}\{Low_t \le q_t\}}{\lambda\theta}\left(-\nabla q_t + \frac{\nabla es_t}{es_t}q_t - \frac{\nabla es_t}{es_t}Low_t\right)\right. \\
&\quad \left.+ \left(\frac{\nabla es_t}{es_t}(es_t - q_t) + \nabla q_t\right)\right)\Bigg] \\
&= \mathbb{E}\Bigg[\frac{1}{es_t^2}\Bigg(\frac{\mathbb{1}\{Low_t \le q_t\}}{\lambda^2\theta^2}\left(-\nabla q_t + \frac{\nabla es_t}{es_t}q_t - \frac{\nabla es_t}{es_t}Low_t\right)^T \\
&\quad \left(-\nabla q_t + \frac{\nabla es_t}{es_t}q_t - \frac{\nabla es_t}{es_t}Low_t\right)\Bigg) \\
&\quad + \frac{1}{es_t^2}\left(\frac{\nabla es_t}{es_t}(es_t - q_t) + \nabla q_t\right)^T\left(\frac{\nabla es_t}{es_t}(es_t - q_t) + \nabla q_t\right) + \frac{\mathbb{1}\{Low_t \le q_t\}}{\lambda\theta es_t^2} \\
&\quad \left(\left(-\nabla q_t + \frac{\nabla es_t}{es_t}q_t - \frac{\nabla es_t}{es_t}Low_t\right)^T\left(\frac{\nabla es_t}{es_t}(es_t - q_t) + \nabla q_t\right)\right. \\
&\quad \left.+ \left(\frac{\nabla es_t}{es_t}(es_t - q_t) + \nabla q_t\right)^T\left(-\nabla q_t + \frac{\nabla es_t}{es_t}q_t - \frac{\nabla es_t}{es_t}Low_t\right)\right)\Bigg]
\end{aligned}
$$

One may note that

$$
\mathbb{E}\left[\frac{1}{es_t^2}\left(\frac{\nabla es_t}{es_t}(es_t - q_t) + \nabla q_t\right)^T\left(\frac{\nabla es_t}{es_t}(es_t - q_t) + \nabla q_t\right)\right]
$$

is independent of λ, and one may also notice the following relationship:

$$
\begin{aligned}
\mathbb{E}[\mathbb{1}\{Low_t \le q_t\}h(Low_t, q_t, es_t)] &= \mathbb{E}[\mathbb{E}[\mathbb{1}\{Low_t \le q_t\}h(Low_t, q_t, es_t)|\mathscr{F}_{t-1}]] &(4.8)\\
&= \mathbb{E}[\lambda\mathbb{E}[\mathbb{1}\{y_t \le q_t\}h(y_t, q_t, es_t)|\mathscr{F}_{t-1}]] &(4.9)\\
&= \lambda\mathbb{E}[\mathbb{1}\{y_t \le q_t\}h(y_t, q_t, es_t)] &(4.10)
\end{aligned}
$$

for any function h, which can be easily derived from Assumption 4.1 (1) using the fact that q_t and es_t are completely determined by $\mathscr{F}_{t-1}$. This implies that

$$
\begin{aligned}
\mathbb{E}\Bigg[\frac{\mathbb{1}\{Low_t \le q_t\}}{\lambda\theta es_t^2}\Bigg(&\left(-\nabla q_t + \frac{\nabla es_t}{es_t}q_t - \frac{\nabla es_t}{es_t}Low_t\right)^T\left(\frac{\nabla es_t}{es_t}(es_t - q_t) + \nabla q_t\right) \\
&+ \left(\frac{\nabla es_t}{es_t}(es_t - q_t) + \nabla q_t\right)^T\left(-\nabla q_t + \frac{\nabla es_t}{es_t}q_t - \frac{\nabla es_t}{es_t}Low_t\right)\Bigg)\Bigg]
\end{aligned}
$$

is independent of λ. Thus all the other terms apart from

$$\mathbb{E}\left[\frac{\mathbb{1}\{Low_t \leq q_t\}}{\lambda^2\theta^2 es_t^2}\left(-\nabla q_t + \frac{\nabla es_t}{es_t}q_t - \frac{\nabla es_t}{es_t}y_t\right)^T\left(-\nabla q_t + \frac{\nabla es_t}{es_t}q_t - \frac{\nabla es_t}{es_t}y_t\right)\right]$$

are independent of λ. In $\mathbb{E}\left[g_t g_t^T\right] - \mathbb{E}\left[\tilde{g}_t \tilde{g}_t^T\right]$, these terms are canceled out. One can further obtain the following by the relationship in expressions (4.8)-(4.10),

$$\begin{aligned}
&\mathbb{E}\left[\frac{\mathbb{1}\{Low_t \leq q_t\}}{\lambda^2\theta^2 es_t^2}\left(-\nabla q_t + \frac{\nabla es_t}{es_t}q_t - \frac{\nabla es_t}{es_t}Low_t\right)^T\left(-\nabla q_t + \frac{\nabla es_t}{es_t}q_t - \frac{\nabla es_t}{es_t}Low_t\right)\right] \\
&= \frac{1}{\lambda}\mathbb{E}\left[\frac{\mathbb{1}\{y_t \leq q_t\}}{\theta^2 es_t^2}\left(-\nabla q_t + \frac{\nabla es_t}{es_t}q_t - \frac{\nabla es_t}{es_t}y_t\right)^T\left(-\nabla q_t + \frac{\nabla es_t}{es_t}q_t - \frac{\nabla es_t}{es_t}y_t\right)\right]
\end{aligned}$$

Therefore, we have

$$\begin{aligned}
\mathbb{E}\left[g_t g_t^T\right] - \mathbb{E}\left[\tilde{g}_t \tilde{g}_t^T\right] = \frac{\lambda - 1}{\lambda}\mathbb{E}\Bigg[&\frac{\mathbb{1}\{y_t \leq q_t\}}{\theta^2 es_t^2}\left(-\nabla q_t + \frac{\nabla es_t}{es_t}q_t - \frac{\nabla es_t}{es_t}y_t\right)^T \\
&\left(-\nabla q_t + \frac{\nabla es_t}{es_t}q_t - \frac{\nabla es_t}{es_t}y_t\right)\Bigg]
\end{aligned}$$

which is positive semi-definite as $\lambda > 1$ by Assumption 4.1 (1). Consequently, $\boldsymbol{A}_T - \tilde{\boldsymbol{A}}_T$ is positive semi-definite, since

$$\begin{aligned}
\boldsymbol{A}_T - \tilde{\boldsymbol{A}}_T &= \mathbb{E}\left[\sum_{t=1}^{T} g_t g_t^T\right] - \mathbb{E}\left[\sum_{t=1}^{T} \tilde{g}_t \tilde{g}_t^T\right] \\
&= \sum_{t=1}^{T}\left(\mathbb{E}\left[g_t g_t^T\right] - \mathbb{E}\left[\tilde{g}_t \tilde{g}_t^T\right]\right)
\end{aligned}$$

We conclude the proof by observing that

$$(\boldsymbol{D}_T)^{-1}\boldsymbol{A}_T(\boldsymbol{D}_T)^{-1} - (\tilde{\boldsymbol{D}}_T)^{-1}\tilde{\boldsymbol{A}}_T(\tilde{\boldsymbol{D}}_T)^{-1} = (\boldsymbol{D}_T)^{-1}\left(\boldsymbol{A}_T - \tilde{\boldsymbol{A}}_T\right)(\boldsymbol{D}_T)^{-1}$$

is also positive semi-definite, since $\boldsymbol{D}_T = \tilde{\boldsymbol{D}}_T$ and is symmetric by its definition. □

Chapter 5

Probabilistic Forecasting of Extreme Daily Temperature using Bivariate Models of the Daily Minimum and Maximum

5.1 Introduction

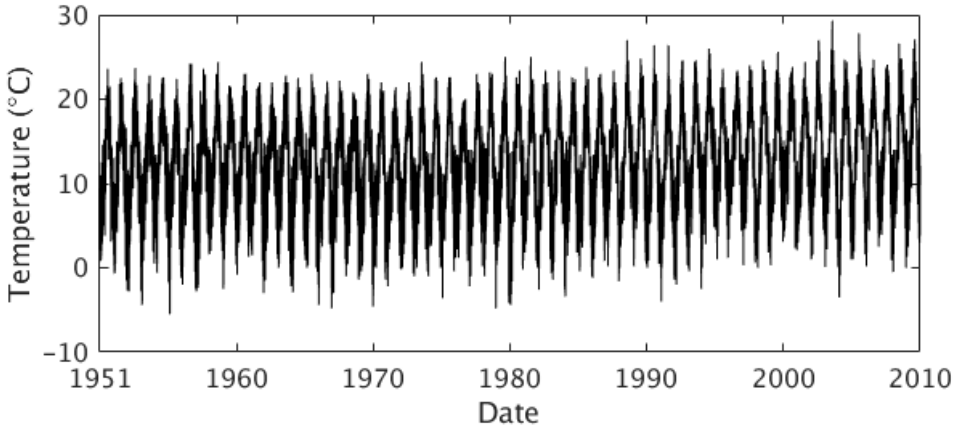
Extreme weather events can have negative impact on society and the economy. For example, heat waves and cold snaps can lead to serious health problems and increased morbidity (Wang et al. 2013; Dupuis 2012, 2014). The exposure of a variety of businesses to weather extremes has prompted the development of weather derivatives (Bertrand et al. 2015; Gülpınar and Çanakoğlu 2017). With climate change, the frequency and severity of extreme weather is increasing (Fildes and Kourentzes 2011; Meehl and Tebaldi 2004). Forecasts of extreme weather events are of great importance, as they enable warnings to be made to the public, and allow services, such as hospitals, to prepare. Although accurate point forecasting is an aim, the inevitable uncertainty in the weather means that probabilistic forecasting is often more appropriate. In this paper, we focus on the

probabilistic forecasting of heat waves. The definition of a heat wave is often based on the daily maximum temperature exceeding a specified threshold. However, heat waves are also sometimes defined as the simultaneous exceedances of the daily minimum and maximum temperature over specified thresholds (see, for example, Keellings and Waylen 2014). This has been the definition adopted by the UK Met Office in relation to their work for the UK Department of Public Health. In this paper, we aim to improve the prediction of both the marginal and joint distributions of the daily minimum and maximum temperature.

Probabilistic temperature forecasts can be obtained from a statistical time series model or a physical model, such as an NWP system. By running a NWP model with multiple different initial conditions, an ensemble of predictions can be produced, and these can be used as the basis of probabilistic prediction. However, a procedure is required to convert the ensemble into a probabilistic forecast (see, for example, Taylor and Buizza 2004). A further obstacle to their use is that their derivation is computationally intensive, which has implications for their cost, and geographical coverage. By contrast, historical observations of daily minimum and maximum temperature are readily available for many locations, which allows statistical models to be produced. Such models are the focus of this paper. These models must be able to capture the relatively complex dynamics in daily temperature data. Indeed, a careful modelling of a long time series of temperature data provides an opportunity to confirm, or perhaps improve, our understanding of how temperature extremes have been changing.

Daily temperature data has been modelled using univariate times series models, such as AR and GARCH models (see, for example, Taylor and Buizza 2004; Campbell and Diebold 2005; Dupuis 2012, 2014; Erhardt et al. 2015). For the application in this paper, a further modelling stage must be employed to produce the joint distribution, which is needed if the definition of a heat wave is based on the simultaneous exceedance of the minimum and maximum temperature above specified thresholds. The alternative is a joint model of the two temperature variables, which has the appeal of simplicity and efficiency,

Figure 5.1: Seville daily minimum temperature (TN) time series (in degrees Celsius) for the first 60-year estimation sample.



and allows a richer model structure for each variable. Indeed, even if only the marginal distribution of one variable is required, a joint model can enable improved estimation because it makes better use of the available information. Another advantage of a joint model is that it can be used to produce improved density forecasts of the mean of the daily minimum and maximum, which is often studied under the nomenclature of *average daily temperature* (see, for example, Campbell and Diebold 2005), and also the density forecast of the DTR, which has been used as an index of climate change (see, for example, Qu et al. 2014).

Although multivariate time series models, such as the VARMA and MGARCH models, have been applied successfully in various applications (see, for example, Bauwens et al. 2006; Taylor and Jeon 2018), they have not been used to model daily maximum and minimum temperature. In this paper, we contribute to the literature on the modelling and probabilistic forecasting of heat waves by comparing bivariate and univariate models of the daily minimum and maximum temperature. We use Spanish data, which is of particular interest because Southern Europe has experienced an increased number of heat waves in recent years.

Section 5.2 introduces the Spanish data, Sections 5.3 and 5.4 describe candidate time series models, Section 5.5 presents empirical results, and Section 5.6 summarises the paper.

5.2 Data

We consider historical data for the following four cities in Spain: Albacete, Seville, Cáceres and Madrid. Our dataset consists of daily minimum and maximum observations, measured in degrees Celsius, for the 65-year period from 1951 to 2015, inclusive. For each day, the values recorded are the minimum and maximum temperature occurring on that day. In this study, we use the common notation TN, TX and TG as abbreviations for the daily miNimum, maXimum and averaGe temperature, respectively. We obtained the data from the website of the European Climate Assessment and Dataset project (http://www.ecad.eu/). As has become standard in the literature, the observations for 29 February, occurring in each leap year, are removed from the series in order to have a constant 365 days in each year (see, for example, Campbell and Diebold 2005; Dupuis 2012). In our empirical work, we used the first 60 years of data to specify models, and then used rolling windows of the same length to estimate model parameters. Heat waves are not universally defined as is discussed in Section 5.1. We adopt a definition where heat waves are defined as the exceedances of temperatures for 3 consecutive days, as it has been found that people's health is sensitive to the exposure to high temperatures for 3 consecutive days (Pattenden et al. 2003; Le Tertre et al. 2006; Hajat et al. 2006; Perkins and Alexander 2013). Based on this definition, 1-3 day-ahead density forecasts are required to obtain heat wave forecasts in practice. Therefore, in this study, we used the final 5 years of data to evaluate post-sample forecasts for lead times up to 3 days.

Figure 5.1 is a plot of TN at Seville for the first estimation period 1951 to 2010, and Figure 5.2 is the corresponding plot of TX. The plots provide some indication of an overall rising trend over the latter half of the 60-year series, which is consistent with the widely discussed rise in global temperatures since the 1970s. Naturally, the series possess annual seasonality, which accounts for the repeating periodic spikes. The seasonality is clear from Figure 5.3, which plots TN and TX observations for Seville against the day of the year for the 60-year period 1951 to 2010. Note that, at least in the TN observations, there is an annual cycle in both the mean and the variance, with the mean generally at its highest in

Figure 5.2: Seville daily maximum temperature (TX) time series (in degrees Celsius) for the first 60-year estimation sample.

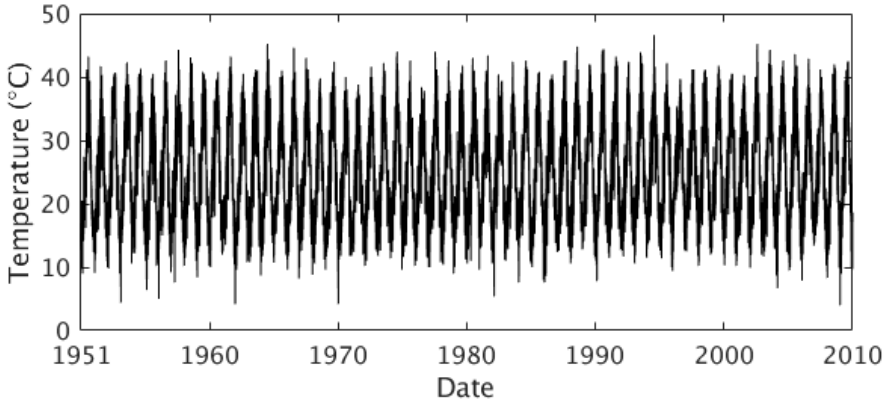
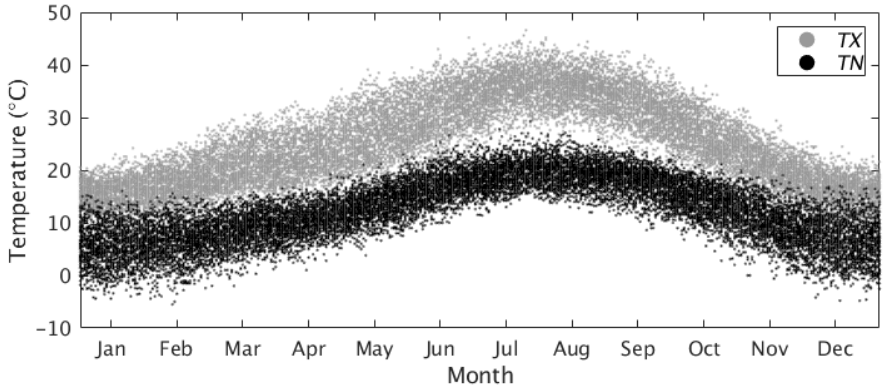


Figure 5.3: Seville daily minimum (TN) and maximum temperature (TX) (in degrees Celsius) plotted against the day of the year for the first 60-year estimation sample.



the summer months, while the variance is at its highest in the winter.

5.3 Univariate ARMA-GARCH Models

5.3.1 Model Specification

As we stated in Section 5.1, there is no existing literature considering the joint modelling of TN and TX. There are, however, several studies that have modelled either TN, TX or TG using univariate models. It has been shown that daily series of these variables exhibit the following features: a trend in the mean and variance; a seasonal pattern in the mean and variance; and both large and small absolute deviations from the mean tend to

cluster (Taylor and Buizza 2004; Dupuis 2012, 2014). These characteristics are particularly suitable for autoregressive modelling of the mean and variance.

Tol (1996) and Taylor and Buizza (2004) use AR-GARCH models to estimate the distribution of TG. They model the mean and variance using seasonal terms based on quadratic functions of a counter for the day of the year. Campbell and Diebold (2005) also implement an AR-GARCH model for TG, but their seasonal terms are based on Fourier terms, which are sums of sine and cosine functions of different cycles. In fitting an ARMA model to TX data, Wong (2015) uses a relatively complex model for the mean, but does not consider a model for the variance. Dupuis (2012; 2014) uses a multi-step approach in the univariate modelling of the TN and TX; the series are preprocessed by an AR-GARCH model before an EVT approach is used to estimate the tails of the resultant residuals. Erhardt et al. (2015) uses a copula approach to estimate the joint distribution of TG at different locations, with marginal distributions estimated by a multi-step AR model.

Empirical results have provided support for the AR-GARCH models in comparison with a variety of simpler alternatives. In this study, we extend the literature on univariate modelling by considering ARMA structures for the mean, rather than just AR, and by including novel trend features. The univariate model that we consider is the ARMA-GARCH model of expressions (5.1)-(5.5). In our empirical work, we applied the model separately to TN, TX, TG and DTR. ARMA-GARCH models have not previously been used for DTR. Our model has the following expression:

$$
\begin{aligned}
T_t &= S_1(\boldsymbol{\rho}_1,t) + \psi_1 Trend_t + S_2(\boldsymbol{\rho}_2,t)Trend_t + r_t && (5.1)\\
r_t &= \sum_{i=1}^{m} \phi_i r_{t-i} + \sum_{i=1}^{n} \delta_i \varepsilon_{t-i} + \varepsilon_t && (5.2)\\
\varepsilon_t &= h_t^{1/2}\eta_t && (5.3)\\
h_t &= S_3(\boldsymbol{\omega}_1,t) + \psi_2 Trend_t + S_4(\boldsymbol{\omega}_2,t)Trend_t \\
&\quad + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{o} \gamma_i \varepsilon_{t-i}^2 \mathbb{1}(\varepsilon_{t-i}<0) + \sum_{i=1}^{p} \beta_i h_{t-i} && (5.4)\\
S_i(\boldsymbol{\lambda},t) &= \lambda_0 + \sum_{j=1}^{J_i} \lambda_{1,j}\sin\left(2j\pi\frac{d(t)}{365}\right) + \lambda_{2,j}\cos\left(2j\pi\frac{d(t)}{365}\right) && (5.5)
\end{aligned}
$$

T_t is the temperature variable; $\boldsymbol{\rho}_i$ and $\boldsymbol{\omega}_i$ are vectors of parameters; ψ_i, ϕ_i, δ_i, α_i, γ_i, β_i, λ_0 and $\lambda_{i,j}$ are scalar parameters; $\boldsymbol{\lambda}$ is a vector with entries λ_0 and $\lambda_{i,j}$; $S_i(\boldsymbol{\lambda},t)$ are the seasonal terms involving the sum of pairs of Fourier terms; $d(t)$ is a repeating step function that numbers the days of the year from 1 to 365 within each year; J_i denotes the number of pairs of Fourier terms; r_t is the stochastic part of T_t; ε_t is the error term in the ARMA process for r_t; h_t is the variance of ε_t; η_t is an i.i.d. distribution with mean 0 and variance 1; and m, n, q, o, and p are the orders of the ARMA and GARCH components. Following Franses et al. (2001), Taylor and Buizza (2004) and Dupuis (2014), we included an asymmetric GARCH term in expression (5.4) to accommodate the effect that the impact of temperatures lower than expected on conditional volatility tends to be different from the impact of temperatures higher than expected.

In the proposed formulation, ε_t can be any i.i.d. distribution with zero mean and unit variance. We consider two different distributions: the Gaussian distribution and the generalised asymmetric skew-t distribution proposed by Zhu and Galbraith (2010). In addition to a skewness parameter, this distribution allows the degrees of freedom to be different for each side of the mean. The Gaussian distribution, Student's t distribution, and the skew-t are special cases of this distribution. In addition to our two distributional assumptions, we consider the EVT approach of McNeil and Frey (2000). This involves the peaks over threshold EVT approach being applied to the standardised residuals obtained from the ARMA-GARCH model. In this approach, we set the threshold as the 90% quantile of the assumed distribution, and fitted a generalised Pareto distribution to the standardised residuals that exceed this threshold.

5.3.2 Model Discussion

In Section 2, we discussed how our TN and TX time series exhibited an apparent trend, which is consistent with the literature on climate change. Unit root tests rejected the hypothesis of the trend being stochastic. We considered a variety of approaches to modelling a deterministic trend in the mean and variance of the series, including linear

and quadratic functions of time, but these delivered relatively poor fit. Visual inspection of the time series suggests a steady rise only from around the 1970s. This led us to consider a trend defined as being zero up until the start of a chosen year, and linear thereafter. We chose the starting year using the SBC. We did this for the TN and TX series for all four locations, and found that the optimal starting point for the linear trend was close to 1974. In view of this, we defined the variable $Trend_t$ in expressions (5.1)-(5.5) as being a linear trend starting on 1 January 1974. In Figure 5.4, we show a trend of this type fitted to the TN time series for Seville. Although we feel this simplistic modelling of the trend is reasonable for our study, which has its emphasis on short-term probabilistic prediction, we acknowledge that there is potential for the incorporation of a more sophisticated approach, such as the semiparametric panel model used by Atak et al. (2011) for monthly data.

A novel feature of our model is that we have included, in expressions (5.1) and (5.4) for the mean and variance, an interaction term, which is the product of the trend and seasonality. This allows the model to accommodate a different trend for each day of the year, which would imply that climate change does not have a uniform affect on the different seasons of the year. Figure 5.5 shows how the resulting estimated trend differs across different days of the year. For clarity of presentation, the figure focuses on just the first day of each month of the year. Allowing the trend to differ across the days of the year was motivated by the work of Proietti and Hillebrand (2017) with monthly temperature data.

In addition to the annual cyclicality, we also experimented with the inclusion of Fourier terms to model a possible intraweek cycle, which has been observed in the DTR for a variety of locations (see, for example, Forster and Solomon 2003). However, these terms were not statistically significant in our models using the asymptotic theory developed in Han and Kristensen (2014), and so we do not consider them further.

The lag orders, the numbers of Fourier terms, and the specification of the trend term were chosen using the SBC. For the mean, this led to the use of three pairs of Fourier terms,

interaction terms involving the trend and three pairs of Fourier terms, and an ARMA(3,1) model. For the variance, this led to two pairs of Fourier terms, the same trend variable as in the model for the mean, interaction terms involving this trend and two pairs of Fourier terms, and an asymmetric GARCH term.

Tables 5.1 and 5.2 present the parameters for the univariate ARMA-GARCH model with Gaussian assumption, estimated using the first 60-year rolling window of TN and TX recorded at Seville. The insight provided by the model is consistent with previous studies. The coefficients for the trend terms in the ARMA parts of both models are positive, which indicates that the levels of TN and TX are rising. The coefficient of the trend term in the ARMA part for TN is larger than that for TX, which suggests a decrease in the DTR. These findings are consistent with those in previous work (see, for example, Dupuis 2014; Qu et al. 2014). One term of each interaction pair is significant in both the ARMA and GARCH parts of the model, with the GARCH part for TX being the only exception, implying that the impact of the trend is not the same across the seasonal cycle. This was the finding of Proietti and Hillebrand (2017) for monthly data, while our work has confirmed the existence of the effect in the mean and variance of daily data. In the rest of the study, we revert from presenting the parameters for other models, because the insight was similar, and because, for the bivariate models, the number of parameters is relatively large.

Figure 5.4: Seville daily minimum temperature (TN) (in degrees Celsius) for the first 60-year estimation sample, with a linear trend starting from the beginning of 1974.

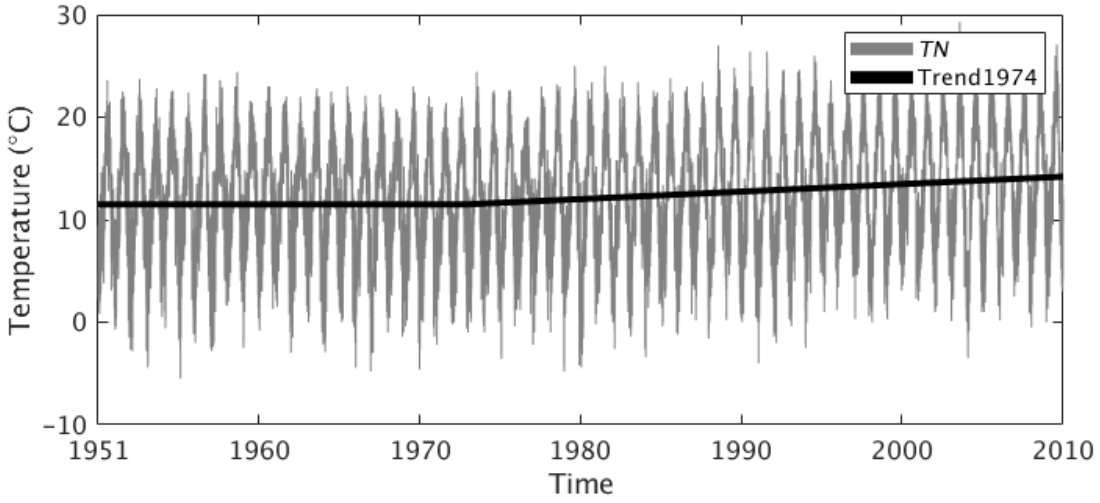


Figure 5.5: For the first day of each month, Seville daily minimum temperature (TN) (in degrees Celsius) plotted for the first 60-year estimation sample, along with the trend estimated by the univariate ARMA-GARCH model with Gaussian assumption.

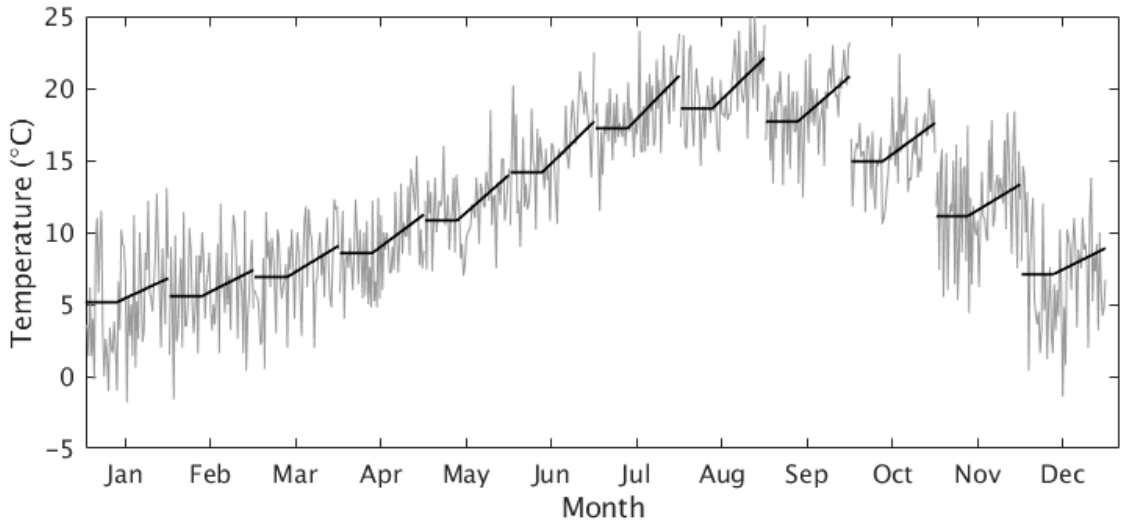


Table 5.1: Parameter estimates and standard errors for the mean component of the univariate ARMA-GARCH model with Gaussian assumption, derived using the first 60 years of daily minimum (TN) and maximum (TX) temperature observations for Seville.

	TN	TX
$(\rho_1)_0$	115(0.86)	245(1.08)
$(\rho_1)_{1,1}$	-31.4(1.12)	-40.3(1.43)
$(\rho_1)_{2,1}$	8.11(0.773)	17.3(0.958)
$(\rho_1)_{1,2}$	0.578(0.695)	-2.19(0.8)
$(\rho_1)_{2,2}$	-58.2(1.19)	-87.9(1.39)
$(\rho_1)_{1,3}$	-2.77(0.774)	-2.02(0.94)
$(\rho_1)_{2,3}$	-2.29(0.705)	-2.36(0.792)
ψ_1	0.002(0.000124)	0.00104(0.000175)
$(\rho_2)_{1,1}$	0.0000171(0.000163)	0.000816(0.000233)
$(\rho_2)_{1,2}$	-0.000739(0.000171)	-0.000296(0.00023)
ϕ_1	1.48(0.0218)	1.71(0.0159)
ϕ_2	-0.513(0.0162)	-0.797(0.0164)
ϕ_3	-0.00311(0.00803)	0.073(0.00737)
δ_1	-0.863(0.0208)	-0.932(0.0142)

Table 5.2: Parameter estimates and standard errors for the variance component of the univariate ARMA-GARCH model with Gaussian assumption, derived using the first 60 years of daily minimum (TN) and maximum (TX) temperature observations for Seville.

	TN	TX
$(\omega_1)_0$	101(15.8)	234(22.2)
$(\omega_1)_{1,1}$	4.9(2.38)	22(5.02)
$(\omega_1)_{2,1}$	0.651(1.33)	-7.92(3.47)
$(\omega_1)_{1,2}$	51.9(8.29)	-65.5(7.5)
$(\omega_1)_{2,2}$	8.64(1.83)	-8.24(3.37)
ψ_2	-0.00301(0.000507)	0.0012(0.000579)
$(\omega_2)_{1,1}$	-0.000245(0.000245)	0.000296(0.000745)
$(\omega_2)_{2,1}$	-0.00211(0.000414)	0.00125(0.000765)
α_1	0.028(0.00613)	0.124(0.0117)
γ_1	0.0384(0.00875)	-0.0685(0.0119)
β_1	0.753(0.0349)	0.447(0.0473)

5.4 Bivariate VARMA-MGARCH Models

5.4.1 A Gaussian Bivariate VARMA-MGARCH Model

In this section, we present a bivariate model for TN and TX. We use a vector ARMA (VARMA) model for the mean of a vector containing the two variables, and a multivariate GARCH (MGARCH) model for the covariance matrix. There exist different MGARCH specifications, such as the VEC model (Bollerslev et al. 1988), the dynamic conditional correlation model (Engle 2002), and the BEKK model (Engle and Kroner 1995). In this study, we use the VEC model, which has a very flexible and general specification. In comparison with the univariate model of Section 5.3, our bivariate model has the same essential features, with the addition of a model for the conditional covariance. Having said that, it is important to note that the lag structure is far richer in the bivariate model, as the two temperature variables are also modelled in terms of lags of each other. The proposed

model is expressed as follows:

$$\begin{aligned}
\boldsymbol{T}_t &= \boldsymbol{S}_1(\boldsymbol{\rho}_1,t)+\boldsymbol{\lambda}_1 Trend_t+\boldsymbol{S}_2(\boldsymbol{\rho}_2,t)Trend_t+\boldsymbol{r}_t && (5.6)\\
\boldsymbol{r}_t &= \sum_{i=1}^{m}\boldsymbol{B}_{1,i}\boldsymbol{r}_{t-i}+\sum_{i=1}^{n}\boldsymbol{B}_{2,i}^2\boldsymbol{\epsilon}_{t-i}+\boldsymbol{\epsilon}_t && (5.7)\\
\boldsymbol{\epsilon}_t &= \boldsymbol{H}_t^{1/2}\boldsymbol{\eta}_t && (5.8)\\
\text{vech}(\boldsymbol{H}_t) &= \boldsymbol{S}_3(\boldsymbol{\rho}_3,t)+\boldsymbol{\lambda}_2 Trend_t+\boldsymbol{S}_4(\boldsymbol{\rho}_4,t)Trend_t\\
&\quad+\sum_{i=1}^{q}\boldsymbol{B}_{3,i}\text{vech}(\boldsymbol{\epsilon}_{t-i}\boldsymbol{\epsilon}_{t-i}^{T})\\
&\quad+\sum_{i=1}^{o}\boldsymbol{B}_{4,i}\text{vech}(\boldsymbol{\epsilon}_{t-i}\boldsymbol{\epsilon}_{t-i}^{T})\mathbb{1}\left(\text{vech}(\boldsymbol{\epsilon}_{t-i}\boldsymbol{\epsilon}_{t-i}^{T})<\boldsymbol{0}\right)\\
&\quad+\sum_{i=1}^{p}\boldsymbol{B}_{5,i}\text{vech}(\boldsymbol{H}_{t-i}) && (5.9)\\
\boldsymbol{S}_i(\boldsymbol{\rho}_i,t) &= (\boldsymbol{\rho}_i)_1+\sum_{j=1}^{J_i}\left((\boldsymbol{\rho}_i)_{2j}\sin\left(2j\pi\frac{d(t)}{365}\right)+(\boldsymbol{\rho}_i)_{2j+1}\cos\left(2j\pi\frac{d(t)}{365}\right)\right) && (5.10)
\end{aligned}$$

where $\boldsymbol{T}_t=(TN_t,\ TX_t)$; vech denotes the operator that stacks the lower triangular portion of a matrix as a column vector; $\boldsymbol{\lambda}_i$ are vectors of parameters; $(\boldsymbol{\rho}_i)_k$ are scalar parameters; $\boldsymbol{\rho}_i$ are vectors consisting of the concatenation of $(\boldsymbol{\rho}_i)_k$ for for $k=1,2$; $\boldsymbol{B}_{k,i}$ are parameter matrices for $k=1,2,...5$; $\boldsymbol{S}_i(\boldsymbol{\lambda},t)$ is a deterministic seasonal vector; $\boldsymbol{r}_t$ contains the stochastic parts of $\boldsymbol{T}_t$; J_i denotes the number of pairs of Fourier terms; $\boldsymbol{\epsilon}_t$ is the error term of $\boldsymbol{r}_t$; $\boldsymbol{H}_t$ is the covariance matrix of $\boldsymbol{\epsilon}_t$; and m, n, q, o and p are the VARMA and MGARCH orders. We assume $\boldsymbol{\eta}_t$ is a standard multivariate Gaussian distribution, which is a common assumption in the MGARCH literature (Bauwens et al. 2006). $d(t)$ and $Trend_t$ are defined as in Section 5.3. For practicality, in specifying the bivariate model, we used the analogous terms that we had selected for the univariate model.

5.4.2 Bivariate VARMA-MGARCH with a Dynamic Copula

In addition to a multivariate Gaussian, multivariate Student's t distributions and various multivariate skew-t distributions have also been considered in MGARCH applications (Bauwens and Laurent 2005; Lee and Long 2009). In these distributions, the marginal

distributions have the same degrees of freedom. For us, this was a limitation because our univariate modelling had shown that the marginal distributions of TN and TX have very different degrees of freedom. To overcome this, and to enable a flexible modelling of the dependence structure between marginal distributions, we propose the use of a copula. Dependence structure refers to everything that can't be captured by the marginal distributions of a multivariate distribution, such as correlation and correlation in higher moments.

Sklar's theorem enables the decomposition of a joint distribution into its marginal distributions and a dependence function, which is the copula (Kolev et al. 2006; Patton 2012). We consider the Gaussian copula and the Student's t copula, which are standard in the literature (Patton 2006, 2012). The Gaussian copula has one parameter, which is a correlation parameter, and the Student's t copula has two parameters, which are a correlation parameter and a parameter for the degrees of freedom. Note that the degrees of freedom of the copula relate only to the dependency structure between the variables, and not to their marginal distributions. Our proposed copula-based VARMA-MGARCH model can be written as follows:

$$\begin{aligned}
\boldsymbol{T}_t &= \boldsymbol{S}_1(\boldsymbol{\rho}_1,t) + \boldsymbol{\lambda}_1 Trend_t + \boldsymbol{S}_2(\boldsymbol{\rho}_2,t) Trend_t + \boldsymbol{r}_t && (5.11)\\
\boldsymbol{r}_t &= \sum_{i=1}^{m} \boldsymbol{B}_{1,i}\boldsymbol{r}_{t-i} + \sum_{i=1}^{n} \boldsymbol{B}_{2,i}\boldsymbol{\epsilon}_{t-i} + \boldsymbol{\epsilon}_t && (5.12)\\
\boldsymbol{\epsilon}_t &= (\varepsilon_{1,t}, \varepsilon_{2,t})^T && (5.13)\\
F_{\boldsymbol{\epsilon}_t}(z_1,z_2) &= C_{\boldsymbol{R}_t,\nu}(F_{\varepsilon_{1,t}}(z_1), F_{\varepsilon_{2,t}}(z_2)) && (5.14)\\
\mathrm{var}(\varepsilon_{j,t}) &= h_{j,t} && (5.15)\\
\boldsymbol{H}_t &= \begin{pmatrix} h_{1,t} & R_t\sqrt{(h_{1,t}h_{2,t})} \\ R_t\sqrt{(h_{1,t}h_{2,t})} & h_{2,t} \end{pmatrix} && (5.16)\\
\mathrm{vech}(\boldsymbol{H}_t) &= \boldsymbol{S}_3(\boldsymbol{\rho}_3,t) + \boldsymbol{\lambda}_2 Trend_t + \boldsymbol{S}_4(\boldsymbol{\rho}_4,t) Trend_t \\
&\quad + \sum_{i=1}^{q} \boldsymbol{B}_{3,i}\mathrm{vech}(\boldsymbol{\epsilon}_{t-i}\boldsymbol{\epsilon}_{t-i}^T) \\
&\quad + \sum_{i=1}^{o} \boldsymbol{B}_{4,i}\mathrm{vech}(\boldsymbol{\epsilon}_{t-i}\boldsymbol{\epsilon}_{t-i}^T)\mathbb{1}\left(\mathrm{vech}(\boldsymbol{\epsilon}_{t-i}\boldsymbol{\epsilon}_{t-i}^T) < \mathbf{0}\right)
\end{aligned}$$

$$+\sum_{i=1}^{p} \boldsymbol{B}_{5,i}\text{vech}(\boldsymbol{H}_{t-i}) \tag{5.17}$$

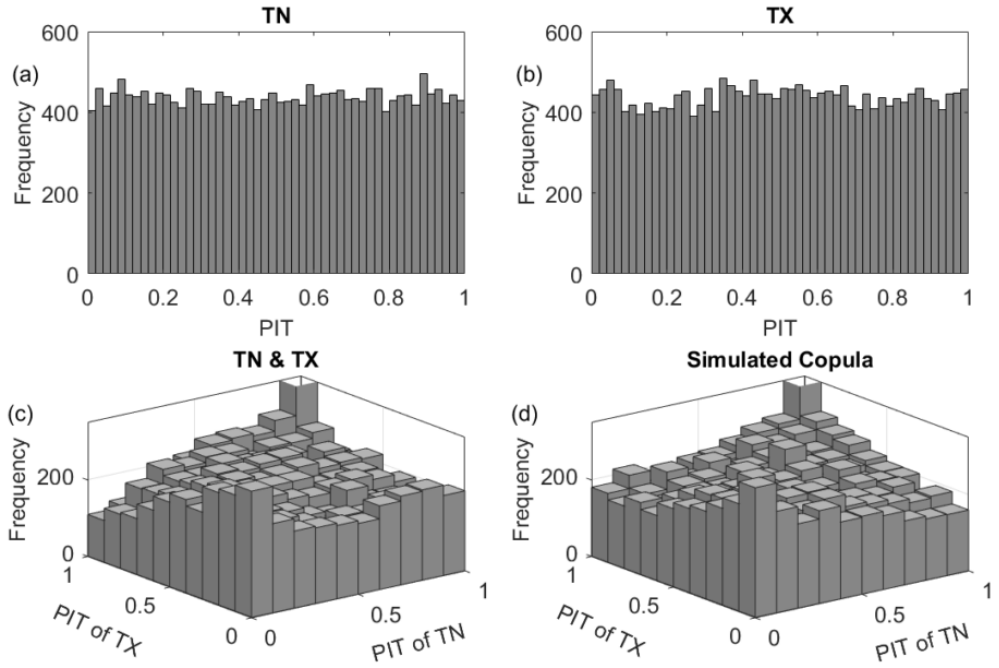
$$\boldsymbol{S}_i(\boldsymbol{\rho}^i,t) = \boldsymbol{\rho}_1^i + \sum_{j=1}^{J_i}\left(\boldsymbol{\rho}_{i,2j}\sin\left(2j\pi\frac{d(t)}{365}\right)+\boldsymbol{\rho}_{i,2j+1}\cos\left(2j\pi\frac{d(t)}{365}\right)\right) \tag{5.18}$$

where $h_{j,t}$ is the variance of the error term $\varepsilon_{j,t}$; $C_{\boldsymbol{R}_t,\nu}(\cdot,\cdot)$ is the Gaussian or Student's t copula; R_t is the correlation parameter in the copula; $\boldsymbol{R}_t$ is the correlation matrix corresponding to $\boldsymbol{H}_t$; ν is the degrees of freedom for the Student's t copula and it is ∞ for the Gaussian copula; $F_{\boldsymbol{\epsilon}_t}(\cdot,\cdot)$ is CDF of $\boldsymbol{\epsilon}_t$; $F_{\varepsilon_{i,t}}(\cdot)$ is the distribution function of $\varepsilon_{i,t}$; and the other notations are the same as in Section 5.4.1.

To obtain this model, we have simply replaced the covariance matrix decomposition in expressions (5.8) of Section 5.4.1 by the copula decomposition in expressions (5.13)-(5.14). Consequently, we do not use the VEC approach in the standard way to model the dynamics of the covariance matrix of $\boldsymbol{\epsilon}_t$, but instead we use the VEC approach to model the variances of $\varepsilon_{j,t}$ and the copula's correlation parameter simultaneously in expressions (5.15)-(5.16). Note that, as the copula's correlation parameter is modelled within the VEC structure, the dependency between the error terms is modelled using a dynamic copula. Other dynamic copulae are considered by Patton (2006; 2012). A feature of our work is that the variances of the marginal distributions and the copula parameter(s) are modelled simultaneously, while other studies have tended to estimate the copula parameter(s) after first modelling the individual marginal distributions.

In Figure 5.6, we illustrate the in-sample model fit for the VARMA-MGARCH model with generalised asymmetric skew-t distribution and Student's t copula. The upper plots are the histograms of the PIT for TN and TX under the marginal distributions estimated using the VARMA-MGARCH model with generalised asymmetric skew-t distribution and Student's t copula. Each histogram is approximately uniformly distributed between 0 and 1, which supports the hypothesis that the model is correctly specified (Diebold et al. 1998). The bottom left plot is the histogram of the bivariate PIT for the joint distribution of TN and TX. The lack of bivariate uniformity in this plot reflects the existence of dependence

Figure 5.6: Histograms of PIT with for the marginal distributions of TN (a) and TX (b), the joint distribution of TN and TX (c), and a sample simulated from the estimated Student's t copula (d), based on the bivariate VARMA-MGARCH model with generalised asymmetric skew-t distribution distribution and Student's t copula, derived using the first 60 years of daily minimum (TN) and maximum (TX) temperature observations for Seville.



structure. Finally, the bottom right plot is the histogram of the bivariate PIT resulting from a sample simulated from the estimated Student's t copula. The similarity between the lower two plots of Figure 5.6 supports the use of this copula to capture the dependence structure between TN and TX.

5.5 Empirical Study

5.5.1 Models

We implemented the univariate ARMA-GARCH model of Section 5.3 for TN and TX. We used four different distributional assumptions: Gaussian distribution, generalised asymmetric skew-t distribution, Gaussian distribution with EVT for the tails, and generalised asymmetric skew-t distribution with EVT for the tails. To obtain the bivariate distribution of TN and TX, we took the simplistic approach where we assumed the

independency between the univariate distributions. We also fitted the univariate ARMA-GARCH model of Section 5.3 to the TG and DTR series. We considered the same four distributional assumptions and assumed the univariate distributions were independent. To fit the models, we used the interior point optimisation algorithm in Matlab (Byrd et al. 1999; Waltz et al. 2006). The initial parameters were derived sequentially using ordinary least square (OLS). (In a first step, initial values of ρ_1, ψ_1 and ρ_2 were derived by using OLS applied to T_t. In a second step, the resulting residuals were used with OLS to obtain initial ARMA parameters ϕ_i and δ_i. In a third step, the rest of the initial parameters were derived by applying OLS to the squared residuals from the first step.)

We then implemented the Gaussian bivariate VARMA-MGARCH model for TN and TX, which was described in Section 5.4. In terms of the bivariate VARMA-MGARCH models for TN and TX, described in Section 5.4.2, we considered four different distributional assumptions: Gaussian marginal distributions with no dependence structure; generalised asymmetric skew-t marginal distributions with no dependence structure; generalised asymmetric skew-t marginal distributions with a Gaussian copula; and generalised asymmetric skew-t marginal distributions with a Student's t copula. In the VEC approach that we used within the VARMA-MGARCH models, we ensured the positive definiteness of the matrix $\boldsymbol{H}_t$ by adding a large penalty term to the likelihood in the case of a violation occurring. For the optimisation algorithm, we set the initial parameter values to be those estimated using the univariate Gaussian models for TN and TX, with the remaining parameters set to zero. We then passed this initial parameter vector to the interior point optimisation algorithm in Matlab.

5.5.2 Using Simulation to Generate Forecasts up to 3 Days-ahead

The ARMA-GARCH and VARMA-MGARCH models have closed form 1 day-ahead distributional forecasts, but such analytic solutions do not exist for multi-step-ahead prediction. To address this, we used simulation from each model with 1500 sample paths

simulated over the forecast horizon of 3 days. The marginal and joint distributions were constructed from these sample paths. We make two remarks here. Firstly, in practice, TX should obviously be no less than TN, or equivalently, DTR should be non-negative. We do not address this issue explicitly in our models because such temperature crossing occurs very infrequently due to the distributions of TN and TX being reasonably far apart throughout the year. In our simulation study, the frequency of crossing occurred in fewer than 1% of the simulated sample paths, and so we took the pragmatic approach of setting TN and TX to be their average when crossing occurred. As for the ARMA-GARCH models for TG and DTR, we set DTR to be zero if a sample path gave a negative value. Secondly, the matrix $\boldsymbol{H}_t$ can violate the assumption that it is positive definite. The frequency of this violation was less than 0.01%, and so we simply re-simulated a sample path if the violation occurred.

5.5.3 Evaluating Forecasts of Marginal and Joint Distributions

To evaluate each model's forecasts of the mean, we used the RMSE. To evaluate univariate and multivariate density forecasts, we used the CRPS and energy score, respectively. The CRPS is widely used to evaluate univariate density forecasts (Gneiting and Raftery 2007). It is said to be *proper* because its expectation is minimised by the correct forecast. This property is important, as it encourages honest assessments (Garthwaite et al. 2005). The CRPS is expressed as follows:

$$CRPS(F_t, y_t) = \int_{-\infty}^{\infty} (F_t(x) - \mathbb{1}(y_t \geq x))^2 dx \tag{5.19}$$

$$= -\frac{1}{2}\mathbb{E}_{F_t}|Z_t^1 - Z_t^2| + \mathbb{E}_{F_t}|Z_t^1 - y_t| \tag{5.20}$$

where y_t is the actual observation; F_t is the predicted CDF; Z_t^1 and Z_t^2 are two independent copies of a random variable with CDF F_t; and $\mathbb{E}_{F_t}$ is the expectation with respect to the distribution with CDF F_t. Expression (5.20) is a potentially useful way to calculate the CRPS in practice (Gneiting and Raftery 2007). The score is averaged over the post-sample

period to produce an overall measure of accuracy.

Gneiting and Raftery (2007) introduce the energy score as a generalisation of the CRPS to evaluate multivariate distributions. The energy score is expressed as follows:

$$Energyscore(F_t, \boldsymbol{y}_t) = -\frac{1}{2}\mathbb{E}_{F_t}\|\boldsymbol{Z}_t^1 - \boldsymbol{Z}_t^2\| + \mathbb{E}_{F_t}\|\boldsymbol{Z}_t^1 - \boldsymbol{Z}_t^2\| \tag{5.21}$$

where F_t is the estimated multivariate CDF, $\boldsymbol{y}_t$ is the actual observation; $\|\cdot\|$ is the Euclidean norm; $\boldsymbol{Z}_t^1$ and $\boldsymbol{Z}_t^2$ are two independent copies of a random vector drawn from the distribution with CDF F_t. The energy score is not as well established as the CRPS, which reflects the tendency for density forecasting studies to focus only on univariate densities. However, with the increasing availability of large datasets, the importance and feasibility of forecasting multivariate distributions is increasing, and this is leading to increased interest in measures for evaluating joint density forecasts, including the energy score.

As the values of the RMSE, CRPS and energy score do not have an intuitive interpretation, we calculated the ratio of the score for each method to the score for a benchmark, which we chose as the univariate ARMA-GARCH model with Gaussian distribution. We averaged the resulting ratios over the four Spanish series, and then subtracted this from 1 to deliver a measure that reflects the percentage by which each method is better than the benchmark method. This measure is known in the probabilistic forecast evaluation literature as a skill score. Positive values imply greater accuracy than the benchmark method, and higher values indicate superior accuracy.

For 1 day-ahead prediction, Table 5.3 presents the RMSE skill scores and Table 5.4 presents the skill scores for the CRPS and energy score. We do not report the results for 2 and 3 day-ahead prediction, as the relative performances of the models was similar to that for 1 day-ahead. In Tables 5.3 and 5.4, the first four columns of values present the skill scores for the four temperature variables, TN, TX, TG and DTR. Each row in the tables can be viewed as representing the forecasts produced by a pair of the temperature variables. For example, the first four rows in each table were produced by the univariate

models of TN and TX. For these rows, the TN forecasts were produced using just the model for TN; the TX forecasts were produced using just the model for TX; and the TG and DTR forecasts were produced using the models for both TN and TX.

For the univariate models, there is some indication in the tables that using the generalised asymmetric skew-t is more accurate for TX and TG than employing a Gaussian assumption. With regard to the EVT approach, it did not lead to improved forecast accuracy. It is interesting to see that generating forecasts for TN from the univariate models for TG and DTR led to greater accuracy than simply using a univariate model for TN. This was not the case for TX. This reflects the view of Dupuis (2014) that TN is more challenging to model than TX.

Turning to the bivariate VARMA-MGARCH models, it is interesting to note from Tables 5.3 and 5.4 that each bivariate model produced better RMSE and CRPS results than the univariate models for TN, TG and DTR, and that the bivariate models were no worse than the univariate models for TX. Given that TG is the focus in some important applications, such as weather derivatives (see, for example, Elias et al. 2014; Gülpınar and Çanakoğlu 2017), it is useful to know that a bivariate model of TN and TX can outperform a univariate model of TG.

The final column of Table 5.3 presents the energy score for forecasts of the joint distribution of TN and TX. As this distribution can be used to generate a forecast of the joint distribution of TG and DTR, the energy score obtained for these two joint distributions will be identical. The energy scores in Table 5.3 show that the bivariate modelling led to the greatest forecast accuracy, with particularly good results for the generalised asymmetric skew-t distribution with dynamic Gaussian copula.

Figures 5.7-5.9 show the CRPS, for TN and TX, and the energy score calculated for each month. To prevent the figures becoming overwhelming, we select only three models, the univariate ARMA-GARCH model with generalised asymmetric skew-t distribution, the bivariate VARMA-MGARCH model with generalised asymmetric skew-t distribution and no dependency, and the bivariate VARMA-MGARCH model with generalised asymmetric skew-t distribution and Gaussian copula. Firstly, we observe that the

univariate model performs very competitively for TX, but is noticeably outperformed by the bivariate models for TN, and also for the joint distribution of TN and TX. Secondly, the dependency modelling of TN and TX in the bivariate models does not improve forecasting accuracy for the marginal distributions of TN and TX, but does improve a little the prediction accuracy for the joint distribution of TN and TX. Thirdly, the use of a bivariate model would seem to be particularly beneficial in the summer. This is noteworthy because the forecasting of heat waves is an important focus of this study. Fourthly, the skill scores for the bivariate models are sometimes negative for the winter, suggesting that there is still scope for improving these models.

5.5.4 Evaluating Exceedance Probability Forecasts

To evaluate probability forecasts of heat waves, we assessed the accuracy of forecasts of the probability of the event that TN and TX exceed specified thresholds for three consecutive days. There is no universal choice for the thresholds as they are generally location-based. In this study, we considered two pairs of thresholds: 15 and 30 degrees Celsius for TN and TX, respectively, and also thresholds of 17.5 and 35 degrees Celsius. We evaluated the forecasts using the widely used *Brier score*, which is expressed as:

$$BS(p_t, o_t) = (p_t - o_t)^2$$

where p_t is the predicted probability, o_t is the actual outcome of the event, which is 0 if the event occurs, and 1 otherwise.

For each model, we computed the Brier skill score (see, for example, Taylor 2017b). Table 5.5 presents the skill scores for the thresholds 15 and 30. In the table, the first column corresponds to the event that TN exceeds 15 on each of the next three days; the second column reports the results for the event that TX exceeds 30 on each of the next three days; and the third column corresponds to the event that TN exceeds 15 and TX exceeds 30 on each of the next three days. Similarly, Table 5.6 presents the Brier skill scores for the more extreme thresholds of 17.5 and 35. In Table 5.5, the ranking of methods is similar to that

for the RMSE and CRPS in Tables 5.3 and 5.4, with the best overall results corresponding to the bivariate models of TN and TX produced using the generalised asymmetric skew-t distribution with dynamic Gaussian copula. In Table 5.5, we see a slightly different picture, with the models involving the asymmetric distribution not performing so competitively. However, it is noteworthy that the bivariate models still perform well for the second pair of thresholds.

Table 5.3: RMSE for 1-day ahead forecasts of the mean of the marginal distribution of daily minimum (TN), maximum (TX) and average (TG) temperature, and DTR. Skill scores averaged across the four locations. Bold indicates the best performing model in each column.

	TN	TX	TG	DTR
Univariate ARMA-GARCH for TN and TX				
Gaussian	0.0	0.0	0.0	0.0
Gen asym skew-t	0.0	**0.2**	0.4	-0.4
Gaussian with EVT	0.0	0.1	0.0	0.1
Gen asym skew-t with EVT	0.0	**0.2**	0.5	-0.5
Univariate ARMA-GARCH for TG and DTR				
Gaussian	5.5	-1.8	0.7	2.2
Gen asym skew-t	5.5	-1.9	0.7	2.1
Gaussian with EVT	5.5	-1.8	0.7	2.1
Gen asym skew-t with EVT	5.5	-1.8	0.8	2.1
Bivariate VARMA-MGARCH for TN and TX				
Bivariate Gaussian	**6.2**	0.1	**3.1**	2.2
Gaussian with no dependency	**6.2**	0.1	**3.1**	2.0
Gen asym skew-t with no dependency	**6.2**	**0.2**	**3.1**	**2.3**
Gen asym skew-t with Gaussian copula	6.1	**0.2**	3.0	**2.3**
Gen asym skew-t with Student's t copula	6.1	0.1	3.0	2.2

Table 5.4: CRPS skill scores for 1-day ahead forecasts of the marginal distribution of daily minimum (TN), maximum (TX) and average (TG) temperature, and DTR. Energy skill scores for forecasts of the joint distribution of TN and TX. Skill scores averaged across the four locations. Bold indicates the best performing model in each column.

	CRPS				Energy
	TN	TX	TG	DTR	Score
Univariate ARMA-GARCH for TN and TX					
Gaussian	0.0	0.0	0.0	0.0	0.0
Gen asym skew-t	0.0	0.7	0.6	0.0	0.4
Gaussian with EVT	0.0	0.1	0.0	0.2	0.1
Gen asym skew-t with EVT	0.0	0.7	0.6	-0.1	0.4
Univariate ARMA-GARCH for TG and DTR					
Gaussian	5.1	-2.4	1.1	2.6	1.4
Gen asym skew-t	5.3	-2.2	1.3	2.8	1.6
Gaussian with EVT	5.1	-2.4	1.1	2.6	1.4
Gen asym skew-t with EVT	5.2	-2.2	1.4	2.8	1.6
Bivariate VARMA-MGARCH for TN and TX					
Bivariate Gaussian	7.0	0.1	3.7	2.6	3.2
Gaussian with no dependency	7.1	0.1	3.7	1.6	3.0
Gen asym skew-t with no dependency	**7.3**	0.7	3.9	2.2	3.4
Gen asym skew-t with Gaussian copula	7.1	**0.8**	**4.0**	**2.9**	**3.6**
Gen asym skew-t with Student's t copula	6.7	0.4	3.6	2.5	3.2

Figure 5.7: CRPS skill scores for each month for the prediction of the marginal distribution of the daily minimum (TN) temperature using three different models.

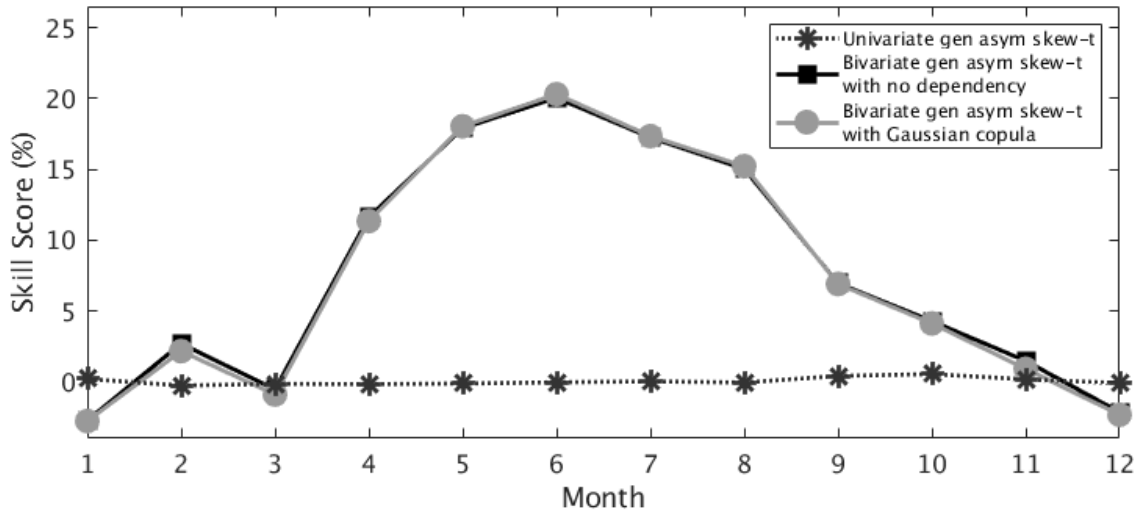


Figure 5.8: CRPS skill scores for each month for the prediction of the marginal distribution of the daily maximum (TX) temperature using three different models.

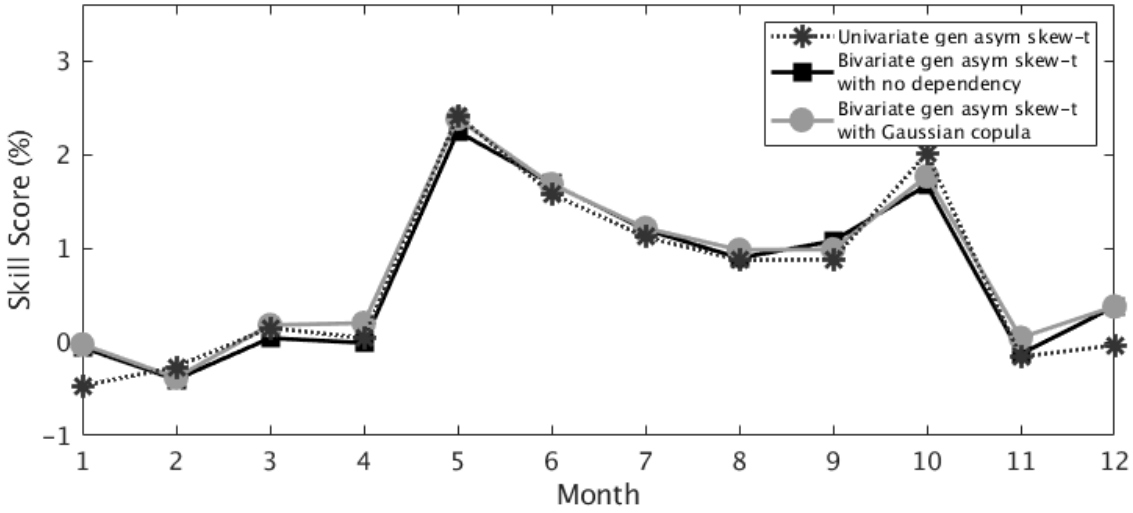


Figure 5.9: Energy skill scores for each month for the prediction of the joint distribution of the daily minimum (TN) and daily maximum (TX) temperature using three different models.

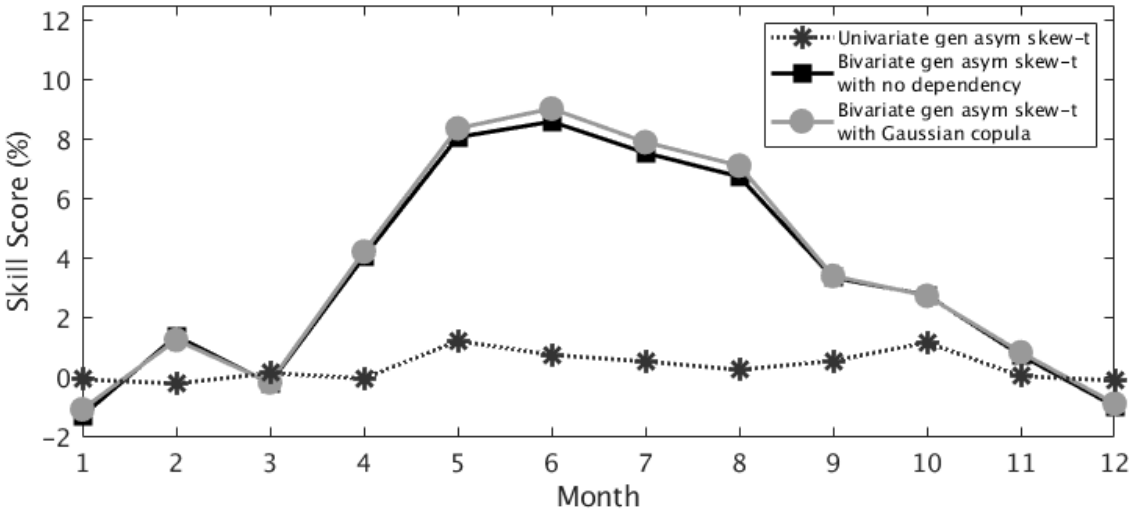


Table 5.5: Brier skill scores for probability forecasts of exceedance over all of the next 3 days for the less extreme pair of thresholds. Skill scores averaged across the four locations. Bold indicates the best performing model in each column.

	TN > 15	TX > 30	TN>15 & TX>30
Univariate ARMA-GARCH for TN and TX			
Gaussian	0.0	0.0	0.0
Gen asym skew-t	-0.2	1.1	1.6
Gaussian with EVT	0.1	0.0	-0.2
Gen asym skew-t with EVT	-0.2	1.3	1.6
Univariate ARMA-GARCH for TG and DTR			
Gaussian	6.1	-1.9	2.5
Gen asym skew-t	6.5	-1.7	3.2
Gaussian with EVT	6.4	-2.1	2.6
Gen asym skew-t with EVT	6.6	-1.8	3.1
Bivariate VARMA-MGARCH for TN and TX			
Gaussian with no dependency	10.6	-0.2	4.4
Bivariate Gaussian	10.5	-0.1	5.2
Gen asym skew-t with no dependency	10.3	1.5	5.6
Gen asym skew-t with Gaussian copula	**10.7**	**1.6**	**6.2**
Gen asym skew-t with Student's t copula	10.4	1.4	5.8

Table 5.6: Brier skill scores for probability forecasts of exceedance over all of the next 3 days for the more extreme pair of thresholds. Skill scores averaged across the four locations. Bold indicates the best performing model in each column.

	TN > 17.5	TX > 35	TN>17.5 & TX>35
Univariate ARMA-GARCH for TN and TX			
Gaussian	0.0	0.0	0.0
Gen asym skew-t	-0.2	-1.3	0.2
Gaussian with EVT	0.0	0.3	0.1
Gen asym skew-t with EVT	-0.2	-1.2	0.4
Univariate ARMA-GARCH for TG and DTR			
Gaussian	6.8	-1.9	1.8
Gen asym skew-t	6.9	-2.7	1.5
Gaussian with EVT	6.8	-2.0	1.9
Gen asym skew-t with EVT	7.0	-2.5	1.6
Bivariate VARMA-MGARCH for TN and TX			
Gaussian with no dependency	9.2	0.0	3.5
Bivariate Gaussian	9.2	**0.4**	**4.3**
Gen asym skew-t with no dependency	8.8	-1.0	2.7
Gen asym skew-t with Gaussian copula	**9.7**	-0.7	3.1
Gen asym skew-t with Student's t copula	9.1	-1.5	2.2

5.6 Summary

In this study, we have considered univariate and bivariate methods for density forecasting of the daily minimum and maximum temperature. We implemented univariate ARMA-GARCH models with different distributional assumptions, and with the use of EVT. We considered bivariate VARMA-MGARCH models with a variety of different distributional and dependency specifications. These included a bivariate normal distribution, with and without an assumption of independence, and a bivariate distribution composed of skew-t marginal distributions and a dynamic copula to capture the conditional dependency.

We evaluated short-term forecasts of the marginal and joint distribution for the minimum and maximum temperature, as well as for average temperature and DTR. We also evaluated the accuracy of probabilistic forecasts of heat waves, defined as the minimum and maximum simultaneously exceeding chosen threshold. In our empirical analysis, a bivariate VARMA-MGARCH model with dynamic dependence structure produced the best results. It was interesting to find that the models with more complex structure produced better empirical results. We note that the most successful model was a VARMA-MGARCH model with 93 parameters. Although this is a reasonably large number of parameters for a model of this type, estimation benefited from the relatively long time series consisting of 21,900 daily observations.

The models revealed some interesting results: the daily minimum and maximum temperature have predictive power for each other; a dynamic dependence structure is useful for modelling the joint distribution of the minimum and maximum temperatures; and this joint distribution produces competitive density forecasts for the marginal distributions of the daily minimum, maximum and average temperature, as well as the DTR. Our models also showed increasing trends in the minimum and maximum temperatures, a decreasing trend in the DTR, and different parts of the seasonal cycle changing at different rates, which are findings that are consistent with those of other researchers. We have discussed the three studies in Chapters 3-5. In the next chapter, we will present a summary of the thesis and discuss some future research areas.

Chapter 6

Conclusion

In this thesis, we have produced three studies. The first study (Chapter 3) contributes to VaR forecasting. VaR is an important market risk measure, and its accurate estimation is of great importance for risk management. The study proposes several quantile regression models based on intra-day range and overnight return in CAViaR and QRHAR models. Intra-day range is a more efficient volatility estimator when compared to squared return, and it also has the advantage of being easy to obtain and less susceptible to market microstructure noise when compared to realised volatility. The empirical study is conducted on 18 market indices and a large set of benchmark methods. The empirical results are promising: the proposed models are able to outperform the benchmark methods, and the models based on the intra-day range and the overnight return outperform the models without the intra-day range and the overnight return.

The second study (Chapter 4) also focuses on financial data. Unlike the first study, which considers the use of the intra-day data in model specification for VaR estimation, we consider ES as well as VaR estimation, and we consider the use of intra-day data in parameter estimation. As VaR has some undesirable features, such as its ignoring the extreme observations beyond it, and a lack of sub-additivity, ES has been proposed for future regulatory frameworks. We consider the use of the CAViaR-FZ models that can be used to estimate VaR and ES simultaneously. We find an interesting observation that the dynamics of a pair of extreme VaR and ES can be well approximated by a pair of less

extreme VaR and ES, which, in essence, provides us more observations in the tail. We show theoretically that such an approach can improve the parameter estimation efficiency for the CAViaR-FZ models. The empirical results support our proposed methods.

The third study (Chapter 5) considers extreme temperature forecasting, which has in common with the first two studies (Chapters 3 and 4), a focus on the use of intra-day data. The daily minimum and maximum temperatures and the intra-day low/high log prices are both daily extreme observations for some intra-day time series. We use parametric univariate ARMA-GARCH and multivariate VARMA-MGARCH models to estimate the marginal and joint distributions of the daily minimum and maximum temperature. We also consider evaluating exceedance probability estimation for the purpose of heat wave forecasting. The empirical results show that the proposed VARMA-MGARCH methods can improve the forecasting accuracy of both the marginal and joint distributions of the daily minimum and maximum temperature, when compared to the benchmark ARMA-GARCH models.

Going forward, we aim to contribute further to the research literature on probabilistic forecasting in both theoretical and applied aspects. As such, one of our future research aims is to work in the area of probabilistic forecasting of multivariate distributions. Understanding the co-movement between multiple time series is particularly important for a wide range of applications (see, for example, Patton 2012; Schefzik et al. 2013). As an example, in finance, one of the causes behind the financial crisis in 2007-2008 is believed to be the poor estimation and understanding of the relationship between certain asset returns, such as the returns related to housing and mortgage defaults (Oh and Patton 2017). Multivariate probabilistic forecasting is particularly intriguing and challenging due to its inherently complex nature and computational difficulty. Although this is a rapidly developing area within the forecasting literature, it is still in its infancy, when compared with the vast literature on non-probabilistic point forecasting and the probabilistic forecasting of univariate distributions. We have several papers in progress. One paper contributes to the literature on forecast evaluation and decision analysis. Evaluation for

univariate distributions has been well-studied in a variety of applications, ranging from meteorology to economics and finance. However, multivariate distributional evaluation, by comparison, is still in its infancy. We present a novel measure for evaluating multivariate distributions based on the work of Gneiting (2007; 2011), which involves different weights being put on different regions of a distribution. In particular, we use the measure to evaluate the tail estimation of the multivariate density forecasts. The proposed evaluation measure is applied to both temperature and financial data. Another paper studies proper scoring rules for evaluating multivariate probabilistic forecasts. The paper considers the properties of two well-known proper scoring rules for multivariate distributions. In particular, we decompose the scoring rules into layer representations, which results in a family of novel proper scoring rules for a certain statistical object.

References

Acerbi, C. and Szekely, B. (2014). Back-testing expected shortfall. *Risk*, 27:76–81.

Ahoniemi, K. and Lanne, M. (2013). Overnight stock returns and realized volatility. *International Journal of Forecasting*, 29(4):592–604.

Aït-Sahalia, Y., Mykland, P. A., and Zhang, L. (2011). Ultra high frequency volatility estimation with dependent microstructure noise. *Journal of Econometrics*, 160(1):160–175.

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

Andersen, L. B. and Piterbarg, V. V. (2007). Moment explosions in stochastic volatility models. *Finance and Stochastics*, 11(1):29–50.

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

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Ebens, H. (2001a). The distribution of realized stock return volatility. *Journal of Financial Economics*, 61(1):43–76.

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

Andersen, T. G., Bollerslev, T., and Meddahi, N. (2011). Realized volatility forecasting and market microstructure noise. *Journal of Econometrics*, 160(1):220–234.

Atak, A., Linton, O., and Xiao, Z. (2011). A semiparametric panel model for unbalanced data with application to climate change in the United Kingdom. *Journal of Econometrics*, 164(1):92–115.

Baillie, R. T., Bollerslev, T., and Mikkelsen, H. O. (1996). Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 74(1):3–30.

Barndorff-Nielsen, O. E. (2002). 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):253–280.

Bauwens, L. and Laurent, S. (2005). A new class of multivariate skew densities, with application to generalized autoregressive conditional heteroscedasticity models. *Journal of Business and Economic Statistics*, 23(3):346–354.

Bauwens, L., Laurent, S., and Rombouts, J. V. (2006). Multivariate GARCH models: A survey. *Journal of Applied Econometrics*, 21(1):79–109.

Berkowitz, J., Christoffersen, P., and Pelletier, D. (2011). Evaluating value-at-risk models with desk-level data. *Management Science*, 57(12):2213–2227.

Bertrand, J.-L., Brusset, X., and Fortin, M. (2015). Assessing and hedging the cost of unseasonal weather: Case of the apparel sector. *European Journal of Operational Research*, 244(1):261–276.

Blair, B. J., Poon, S.-H., and Taylor, S. J. (2001). Forecasting S&P 100 volatility: the incremental information content of implied volatilities and high-frequency index returns. *Journal of Econometrics*, 105(1):5–26.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327.

Bollerslev, T. (1987). A conditionally heteroskedastic time series model for speculative prices and rates of return. *Review of Economics and Statistics*, 69(3):542–547.

Bollerslev, T., Engle, R. F., and Wooldridge, J. M. (1988). A capital asset pricing model with time-varying covariances. *Journal of Political Economy*, 96(1):116–131.

Bollerslev, T., Patton, A. J., and Quaedvlieg, R. (2016). Exploiting the errors: A simple approach for improved volatility forecasting. *Journal of Econometrics*, 192(1):1–18.

Brandt, M. W. and Jones, C. S. (2006). Volatility forecasting with range-based EGARCH models. *Journal of Business & Economic Statistics*, 24(4):470–486.

Brooks, C. (2014). *Introductory Econometrics for Finance*. Cambridge University Press.

Brownlees, C. T. and Gallo, G. M. (2010). Comparison of volatility measures: A risk management perspective. *Journal of Financial Econometrics*, 8(1):29–56.

Byrd, R. H., Hribar, M. E., and Nocedal, J. (1999). An interior point algorithm for large-scale nonlinear programming. *SIAM Journal on Optimization*, 9(4):877–900.

Campbell, S. D. and Diebold, F. X. (2005). Weather forecasting for weather derivatives. *Journal of the American Statistical Association*, 100(469):6–16.

Chen, C. W., Gerlach, R., Hwang, B. B., and McAleer, M. (2012). Forecasting value-at-risk using nonlinear regression quantiles and the intra-day range. *International Journal of Forecasting*, 28(3):557–574.

Clements, M. P., Galvão, A. B., and Kim, J. H. (2008). Quantile forecasts of daily exchange rate returns from forecasts of realized volatility. *Journal of Empirical Finance*, 15(4):729–750.

Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, 7(2):174–196.

Corsi, F., Mittnik, S., Pigorsch, C., and Pigorsch, U. (2008). The volatility of realized volatility. *Econometric Reviews*, 27(1-3):46–78.

Diebold, F. X., Gunther, T. A., and Tay, A. S. (1998). Evaluating density forecasts with applications to financial risk management. *International Economic Review*, 39(4):863–883.

Diebold, F. X. and Mariano, R. S. (2002). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 20(1):134–144.

Dupuis, D. J. (2012). Modeling waves of extreme temperature: the changing tails of four cities. *Journal of the American Statistical Association*, 107(497):24–39.

Dupuis, D. J. (2014). A model for nighttime minimum temperatures. *Journal of Climate*, 27(19):7207–7229.

Dupuis, D. J. (2017). Assessing the role of hourly changes in the occurrence of daily extreme temperatures. *Journal of Climate*, 30(20):8393–8403.

Elias, R., Wahab, M., and Fang, L. (2014). A comparison of regime-switching temperature modeling approaches for applications in weather derivatives. *European Journal of Operational Research*, 232(3):549–560.

Embrechts, P., Puccetti, G., Rüschendorf, L., Wang, R., and Beleraj, A. (2014). An academic response to basel 3.5. *Risks*, 2(1):25–48.

Engle, R. (2002). Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business and Economic Statistics*, 20(3):339–350.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica: Journal of the Econometric Society*, 50:987–1007.

Engle, R. F. and Kroner, K. F. (1995). Multivariate simultaneous generalized ARCH. *Econometric Theory*, 11(01):122–150.

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

Erhardt, T. M., Czado, C., and Schepsmeier, U. (2015). R-vine models for spatial time series with an application to daily mean temperature. *Biometrics*, 71(2):323–332.

Fildes, R. and Kourentzes, N. (2011). Validation and forecasting accuracy in models of climate change. *International Journal of Forecasting*, 27(4):968–995.

Fissler, T., Ziegel, J., and Gneiting, T. (2016). Expected shortfall is jointly elicitable with value at risk - implications for backtesting. *Risk*, Janurary:58–61.

Fissler, T. and Ziegel, J. F. (2016). Higher order elicitability and Osband's principle. *The Annals of Statistics*, 44(4):1680–1707.

Forster, P. M. d. F. and Solomon, S. (2003). Observations of a 'weekend effect' in diurnal temperature range. *Proceedings of the National Academy of Sciences*, 100(20):11225–11230.

Franses, P. H., Neele, J., and van Dijk, D. (2001). Modeling asymmetric volatility in weekly Dutch temperature data. *Environmental Modelling and Software*, 16(2):131–137.

Fuertes, A.-M. and Olmo, J. (2013). Optimally harnessing inter-day and intra-day information for daily value-at-risk prediction. *International Journal of Forecasting*, 29(1):28–42.

Garthwaite, P. H., Kadane, J. B., and O'Hagan, A. (2005). Statistical methods for eliciting probability distributions. *Journal of the American Statistical Association*, 100(470):680–701.

Gerlach, R. and Wang, C. (2016). Forecasting risk via realized GARCH, incorporating the realized range. *Quantitative Finance*, 16(4):501–511.

Gerlach, R. H., Chen, C. W., and Chan, N. Y. (2012). Bayesian time-varying quantile forecasting for value-at-risk in financial markets. *Journal of Business & Economic Statistics*, 9(4):481–492.

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

Gneiting, T. and Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477):359–378.

Gneiting, T. and Ranjan, R. (2011). Comparing density forecasts using threshold- and quantile-weighted scoring rules. *Journal of Business and Economic Statistics*, 29(3):411–422.

Gülpınar, N. and Çanakoğlu, E. (2017). Robust portfolio selection problem under temperature uncertainty. *European Journal of Operational Research*, 256(2):500–523.

Hajat, S., Armstrong, B., Baccini, M., Biggeri, A., Bisanti, L., Russo, A., Paldy, A., Menne, B., and Kosatsky, T. (2006). Impact of high temperatures on mortality: Is there an added heat wave effect? *Epidemiology*, 17(6):632–638.

Han, H. and Kristensen, D. (2014). Asymptotic theory for the QMLE in GARCH-X models with stationary and nonstationary covariates. *Journal of Business & Economic Statistics*, 32(3):416–429.

Hansen, P. R., Huang, Z., and Shek, H. H. (2012). Realized GARCH: A joint model for returns and realized measures of volatility. *Journal of Applied Econometrics*, 27(6):877–906.

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

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

Heber, G., Lunde, A., Shephard, N., and Sheppard, K. (2009). *Oxford-Man Institute's Realized Library, Version 0.2*. Oxford-Man Institute, University of Oxford.

Hounyo, U., Gonçalves, S., and Meddahi, N. (2017). Bootstrapping pre-averaged realized volatility under market microstructure noise. *Econometric Theory*, 33(4):791–838.

Hua, J. and Manzan, S. (2013). Forecasting the return distribution using high-frequency volatility measures. *Journal of Banking & Finance*, 37(11):4381–4403.

Jeon, J. and Taylor, J. W. (2013). Using CAViaR models with implied volatility for value-at-risk estimation. *Journal of Forecasting*, 32(1):62–74.

Keellings, D. and Waylen, P. (2014). Increased risk of heat waves in Florida: Characterizing changes in bivariate heat wave risk using extreme value analysis. *Applied Geography*, 46:90–97.

Kolev, N., Anjos, U. d., and Mendes, B. V. d. M. (2006). Copulas: a review and recent developments. *Stochastic Models*, 22(4):617–660.

Komunjer, I. and Vuong, Q. (2010). Semiparametric efficiency bound in time-series models for conditional quantiles. *Econometric Theory*, 26(2):383–405.

Kou, S. G. (2002). A jump-diffusion model for option pricing. *Management Science*, 48(8):1086–1101.

Kumar, D. and Maheswaran, S. (2015). Long memory in Indian exchange rates: an application of power-law scaling analysis. *Macroeconomics and Finance in Emerging Market Economies*, 8(1-2):90–107.

Le Tertre, A., Lefranc, A., Eilstein, D., Declercq, C., Medina, S., Blanchard, M., Chardon, B., Fabre, P., Filleul, L., Jusot, J.-F., et al. (2006). Impact of the 2003 heatwave on all-cause mortality in 9 French cities. *Epidemiology*, 17(1):75–79.

Lee, T.-H. and Long, X. (2009). Copula-based multivariate GARCH model with uncorrelated dependent errors. *Journal of Econometrics*, 150(2):207–218.

Linsmeier, T. J. and Pearson, N. D. (1997). Quantitative disclosures of market risk in the SEC release. *Accounting Horizons*, 11(1):107.

Maheu, J. M. and McCurdy, T. H. (2011). Do high-frequency measures of volatility improve forecasts of return distributions? *Journal of Econometrics*, 160(1):69–76.

Manganelli, S. and Engle, R. F. (2004). Value at risk models in finance. In Szegö, G., editor, *Risk Measures for the 21st Century*, pages 123–143. Wiley, Chichester, UK.

Martens, M., Van Dijk, D., and De Pooter, M. (2009). Forecasting S&P 500 volatility: Long memory, level shifts, leverage effects, day-of-the-week seasonality, and macroeconomic announcements. *International Journal of Forecasting*, 25(2):282–303.

McNeil, A. J. (1999). Extreme value theory for risk managers. *Departement Mathematik ETH Zentrum*.

McNeil, A. J. and Frey, R. (2000). Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. *Journal of Empirical Finance*, 7(3):271–300.

Meehl, G. A. and Tebaldi, C. (2004). More intense, more frequent, and longer lasting heat waves in the 21st century. *Science*, 305(5686):994–997.

Mörters, P. and Peres, Y. (2010). *Brownian Motion*. Cambridge University Press.

Nieto, M. R. and Ruiz, E. (2016). Frontiers in VaR forecasting and backtesting. *International Journal of Forecasting*, 32(2):475–501.

Nolde, N. and Ziegel, J. F. (2017). Elicitability and backtesting: Perspectives for banking regulation. *Annals of Applied Statistics*, forthcoming.

Oh, D. H. and Patton, A. J. (2017). Modeling dependence in high dimensions with factor copulas. *Journal of Business & Economic Statistics*, 35(1):139–154.

Parkinson, M. (1980). The extreme value method for estimating the variance of the rate of return. *Journal of Business*, 53(1):61–65.

Pattenden, S., Nikiforov, B., and Armstrong, B. (2003). Mortality and temperature in Sofia and London. *Journal of Epidemiology & Community Health*, 57(8):628–633.

Patton, A. J. (2006). Modelling asymmetric exchange rate dependence. *International Economic Review*, 47(2):527–556.

Patton, A. J. (2012). A review of copula models for economic time series. *Journal of Multivariate Analysis*, 110:4–18.

Patton, A. J. and Sheppard, K. (2009). Evaluating volatility and correlation forecasts. *Handbook of Financial Time Series*, pages 801–838.

Patton, A. J., Ziegel, J. F., and Chen, R. (2017). Dynamic semiparametric models for expected shortfall (and value-at-risk). *arXiv preprint arXiv:1707.05108*.

Perkins, S. and Alexander, L. (2013). On the measurement of heat waves. *Journal of Climate*, 26(13):4500–4517.

Pritsker, M. (2006). The hidden dangers of historical simulation. *Journal of Banking & Finance*, 30(2):561–582.

Proietti, T. and Hillebrand, E. (2017). Seasonal changes in central England temperatures. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 180(3):769–791.

Qu, M., Wan, J., and Hao, X. (2014). Analysis of diurnal air temperature range change in the continental United States. *Weather and Climate Extremes*, 4:86–95.

Righi, M. B. and Ceretta, P. S. (2015). A comparison of expected shortfall estimation models. *Journal of Economics and Business*, 78:14–47.

Rogers, L. C. and Zhou, F. (2008). Estimating correlation from high, low, opening and closing prices. *Annals of Applied Probability*, 18(2):813–823.

Ross, S., Christensen, M., Drew, M. E., Thompson, S., Westerfield, R., and Jordan, B. (2011). *Fundamentals of Corporate Finance*. McGraw Hill Higher Education.

Schefzik, R., Thorarinsdottir, T. L., Gneiting, T., et al. (2013). Uncertainty quantification in complex simulation models using ensemble copula coupling. *Statistical Science*, 28(4):616–640.

Sheather, S. J. et al. (2004). Density estimation. *Statistical Science*, 19(4):588–597.

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

Taylor, J. W. (2008). Using exponentially weighted quantile regression to estimate value at risk and expected shortfall. *Journal of Financial Econometrics*, 6(3):382–406.

Taylor, J. W. (2017a). Forecasting value at risk and expected shortfall using a semiparametric approach based on the asymmetric laplace distribution. *Journal of Business & Economic Statistics*, forthcoming.

Taylor, J. W. (2017b). Probabilistic forecasting of wind power ramp events using autoregressive logit models. *European Journal of Operational Research*, 259(2):703–712.

Taylor, J. W. and Buizza, R. (2004). A comparison of temperature density forecasts from GARCH and atmospheric models. *Journal of Forecasting*, 23(5):337–355.

Taylor, J. W. and Jeon, J. (2018). Probabilistic forecasting of wave height for offshore wind turbine maintenance. *European Journal of Operational Research*, forthcoming.

Tol, R. S. (1996). Autoregressive conditional heteroscedasticity in daily temperature measurements. *Environmetrics*, 7(1):67–75.

Tsiakas, I. (2008). Overnight information and stochastic volatility: A study of European and US stock exchanges. *Journal of Banking & Finance*, 32(2):251–268.

Waltz, R. A., Morales, J. L., Nocedal, J., and Orban, D. (2006). An interior algorithm for nonlinear optimization that combines line search and trust region steps. *Mathematical Programming*, 107(3):391–408.

Wang, M.-z., Zheng, S., He, S.-l., Li, B., Teng, H.-j., Wang, S.-g., Yin, L., Shang, K.-z., and Li, T.-s. (2013). The association between diurnal temperature range and emergency room admissions for cardiovascular, respiratory, digestive and genitourinary disease among the elderly: a time series study. *Science of The Total Environment*, 456:370–375.

Wang, X., Wu, C., and Xu, W. (2015). Volatility forecasting: The role of lunch-break returns, overnight returns, trading volume and leverage effects. *International Journal of Forecasting*, 31(3):609–619.

Wong, T. (2015). Statistical analysis of heat waves in the state of Victoria in Australia. *Australian and New Zealand Journal of Statistics*, 57(4):463–480.

Zhu, D. and Galbraith, J. W. (2010). A generalized asymmetric student-t distribution with application to financial econometrics. *Journal of Econometrics*, 157(2):297–305.

Žikeš, F. and Baruník, J. (2015). Semi-parametric conditional quantile models for financial returns and realized volatility. *Journal of Financial Econometrics*, 14(1):185–226.