CHARLES UNIVERISTY IN PRAGUE

FACULTY OF SOCIAL SCIENCES

Institute of Economic Studies

Daniel Benčík

Range-based Volatility Estimation and Forecasting

Master Thesis

Prague 2012

Author: **Daniel Benčík**

Supervisor: **PhDr. Jozef Baruník, Ph.D.**

Planned thesis defense: **2012**

Bibliographical record

BENČÍK, Daniel. Range-based Volatility Estimation and Forecasting. Prague, 2012. 69 pages. Master Thesis (M.A.) Charles University, Faculty of Social Sciences, Institute of Economic Studies. Macroeconomics and Econometrics Department.

Thesis supervisor PhDr. Jozef Baruník, Ph.D.

Abstrakt

Tato diplomová práce analyzuje nové možnosti v předpovídání denního rozpětí cen (tj. rozdílu nejvyšší a nejnižší denní ceny instrumentu). Hlavním zaměřením naší práce je zkoumání možných zlepšení stávajících modelů používaných pro modelování denního rozpětí. Jmenovitě zkoumáme přínos použití eficientnějších odhadů denní volatility jakožto prediktorů denního rozpětí. Konkrétní odhady volatility zkoumané v této práci zahrnují range-based estimátory (Parkinson, Garman & Klass, Rogers & Satchell, atd.) a realizované míry denní variance (realizovaná variance, realizované rozpětí). Součástí těchto výzkumů je i empirické porovnání eficience jednotlivých range-based estimátorů denní volatility.

Dalším směrem výzkumu naší práce je analýza přínosů rozdělení obchodního dne do obchodních session na základě aktivity různých obchodních center (např. asijská, evropská, americká session). V tomto ohledu analyzujeme, zda odhady volatility získané z celodenních dat spolehlivě agregují informace pocházející z různých session. Naší intuicí je, že různé obchodní session přináší odlišné informace díky odlišné hloubce trhu. Předpokládáme, že jednotlivé session poskytují užitečné informace, které jsou v agregované míře denní volatility skryté (nevyužitelné).

Dále zkoumáme možnost průběžných aktualizací předpovědí denní volatility pomocí intraday informací dostupných v daném momentě. Konkrétně to znamená, že jakmile obchodní session skončí, míry její volatility a obchodní aktivity jsou zahrnuty do stávajícího modelu pro předpověď dnešní volatility. Tyto průběžně aktualizované předpovědi vykazují významné přínosy týkající se kvality předpovědi. Z toho vyplývá, že intraday obchodníci aktivní v pozdějších hodinách obchodování mají významnou výhodu oproti obchodníkům aktivním na začátku obchodního dne.

Modely uvažované v této práci zahrnují HAR, CARR a modely založené na kointegraci nejvyšší a nejnižší denní ceny. Modely podávající solidní výkon při in-sample modelování jsou porovnány pomocí out-of-sample předpovědí. Na rozdíl od výsledků publikovaných v literatuře, modely využívající kointegračního vztahu nejvyšší a nejnižší denní ceny podávají predikce špatné kvality. Nejlepším modelem pro modelování denních rozpětí se v naší práci ukázal HAR model využívající realizované rozpětí jako prediktor volatility v kombinaci s GARCH komponentou pro modelování volatility denních rozpětí.

Abstract

In this thesis, we analyze new possibilities in predicting daily ranges, i.e. the differences between daily high and low prices. The main focus of our work lies in investigating how models commonly used for daily ranges modeling can be enhanced to provide better forecasts. In this respect, we explore the added benefit of using more efficient volatility measures as predictors of daily ranges. Volatility measures considered in this work include realized measures of variance (realized range, realized variance) and range-based volatility measures (Parkinson, Garman & Klass, Rogers & Satchell, etc). As a subtask, we empirically assess efficiency gains in volatility estimation when using range-based estimators as opposed to simple daily ranges. As another venue of research in this work, we analyze the added benefit of slicing the trading day into different sessions based on trading activity (e.g. Asian, European and American session). In this setting we analyze whether whole-day volatility measures reliably aggregate information coming from all trading sessions. We are led by intuition that different sessions exhibit significantly different characteristics due to different order book thicknesses and trading activity in general. Thus these sessions are expected to provide valuable information concealed in the aggregate volatility measure.

Next, we investigate the possibility to gradually update daily volatility forecasts by incorporating all up-to-date information. That means once a trading sessions ends its volatility and trading activity measures are used for updating the current day's volatility forecast. These updated forecasts exhibit very strong gains in terms of goodness-of-fit and thus short-term traders active in later sessions of the day can gain a significant advantage over traders active early in the day.

The array of models within which we investigate the aforementioned effects include the heterogeneous autoregressive model, conditional autoregressive ranges model and a vector error-correction model of daily highs and lows. Models performing well in terms of in-sample fit are challenged on out-of-sample, one-day-ahead forecasting. Contrary to results presented in literature, models based on co-integration of daily highs and lows fail to produce good quality forecasts. When one strives for the best one-day-ahead daily ranges forecasts a HAR model using realized ranges as predictors with a GARCH volatility-of-volatility component is the preferred option.

Klíčová slova

volatilita, výnosy, futurita, dlouhá paměť, kointegrace, predikce

Keywords

volatility, returns, futures contracts, long memory, cointegration, prediction

Pages: 138000 characters of the main text including spaces

Prohlášení

1. Prohlašuji, že jsem předkládanou práci zpracoval/a samostatně a použil/a jen uvedené prameny a literaturu.
2. Prohlašuji, že práce nebyla využita k získání jiného titulu.
3. Souhlasím s tím, aby práce byla zpřístupněna pro studijní a výzkumné účely.

V Praze dne … Daniel Benčík

Declaration

1. I hereby declare that I have compiled this master thesis independently using only the listed literature and resources.
2. I hereby declare that this master thesis was not used for acquiring a degree different from that at the Institute of Economic Studies.
3. I hereby consent for this thesis to be made available for study and research purposes.

Prague, ... Daniel Benčík

Acknowledgements

I would like to express deep gratitude to my supervisor PhDr. Jozef Baruník, Ph.D. for his patience and continuous support. His willingness to help and continuous encouragement made the writing of this thesis a pleasant experience. My thanks also belong to Allin Cottrell and Riccardo Lucchetti for their help with gretl scripting. Lastly, I wish to express gratitude to my parents for supporting me in my university studies.

Master Thesis Proposal

Institute of Economic Studies
Faculty of Social Sciences
Charles University in Prague

Author:	**Bc. et Bc. Daniel Benčík**	Supervisor:	PhDr. Jozef Baruník
E-mail:	eubie@seznam.cz	E-mail:	barunik@utia.cas.cz
Phone:	777 83 11 42	Phone:	-
Specialization:	FFTaB	Defense Planned:	June 2012

Proposed Topic:

Predicting a range-based volatility measure for futures markets

Topic Characteristics:

The difference between a day's highest and lowest price, i.e. the daily range, is an estimator of volatility that is shown by research to be of same usefulness as other usually considered measures of volatility (realized variance, standard deviation of prices and others). Some research even suggests the superiority of range-based volatility measures under some circumstances. Despite these, research on either explaining or predicting the values of range-based indicators is very limited as other volatility measures are more popular.

The focus of my work will be an investigation of the possibility to explain/predict one range-based volatility estimator of American stock index futures instruments, namely the daily range. Specific data used in my thesis will be tick data of either EUR/USD Forex or Light Sweet Crude Oil continuous futures contract. The analysis will be carried out on more instruments, if data is available. Envisaged modeling techniques are expected to range from classical linear models through non-linear models as well as models incorporating various trading-related variables that are not directly linked to volatility. The main questions to be addressed are whether there are more possibilities to model daily ranges than just the simple ARIMA-type framework and whether measures of volatility different from range-based ones can help explain/predict these, i.e. whether different measures of volatility are complementary as one would expect.

Hypotheses:

1. Exploitable linkages between various volatility measures (realized variance, standard deviation) that can be used in daily ranges prediction.
2. Exploitability of implied volatility for daily ranges predictions
3. Exploitability of various daily range characteristics (gamma distribution, long memory, co-integration of daily high and low prices)
4. Existence and exploitability of non-linear dependencies in daily ranges

Methodology:

As economic models pertaining to volatility are generally lacking, the empirical approach is the only considerable one for my thesis. Envisaged methodology will be first adopted from existing research. As different research papers deal with different instruments (e.g. mid-cap versus large-cap stock indices), the results of the same model applied on different instruments can differ substantially. For the investigation of clustering typical for financial markets volatility, GARCH-type models will be employed. As there are several variables possibly influencing range-based estimators, regressive models will be used at first to distinguish between those with and without explanatory power. Flowingly, non-linear models (e.g. TAR, STAR) will be used on regressors with proven explanatory power. The usefulness of Bayesian econometric models will be tested when assessing the exploitability of the fact that daily ranges seems to follow a gamma distribution. Finally, ARFIMA modeling will be used to infer the usefulness of long-memory property of daily ranges.

Outline:

1. Introduction, Literature overview
   a. Measures of volatility
   b. Existing research on relationships of different measures of volatility
   c. Existing research on modeling daily ranges
2. Methodology
   a. Basic analysis of the nature of daily ranges and its implications
   b. Investigation of possible determinants of daily ranges
3. Empirical part
   a. Data description
   b. Regression Model
   c. Discussion of the results, hints for future research

Core Bibliography:

Corrado, C., Truong, C.(2004): Forecasting Stock Index Volatility: The Incremental Information in the Intraday High-Low Price Range, November 2004

Vipul, J. J. (2008): Estimation of Forecasting of Stock Volatility With Range-Based Estimators, The Journal of Futures Markets, Vol. 28, 2008

Cheung, Y. W. (2007): An Empirical Model of Daily Highs and Lows, International Journal of Finance and Economics, 12, 1-20

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

Ghysels, E., Santa-Clara, P., Valkanov, R. (2006) Predicting Volatility: getting the most out of return data sampled at different frequencies, Journal of Econometrics 131, 2006, 59-95

Author

Supervisor

Contents

Introduction

Volatility of asset prices plays a critical role in finance. Value-at-risk estimation, option pricing and other activities common in the financial industry rely on its correct prediction. Recently, a huge body of research focused on volatility emerged. Nobel Prize awarded ARCH type models and their generalizations are examples of such advances (Engle 1982, Bollerslev 1986).

Since the true price generating process of real-world financial data is unknown, researchers and practitioners can merely estimate the characteristics of this process, including its volatility. In this respect, most research focuses on modeling and forecasting of returns' standard deviation, as standard deviation is the most popular volatility measure. However, other volatility measures exist and can be more useful, especially for short-term investment.

The work presented in this thesis views volatility modeling from the standpoint of a short-term investor or speculator whose investment horizon does not exceed one trading day. A crucial question for such an investor is how large a move is to be expected once a position is open. For this purpose, predictions of different volatility measures provide different levels of usefulness. An above-average standard deviation prediction indicates higher volatility, however it is difficult to assess the exact extent of future price movement, as there is no clear connection between standard deviation and ranges (differences between highest and lowest daily prices). A proper prediction of the day's range is, however, helpful as it can be directly translated into profit targets, stop losses, etc., and thus can be used for the management of an open position.

Contributions of this thesis to the existing body of volatility related literature are numerous. Firstly, we focus on predicting daily ranges using daily ranges themselves (Chou 2005) as well as different measures of volatility as predictors. While exploiting linkages between different volatility measures has already been published in some papers (Engle & Gallo 2003) neither of these papers focus specifically on daily ranges prediction. On top of that, in the existing literature linkages between volatility measures of comparable efficiency are discussed. In this work, however, we investigate linkages between volatility measures with sharply different efficiencies. More specifically, we investigate how well we can forecast precise volatility measures by noisy ones (daily ranges) and vice versa. This gives us some information on the cost-benefit tradeoff of obtaining pricey intraday data versus daily data obtainable free of charge.

Further contributions include measuring empirical efficiency gains of different range-based volatility estimators. In the second half of the 20th century several range-based volatility estimators were proposed, with each new estimator either improving efficiency or relaxing crucial assumptions of the previous ones (Parkinson 1980, Garman & Klass 1980, Rogers & Satchell 1990). In our work, we empirically test whether efficiency gains reported in theory are observable in practice. Had there been significant efficiency gains in daily volatility estimation for one specific range-based estimator, its use as a regressor for predicting daily ranges would be indicated. This stems from the intuitive idea that more precise measurement of volatility should act as better regressors/predictors.

Next, we investigate the role of different trading sessions, i.e. periods of a trading day defined by geographical location of traders predominantly active in the market (e.g. Asian, European and American sessions). Effects investigated are, for example, whether volatility measured on just the main sessions (main in terms of trading activity) provides a better measure of volatility than volatility measured on the whole trading day. If the former were the case, trading during less active sessions would be introducing noise into volatility measuring and better volatility estimates could be obtained by focusing solely on the most active sessions. Also, we investigate whether it pays off to slice up a day into trading sessions and then predict daily volatilities using volatilities and trading intensity variables measured on these separate sessions. As different trading sessions are characterized by different order book thicknesses and different traded volumes (resulting in different volatility-volume relationships), each session provides unique information. Since daily volatility/trading activity measures conceal this diversity across sessions, it is possible that models employing volatility/trading activity measured on separate lagged sessions as opposed to whole lagged days might bring gains in terms of goodness of fit. Next, using current day's session data, we create gradually updated daily volatility forecasts. As a session finishes, its volatility/trading activity measures are inserted into a model and an updated forecast is generated. Cumulatively adding sessions as they finish during the day provides for an opportunity to significantly increase forecasting performance of all models presented in this work. As the idea of updating daily volatility forecasts throughout the day is not present in current academic literature, it is one of the strongest contributions of our work.

Lastly, we combine the results of our investigations described above with several existing models used for daily volatility modeling/forecasting. The work finishes by picking the best models for daily ranges prediction based on out-of-sample forecasting performance. Out results provide new insights into volatility forecasting and better forecasts of daily ranges as well as other volatility measures are thus made possible.

The work is organized as follows. In Section 1, we provide a reader with our motivation for daily ranges modeling. Section 2 describes out dataset, while Section 3 continues with initial data analysis as well as an extensive correlation analysis. In Section 4, we compare various daily variance estimators in terms of efficiency and usefulness for daily ranges prediction. In Section 5, we empirically investigate our hypotheses on three different models designed for daily ranges prediction. The best models are then compared in an out-of-sample forecasting exercise in Section 6. A description of our findings concludes this thesis.

1. Motivation

In this section, we provide motivation for our interest in predicting daily ranges. For clarity of explanations, we firstly turn to technicalities related to naming conventions, as existing literature is not united in the matter of notations. Let us denote the price of a financial asset measured at time $0 \le t \le T$ on day D as $P_{t,D}$. Then let us assume that log-price

$$p_{t,D} = \log_e\left(P_{t,D}\right) \tag{1}$$

evolves according to a diffusion process

$$dp_{t,D} = \mu_{t,D}dt + \sigma_D dW_{t,D} + c_D dJ_{t,D} \tag{2}$$

where $\mu_{t,D}, \sigma_D, c_D$ correspond to the drift, volatility and jump terms and $W_{t,D}, J_{t,D}$ are Wiener and constant-intensity Poisson processes[1]. The daily price range is defined as

$$R_D = \sup_{0\le t\le T} P_{t,D} - \inf_{0\le t\le T} P_{t,D} \tag{3}$$

the daily log-range is defined as

$$R_D^{\log} = \sup_{0\le t\le T} p_{t,D} - \inf_{0\le t\le T} p_{t,D} \tag{4}$$

and the daily log-return is denoted

$$r_D = p_{T,D} - p_{T,D-1} \tag{5}$$

[1] Taking the variance and jump terms constant for the whole day is misleading in practice, as volatility exhibits intraday seasonality and jumps occur mostly around news announcements which are not distributed evenly in time. However, in the above context we chose to use this notation for simplifying the introduction.

Lastly, let us denote high, low, open and close prices observed during the day

$$h_D = \sup_{0 \le t \le T} p_{t,D} \qquad l_D = \inf_{0 \le t \le T} p_{t,D} \tag{6}$$

$$o_D = p_{f,D} \qquad c_D = p_{T,D} \tag{7}$$

where f is a portion of a trading day during which trading activity is minimal, i.e. in practice corresponds to postmarket of previous day combined with the premarket of the current day. From this explanation it is clear that T does not necessarily represent the end of day. Instead, it represents a time at which trading activity halts (most commonly caused by exchange closure).

Before continuing to describe out motivation for daily ranges prediction, we give a short introduction into the matter of range-based volatility estimation, as this will facilitate easier explanation. As stated before, volatility of any real-world price process is unknown and thus we can only rely on its estimates. Fortunately, many possibilities exist in estimating volatility. Volatility estimates differ in their data intensiveness (using intraday versus end-of-day data), efficiency (precision of volatility estimation) and, surprisingly, popularity. Another factor that can serve as a distinguishing factor among volatility estimates is the length of time such a measure needs to produce an estimate.

For example, the most common measure of volatility, the standard deviation of returns, cannot be calculated from one daily return only. This inevitably leads to only being able to measure and forecast average volatility and commonly known features of volatility can be made less distinct or even disappear (e.g. the well-documented property of volatility clustering where periods with high levels of volatility are followed by periods of high volatility and vice versa). On the other hand, there are a number of estimates which are capable of estimating volatility for just one day using data provided by that day's trading. With such estimates of volatility it is possible to exploit any volatility related feature for forecasting. Two of the most popular measures used for the inference of most recent (one-day) variance, i.e. of the squared diffusion coefficient σ_D, are the squared daily return

$$r_D^S = r_D^{\ 2} \tag{8}$$

and the absolute daily return

$$r_D^A = |r_D| \tag{9}$$

However, already Parkinson (1980) showed that under the assumption of $\mu_{t,D} = 0$, we can estimate σ_D^2 by

$$\left(\hat{\sigma}_D^{Park}\right)^2 = [4\ln(2)]^{-1}\left(R_D^{\log}\right)^2 \tag{10}$$

and achieve approximately five times higher efficiency of variance estimation compared to squared daily returns. The efficiency gain can be intuitively attributed to the fact that an estimate which incorporates extreme price values takes into account the whole day's evolution of price while estimates based solely on close prices only utilize prices measured at one predetermined point during each day.

Still keeping the assumption of $\mu_{t,D} = 0$ and following this idea further, Garman & Klass (1980) suggest an estimator

$$\left(\hat{\sigma}_D^{GK}\right)^2 = 0.12\frac{[o_D - c_{D-1}]}{f} + 0.78\frac{0.5(h_D - l_D)^2 - (2\ln(2)-1)[c_D - o_D]}{1-f} \tag{11}$$

and claim that the efficiency gain compared to (8) is approximately 7.4 regardless of f.

The disadvantage of aforementioned range-based estimates of daily volatility is the restrictive assumption of zero drift. As $\hat{\sigma}_D^{GK}, \hat{\sigma}_D^{Park}$ become biased with $\mu_{t,D} \neq 0$, Rogers & Satchell (1990) relax this assumption and propose

$$\left(\hat{\sigma}_D^{RS}\right)^2 = (l_D - o_D)(l_D - c_D) + (h_D - o_D)(h_D - c_D) \tag{12}$$

which has only slightly lower efficiency compared to $\hat{\sigma}_D^{GK}$.

A recent work of Yang & Zhang (2000) provides a drift-independent estimator which allows for the presence of jumps occurring during exchange opening. Denoting n the number of days used for the estimate, we have

$$\left(\hat{\sigma}_D^{YZ}\right)^2 = V_{O,D} + kV_{C,D} + (1-k)V_{RS,D}^n$$

$$V_{RS,D}^n = n^{-1}\sum_{i=0}^{n}\left(l_{D-i} - o_{D-i}\right)\left(l_{D-i} - c_{D-i}\right) + \left(h_{D-i} - o_{D-i}\right)\left(h_{D-i} - c_{D-i}\right)$$

$$V_{o,D} = (n-1)^{-1}\sum_{i=0}^{n}\left(o_{D-i} - \bar{o}_D\right)^2 \quad V_{c,D} = (n-1)^{-1}\sum_{i=0}^{n}\left(c_{D-i} - \bar{c}_D\right)^2 \tag{13}$$

$$\bar{c}_D = n^{-1}\sum_{i=0}^{n} c_{D-i} \qquad \bar{o}_D = n^{-1}\sum_{i=0}^{n} o_{D-i}$$

To obtain minimum variance of $\hat{\sigma}_D^{YZ}$ set

$$k = \frac{0.34}{1.34 + \frac{n+1}{n-1}} \tag{14}$$

Lastly, Brunetti & Lindholdt (2002) show that the unbiased estimator of σ_D given by

$$\hat{\sigma}_D^R = \sqrt{\frac{\pi}{8}} R_D^{\log} \tag{15}$$

is approximately 6.5 times more efficient that the unbiased estimator

$$\hat{\sigma}_D^A = \sqrt{\frac{\pi}{2}} r_D^A \tag{16}$$

Hence, the inclusion of extreme prices into variance estimates is capable of producing significant efficiency gains which is of vital importance for all applications relying on volatility. At the same time, range-based estimators, i.e. (10) - (16), do not require tick by tick data needed for the construction of finely spaced intraday returns.

In spite of high efficiency of several aforementioned volatility measures over $R_D^{\log}$ the daily range offers a unique property which is our main motivation for its prediction. While for long-term investment or option pricing a correct assessment of σ_D is crucial, short-term investors/day-traders are more likely to benefit from a precise prediction of $R_D^{\log}$ itself rather than σ_D. To illustrate this point, let an imaginary agent open a position (by assumption in the direction coinciding with the future market direction) and let us investigate the best strategy for exiting such a position. Having a perfect prediction of the day's range allows the agent to set a reasonable value of profit target, as daily range relates directly to the extent of price movement[2]. Other volatility estimates, including the previously defined range-based ones, cannot be used in such a manner. Thus, we focus solely on daily ranges prediction. Other range-based estimators were not defined in vain, however. Using the above defined range-based estimators, we will investigate whether it is possible to benefit from higher precision of past volatility measurement (by using $\hat{\sigma}_D^{GK}$ or $\hat{\sigma}_D^{RS}$) for creating better daily range forecasts. Intuitively, forecasts produced by less noisy predictors should be superior.

Even though our main interest in predicting daily ranges stems from the desire for better money management in high-frequency trading, there is also a more general reason for which we should care about daily ranges prediction. Precise predictions of daily ranges can be useful in predicting other measures of volatility, which has been demonstrated by several authors. For example, Engle & Gallo (2003) assume a multivariate MEM-GARCH process of daily ranges, daily realized volatilities and absolute daily returns. Estimation results show a significant level of interaction between these three volatility measures. By enriching the usual MEM-GARCH model for each measure by lagged values of the other measures, model forecasts match well those obtained from implied volatility indices. On a similar note, Corrado & Truong (2007) investigate the usefulness of adding squared daily ranges and implied volatility levels into a GJR-GARCH model for the variance of residual term and find both variables useful for improving forecast quality. Hence, having a precise expected value of the next day's range can be used per se as well as an input for the prediction of other volatility measures related to the next day.

To sum up, daily ranges have potential for practical uses in day-trading and they provide for improving forecasts of other volatility measures while placing minimal requirements on historical data. This explains our motivation.

---

[2] Herein, a profit target is a predetermined amount of price movement in one direction after which a position is exited as further market movements in the direction of trade are considered unlikely. The exact target setting (for a long position) involves subtracting the lowest price for the day from the entry price and then subtracting this difference from the predicted range, arriving at an exit point nearing the highest price for the day.

2. Data Description

Having described the motivation of our work, we now turn to familiarizing the reader with our dataset. Data used throughout this thesis relate to the EUR/USD Forex futures contract traded on Chicago Mercantile Exchange from Nov, 9 2007 to Nov, 9 2011. Between these dates contracts with several different delivery months were traded, namely deliveries from 12-07 to 12-11 (MM-YY). To allow for a study of the whole dataset at once without having to create a model separately for each delivery month, a continuous contract was created from available deliveries based on the maximum volume rule. Following this rule, data from a contract with specific delivery is taken into the continuous contract for a specific day if this delivery was the most heavily traded of all deliveries on that particular day. The list of cut-off dates[3] is provided in the Appendix as Table A.1. Despite bid-ask data were available, for the work presented in this thesis we decided to work with prices defined by traded prices and to neglect the effect of market microstructure, i.e. no bid-ask smoothing was employed. Our reasons for this choice were several fold. Firstly, there is no clear consensus in literature as to which method of ridding data of the bid-ask bounce is the best one. Secondly, the main aim of this work is to provide for a general assessment of new possibilities in daily ranges forecasting. Even though we acknowledge there is a measurement error in very precise volatility measures induced by the presence of the bid-ask bounce, its magnitude is hardly significant enough to bias our results to a strong degree.

Next, we discuss timing conventions used throughout this thesis. Connected to the nature of Forex futures contracts is the concept of Electronic Trading Hours (ETH). Forex futures are usually traded in several trading sessions depending on the activity of different trading centers (East Asia, Europe and America). For this reason trading sessions, as recorded by the Exchange, do not coincide with calendar dates. Instead, a trading session dated Oct 15th starts on Oct 14th at 17:00 CST[4] (start of East Asian session[5]) and ends on Oct, 15 at 16:00 (end of the U.S. trading session). In the period

---

[3] i.e. dates between which a given delivery contract was the most heavily traded of all delivery contracts and thus (for the given period) is included in the continuous contract.

[4] Central Standard Time, time zone of the Exchange location. Throughout this thesis, all quotations of time are expressed in CST.

[5] To prevent misleading the reader, let us note that Asian trading centers (Singapore, Shanghai, Tokyo) as well as Australian centers (Sydney Futures Exchange, Australian Stock Exchange) are included in the East Asian session.

between 16:00 and 17:00 no trading center is open and thus trading activity is minimal. In order to have the sessions of all financial centers within one day of trading, we followed the ETH standard.

The last point related to data description is connected to the fact that during some days of year, trading is halted. The reasons might be national holidays or Exchange imposed restrictions on trading. Missing observations for such trading-free dates were not reconstructed artificially. Also, on several occasions the Exchange accepts orders only during a certain part of the day or trading activity is generally lower than normally (typically around Christmas and New Year's Eve). As readings of traded volume and volatility from these days could distort our results, we decided to omit them[6]. After removing holidays and days with illiquidity present in the market, our data sample consisted of 999 trading days.

[6] Specifically, we omitted all dates between Dec 23rd and Jan 2nd from the continuous contract as well as any date for which total traded volume was less than 80,000 contracts.

3. Initial Data Analysis

In this section, we provide basic statistical analysis of data at hand. Our aim is to obtain preliminary insights into data behavior and the identification of patterns that could later be used for proper model construction, e.g. the distribution of error terms. Specific investigated features include the unconditional distribution of volatility coupled with time patterns present in volatility and trading activity measures (on a daily as well as intraday basis). Persistence of log-returns and log-ranges sampled at different frequencies is presented.

3.1 Definition of ranges

Before moving forward, let us briefly investigate the type of daily ranges that will be modeled throughout this thesis. The existing literature does not address whether R_D or $R_D^{\log}$ is the correct specification of range. The general preference of log-prices in quantitative finance coupled with higher computation precision obtained when using values close to zero speak in favor of $R_D^{\log}$. However, our previously stated money management "technique" depends on a prediction of R_D. Hence our first task is to determine whether predicting $R_D^{\log}$ is sufficient for our goal. For this purpose, we estimate a model

$$R_D = \alpha + \beta R_D^{\log} + \varepsilon_D \tag{17}$$

Estimation results involving daily data from the whole sample are listed in the Appendix as Table A.2. Heteroskedasticity is present and thus inference based on AdjR^2[7] should be taken into account only grossly. Treating heteroskedasticity via GARCH modeling could solve the problem only artificially, as we are estimating

$$H_D - L_D = \alpha + \beta \log\left(\frac{H_D}{L_D}\right) + \varepsilon_D \tag{18}$$

[7] Adjusted R-Square

and it is natural that the difference of two variables will not depend linearly on the logarithm of their ratio.

Taking into account $E(R_D) = 0.01684$ and the fact that more than 70% of residuals lie within 0 ± 0.001 (in other words, the AdjR^2 is very high despite possible bias), we chose $R_D^{\log}$ as the range variable since distortion is minimal[8].

3.2 Time Patterns

Having properly defined the daily range, in the following section we investigate day-of-the-week, hour-of-the-week and high-frequency intraday patterns. From these analyses we can infer hints that could help us in model design. For example, if significant differences in daily ranges are observed across days of week, enriching a model by dummies for separate days of the week might prove fruitful. Investigating intraday patterns of volatility and trading activity will enable us to determine the starting points of different trading sessions as well as visually assess the basic characteristics of the volume-volatility relationship.

3.2.1 Intraday Patterns

Intraday patterns observed in financial data include, for example, volume spikes around exchange opening and regular news releases, a die-off of trading activity during premarket and postmarket periods etc. For the analysis of intraday patterns, one needs to assume trading windows which are spaced more finely than in intervals of whole days. Shortening this sampling period grants the researcher an ability to see more detailed information, on the other hand measurement errors stemming from microstructure noise as well as information overwhelm might become a problem - we might stop seeing the forest for the trees. In accordance with standards used in high-frequency trading and academic research focused on intraday data analysis, we arrived at the sampling period of 5 minutes and thus created 5-minute trading bars for the whole 4 year data sample. For each bar traded volume and range were recorded. All 5-minute bars with the same timestamp (i.e. trading bars representing the same 5 minute interval in all days contained in our dataset) were collected and the recorded variables were averaged to

[8] From here on, we use the term daily range for the range of daily extreme log-prices.

obtain a general idea of how the instrument at hand behaves during that specific 5-minute window of the day. In spite of this procedure's simplicity, its results presented below were not published in academic literature yet.

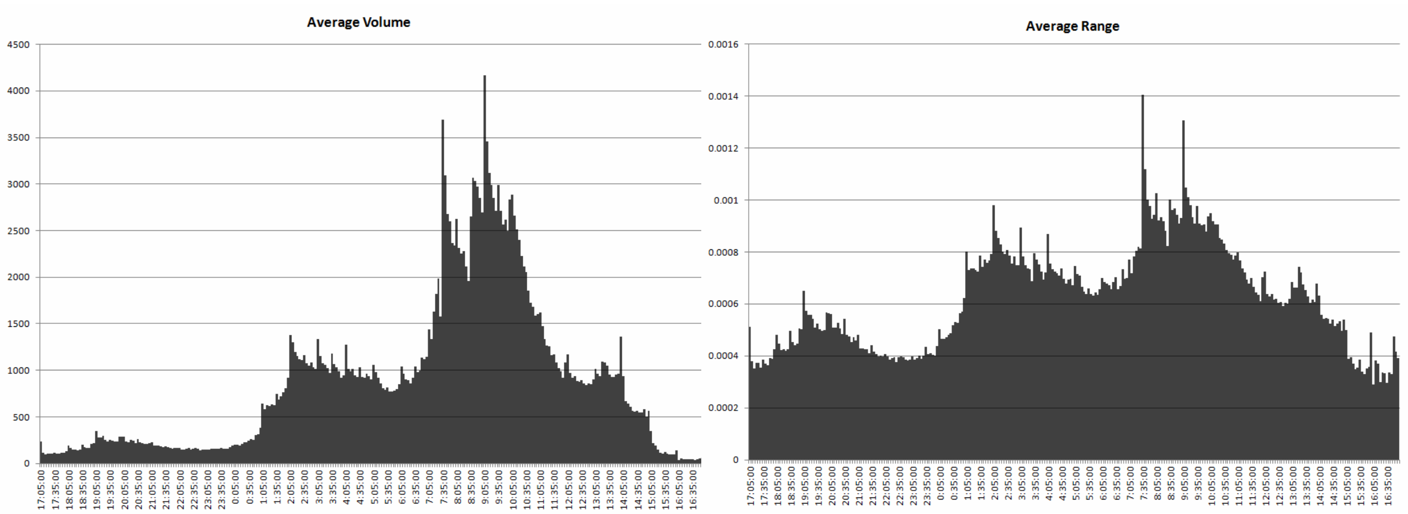


Figure 1: Intraday volume/volatility patterns. Each bar represents the whole-sample average of volume/volatility for the corresponding 5 minute trading interval.

In accordance with e.g. Dacorogna et al (1993), the distribution of observed averages is indicative of different trading activity depending on time of day, corresponding to different geographical distribution of market participant throughout the day.

In the first hours of trading, East Asian traders are active. The overall calm of this period as compared with the overall behavior is caused by three factors. Firstly, it is natural to expect that European and American investors, exporting companies, etc. will be attracted to EUR/USD futures contract more than East Asian traders. Secondly, opening times of different East Asian exchanges are not synchronized and thus surges in traded volume resulting from synchronized commencement of trading do not take place. Lastly, while in Europe and America we observe daylight saving time (DST), East Asian countries do not observe DST. Thus, the effect of e.g. Shanghai Stock Exchange opening affects different 5-minute bars depending on the actual DST observed in Chicago. This causes a possible effect of large East Asian stock exchanges openings to be diluted[9]. As the presence of European and American traders in this early time of the day cannot be ruled out and as we have no information related to whether a particular

[9] As can be seen from two distinctly identifiable small surges in volume at 19:00 and 20:00

trade resulted from interactions of Asian traders exclusively[10], we have no error-proof procedure for removing the impact of this DST effect. Considering, however, the average volume traded volume between 16:00 and 17:00 (non-trading period), we can see that the average volume attributable to East Asian traders is rather low as it is on average only quadruple of the average volume traded in this non-trading period. Hence, any attempts to correct for the DST effect would bring only marginal impact on our results. In other words, we need not worry about imprecision of data measurements stemming from DST mismatch between Asian traders and traders from Europe and America.

On the contrary, in the period between 1:00 and 2:00 a surge in trading volume attributable to the growing presence of European traders occurs. Another distinct surges at 2:00 (bar with timestamp 2:05) and 3:00 attributable to openings of some European exchanges[11] occur. The last surge connected to Europe at 4:00 presents a puzzle. Minor volume surges occurring each 30 minutes can be attributed to regular news releases. As opposed to the East Asian case, when dealing with data created by mostly American and European trading, problems with DST are only minor, as time shifts in Europe and the United Stated are separated by two weeks. Thus, for only two weeks each year are data influenced by different DST zones in Europe and America. In the remaining weeks of the year, both continents are in the same DST zone.

The American session presents several distinct surges of trading activity at 7:30 8:30 and 9:00 which are most likely related to opening of different exchanges. A small surge of volume at 12:00 is likely related to the closing of London session and 14:00 marks the official end of U.S. trading. Position traders usually make trading decisions at the end of a trading day and if enough profit or loss is accumulated, positions are terminated. This is most likely the reason for volume surges accompanying ends of different sessions.

The heavily researched relationship between market activity and volatility does not seem to hold for our instrument. Comparing the European and American sessions, despite the average volume and transaction count nearly double in the American session, the increase in ranges is small. Moreover, comparing the East Asian session with the other two, we observe nearly identical levels of volatility on significantly lower

---

[10] This would be feasible if the investigated futures contract was traded simultaneously on more exchanges as for example in Dacorogna et al (1993)

[11] Most importantly the London Stock Exchange and the Frankfurt Stock Exchange.

market activity. Lastly, taking into account the minimum trading activity during non-trading period and only a minute decline in volatility compared to overall level of volatility, we claim there is no relationship between average volatility and average market activity when talking about whole trading days. However, for different parts of the day there might be a relationship between market activity and volatility. This relationship seems to be different for each session, most likely due to different order book thicknesses Not over different sessions.

Several papers (e.g. Chu & Lam 2008) indicate the usefulness of information provided by the market during illiquid periods (after close and/or before opening). Also, it is likely that general trading conditions (news releases, for example) affect the market for a period of several hours as opposed to just momentarily[12]. Hence we expect some relationships to hold between market behaviors of succeeding sessions. For this purpose, we divide the trading day into several time periods marked by volume surges, i.e. trading sessions. This division is more or less arbitrary and only in some cases reflects the opening times of exchanges around the world as a surge in trading activity is more informative than an opening of an Exchange. The specific periods are

Session	End Time
preAsian	19:00
Asian	1:00
preEuropean	2:00
European	6:00
preAmerican	7:30

Session	End Time
American - 1	9:00
American - 2	14:00
postAmerican - 1	15:00
postAmerican - 2	16:00
Non-Trading	17:00

Table 1: Intraday session time delimiters.

and Figure A.3 in the Appendix depicts this division visually.

Out of the previously mentioned 999 trading days that constitute our dataset, only 66 had some activity in the non-trading session, for which reason we omitted this session completely and never included it at any later stage of work. The remaining sessions were traded commonly and we observed only three missing observations for preAsian session and one missing observation for the European and preAmerican

[12] Imagine, e.g. an ECB announcement with a strong impact on the EURUSD currency pair. European traders operate according to the announcement immediately, while American traders will react to this announcement only when they become active in the market, i.e. with a delay or several hours.

session. In these cases, missing observations were replaced with averages of given sessions' data of the preceding and succeeding day.

3.2.2 Hour of the Week and Day of the Week Analysis

The same procedure as in the previous subchapter was used for a visual evaluation of the day of the week effect as well as hour of the week effect (as in Dacorogna et al 1993).

Investigating different hours of week in terms of market activity leads to several conclusions. Firstly, volume peaks in the middle of the week. All days show the same structure of calm East Asian session, a steep rise of activity with the commencement of European trading with a following setback. A surge of activity with American session start follows. Thursday and Friday seem to have significantly more active preAmerican and American1 sessions. Owing to the high number of observations (roughly 140 for each hour of week) these findings can be considered robust.

In terms of volatility, intraday patterns differ across days of the week. Focusing on ranges, the volatility of American session is significantly higher then for the rest of the day during Thursday and Friday, while during start of the week, volatilities of European and American session are comparable. The presented effects are documented in Figure 2.

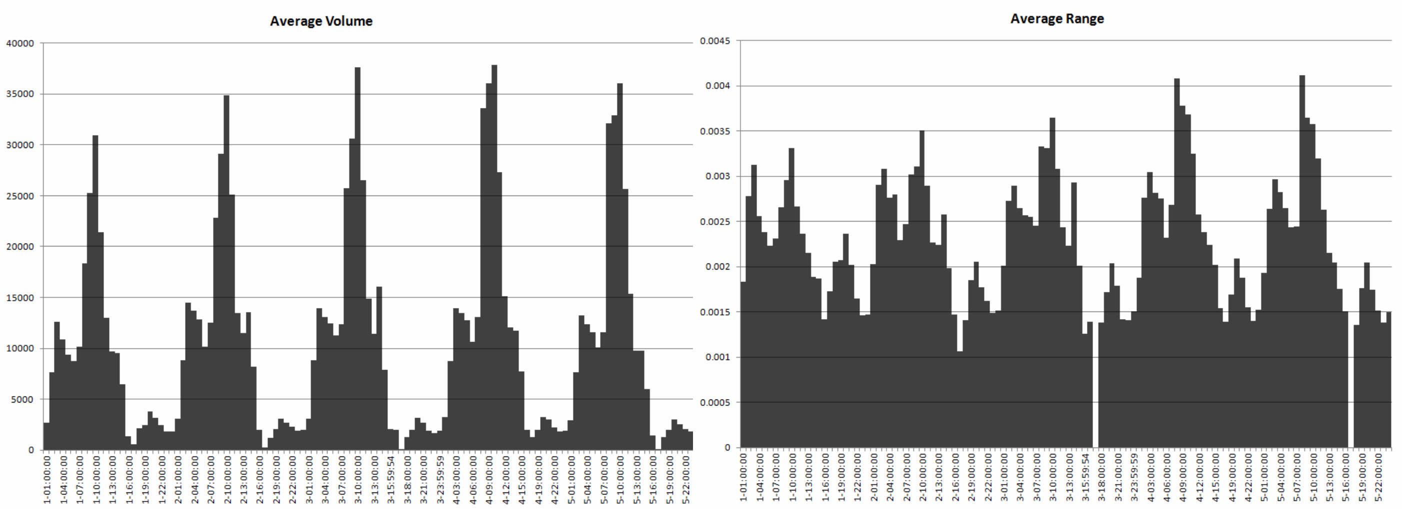


Figure 2: Hour of the week patterns in volume/volatility. A timestamp of 1-13:00:00 represents Monday from 12:00 to 13:00. Gaps in the middle and right side of the Average Range chart capture the Non-Trading session.

Investigations of day of the week effect reveal a lower trading activity on Mondays while the remaining days seems to be equally active. Comparing daily ranges, one observes generally higher volatility on Tuesdays and Thursdays. The effects are presented in Figure 3.

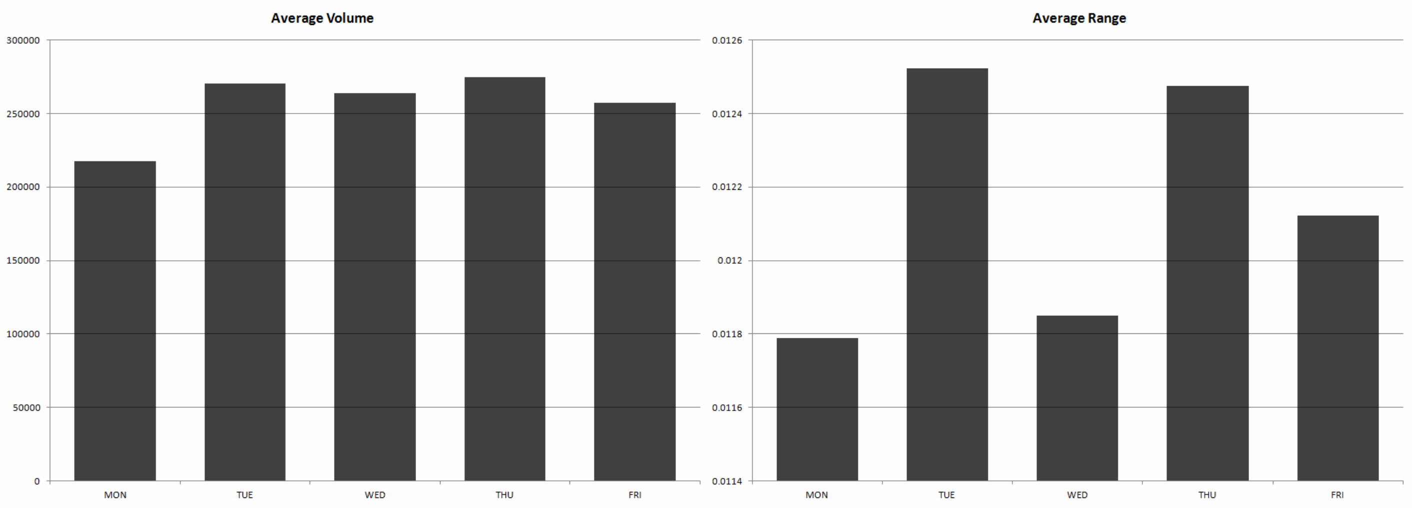


Figure 3: Day of the week patterns in volume/volatility.

3.3 Descriptive Statistics

Having discussed time patterns present in our data, we now turn to providing the reader with a statistical notion of our dataset. For that purpose, descriptive statistics of data sampled at different frequencies are shown below.

	Mean	St. Dev	Skew	Kurt	GHE
r_{5M}	-0.0000003	0.0004894	0.3268	26.3631	0.493
r_{1H}	-0.0000035	0.0016567	0.1055	8.2748	0.502
r_{12H}	-0.0000277	0.0044624	0.0660	4.6478	0.511
r_{1D}	-0.0000809	0.0079047	0.0473	1.1128	0.512
$R_{5M}^{\log}$	0.0006353	0.0004806	3.2356	31.3406	0.974
$R_{1H}^{\log}$	0.0023254	0.0015485	2.3790	11.5503	0.966
$R_{12H}^{\log}$	0.0063174	0.0045183	1.8154	5.0818	0.981
$R_{1D}^{\log}$	0.0121566	0.0055872	1.6842	4.1943	0.990

Table 2: Descriptive statistics of data sampled at 5M, 1H, 12H and daily frequencies.

Growing mean values and standard deviations of returns with less frequent sampling are expected. Skewness as well as excess kurtosis of log-returns approaches zero with lowering sampling frequency. The growing similarity of log-returns' distribution to Gaussian distribution as sampling frequency is being lowered is a well-documented feature of financial markets. However, when we check for normal distribution of log-returns, the null hypothesis is rejected regardless of sampling frequency and test type (Jarque-Bera, Kolmogorov-Smirnov) due to the well-known heavy-tailed feature of financial data[13]. The Hurst exponent of log-returns measured using Generalized Exponent method stays very close to 0.5 which is a value corresponding to random walk[14].

Turning to ranges, growing mean values and standard deviations with lower sampling frequencies are an expected outcome as well. Both skewness and excess kurtosis fall in magnitude with lower sampling frequencies, most likely due to a relatively lower impact of bid-ask bounce. The Hurst Exponent is extremely high for ranges measured at all frequencies which indicates very strong persistence.

Speaking of distributional properties of ranges, Locke's nonparametric test rejects the null hypothesis of gamma distributed data at all feasible significance levels, which was contrary to our expectations. A frequency distribution of daily ranges, coupled with the best fitting gamma distribution p.d.f. is included in the Appendix as Fig. A.4. This unexpected behavior could be disentangled by considering the distributions of separate sessions' ranges. However, for these the null hypothesis of gamma distribution is rejected as well. The histograms of separate sessions' ranges are included in Fig. A.4 as well. Alizadeh, Brandt & Diebold (2001) argue that logs of ranges are approximately normally distributed. Checking for this, the null hypothesis of normal distribution was rejected both for daily log-ed ranges as well as for sessions' log-ed ranges (except for the European session, results not presented here for brevity reasons). Hence, distributional properties of daily ranges as well as session ranges remained unknown.

Lastly, turning to stationarity checking as a prerequisite for time series modeling of daily ranges, we ran the Augment Dickey-Fuller test. An ADF test with five lags and a constant included rejects the null hypothesis of unit-root in daily ranges with practically zero p-value, hence stationarity is claimed.

A time plot of daily ranges, included in the Appendix as Figure A.5, reveals a period Q2 2008 - Q1 2009 of high volatility related to the financial crisis. Apart from this period, stationarity of the series is evident even visually.

---

[13] For brevity purposes, p-values of Jarque-Bera and Kolmogorov-Smirnov tests were omitted, but in each case the null hypothesis of normal distribution was rejected on all feasible significance levels.

To sum up this section, the dataset at hand does not exhibit any irregular patterns. Both log-returns as well as log-ranges behave in accordance with expectations (random walk of non-normally distributed log-returns, extremely high Hurst exponent of log-ranges). Neither gamma nor lognormal distribution seems to fit daily ranges in our data and based on the results of an ADF test stationarity of daily ranges can be claimed. Investigated intraday patterns reveal different market behavior during the trading day and for this reason we divide the day into several trading sessions distinguished by volume surges.

3.4 Autocorrelations and cross correlations

Having described the basic characteristics of our dataset in the previous section, we continue with the issue of correlation analysis. Firstly, we investigate autocorrelations in daily ranges per se, as this will guide us in the selection of a proper ARMA-type model for later parts of this work. Next, we investigate correlations between daily ranges and trading activity variables (trading volume, average trade size and transaction count) measured both on whole days as well as separate sessions. These correlations will shed light on whether the inclusion of trading activity variables into models for daily ranges can prove fruitful.

Autocorrelations of daily ranges (Figure A.6) start at values of approximately 0.4 and decay very slowly. The first autocorrelation inside the Barlett test critical interval is located beyond the 70th lag and hence we can assume an existence of a long memory process governing ranges. This conclusion is derived from highly significant autocorrelations at distant lags and an extremely large value of the Hurst exponent. The shape of PACF and ACF hint towards an AR(7) process governing ranges. We will, however, postpone model specification until Section 5.

Next, we investigate relationships between volatility and variables capturing trading activity, both on whole days as well as on separate intraday sessions. Correlations of trading activity variables (average trade size, traded volume and transaction count) and ranges within separate sessions as well as on whole days are reported in Table A.7. Contrary to the popularly held belief, whole day ranges depend (linearly) much more on transaction count and average trade size than on traded volume.

---

[14] GHE is the most efficient estimator of the Hurst exponent (see a recent comparison of all Hurst exponent estimators under heavy-tailed distributions in Baruník & Krištoufek 2010)

This result is in line with findings of Jones et al (1994) and Chan & Fong (2000). Focusing on Figure 1, one could suspect different range-volume relationships among intraday sessions, as these sessions differ hugely in trading activity with ranges differing only slightly. This feature can simply be attributed to different order book thicknesses during various sessions. The data confirm this as correlations of ranges and trading activity variables differ vastly across trading sessions. While for most sessions transaction count correlates with range more than volume, for the preEuropean session the opposite holds. The variance of correlations between ranges and average trade sizes across sessions is also high. To sum up, all trading sessions exhibit remarkably different volatility-trading relationship. Note that all correlations are of expected signs.

Correlations of daily ranges and sessions' trading activity/volatility are reported in Table A.8. High correlations among daily ranges and ranges/trading activity of early sessions hold promise of possible daily ranges prediction updates conditional on early trading activity/volatility. As opposed to previous results (Table A.7), transaction count of current day's sessions has a significantly lower correlation with current day's range, albeit still non-trivial. The same applies for traded volume, only to a larger extent.

Since daily ranges exhibit both strong autocorrelation and strong correlations with trading activity variables of separate current day sessions, we turn to the investigation of correlations between current daily ranges and variables (volatility and trading activity) measured on lagged sessions. The results are presented in Table A.9 and are indicative of strong dependence of daily ranges on ranges, transaction counts and average trade sizes of different lagged sessions. Hence, the question arises what kind of dependence governs daily ranges. On one hand, current daily range could depend on past realizations of daily ranges provided that these lagged daily ranges embed all information contained in different past sessions. Or current daily range might depend on lagged variables measured over separate sessions and as these are also related to lagged daily ranges, the autoregressive property of daily ranges might only hide the true dependence of daily ranges on different lagged sessions' variables. The investigation of whether or not does lagged daily range contain all useful information for daily ranges forecasting will be carried out in Section 5.1.

Next, we briefly investigate mutual correlations among volatility and trading activity variables of intraday sessions. For example, we measure the correlation between preAsian volume and European range. These session variables have been shown to exhibit high correlations with daily ranges. However, if these session variables are also highly mutually correlated, they do not convey unique information and thus their inclusion in any model will only lead to a cosmetic increase in explanatory power as well as a reduction in degrees of freedom. By checking mutual correlations between these trading variables measured on separate sessions, we check the uniqueness of information they carry. High correlations between realized variance, realized range and range measured on the same session are expected, hence we only investigate correlations between realized ranges, average trade sizes and transaction counts of all intraday sessions. The resulting 27x27 (three variables for nine sessions) matrix is too large, hence we do not provide it here but the whole matrix can be obtained upon request. Out of 351 unique correlations, only 33 are above 0.6, only 8 are above 0.75 and none is above 0.9. Hence, even though there are some session variables which exhibit significant mutual correlations, their number is not as high so as to invalidate the inclusion of sessions' trading variables into models for daily ranges.

Lastly, Dacorogna (1997) investigates causality between coarsely grained and finely grained (i.e. long-term and short-term) volatility and finds that long-term volatility causes short term volatility. This result is arrived at by studying asymmetric cross-correlations between long and short term volatilities. As this question lies out of the focus of this work, cross correlations with negative lags are not presented in Table A.9 but can be provided by authors upon request.

To sum up, daily ranges are strongly autocorrelated. Correlations between daily ranges, volatility and trading activity variables measured over separate sessions of the same day are high. Combining these two results led us to the investigation of correlations between daily ranges and lagged volatility and trading activity variables measured over separate sessions. These correlations proved to be high as well. Further investigations proved that different session variables of one trading day convey unique information. Thus we arrive at the question of whether strong autocorrelation in daily ranges is indeed a dependence of daily volatility on lagged daily volatility, or whether today's range depends on previous day's session variables. If the latter was the case then the dependence of today's range on yesterday's range would be only indirect, as both would be predominantly driven by yesterday's volatility and trading activity variables measured over different sessions.

4. Comparing volatility measures

In the preceding Section, we provided the first building blocks for our later work focused on modeling daily ranges. After showing the possible benefits of slicing up a day into trading session, we continue by analyzing different venues of enhancing models commonly used for daily ranges modeling. Specifically, we investigate the usefulness of regressing daily ranges on other measures of volatility.

Thus, in this section, we firstly introduce the concept of realized measures of variance (realized variance and realized range) which use intraday data for the calculation of daily variance. We are motivated by high efficiency gains compared to range-based estimators. After familiarizing the reader with realized measures of variance, we investigate whether range-based volatility estimators differ in their efficiency even on real-world data or whether their differing efficiency is confined to simulated processes. Our motivation is to infer which range-based volatility estimator provides the best efficiency on real-world data. A range-based estimator with the lowest measurement error should be used as a regressor in daily ranges modeling to obtain the best fit in case intraday returns for realized range/variance are not available (data costs, illiquid markets[15]). Moreover, as our dataset provides intraday returns, we assess the imprecision in daily variance estimation when using daily ranges/range-based estimators instead of realized ranges/realized variance.

4.1. Realized measures of variance

In their influential paper, Andersen et al (2001) introduce the concept of realized variance for the estimation of daily variance, where realized variance is simply a sum of squared intraday returns. By using high frequency data, volatility measured by this approach can be considered observed rather than latent (as in e.g. ARMA-GARCH models of log-returns) and by increasing sampling frequency, one can theoretically approach true volatility of the underlying process with arbitrary precision. In practice, however, increasing sampling frequency brings increased precision of measurement and at the same time increased bias induced by market microstructure. As with higher sampling frequencies asset returns diminish while microstructure effects remain of

[15] For example selected bond markets or energy markets with very distant delivery dates.

relatively constant size, extremely frequent sampling induces a strong bias into volatility measurements.

Martens & van Dijk (2007) and Christensen & Podolskij (2007) in independent studies build upon the work Parkinson (1980) who shows that squared daily ranges provide efficiency gains compared to squared daily returns. These two studies take Parkinson's insights and apply it to the topic of measuring daily variance on intraday data. Specifically, they propose the replacement of squared intraday returns in the calculation of realized variance by squared intraday ranges. Daily variance estimate obtained by this procedure was coined realized range and according to empirical sections in Marten & van Dijk and Christensen & Podolskij realized range provides efficiency gains over realized variance. As the observed (high frequency) range is likely to overestimate the true range of underlying price process due to market microstructure (the period's highest price occurred more likely on ask and vice versa for the lowest price), both papers propose bias-correcting constants. While Martens & van Dijk suggest normalizing each square of intraday range by $[4\ln(2)]$ (as Parkinson 1980), Christensen & Podolskij alter the normalizing constant depending on sampling frequency. In our measurements, we followed the former approach.

The question of optimal sampling frequency bringing the best precision-bias tradeoff for measuring realized variance and realized range has been investigated by several authors. In our work, we chose to sample returns at 5-minute intervals, as this sampling frequency was firstly proposed by Andersen & Bollerslev (1998) for measuring ex post daily foreign exchange volatility. Also this sampling frequency was originally used in Andersen et al (2001). Thus, in our work realized variance was calculated as the sum of squared differences between the log-close and log-open prices for each 5-minute bar of the ETH (whole day) session. Daily realized ranges were calculated as summed squared differences of log-high and log-low prices of each 5-minute bar normalized by $[4\ln(2)]$.

Time plots of squared daily log-ranges divided by $[4\ln(2)]$ (i.e. Parkinson's measure of daily variance), daily realized variance and daily realized ranges are shown in Figure A.10. Owing to similarity of realized range/realized variance construction it is not surprising to find nearly identical development of these measures. Comparing squared daily ranges and realized variables we observe a good match, however, some differences are present. As we will see in succeeding analysis, the differences between realized ranges and daily ranges are significant. Correlations of daily ranges and lagged values of ranges, realized variances and realized ranges (daily as well as sessions) can be found in Table A.11[16]. Other volatility measures (standard deviation of returns, squared or absolute daily returns, etc.) were omitted from this exercise as they are significantly less efficient than realized variables. Reported correlations reveal an expected pattern - ETH range is more correlated with lagged realized variables than with its own lagged values[17]. Hence, realized variables seem to be higher quality predictors of ETH ranges. Comparing realized variance and realized range in a similar manner, we find that correlations of realized range and ETH range are higher in 63 out of 80 cases than correlations of ETH ranges and realized variance, indicating superiority of realized range for ETH range forecasting. The question of whether to augment an autoregressive model of ETH ranges by lagged values of realized ranges or whether to replace lagged daily ranges by lagged realized ranges remains to be investigated in subsequent sections.

To sum up, in this section we investigated which realized measures of daily variance is a better predictor of daily ranges. Theoretical results favor realized ranges due to higher efficiency compared to realized variance. Our empirical results are in accordance with this conclusion, as daily and sessions ranges exhibit higher correlations with daily and sessions realized ranges than with daily and sessions realized variances.

4.2. Assessing efficiency gains of range-based estimators

In this section, we draw upon conclusions of the previous section. We firstly investigate the general relationship between daily ranges and realized ranges, including efficiency gains of using realized ranges as compared to daily ranges. Secondly, we are interested in whether differing efficiencies of range-based estimators presented in Section 1 are measurable on our dataset.

The first exercise (assessing the usefulness of realized ranges for daily ranges modeling) complements our previous investigations of correlations. In volatility related

[16] Correlations of daily ranges and lagged daily ranges are reported redundantly (already contained in Table A.9). In Table A.11 we state them for comparison purposes with other volatility measures.
[17] In 71 cases out of 80 cases in total (8 lags for 10 trading sessions), correlations of ETH range with lagged realized variances of different sessions are higher than correlations of ETH range with (normal) range of different sessions. For lagged realized range, this result is 74 cases out of 80.

literature, Mincer-Zarnowitz regressions[18] are used to assess a performance of one measure of volatility against a benchmark (usually realized variance). We use the same methodology for comparative quality assessment of daily ranges and realized ranges/realized variance. For this purpose we estimate

$$RR_D = \alpha + \beta\left[\left(R_D^{\log}\right)^2 / 4\ln(2)\right] + \varepsilon_D$$
$$RV_D = \alpha + \beta\left[\left(R_D^{\log}\right)^2 / 4\ln(2)\right] + \varepsilon_D \qquad (19)$$

i.e. how well daily squared daily ranges match realized measures of variance. Estimation results of (19) are shown in Table A.12. If squared daily ranges and realizes ranges/variance were comparable in terms of accuracy, residuals would show no heteroskedasticity, α would be close to zero and β would be close to unity. Our results suggest that squared daily ranges cannot replace realized measures of variance, as residuals exhibit strong heteroskedasticity (mostly pronounced in the time period related to the onset of crisis). Moreover, Chow tests carried out on both halves of the whole set and later on quarters of the set reject the null hypothesis of time-invariant values of α and β parameters. Intuitively, this can be attributed to the fact that daily ranges neglect a large part of intraday information. For example, if we consider a "V" and a "W" shaped evolution of prices within a day, daily ranges for these days might be the same but in that case both realized measures will be much larger for the second day. This changing daily structure uncaptured by the daily range is the reason for a failing time-invariance of parameter estimates. The general strong decrease of AdjR^2 for both models in the second half of our dataset indicates that in calmer times, differences in measures calculation pronounce the differences in measured volatilities.

Having investigated parameter stability and obtaining a general idea of that daily ranges and realized variance measures can not be interchanged, we now turn to an investigation of how well range-based estimators compete with each other and with realized measures (in terms of efficiency). For this purpose, equations resembling (19) were estimated with HAC standard errors (to account for heteroskedasticity).

---

[18] Mincer, J. A., Zarnowitz, V. (1969) The Evaluation of Economic Forecasts: Analysis of Forecasting Behavior and Performance, NBER Books, National Bureau of Economic Research

The only difference compared to (19) is that in this exercise, we regress all range-based measures on both realized measures and vice versa. That is, we estimate

$$
\begin{aligned}
\left(R_D^{\log}\right)^2/4\ln(2) &= \alpha+\beta RV_D+\varepsilon_D & \left(R_D^{\log}\right)^2/4\ln(2) &= \alpha+\beta RR_D+\varepsilon_D \\
\left(\hat{\sigma}_D^{GK}\right)^2 &= \alpha+\beta RV_D+\varepsilon_D & \left(\hat{\sigma}_D^{GK}\right)^2 &= \alpha+\beta RR_D+\varepsilon_D \\
\left(\hat{\sigma}_D^{RS}\right)^2 &= \alpha+\beta RV_D+\varepsilon_D & \left(\hat{\sigma}_D^{RS}\right)^2 &= \alpha+\beta RR_D+\varepsilon_D
\end{aligned} \tag{20}
$$

$$
\begin{aligned}
RV_D &= \alpha+\beta\left[\left(R_D^{\log}\right)^2/4\ln(2)\right]+\varepsilon_D & RR_D &= \alpha+\beta\left[\left(R_D^{\log}\right)^2/4\ln(2)\right]+\varepsilon_D \\
RV_D &= \alpha+\beta\left(\hat{\sigma}_D^{GK}\right)^2+\varepsilon_D & RR_D &= \alpha+\beta\left(\hat{\sigma}_D^{GK}\right)^2+\varepsilon_D \\
RV_D &= \alpha+\beta\left(\hat{\sigma}_D^{RS}\right)^2+\varepsilon_D & RR_D &= \alpha+\beta\left(\hat{\sigma}_D^{RS}\right)^2+\varepsilon_D
\end{aligned} \tag{21}
$$

Recalling the definition of Garman & Klass volatility measure in (11), the value of parameter f needs to be chosen. As f represents the portion of a day during which trading is halted, we set $f = 1/24$ since trading is practically absent only in the Non-Trading session between 16:00 and 17:00. In our computations daily open price is the first price traded in the preAsian session while the close price is the last price traded in the postAmerican2 session. Lastly, we neglected the Yang-Zhang measure of variance (13) as this measures average past variance rather than daily variance[19].

The results of (20), (21) are presented in Table A.13. As all five variables are measures of the same quantity (variance) and RR_D, RV_D estimates border on the true value of volatility, we can loosely interpret the results as follows: $\alpha = 0 \wedge \beta = 1$ in the upper pane of Table A.13 imply that RR_D, RV_D are best predictors of all range-based variance estimates and can explain all variance related information captured in these range-based estimates.

On the other hand, β estimates in the lower pane of Table A.13 indicate a decomposition of range-based volatility measures into information on variance and noise. These β estimates as well as AdjR^2 of all models indicate that approximately 57% of information in range-based variance estimates is related to variance of the underlying process and the remaining share of information is noise. From this we could roughly infer efficiency gains of using realized ranges/variance instead of range-based estimates. We need to bear in mind that microstructure noise and possible jumps were not removed from intraday data, hence values of realized ranges/variances might be biased. Even though this bias is most likely small, for exact assessment of efficiency gains we cannot neglect it.

Lastly, for range-based variance estimators we observe a clash of theoretical and empirical results. While in theory, Garman & Klass as well as Rogers & Satchell estimators promise significantly higher efficiency over Parkinson's measure, our results do not confirm this. β estimates as well as AdjR^2 in the lower-pane of Table A.13 are nearly identical for all three range-based estimators. Hence we observe no significant efficiency gain. We thus do not have any reason to use the RS or GK measure as predictors in daily ranges modeling instead of daily ranges themselves.

To sum up, in this section we showed that linkages between realized measures of variance and daily ranges are strongly time-variant. Neither range-based estimator can match the precision of realized ranges/variance. Only roughly 57% of information provided by range-based estimators is related to variance, the remaining part is error measurement. Even though Garman & Klass and Rogers & Satchell estimates promise significant efficiency gains compared to Parkinson's measure, our results indicate roughly equal efficiency of all three range-based volatility estimators.

[19] Rogers & Satchell (1991) propose a procedure for correcting the bias of both RS and GK measures resulting from infrequent trading. As our data have at least 80,000 contracts traded each day, the bias correction terms would be minuscule and hence we chose their omission in RS, GK construction.

5. Modeling daily ranges

In this section, we take the results of preceding sections and try to exploit these results for improving several existing approaches to volatility modeling. This section provides the bulk of our contribution to volatility modeling literature. At first, we provide an overview of plausible models for daily ranges prediction present in current academic literature. Next, we focus on several models which were chosen for this thesis either due to their general properties (goodness of fit) or their specific focus on daily ranges modeling. For each of these models we investigate the added benefit of using more precise volatility measures for daily ranges prediction. Also, we investigate the added benefit of slicing up historical data into sessions. Lastly, we evaluate the possibility to obtain updated daily volatility forecasts of increasing quality as time passes during a trading day.

Modeling daily ranges is in a certain way special. Unlike many other variables in economics or finance for which usually one approach is used dominantly (e.g. VAR for monetary economics variables), volatility and daily ranges especially can be modeled using various approaches. Generally, these can be separated into two groups depending on whether or not the assumption of long memory is exploited. In this thesis, we only focus on models which do not incorporate long memory. Our reasons for this choice are comparative ease of estimation, high prevalence in published papers and better economic interpretation of models that do not assume long memory.

When modeling volatility and neglecting the assumption of long memory, simple AR (ARCH type, see Engle 1982) or more refined GARCH models can be used. However, several papers question the validity of GARCH-type models. For example, Starica (2003) finds that a GARCH-type model can only be used for short-term volatility prediction on the most commonly known market indices (S&P, DJIA). Moreover, misspecification of the GARCH process (e.g. omitting the IGARCH effect when it is in place) dramatically decreases the forecasting performance of the model. As GARCH models in general incorporate only lagged values of different measures of volatility, it is possible that such omission of an important explanatory variable is common.

Recent developments in volatility modeling applicable to modeling daily ranges are, for example, mixtures of long, medium and short-term volatilities (HAR of Corsi 2004) or one can exploit previously mentioned linkages between various measures of volatility for predicting daily ranges, as in Engle & Gallo (2003). Following the latter approach, having a good prediction of daily range is not beneficial only per se, but also has a side benefit of providing for better predictions of other measures of volatility. Other daily ranges modeling approaches that can be mentioned are vector error-correction models (Cheung et al 2007) or models incorporating data sampled at different frequencies (Ghysels 2003). Lastly, when considering predictions related specifically to futures markets, models rooted in cost-of-carry theory of futures contracts (e.g. Zeng & Swanson 1998) or models based on readings of implied volatility (Jorion 1995) can be used.

Apart from the aforementioned drawbacks of current volatility modeling/forecasting techniques, another type of drawback present in this area is the error-minimization approach. According to this approach, the best model chosen from a group is the one with the lowest mean squared error or similar error-based measure of goodness of fit. However, Leitch and Tanner (1999) provide arguments against this approach. Their findings confirm that models performing well in such error-minimization need not be the ones with greatest real-life applicability. On the contrary, their findings suggest that sometimes models with the worst mean squared error are the ones with greatest real life value (as measured by e.g. profitability of a trading strategy utilizing the model's forecast). Part of research suggests that ranges prediction can indeed result in profitable trading strategies (Cheung 2007, Cheung et al 2010). Including the topic of error-minimizing versus profit-maximizing daily range/volatility forecasts would make this thesis too large a body of possibly incomplete research. Hence, in this thesis, our aim is to assess possibilities and limitations of the error-minimizing approach of volatility forecasting only.

In the next sections, we discuss and estimate the models chosen for this thesis.

5.1 ARMA-GARCH

As the first approach for daily ranges modeling, we chose the prevalent method of time-series analysis in economics, the ARMA-GARCH model. Our motivation for this exercise is to obtain a base model to which we can compare other models designed specifically for volatility, or even better, for daily ranges. The application of ARMA-GARCH type models to volatility modeling is not uncommon. Examples of this approach can be found in, for example, Pong et al (2003) who use an ARMA model in predictions of realized volatility or Ahoniemi (2009) where ARMA models are used for modeling and forecasting of various instruments' implied volatility.

Estimation of ARMA-GARCH models was carried out by jointly considering both the mean-of-volatility as well as volatility-of-volatility component. Selection of the best specification of both components was carried out with three criteria in mind: parsimony of both components, absence of autocorrelations in both residuals and their squares and the match of expected and true distribution of model residuals. The presence of "bad-new" or "leverage" effect, i.e. the effect of increased volatility following a drop in market prices, was investigated by including a dummy for previous day's price decrease in the mean-of-volatility equation. Following ACF/PACF of daily ranges (Figure A.6), an AR(7) model was suspected. The mean-of-volatility component was capable of removing autocorrelations in residuals, but suffered from an insignificant 7^{th} AR lag. However, upon switching to AR(6) model for mean-of-volatility a strong autocorrelation in residuals on 7^{th} lag appeared. Hence, the AR(7) specification was kept.

Squared residuals of a pure AR(7) process exhibited significant autocorrelations on all lags, hence a volatility-of-volatility component was needed. Here, GARCH(1,1) with T-distributed residuals of the AR(7) component was found to be the best specification as it removed autocorrelations in squared residuals on nearly all lags. Both ARCH and GARCH terms were significant. The estimated degrees of freedom pertaining to residuals' Student distribution was significantly different from 2 (normal distribution). Added second ARCH or GARCH terms turned out insignificant, ruling out the need for a more complex GARCH component. Thus, we considered the GARCH(1,1)- t specification well justified.

ARCH models used for modeling log-return volatility are known to suffer from a necessity to include many lags of the ARCH term to remove autocorrelations in squared residuals. GARCH models solve this by allowing for an MA term in volatility prediction, which is commonly able to replace many ARCH terms. As an AR(7) model of volatility is rather complex, attempts were made to reduce the number of terms by estimating an ARMA model of daily ranges. Both ARMA(1,1)-GARCH(1,1)-t and ARMA(2,2)-GARCH(1,1)-t models were incapable of removing autocorrelations in residuals themselves. Specifying more complicated ARMA models was unfeasible due to two reasons. Firstly, our motivation was to obtain a model parsimonious compared to AR(7). Secondly, estimations of ARMA(2,2) contained two pairs of common roots indicating that ARMA(2,1) model was the most complex identifiable ARMA model on our dataset. More complicated models would have a non-invertible MA component.

For this reason, attempts to include MA terms into the mean-of-volatility component were halted and AR(7)-GARCH(1,1)-t was taken as the best representative of the ARMA-GARCH class of models. On a last note, all parameter estimates in the AR(7)-GARCH(1,1)-t model were positive, which is a desired property in a volatility forecasting model. This, by preventing volatility forecasts to be negative, ensures that a volatility forecast will always be at the practitioner's disposal. Surprisingly, the previous day's "bad-news" effect is insignificant.

To sum up, the first model for daily ranges was drawn from classical ARMA-GARCH approach. AR(7) model jointly estimated with GARCH(1,1)-t component was chosen as the best specification and in-sample estimation results are presented in Table A.14. Table A.15 contains Q-Q plots of AR(7)-GARCH(1,1) residuals with normally and T-distributed disturbances. Albeit the Q-Q plot of residuals using T-distribution is not perfect, it is much better than if we had used normal distribution.

5.2 Heterogeneous autoregressive model

The idea that markets transform information into prices efficiently had been the cornerstone of academic thinking in finance for the last 50 years even since Eugene Fama introduced the concept of the Efficient Market Hypothesis in the 1960s. Implications of this hypothesis range from impossibility to make a consistent profit in any market to prices being constantly at their equilibrium levels, reflecting all the fundamental information available. One of the assumptions that allowed academics to arrive at such strong conclusions is the homogeneity of market participants. That means traders are expected to share the same opinion, to be capable of assessing available information in the same way (thus differences in market actions are driven solely by different information sets) and also, their decision making horizons are expected to be equal.

The last assumption was questioned by Muller et al (1993) where the collective of authors propose the Heterogeneous Market Hypothesis. This hypothesis expects participants to differ in their investment horizon. Retail traders can be divided into intraday traders and position traders who hold their positions for several days. Active portfolio managers update the composition of held assets weekly or bi-weekly. Large institutions with long-term investment horizons, such as pension funds, update their portfolios on a basis of months, whereas central banks are likely to intervene in the FX markets etc. on a quarterly basis. Differences in the length of investment horizon are given by differing capabilities to withstand losses, different amounts of capital invested and different motives to trade. While retail traders can usually withstand only small adverse movements before terminating a position with a loss, large institutions rely more on long-term growth and thus short-term fluctuations are not a reason for quitting an open trade. Lastly, central banks are predominantly interested in protecting the country's price stability or exchange rate and the profitability motive in trading is absent in this case. As reaction time of economies to central banks' moves is long, it is not surprising that central banks are not interested in weekly or even monthly market changes.

In subsequent work, Dacorogna et al (1997) use the insights of HMH to propose an extension of GARCH models, namely HARCH model (heterogeneous GARCH). The basic idea in HARCH is to combine volatility views of differing time horizons so as to capture the views of more types of market participants. In empirical applications, the HARCH effect is significant and thus validates the view of HMH.

Corsi (2003) follows up to HARCH modeling by proposing his own model. Firstly Corsi comments on mainstream methods of volatility modeling. On one hand, long memory models suffer from difficult estimation procedures, dubious economic interpretation and need a long buildup period. On the other hand, parsimonious (G)ARCH type models are unable to fully replicate stylized facts related to volatility. For example, returns normalized by volatility forecasts depart from normal distribution and autocorrelations of volatilities exhibit an exponential as opposed to hyperbolic decline observed in reality. To correct for these shortages, Corsi draws upon the conclusions of HARCH and argues that a correctly specified volatility model should incorporate different market views by focusing on volatilities measured over periods of different lengths, i.e. short-term, medium-term and even long-term volatilities (to capture the long-memory property). Specifically, for the modeling of realized variance, Corsi proposes a model of the following specification

$$RV_D = \alpha_0 + \alpha_1 RV_{D-1} + \alpha_2 RV_{D-1}^{(5)} + \alpha_3 RV_{D-1}^{(22)} + \varepsilon_t \tag{22}$$

where $RV_D^{(a)}$ is a simple average of realized variances of days $(D-a;D]$, i.e. $RV_D^{(1)} = RV_D$. In other words, a HAR model intuitively combines volatility of the previous trading day and average volatilities of the last week's and last month's trading. Estimation can be performed using OLS and in empirical work, HAR volatility predictions attain great in-sample fit and out-of-sample forecasts while replicating volatility related stylized facts.

In a succeeding work, Corsi & Reno (2009) assess the "bad-news" effect on volatility known from the family of T-GARCH models. The "bad-news" (or leverage) effect draws upon the idea that bear markets are usually accompanied by larger volatility than bull markets. The specification enriched by leverage effect

$$RV_D = \alpha_0 + \alpha_1 RV_{D-1} + \alpha_2 RV_{D-1}^{(5)} + \alpha_3 RV_{D-1}^{(22)} + \beta_1 I_{D-1}^{(1)} r_{D-1}^{(1)} + \beta_2 I_{D-1}^{(5)} r_{D-1}^{(5)} + \beta_3 I_{D-1}^{(22)} r_{D-1}^{(22)} + \varepsilon_D \qquad (23)$$

includes an indicator variable $I_D^{(a)}$ equal to one in case the average daily return $r_D^{(a)}$ measured over days $(D-a;D]$ is negative. On Corsi & Reno's dataset, all $\beta_1 ... \beta_3$ estimates are significant indicating that volatility has "long memory" not only in itself, but also remembers short-term, medium-term and long-term market declines. In in-sample modeling and out-of-sample forecasting, this new (LHAR - Leveraged HAR) model performs better than the original specification (22).

5.1.1. Basic HAR modeling

Having introduced the HAR models in general in the preceding section, we continue by estimating the basic HAR specification on our dataset. In this section, we do not consider data measured on trading sessions yet. Instead, we investigate general dependencies between daily ranges, lagged daily ranges, lagged realized ranges and lagged daily trading activity variables.

Despite the original HAR (22) was proposed for realized variance, there should be no problem applying it to any other volatility/variance measure. Moreover, as opposed to the previous ARMA-GARCH model, HAR models turned out to be suitable for the description of long-memory processes, for example, daily ranges. It is for this reason why, merging the logics of HAR and findings of preceding sections (Sections 1.1.5, 1.1.6), we propose to augment the LHAR model by variables significantly correlated with the daily range. These include lagged average daily, weekly and monthly trade size and transaction count. Moreover, we augment the daily ranges HAR model by realized ranges due to their superior efficiency. As previous results imply, daily and realized ranges do not hold the same information, hence it is possible that the inclusion of both into a HAR model will turn out useful.

Dealing with heteroskedasticity in modeling daily ranges with realized ranges can be accomplished either by robust standard error estimation or via simultaneous estimation of a HAR and GARCH model. We chose to utilize the former approach and allow for HAR-GARCH modeling only for specifications in which we observe a marked improvement of the augmented model over the base HAR specification. We chose not to apply the volatility-of-volatility modeling using GARCH to all models in order to firstly obtain results directly comparable to those of Corsi's. Also, as the number of models estimated in this section is large and there is no automated way of determining the best GARCH specification for volatility of residuals, estimating a HAR-GARCH model for all specifications would be extremely time consuming. Lastly, we model a HAR dependence in both volatility as well as variance measures, i.e. daily ranges and their squares.

The list of augmented HAR models considered in this section, coupled with their specifications, is presented in Table A.16. However, for the sake of clarity, we briefly define the different specification of HAR models used from here on to prevent confusion. The base specification for daily ranges modeling is

$$R_D^{\log} = \alpha_0 + \alpha_1 R_{D-1}^{\log} + \alpha_2 R_{D-1}^{\log,(5)} + \alpha_3 R_{D-1}^{\log,(22)} + \varepsilon_D \tag{24}$$

Including the leverage effect into a HAR model of daily ranges leads to the LHAR specification given by

$$\begin{aligned} R_D^{\log} &= \alpha_0 + \alpha_1 R_{D-1}^{\log} + \alpha_2 R_{D-1}^{\log(5)} + \alpha_3 R_{D-1}^{\log,(22)} \\ &\quad + \beta_1 I_{D-1}^{(1)} r_{D-1} + \beta_2 I_{D-1}^{(5)} r_{D-1}^{(5)} + \beta_3 I_{D-1}^{(22)} r_{D-1}^{(22)} + \varepsilon_D \end{aligned} \tag{25}$$

In order to infer modeling performance gains stemming from using more precise information on volatility, we regress daily ranges on realized ranges solely in the R-HAR specification, i.e.

$$R_D^{\log} = \alpha_0 + \alpha_1 RR_{D-1} + \alpha_2 RR_{D-1}^{(5)} + \alpha_3 RR_{D-1}^{(22)} + \varepsilon_D \tag{26}$$

To investigate the added benefit of using variables representing trading activity (average trade size/transaction count[20]) and their possible long-term influence on volatility, we define models with -S/-C suffixes as

$$R_D^{\log} = \alpha_0 + \alpha_1 R_{D-1}^{\log} + \alpha_2 R_{D-1}^{\log,(5)} + \alpha_3 R_{D-1}^{\log,(22)} + \gamma_1 TS_{D-1} + \gamma_2 TS_{D-1}^{(5)} + \gamma_3 TS_{D-1}^{(22)} + \varepsilon_D \tag{27}$$

$$R_D^{\log} = \alpha_0 + \alpha_1 R_{D-1}^{\log} + \alpha_2 R_{D-1}^{\log,(5)} + \alpha_3 R_{D-1}^{\log,(22)} + \delta_1 TC_{D-1} + \delta_2 TC_{D-1}^{(5)} + \delta_3 TC_{D-1}^{(22)} + \varepsilon_D \tag{28}$$

Lastly, to investigate the effect of mixing both realized ranges and daily ranges into one equation (i.e. to infer whether these measures contain useful and different information), we define a HAR-R specification

$$R_D^{\log} = \alpha_0 + \alpha_1 R_{D-1}^{\log} + \alpha_2 R_{D-1}^{\log,(5)} + \alpha_3 R_{D-1}^{\log,(22)} + \phi_1 RR_{D-1} + \phi_2 RR_{D-1}^{(5)} + \phi_3 RR_{D-1}^{(22)} + \varepsilon_D \tag{29}$$

Thus, the complicated R-LHAR-SC model is nothing but a HAR model where realized ranges are used as regressors but daily ranges are not (Eq. 26)). On top of that, the specification is enriched by information on leverage effect, average trade size and transaction count as described in (25), (27) and (28).

With all the HAR specifications clearly defined, we now turn to empirical estimations. Firstly, we focused on HAR, HAR-R and R-HAR specifications, i.e. we investigated whether realized ranges alone can be used for daily ranges modeling or whether they should at least be added to a standard base HAR specification. The estimated results are presented in Table A.17. The results of HAR-R model indicate that neither for variance nor for volatility modeling does mixing of daily and realized ranges into one model bring gains. Concurrently using both daily ranges and realized ranges as

[20] Logarithms (not levels) of average trade size and transaction count were used in all regressions contained in this thesis.

regressors does not yield any benefit, indicating that daily ranges do not contain any useful information other than information on variance (in line with expectations).

Moreover, results from HAR and R-HAR models indicate that neither variable is strictly superior in terms of goodness-of-fit when modeling daily ranges. The increase in $AdjR^2$ is relatively small and as heteroskedasticity induced a bias into this measure of fit quality, comparisons are limited. For this reason, all further models are evaluated with either daily or realized ranges being the RHS measure of (average) volatility, as neither shows clearly superior to the other.

In-sample estimation results of all specifications are presented in Table A.18 for volatility models and Table A.19 for variance models. In volatility modeling, square roots of realized ranges were used as regressors instead of realized ranges, on the other hand in variance modeling, squared daily ranges were modeled. For relieving hetoroskedasticity, HAC method was used.

Focusing firstly on Table A.18 (volatility modeling), we do not observe any significant increase in $AdjR^2$ for either model. Contrary to Corsi, the "bad-news" effect is not consistently significant in all models. Moreover, only in LHAR-S model were all three "bad-news" terms found to be significant at least on a 95% critical level. In general, however, only the first-lag "bad-news" effect seems to be present in the data. Next, when comparing HAR vs. R-HAR and LHAR vs. R-LHAR (i.e. we compare the benefit of using realized ranges for predicting daily ranges), we see that R-HAR and R-LHAR have a higher count of significant parameter estimates. Most notably, in HAR/LHAR models the lack of autoregressive dependency of order one is rather surprising. Since this anomaly is not present in R-HAR/R-LHAR models, where realized ranges are used as regressors, this can only be caused by the noise included in daily ranges.

Focusing on the added benefit of including transaction count and average trade size (-S/-C specifications), we cannot observe any significant effects stemming from these variables. Comparing the models enriched by these trading activity variables to models without these variables, we find only weak significance of monthly average trade size. When investigating the effect of including transaction count into a HAR model, we find no significance of estimated parameters under any specification.

Following to Table A.19 (variance modeling), we firstly find a consistently significant "bad-news" effect from the previous day (in all specifications), which is in line with the well documented leverage feature of financial returns. However, except for two specifications, weekly and monthly "bad-news" effects are absent, contrary to Corsi & Reno's results. Comparing squared daily ranges and realized ranges for squared daily ranges forecasting, we cannot claim one to be strictly superior to the other. Similarly as for the volatility equation discussed in the previous paragraph, neither transaction count nor average trade size bring significant information related to squared daily ranges. Parameter estimates of these variables are significant only in several specifications. Considering the practically zero increment in AdjR^2, which could also be only due to insufficient penalization of R^2 for parameter count, we find no practical value in augmenting a HAR variance model with these trading activity variables.

The results of previous two paragraphs demonstrate that trading activity variables which are highly correlated with daily ranges do not have any significant modeling power. We are led to the question of whether this conclusion holds generally for all volatility/variance measures or whether it is specific to daily ranges only. For this purpose, we evaluated the same battery of models as in the previous two paragraphs. However, this time the roles of realized and daily ranges were swapped. This means that realized ranges were used as explained variable and in the R-HAR specification, daily ranges were used as regressors. Any significant impact of average trade size or transaction count in these HAR models would indicate a strong difference in the possibility to model daily as opposed to realized ranges via trading activity variables. However, as presented in Tables A.20 and A.21, no systemic impact of these measures of trading activity is observed. In accordance with intuition, in this setting R-HAR specification is inferior to a HAR specification for both volatility and variance. In other words, we cannot gain better realized ranges predictions by using a noisier measure of volatility (daily ranges). Considering, however, the data-intensiveness of realized ranges as opposed to daily ranges, the drop in explanatory power between HAR and R-HAR as judged by AdjR^2, is rather small[21]. Plainly speaking, even realized ranges can be modeled by daily ranges to a great extent. Even though this kind of modeling does not make sense once we have realized ranges at our disposal, the attained AdjR^2 in R-HAR specification is very surprising. Turning to other aspects of the results, a consistent finding is a strongly significant weekly "bad-news" effect (in all models where this term is included) and in all models of the R-LHAR class, we find strong significance of all "bad-news" effects. A noteworthy fact is a consistently significant parameter of previous

[21] From 74.7% to 70.4% in volatility equation and from 72.3% to 66.8% in variance equation, respectively.

day's (realized as well as daily) range in all specifications, which is in contrast to the findings of tables A.18 and A.19 (previous paragraphs), where an AR(1) dependence was missing.

To sum up the results of this section, we conclude that when modeling daily ranges best results are obtained from an R-HAR specification, while for squared daily ranges modeling, LHAR specification seems the most plausible. The added benefit of having a more precise measure of volatility (realized range) brings only a small increase in daily ranges modeling in-sample performance. High noise content of daily ranges as a LHS variable is responsible for worse modeling possibilities by means of daily ranges as opposed to realized ranges and the use of realized ranges can not make up for the high noise content of daily ranges on the LHS. We managed to show that daily ranges do not contain any useful information not captured in realized ranges and that the difference between these two variables in pure information noise. A striking result of our work presented here is the extent to which realized ranges can be explained by daily ranges. Taking into account the low efficiency of daily as compared to realized ranges, we expected the capabilities of daily ranges to predict realized ranges very low. Contrary to Corsi & Reno (2009), there seems to be no connection between the "bad-news" effects measured over different horizons and daily volatility/variance. The significant weekly "bad-news" effect in some specifications is rather an exception than a rule. Lastly, including measures of trading activity into a HAR model does not bring any improvements for either volatility or variance modeling. Thus we can conclude that their information is already fully reflected in lagged volatility/variance measures themselves.

5.1.2. HAR modeling including lagged sessions' information

In Section 1.1.7, possible dependencies between daily ranges and lagged values of trading activity variables and volatility measured over separate trading sessions were suggested. Here, we investigate the possibility of augmenting previously presented HAR specifications by these lagged session variables and whether such augmentations lead to better fits.

As the number of possible regressors is high (5 variables for each of 9 sessions), finding an answer to our question (added benefit of using lagged sessions' information) is tackled using two approaches. In the former one, we augment the base HAR model

(24) by all lagged values of one measure, e.g. we investigate the additional explanatory power of ranges of all lagged sessions. In the latter approach, we augment the base HAR specification by all volatility/trading activity variables related to one specific session. For example, we investigate the additional explanatory power of trading taking place in the Asian session of the preceding day.

Estimation results of the former approach for daily ranges are summarized in Tables A.22/A.23. The results suggest that anything but a few selected variables has an impact on the next day's range. Most notable increases in explanatory power are gained from the inclusion of American2's realized range and realized variance. However, if we have these at hand, we can predict daily ranges by realized ranges which was proven earlier to be the preferred option. Overall, the increases in explanatory power of the models fall behind our expectations and the decomposition of neither variable among separate sessions brings a strong predictive advantage. This conclusion holds both for daily ranges and their squares. To distinguish between features typical for daily ranges and for volatility measures in general, we ran the same battery of models for realized ranges and the results are summarized in Tables A.24/A.25. Variables capturing trading activity have no bearing on the next day's realized range. However, the decomposition of lagged daily realized ranges/variances into realized ranges/variances of separate lagged sessions seems to bring some gains. In the volatility equation all parameter estimates related to sessions' realized ranges/variances are significant and in the variance equation at least some of these parameters are significant, in contrast to other investigated variables. Unfortunately increases in explanatory power stemming from sessions' RR and RV are not significant[22]. Our motivation for delving into this exercise was to infer whether it is some specific session's data which is responsible for the autoregressive property in daily/realized ranges. For example, if the decomposition of realized ranges among sessions produced an increase in $AdjR^2$, today's volatility would most likely depend strongly on the realized ranges of the most important sessions, while the information provided by the least active sessions would just create noise in the realized range. However, this turns out not to be the case and we can say that the all trading sessions (even the least active ones) bear some important information for volatility modeling. In other words, neither session can be disregarded from volatility measurements due to, for example, its low traded volume.

[22] Problems with multicollinearity (as whole days realized measures are a sum of sessions' realized measures) were most likely countered by the exclusion of non-trading session from all our regressions.

Next, we turn to the approach of adding all trading variables related to one specific session of the previous day. Estimation results for daily ranges modeling are to be found in Tables A.26/A.27[23]. Confirming the results of the previous paragraph, neither session has such a strong bearing on the succeeding daily range that we could exploit it in any manner. Despite several terms (mostly realized ranges as in the previous paragraph) are significant, no dramatic increase in the share of explained daily ranges information is observed. Testing the same hypothesis on realized ranges, we obtain estimates in Tables A.28/A.29. While parameters related to some sessions are significant again, $AdjR^2$ values remain basically unchanged.

To sum up this section, we showed evidence of no added benefit of decomposing the previous day into separate sessions. Hence, we conclude that all information relevant for daily volatility prediction is already contained in lagged daily volatility measures. On top of that, we find that even information provided by the least actively traded session is important for the next day's range prediction. Both findings are very strong results of our work.

5.1.3. HAR modeling including non-lagged sessions' information

In the preceding section we investigated the usefulness of information provided by lagged trading sessions. In general, no useful information in past trading sessions was found. In this section, we wish to infer whether it is possible to obtain more precise daily volatility predictions throughout the day as separate sessions end and their volatility/trading activity measures are incorporated into a model on-the-fly. To give an example of the this approach, American investors might create a one-day-ahead forecast of daily or realized ranges based on information available at the end of the previous trading day. Later, these predictions could be made more accurate by including information related to preAsian, Asian and other sessions preceding the American session. Even though these sessions are not likely traded by American investors, they can provide useful information.

To explore this venue, we added non-lagged realized ranges, ranges, trade count and average trade size of different sessions to the basic HAR specification. Firstly, we do this in a non-cumulative manner, i.e. the benefit of adding each session separately is

[23] Due to frequent multicollinearity problems, realized variance was omitted from variables describing each trading session.

investigated. Later we proceed to build forecast updates that utilize all intraday information available at the end of some trading session. For comparison purposes, we estimate the volatility forecast updating possibility both in daily ranges and realized ranges.

Estimation results for the non-cumulative approach are presented in Tables A.30/A.31 for daily ranges and A.32/A.33 for realized ranges. The results for daily ranges are indicative of a possibility to obtain more accurate volatility forecasts using up-to-date information. Each session's information is capable of producing a marked improvement in a volatility forecast, with most importance to be attributed to realized ranges and ranges, which are significant for nearly all sessions. On the other hand, knowing the average trade size seems to have no real benefit as it is only significant in models using information available late in the day. Turning to estimation results for realized ranges forecasts, the same pattern appears. In line with intuition and previous results, sessions' realized ranges are more useful than simple ranges of these sessions.

In the last exercise related to HAR specification, we investigate the cumulative adding of information related to today's sessions. For example, by adding information stemming from the preAsian, Asian and preEuropean session, we put ourselves in the position of traders active during the European session. We investigate, whether European traders can gain significant benefits by using information from already passed sessions. Moreover, we are interested in whether we can obtain better and better daily volatility forecasts as time passes.

We report estimation results in Table A.34/A.35 for daily ranges and Table A.36/A.37 for realized ranges. The results for daily ranges indicate a strong possibility to provide more accurate predictions of daily volatility as more information becomes available. More specifically, entities that start to trade in the American1, American2 sessions can gain nearly twice as precise forecasts of the current day's range compared to traders in the preAsian session. As investigating the possibility of gradual forecast updating it not present in the academic literature yet, this result is an important contribution of our work to the current body of knowledge on volatility modeling. Another nice feature of obtained results is that once a variable is significant in one model, it remains significant even when information from following sessions is added. This is intuitive, as different sessions are expected to provide different information due to the presence and activity of different trading entities. Unfortunately, neither investigated variable (realized range, range, ...) has universal significance in all sessions. Also, it is unclear why in some sessions it is realized range and in other simple range that is significant for the days range prediction. Lastly, focusing on the same set of estimations for realized range forecast updates, we find a strong explanatory power of each sessions' realized range for the whole day's realized range. This result stems naturally from the fact that whole-day realized range is a sum of separate sessions' realized ranges. However, even though the idea is very simple, the practical benefit is tremendous. Just as for daily ranges, traders operating during the American1 and American2 session can predict whole-day realized range with much higher precision using data on already passed sessions, giving them an advantage over their Asian and European counterparts. Turning to other regressors, transaction count and average trade size are significant in only a marginal number of cases, which simplifies the model strongly.

To sum up, our results show that participants in later sessions can benefit much from considering data provided by the market up until the time a trade is taken. This result holds irrespective of whether daily or realized ranges are modeled. Focusing, for example, on traders from American1 and later sessions, gains in predictive power are approximately 20+ percentage points of $AdjR^2$ for daily ranges and 25+ percentage points for realized ranges. Even though the results are somehow mixed (no one variable is significant in all sessions), magnitudes of predictive power increases suggest that these model improvements are not a result of simple curve fitting.

5.1.4. HAR modeling - concluding remarks

Corsi's HAR approach to volatility modeling is suitable both for daily ranges as well as realized ranges modeling. Our results suggest that while some benefits can be gained by predicting daily ranges by realized ranges, these benefits are small. Thus in the absence of intraday data daily ranges can be modeled by average daily ranges of the last day, week and month without having to worry about fit quality.

A HAR model of realized ranges regressed on lagged realized ranges produces a much better fit than a HAR model of daily ranges regressed on lagged daily ranges. This stems naturally from the fact that daily ranges are noisy and having a noisy LHS variable can never result in a great fit. However, when we have an exact volatility measurement as LHS variable, we can attain very good fits by using even noisy proxies, as demonstrated by high $AdjR^2$ of realized ranges regressed on daily ranges.

Our results show no benefit of using lagged trading sessions' information compared to using previous day's information as a whole. From this we conclude that lagged daily volatility reliably aggregates all past information relevant for daily volatility modeling. However, gradual updating of end-of-day daily volatility forecasts by up-to-date information provides significant gains in predictive accuracy and thus works to the benefit of traders active in later sessions of the day.

As both the idea of intraday volatility forecast updating and the result that lagged daily volatility captures all information relevant for daily volatility modeling are missing in current volatility related literature, these findings create a core of this thesis' contribution to volatility related base of knowledge.

5.2 CARR

As stated earlier, several approaches can be used for daily ranges modeling. In this section, we continue by assuming a different than HAR model for the task at hand. Firstly, we briefly introduce the model. We describe the estimation procedure and follow to estimates on our dataset. At the end of the section, we provide comparisons between the new model and HAR model.

One stream of volatility related literature focuses on so called Multiplicative Error Models (MEM), which are suitable for modeling any positive-valued variable. Probably the best-known application of a MEM model is the autoregressive conditional duration (ACD) model proposed by Engle & Russell (1998) for trade durations. Application of MEM modeling to volatility are, for example, the seminal work of Engle (2002) which was followed by a stream of alike literature. Also, the already mentioned work of Engle & Gallo (2006) combines features of VAR and MEM modeling.

An example of a MEM model focused specifically on daily ranges modeling can be found in Chou (2005). Chou combines the idea of GARCH volatility modeling, where volatility predictions are modeled by an ARMA process, and the logics of a MEM model. The newly generated model is coined conditional autoregressive range (CARR) model of order p,q. The specification can be written as

$$R_D^{\log} = \lambda_D \varepsilon_D \qquad \lambda_D = \omega + \sum_{i=1}^{q} \alpha_i R_{D-i}^{\log} + \sum_{i=1}^{p} \beta_i \lambda_{D-i} \tag{30}$$

where ε_D is assumed to follow a positive-valued distribution with unity mean. More specifically, in MLE estimates of the model Chou assumes either Weibull distributed ε_D of Exponentially distributed ε_D (which are a special case of Weibull distributed ones). Investigations carried out on the S&P 500 futures contract reveal superior volatility forecasts of CARR models as compared to GARCH models, the rejection of null hypothesis of exponentially distributed ε_D, presence of a strong "bad-news" effect in the volatility-of-volatility equation as well as a benefit of adding absolute returns (as a complementary measure of volatility) into λ_D equation. Exponential distribution is found to be incorrect for ε_D, while with the assumption of Weibull distributed ε_D, estimated residuals $\hat{\varepsilon}_D = R_D^{\log} / \hat{\lambda}_D$ are nearly Weibull distributed.

In our work, we followed the methodology of Chou and estimated a CARR model with Exponentially and Weibull distributed ε_D term via MLE. Likelihood functions can be written as

$$L_{Exponential}(\alpha_i, \beta_j; R_1, \ldots, R_n) = -\sum_{i-1}^{T} \left[\ln(\lambda_t) + \frac{R_t}{\lambda_t} \right]$$
$$L_{Weibull}(\alpha_i, \beta_j, \theta; R_1, \ldots, R_n) = \sum_{i-1}^{T} \ln\left(\frac{\theta}{R_t}\right) + \theta \ln\left(\frac{\Gamma(1+1/\theta) R_t}{\lambda_t}\right) - \left(\frac{\Gamma(1+1/\theta) R_t}{\lambda_t}\right)^{\theta} \tag{31}$$

In case $\theta = 1$, the Weibull distribution collapses into the Exponential one and this property can be used to evaluate which distribution is more appropriate (by testing $\hat{\theta} = 1$).

Estimations on our dataset were carried out according to (31) using robust standard error estimation techniques (QML covariance matrix). The results for daily ranges prediction are presented in Table A.38 for both $\varepsilon_D \approx Exp(\bullet)$ and $\varepsilon_D \approx Weibull(\theta, \bullet)$ respectively. Table A.49 contains estimation results of the same exercise applied onto square roots of realized ranges. Both for daily and realized ranges only a model for volatility was estimated, as modeling variance resulted in negative variance predictions. An attempt to correct for this situation by rewriting (30) into

$$R_D^{\log} = \lambda_D \varepsilon_D \qquad \lambda_D = \omega + \sum_{i=1}^{q} \exp(\alpha_i) R_{D-i}^{\log} + \sum_{i=1}^{p} \exp(\beta_i) \lambda_{D-i} \tag{32}$$

produced strongly upward-biased predictions of squared daily ranges and realized ranges, hence we opted for not modeling variance further.

Optimal values of lags p, q were determined based on parameter significance and in both cases $p = q = 1$ was found to be optimal. Recalling that the previously investigated HAR model is just a special case of AR model with many lags, we observe a situation which is classic in volatility modeling. Before the onset of GARCH modeling, ARCH models were used for modeling latent volatility and usually many lags were needed to model away all dependencies in volatility. This drawback of ARCH models was resolved by GARCH models, where the added GARCH term (actually an MA term in the ARMA process of volatility predictions) was capable of replacing many distant lags in ARCH models. As a result, GARCH(1,1) is the most frequently found specification of volatility that is capable of capturing all dependencies in volatility. In the case of HAR versus CARR, we observe the same situation. While in a HAR model, the long-term volatility plays a significant role in all specifications, a CARR model only needs one lag of each variable (range and ε_D) to be complete. This result was confirmed when we tried to include medium-term and long-term volatility into the CARR specification and both turned out insignificant (results not presented here for brevity reasons).

Comparing Exponentially and Weibull distributed error terms ε_D, we find strong evidence for $\varepsilon_D \approx Weibull(\theta,\bullet)$ being the correct of these two specifications as we can reject the hypothesis of $\hat{\theta} = 1$ for both daily ranges and square root of realized ranges. However, when comparing other parameter estimates we see that changes in these parameters induced by considering Weibull as opposed to Exponential distribution for ε_D are negligible, which is a result in accordance with those of Chou's. Neither do in-sample fitted values change, as can be seen in Figures A.40/A.41.

Before delving into augmentations of the CARR(1,1) model, we compared the in-sample modeling performance of CARR(1,1) model with Weibull distributed errors to the modeling performance of a base HAR specification (24). In-sample fitted value plots for daily ranges and square roots of realized ranges are depicted in Figures A.42 and A.43. When neglecting the buildup period of MLE estimation, we observe a striking similarity of fitted values indicating a near identity of both models. This is in line with the previous discussion regarding AR models with many lags and corresponding ARMA models, which need only few lags to capture the same information. For these reasons, we did not delve into CARR model augmentations, as estimation results and sessions' information significance would nearly certainly be the same as in case of a HAR model. Hence, we can conclude by saying that CARR and HAR models provide nearly the same level of modeling performance and this is true regardless of whether the MEM error term is assumed to be Exponentially or Weibull distributed. HAR models offer the advantage of avoiding maximum likelihood estimation. MLE in our case turned out to be impossible in the case of variance variables, where a negative prediction of variance measure prevented the evaluation of log-likelihood function. Moreover, a great advantage of HAR models compared to CARR is the possibility to use realized measures of volatility as regressors and their general simplicity.

Summing up, in this section we consider a CARR model for daily ranges. Estimation results show that assuming Weibull distributed residuals is correct, however estimation problems arise when working with variance data (squared daily ranges, realized ranges). Strikingly similar in-sample fits of a base HAR model and a CARR(1,1) model with Weibull disturbances correspond well to the fact that HAR is actually an AR model with a high number of distant lags while CARR(1,1) model is in fact an ARMA(1,1) of daily ranges. Due to their similar modeling performance no further investigations are performed and we move to the next model.

5.3 Cointegration of high and low prices

In this section, we present the last model considered in this thesis. Owing to its different approach to daily ranges modeling, we can investigate the effect of different session variables than in Sections 5.1.2 and 5.1.3. Firstly, we introduce the model theoretically. Next, we provide estimations of the model's basic specification as well as of an augmented specification found in literature. Lastly, we investigate the added benefit of using lagged sessions' trading variables as well as the possibility to obtain updated volatility forecasts throughout the day.

We start the theoretical introduction into co-integration by defining spurious regression. A spurious regression arises in economics when one tries to relate two variables which share a common trend and/or seasonal pattern in an equation of these variables' levels. As an example, we can use an economy's annual gross product Y_t and expenditures on consumption C_t which are known to be a share of the gross product. When we run a regression of the following specification

$$Y_t = a + bC_t + \varepsilon_t \tag{33}$$

the obtained fit is very good, but the conclusion of a relationship between Y_t, C_t other than a common trend/seasonal pattern is flawed, unless these series are co-integrated.

Co-integration was firstly proposed in Granger (1981) and following work on it includes, for example, Granger & Weiss (1983) and Engle & Granger (1987). Without a formal definition, further explanations would be cumbersome. Thus, let us assume two time series x_t, y_t which are both integrated or order one. We call these series co-integrated if for some $a \in \Re$ the linear combination $y_t - ax_t$ is integrated of order zero, i.e. stationary.

As stated, two co-integrated series share a common trend/seasonal pattern and at the same time deviations from this common component are stationary. In order to model such series, one has to consider both long-term information (common component) as well as short-term information (deviation from common component). For this purpose, the class of error-correction models (ECM) was developed and these models have a long tradition in time series econometrics dating back to Sargan (1964). Continuing with our previous example of co-integrated series x_t, y_t, the error correction term is defined as

$$\xi_t = y_t - a x_t \tag{34}$$

and the error correction model of x_t, y_t is then defined as

$$\Delta y_t = \alpha \xi_{t-1} + \gamma \Delta x_t + u_t \tag{35}$$

where u_t is *i.i.d*. The error correction term ξ_{t-1} in (35) can be thought of as an equilibrium error from the previous period. For example, if $\Delta x_t = 0$ and $\xi_{t-1} > 0$ then y_{t-1} was above its equilibrium value. In order to compensate for this, Δy_t needs to be negative to revert to equilibrium. From this it naturally follows that for the system of x_t, y_t to be stable we need $\alpha < 0$. Lastly, from the definition of co-integration all variables of (35) are stationary, hence spurious regression is not present anymore.

Cheung (2007) investigates the usefulness of vector error-correction-models[24] for daily ranges modeling on several stock indices (S&P 500, NASDAQ and DJIA). Since daily high and daily low prices are expected to be integrated of order one while the daily range is stationary (as shown in Section 3.3) an error correction model for changes in daily highs and daily lows could be formulated where daily range would serve as the error correcting term. Cheung's tests for co-integration via ADF testing as well as via Johansen procedure confirms that $\Delta h_D, \Delta l_D \approx CI(1,1)$, the co-integrating vector is found to be approximately $[1, -1.007]$ for all stock indices under investigation. Thus, daily ranges are found to be a close approximation of the stationary sum of daily high and low prices' deviations from their common trend and seasonal pattern. Knowing this, Cheung adopts a VECM model for $\Delta h_D, \Delta l_D$, i.e.

$$\begin{gathered} X_D = (\Delta h_D, \Delta l_D)^T \\ X_D = \alpha + \sum_{i=1}^{p} \beta_i X_{D-i} + \gamma R_{D-1}^{\log} + \xi D + \varepsilon_D \end{gathered} \tag{36}$$

[24] i.e. a set of error-correction-models applied jointly onto co-integrated variables

and the error correcting term (daily range) turns out significant. As common in the finance literature, Cheung checks for the effect of day-of-the-week dummies (vector D). However, the effect is negligible both in terms of parameter estimate significance and in terms of added explanatory power, hence the base specification is (36) with day-of-the-week dummies excluded. The best value of lag parameter p is found by Cheung to be 6 - 7 depending on instrument and generally the predictive power of the model is low, specifically from 8% to 17% in terms of AdjR^2 depending on instrument and variable (Δh_D or Δl_D).

In order to improve the model's predictive power, Cheung tries to include several exogenous variables. Firstly, the added benefit of using lagged and contemporaneous de-trended traded volume is investigated. Despite strong parameter significance improvements in AdjR^2 are only minor. Trying to improve the model in an alternative way, Cheung draws upon the intuitive idea that more price observations for each instrument should give more information about the characteristics of its evolution. Hence he adds changes of daily open and daily close prices as well as daily returns into (36). In this augmented model, the vast majority of added variables are significant and the model's predictive power rises dramatically (to levels of 37.6% to 48.9% in terms of AdjR^2).

In a follow-up work, Cheung et al (2010) investigate the possible profitability of daily ranges predictions obtained via a VECM model like (36). These predictions are tied exclusively to exotic options traded on the Hong Kong market, namely the so called Callable bull/bear contracts. Even though only the base specification (36) is used to produce daily ranges forecasts, the performance of most strategies is good, i.e. they are profitable even net of transaction and interest costs. Strategy results vary depending on parameters settings and the issuer of CBBCs, however, taking into consideration the simplicity of these strategies, the results are very encouraging.

Lastly, He, Kwok & Wan (2010) investigate the possibilities of modeling changes of daily highs and lows using various techniques, including the random walk model, ARIMA-type forecasting and VECM modeling. In-sample fit compared by traditional means of MAD, MSE[25], percentage of correct directional changes of ranges and trading strategy profitability speak clearly in favor of VECM modeling.

---

[25] Mean Absolute Deviation and Mean Squared Error, respectively.

In this section, we delve into ranges modeling via VECM of daily high and low prices because in no other model can we exploit their co-integration. As daily ranges are stationary and they are the difference of daily high and low prices, this problem is a textbook example of what VECMs should be employed on. Further, the use of a VECM is desired due to the difference of this approach compared to those so far investigated. HAR, CARR and VECM differ firstly in their specific applications - HAR and CARR models are nearly exclusively used for volatility modeling, while VECMs are used in many fields of finance (e.g. the modeling of consumers' consumption changes as a response to changes in income, spreads between long-term and short-term interest rates, etc). An interesting consequence of this difference is the fact that in a VECM of highs and lows, modeling volatility is not the topic of interest per se, instead daily ranges modeling comes out of the model as a by-product.

Considering the high impact of enriching the base VECM specification by variables such as changes of daily open and daily close prices, we evaluate the added benefit of using these variables following Cheung. As intuition behind adding these variables is that more price observations should bring more information exploitable for modeling, we take this idea a step further. Namely, we also incorporate changes in open and close prices measured over different trading sessions. While there might be no effect of lagged sessions' variables on ranges per se (as in HAR modeling), there might be an effect of these variables on changes of ranges. Since the added benefit of regressing on changes in sessions' open and close prices cannot be inferred in a HAR/CARR model, using these changes in a VECM is the only choice of evaluating their usefulness.

In this section, we proceed exactly as we proceeded in Section 4.1 on HAR modeling. Firstly, we formally investigate the assumptions of a VECM model, secondly we assess the predictive power and proper specification of (36) on our dataset. Next, we investigate the usefulness of Cheung's augmentations (changes in daily open, close prices and daily returns) and lastly, we turn to improving the model with sessions-related variables.

For the identification of a co-integration relationship between daily highs and lows, we follow the Engle-Granger test. As the ADF-test for daily highs and daily lows confirms the $I(1)$ property[26], we run a regression of

[26] ADF test with five lags and a constant, obtained p-values for lows and highs were 0.2236 and 0.2195.

$$h_D = \alpha + \beta l_D + \varepsilon_D \tag{37}$$

to obtain the co-integrating vector. Estimation results are presented in Table A.44. As ADF testing for the presence of unit-root in residuals of (37) returns a p-value of practically zero, we can conclude that daily highs and lows are co-integrated. The estimated co-integrating vector is $[1,-0.967]$ and taking into account the standard error of β, we can reject the hypothesis of $\beta = 1$. For this reason we chose not to approximate the error correcting term by lagged daily range but instead to calculate it accurately as $EC_D = h_D - 0.967 l_D$. As the next step, we investigate the appropriate number of lags p based on parameter significance. Following Cheung, we set the number of lags equal for both equations (as in eq. 36). Estimation results presented in Table A.45 show that in general, changes in daily low prices have a much stronger and longer memory of preceding daily high/low changes. The lagged error-correction term is insignificant in the Δl_D equation. For changes in daily highs, on the other hand, a clear dependence is present only on previous change of daily low and the error-correction term. Signs of significant parameters in both equations are according to expectations and speak in favor of an mean-reverting process in daily ranges. For example, in daily lows equation, lagged increases in high prices are followed by an increase in daily low prices, so as to keep the daily range in bounds (positive parameters of $\Delta h(-i)$), whereas lagged drops in daily lows are followed by an increase in lows, which again presses the daily range towards its mean value (negative parameters of $\Delta l(-i)$). The daily ranges process would be explosive it there wasn't for these properties. The significant error-correction term parameter negativeness in daily high equation confirms this behavior as an increase in daily range (which is an approximation to the error-correction term) pushes down the next day's high price, hence likely decreasing the range.

Lastly, the explanatory power of the model for both equations is rather high, considering on one hand the low number of significant lags for the daily high equation and on the other hand the explanatory power of base model in Cheung's paper. This high reading of AdjR^2 might be caused by the crisis period, which in general increases the explanatory power of volatility models. However, estimating the base specification on the second half of our dataset (no volatility outliers), the explanatory power drops only by one percentage point (estimation results not presented for brevity reasons).

Having established the base VECM model, we turn to the investigation of day-of-the-week effect and for correctness purposes, we also try to enrich the model by lagged traded volume, average trade size and transaction count. Similar to Cheung's results, we do not find any support for the presence of day-of-the-week effect. Neither dummy has a significant parameter estimate, at the same time no increase in AdjR^2 takes place[27]. The same conclusion applies to lagged trading activity variables, hence we omit all of these variables and day-of-the-week dummies in further models. Estimation results of the enriched model are listed in Table A.46.

Next, following Cheung's approach, we enrich the base VECM by $\Delta o_{D-i}, \Delta c_{D-i}, co_{D-i}$ which stand for the lagged changes in daily open and daily close prices and lagged daily returns, respectively. Model specification thus changes to

$$X_D = \alpha + \sum_{i=1}^{p} \beta_i X_{D-i} + \sum_{j=0}^{q} \gamma_j \Delta o_{D-j} + \sum_{k=1}^{r} \delta_j \Delta c_{D-k} + \sum_{m=1}^{s} \phi_m co_{D-m} + \varphi EC_{D-1} + \varepsilon_D \tag{38}$$

The Δo_{D-m} terms are taken from $m = 0$ as we can utilize today's open price in our predictions. Optimal parameters of q, r, s were chosen based on parameter significance while keeping $p = 6$ as in the base specification. Estimation results are presented in Table A.47 for $q = 4, r = 2, s = 1$. Several points deserve mentioning:

- Significance of the daily highs equation parameters changes dramatically. After adding $\Delta o_{D-k}, \Delta c_{D-l}, co_{D-m}$, lagged values of changes in daily highs become strongly significant and their values change dramatically. Also, the 2nd and 3rd lags of changes in daily lows are significant. From the added variables, changes in daily opens seem to be the most important for the daily high equations, as lags up to the fifth one are strongly significant. AdjR^2 for the daily high equation enjoys a more than two-fold increase, hence explanatory power is significantly better.
- In the daily lows equation significance of parameters changes as well. It turns out that normalization of lagged changes in daily highs and daily lows by lagged changes in daily opens produces significant results in both the Δh_D and Δl_D equation and in both equations the model's predictive power enjoys a significant rise.

[27] This is an interesting result if we recall the distribution of daily ranges by days of the week in Figure 3.

- The significance of error-correction term remains unaffected in both equations, despite heavy changes in significance of other parameters. The null hypothesis of no autocorrelations in residuals is not rejected just as in the base specification, hence we consider the augmented model well specified.

So far, modeling endeavors were inspired by the work of Cheung's. Our results are very similar in terms of significance of added parameters as well as in terms of increased modeling power. In the next paragraphs, we follow up by incorporating information related to different trading sessions. We follow the methodology implemented in Section 4.1. It means that we firstly try to enrich specification (38) by separately adding a certain variable measured over lagged sessions (e.g. include lagged changes in opens of all sessions) and then by separately adding changes of all variables related to one specific lagged session. As a VECM is a model in changes, session variables of interest in this exercise will be changes of volume, changes of average trade size, changes of transaction count, changes of realized range and changes of OHLC prices of each session. Estimation results for exercises in this section are voluminous[28], hence we do not present detailed estimation results here. These detailed estimation results can be obtained from authors upon request.

Turning to the former approach, we find that in general, adding lagged variables of one kind measured over separate sessions does not bring any dramatic increase in explanatory power of the model. The extent of model improvements are captured in the table below[29].

	Specification								
	base	**+Volume**	**+Trans**	**+TradeSize**	**+RR**	**+Open**	**+High**	**+Low**	**+Close**
AdjR^2 (ΔH)	0.547	0.551	0.554	0.547	0.566	0.558	0.561	0.577	0.553
AdjR^2 (ΔL)	0.495	0.495	0.493	0.495	0.497	0.500	0.507	0.558	0.499
AIC	-15.96	-15.94	-15.93	-15.96	-15.97	-15.96	-15.95	-16.09	-15.97
BIC	-15.73	-15.55	-15.48	-15.73	-15.54	-15.44	-15.39	-15.65	-15.64
HQC	-15.87	-15.79	-15.76	-15.87	-15.81	-15.76	-15.74	-15.92	-15.84

Table 3: VECM improvements after adding changes of variables measured over lagged sessions.

[28] 8 variables per session, 5 lags per variable.
[29] A maximum of 5 lags were used in these models and models presented in Table. A.4. For example, in the *+Volume* column changes in session trading volumes over the preceding five days were added to the base specification for each of 9 sessions.

Drawbacks of this exercise's results are two-fold. For neither added variable were we able to obtain significance in any session, that would be strong and across all lags. Instead, dependencies on, e.g. 2nd lag of change of preAsian volume and 4th lag of change in European volume were found, which are both counter-intuitive and most likely a result of curve-fitting. Secondly, increases in explanatory power are negligible compared to the number of added parameters (only slight improvement of ICs). Before closing this exercise, we applied the logics of (38) onto separate sessions, i.e. instead of expecting changes of daily high/lows to depend on lagged sessions' high/low changes, we investigated the dependence of daily high/lows changes on lagged sessions' normalized high/low changes. Estimation results, however, do not support this hypothesis, as can be seen from the table below. Due to estimation output size, we omit it again.

	Specification		
	base	**+Open+Low**	**+Open+High**
AdjR^2 (ΔH)	0.547	0.568	0.565
AdjR^2 (ΔL)	0.495	0.509	0.505
AIC	-15.96	-15.92	-15.89
BIC	-15.73	-15.00	-14.94
HQC	-15.87	-15.57	-15.53

Table 4: Model improvements after adding changes of high & open and low & open prices of different lagged sessions.

Next, we investigate whether a certain lagged session is important for the evolution of daily high and low prices. Thus we add lagged changes of all variables tied to one session and investigate the model's improvement in modeling power. The obtained results are presented in the table below.

	Specification									
	base	**+preAs**	**+As**	**+preEu**	**+Eu**	**+preAm**	**+Am1**	**+Am2**	**+postAm1**	**+postAm2**
AdjR^2 (ΔH)	0.547	0.546	0.556	0.552	0.552	0.550	0.545	0.551	0.551	0.551
AdjR^2 (ΔL)	0.495	0.498	0.495	0.507	0.500	0.503	0.490	0.490	0.505	0.492
AIC	-15.96	-15.90	-15.93	-15.92	-15.92	-15.93	-15.88	-15.90	-15.93	-15.90
BIC	-15.73	-15.31	-15.33	-15.32	-15.33	-15.34	-15.29	-15.32	-15.33	-15.31
HQC	-15.87	-15.68	-15.70	-15.69	-15.69	-15.70	-15.65	-15.68	-15.70	-15.68

Table 5: VECM improvements after adding changes of all variables related to lagged sessions[30].

Unfortunately, the addition of neither session's variables brings a model improvement. Hence, we conclude that there is no benefit in using historical session variables for improving modeling performance in a VECM approach.

As a last exercise connected to VECM modeling of daily ranges, we assess the possibility to improve forecasts of changes of daily high and daily low prices by utilizing session variables as they become available during the day, as in Section 4.1.3. For this purpose, we cumulatively add the same variables as in the previous exercise to the base model (38), i.e. we start with variables of the current day's preAsian session. Then we proceed to adding variables of the Asian session, etc. Increases in explanatory powers with additional sessions are expressed in Table 6 below. As estimation results consist of two tables with nearly 100 rows each, we do not report them here.

	Specification									
	base	**+1 sess**	**+2 sess**	**+3 sess**	**+4 sess**	**+5 sess**	**+6 sess**	**+7 sess**	**+8 sess**	**+9 sess**
AdjR^2 (ΔH)	0.547	0.566	0.650	0.661	0.757	0.780	0.828	0.912	0.919	0.923
AdjR^2 (ΔL)	0.495	0.548	0.618	0.641	0.758	0.779	0.830	0.919	0.923	0.925
AIC	-15.96	-16.07	-16.31	-16.38	-16.93	-17.07	-17.52	-19.03	-19.19	-19.27
BIC	-15.73	-15.74	-15.90	-15.89	-16.35	-16.40	-16.85	-18.27	-18.34	-18.33
HQC	-15.87	-15.94	-16.15	-16.19	-16.71	-16.82	-17.26	-18.74	-18.87	-18.91

Table 6: VECM improvements obtained by cumulatively adding all variables related to current day's sessions.

A trend of increases in explanatory power with passage of time is obvious from the results and is in accordance with the results of the same exercise when utilizing a HAR model (Section 4.1.3). However, the conclusions of this exercise differ from the conclusions of Section 4.1.3, where the current day's session ranges and realized ranges were the most useful for forecast updating. Most notably it seems that for modeling changes in daily high/low prices, the most actual change in close price is the most relevant variable, as in nearly all cases the change in close price of the just-added session is strongly significant, many times being the only significant variable of current day's trading. Just as in the results of HAR modeling, once a variable becomes significant upon being added to the model, it remains significant for the remainder of the day (except for close prices of sessions for aforementioned reasons). Unfortunately, variables that are significant in some sessions need not be significant in others (i.e. significance of realized range of preEuropean session but no significance of realized

[30] These variables were lagged changes in: session range, session realized range, session average trade size, session transaction count, session traded volume and session OHLC prices. Maximum number of lags used was 4.

range measured over other sessions), hence we again face the possibility of curve fitting driving our results. The usefulness of evolving close price for the prediction of changes in highs and lows is, however, a nice result.

As the feature of dependence of daily high/low changes on most up-to-date readings of close price changes is rather striking, we decided to simplify the cumulative model. Instead of utilizing changes of all trading variables related to the just-finished session, we focus on the added benefit of using the change in close price of each session only. By this, we can infer the relative importance of current sessions' close price change compared to the importance of other variables related to the same session. Thus, for each session we add only the change of its close price and compare the goodness-of-fit measures with those obtained in the previous Table 6.

	Base	+1 sess	+2 sess	+3 sess	+4 sess	+5 sess	+6 sess	+7 sess	+8 sess	+9 sess
$AdjR^2$ (ΔH)	0.547	0.564	0.640	0.650	0.738	0.758	0.798	0.858	0.869	0.870
$AdjR^2$ (ΔL)	0.495	0.545	0.617	0.635	0.746	0.764	0.812	0.878	0.878	0.881
AIC	-15.96	-16.07	-16.29	-16.34	-16.82	-16.93	-17.30	-18.29	-18.43	-18.49
BIC	-15.73	-15.73	-16.04	-16.08	-16.55	-16.66	-17.01	-18.00	-18.13	-18.17
HQC	-15.87	-15.87	-16.19	-16.24	-16.71	-16.83	-17.19	-18.18	-18.32	-18.37

Table 7: VECM improvements obtained by cumulatively adding changes in closes related to current day's sessions.

In general, results in Table 6 contain higher $AdjR^2$ measures for both the high and low equation. Despite the fact that these better models use many more exogenous variables, adding them is justified as the values of all information criteria are lower for these more complicated models, as compared to models with only the sessions' change of close. On the other hand, if we focus on $AdjR^2$ in both tables, we see that the added benefit of variables different than changes of session closes is not that high, as

- adding the progressing information on changes of session closes is capable of significantly improving the forecasts during the day
- the maximum difference in $AdjR^2$ between the two specifications is roughly 5 percentage points, which is rather small compared to differences between regressors count of both specifications (9 regressors per session in Table 6 versus one regressor per session in Table 7)

To conclude this section, we can state that the enrichment of a classic VECM model of daily highs and daily lows by daily open and daily close price changes indeed improves the modeling performance significantly, in accordance with previous research.

However, the inclusion of lagged sessions' variables turns out to be of no value for modeling. The only benefit obtainable from session information are session variables pertaining to the current day. Of these variables, changes in sessions' close prices play a role more significant than all other session variables combined. A possibility to obtain forecast of rising quality with passage of time is demonstrated, as in the case of HAR modeling.

6. Out-of-sample Forecasting Exercise

In the preceding sections we investigated the possibility to fit daily ranges in-sample, i.e. on data which were used for estimation. Even though some models performed better than others in-sample, an evaluation of out-of-sample forecasting is necessary to discern real-life qualities of different models. Usually, an over-fitted model (due to large number of parameters) or a model with incorrect specification can provide superb in-sample forecasts, while out-of-sample forecasting quality can be poor.

The viability of HAR models for out-of-sample volatility forecasting is a well established fact in related literature. VECM models are believed to provide good forecasts as well since their predictions are often used as inputs into trading strategies. Our motivation in this section is thus two-fold. Firstly, we wish to infer in general whether HAR/VECM models of daily ranges bring forecasts of significantly higher quality than selected benchmark models. Secondly, we wish to infer whether there are gains in using realized ranges for the prediction of daily ranges. Despite the in-sample modeling difference between HAR and R-HAR models was small, an assessment of their out-of-sample forecasting performance is what truly indicates the use of realized ranges for daily ranges prediction.

In finance literature, several methods for out-of-sample forecasting exist. For example, one step ahead (next day), five steps ahead (next week) or twenty steps ahead (next month) forecasts can be evaluated, depending on the desired use of these forecasts. Longer-term forecasts will be interesting for longer-term investors, while next day forecasts will suffice for intraday traders.

Another type of distinction is whether to use rolling-window forecasting or anchored forecasting. In the former method a certain number of observations is included in the estimation set and after making a forecast, this set moves by one observation, dropping the oldest observation and adding the one for which a forecast was being previously done. In the latter approach, the estimation dataset grows over time as no observation is dropped and only the most recent observation is added after the forecast is generated. Using this method, the estimation dataset coincides with the whole dataset, once all forecasts have been made.

In this work, we focus solely on one-step-ahead forecasts of daily ranges using a rolling window method. Firstly, our motivation for modeling daily ranges stems from a desire to provide for better money management to intraday traders, who only care about the next day's volatility. Secondly, when using anchored forecasting, quality of forecasts obtained early in the dataset and late in the dataset are incomparable, as more observations for estimation should provide for a better model. Hence, with anchored forecasting it is difficult to infer whether better forecasting performance in the late part of dataset is caused by better estimation due to a larger estimation set or by a generally better fit of the model. For this reason, we use solely rolling-window forecasting. Window length was set to the minimum number of observations allowing for a MLE estimation of all models on all rolling-windows, i.e. 400 observations, yielding 599 one-step-ahead forecasts.

For assessing differences in forecast quality several loss functions can be used. Some of the most popular loss functions are RMSE, MAE and Q-LIKE. These loss functions, for a difference series $\{e_t\}_{t=1}^{n}$ of target volatility series、and volatility forecast series $\{h_t\}_{t=1}^{n}$, are defined as

$$MSE = n^{-1}\sum_{i=1}^{n} e_t^2 \qquad MAE = n^{-1}\sum_{i=1}^{n} |e_t| \qquad QLIKE = n^{-1}\sum_{i=1}^{n} \left(\ln h_t + \sigma_t / h_t\right) \qquad (39)$$

Forecasts of better models produce lower values of loss functions. Drawbacks of these measurements are firstly the sensitivity of MSE to outliers and generally, neither squared nor absolute error needs to correspond to the forecast user's loss function. However, as loss functions of different economics agents are not identical, no error measure can express exactly which model is better under all circumstances. Patton (2011) compares the use of different loss functions in the framework of latent variable forecasting, where forecast error of the latent variable is composed of the model dependent error as well as an error introduced by observing the latent variable via some proxy (e.g. daily ranges are a proxy for true volatility of the underlying price process, i.e. by using daily ranges we commit to a measurement error). As Patton discusses, when picking an optimal loss function for proxy variable forecasting two loss function characteristics need to be fulfilled. Firstly, a model producing an optimal forecast of $h_t = \sigma_t$ needs to be identified by the loss function as the best model. Although requiring this property seems intuitive, MAE commonly used in research does not exhibit this behavior. A second desired property is loss function robustness. A loss function is called robust if the ranking of two possibly imperfect (volatility) forecasts by expected (average) loss is the same whether the ranking is done using the true conditional volatility or some conditionally unbiased volatility proxy. In other words, a robust loss function needs to be immune to the fact that forecasts errors contain measurement error of the latent variable. Only MSE and Q-LIKE functions satisfy both of these conditions, hence we only use these when comparing forecasts in our work.

Lastly, simply comparing magnitudes of RMSE or Q-LIKE does not give us information about the significance of these variables' differences across models. In order to solve for this last drawback, one has to use different methods. In this work, we used the same test as Patton, i.e. a volatility forecast comparison test based on the work of Diebold & Mariano (1995) and West (1996).

Following literature on volatility forecasting, we decided to compare the forecasting power of models discussed in Sections 5.2 and 5.4 to several benchmark models. The first considered benchmark is the random-walk model yielding the well-known naive forecast. Another benchmarks are models popular in retail trading, where the average weekly and monthly volatility are considered "good" predictors of next day's volatility. We follow two approaches with these average volatilities. Namely, forecasts equal to previous day's average weekly and monthly volatilities[31] were used as well as forecasts obtained from the following models

$$R_D^{\log} = \alpha_0 + \alpha_1 R_{D-1}^{\log(5)} + \varepsilon_D \qquad R_D^{\log} = \alpha_0 + \alpha_1 R_{D-1}^{\log,(22)} + \varepsilon_D \tag{40}$$

$$R_D^{\log} = \alpha_0 + \alpha_1 RR_{D-1}^{(5)} + \varepsilon_D \qquad R_D^{\log} = \alpha_0 + \alpha_1 RR_{D-1}^{(22)} + \varepsilon_D \tag{41}$$

with GARCH(1,1) governing the volatility of Student-distributed ε_D.

A more sophisticated AR(7)-GARCH(1,1)-t discussed in Section 5.1 was used to obtain the best of benchmark forecasts. By adding these benchmark models to our forecasting exercise, we can firstly infer whether HAR/VECM models outperform the simplest of models before turning to comparing these complicated models one with another. HAR models used herein were enriched by a GARCH(1,1)-t to make comparisons between AR(7), (40), (41) and HAR models comparable.

Lastly, only models for daily ranges forecasting were considered, as these are the only ones obtainable from a VECM of highs and lows. As the number of models estimated in this thesis is large, only base specifications and specifications with a clear increase in explanatory power stemming from additional variables were used for forecasting. Moreover, as the viability of enriching models by current day's data was clearly illustrated in both HAR and VECM sections, we do not provide out-of-sample forecasts for these specifications, as their number would be large. Hence, the list of models chosen for out-of-sample forecasting is the following:

Model	Description	Reference
RW	Random walk - volatility forecast is equal to previous day's volatility.	
SMA5	Average weekly volatility - forecast is a SMA of last five volatilities	
SMA22	Average monthly volatility - forecast is a SMA of last 22 volatilities	
SMA5GARCH	SMA5 with a constant and GARCH modeling included	Eq. (40)
SMA22GARCH	SMA22 with a constant and GARCH modeling included	Eq. (40)
RSMA5GARCH	SMA5GARCH using average realized range as predictor	Eq. (41)
RSMA22GARCH	SMA22GARCH using average realized range as predictor	Eq. (41)
AR7	AR(7)-GARCH(1,1)-t model derived forecasts	A.14
HARGARCH	Forecasts of a HAR model with leverage effect from the previous trading day, GARCH modeling included	A.18
RHARGARCH	Forecasts of a R-HAR model without any leverage effect, GARCH modeling included	A.18
VECM	Forecasts of VECM of Highs and Lows with 6 lags	A.45
VECMAUG	Forecasts of VECM of Highs and Lows with 6 lags enriched by information on changes of closing and opening prices	A.47

Table 8: List of models considered for out-of-sample forecasting evaluation.

An overview of forecasting performance as measured by RMSE and Q-LIKE is listed in the table below.

Model	RMSE	Q-LIKE	Model	RMSE	Q-LIKE
RW	0.000027	-3.452526	RSMA22GARCH	0.000016	-3.500115
SMA5	0.000017	-3.496000	AR7	0.000015	-3.500018
SMA22	0.000016	-3.499393	HARGARCH	0.000015	-3.500247
SMA5GARCH	0.000015	-3.499870	RHARGARCH	0.000015	-3.502932
SMA22GARCH	0.000016	-3.499282	VECM	0.000016	-3.498304
RSMA5GARCH	0.000015	-3.502441	VECMAUG	0.000016	-3.497844

Table 9: Average RMSE and Q-LIKE of one-step-ahead rolling-window forecasts.

As expected, naive forecasts perform the worst of all models. Mutual comparisons of other models are impossible due to small differences in both MSE and

[31] i.e. $\hat{R}_D^{\log} = R_{D-1}^{\log(5)}$ and $\hat{R}_D^{\log} = R_{D-1}^{\log(22)}$, respectively.

Q-LIKE. An apparently puzzling feature of the results is the nearly identical forecasting performance of VECM and VECMAUG, despite the latter having 3x higher $AdjR^2$ in-sample.

To assess differences in forecasting accuracy statistically, we performed Diebold-Mariano-West test for both MSE and Q-LIKE loss functions. The resulting matrix is listed below as Table 10. A negative test statistic in row *A* and column *B* indicates that model *B* provides better forecasts than model *A*.

	SMA5	SMA22	SMA5GARCH	SMA22GARCH	RSMA5GARCH	RSMA22GARCH	AR7	HARGARCH	RHARGARCH	VECM	VECMAUG
RW	-6.41	-6.02	-6.97	-6.14	-7.03	-6.28	-7.21	-6.79	-7.31	-6.82	-6.51
	-6.92	-6.79	-7.46	-6.72	-7.51	-6.83	-7.68	-7.31	-7.76	-7.43	-7.15
SMA5		-1.07	-3.67	-1.28	-3.86	-1.57	-3.49	-2.99	-4.49	-1.47	-1.00
		-1.55	-3.84	-1.40	-4.01	-1.75	-3.17	-3.07	-4.29	-1.36	-0.95
SMA22			-0.85	-0.48	-2.08	-0.96	-1.05	-1.41	-2.68	-0.06	0.20
			-0.25	0.11	-1.87	-0.60	-0.32	-0.65	-2.33	0.47	0.65
SMA5GARCH				0.67	-2.02	0.26	-0.51	-0.44	-2.99	0.91	1.22
				0.34	-2.60	-0.14	-0.16	-0.49	-2.94	1.03	1.21
SMA22GARCH					-2.28	-1.41	-0.82	-1.31	-2.60	0.14	0.38
					-2.24	-1.47	-0.37	-0.85	-2.60	0.39	0.56
RSMA5GARCH						1.85	1.20	1.92	-1.17	1.64	1.87
						1.79	1.87	2.43	-0.85	2.17	2.28
RSMA22GARCH							-0.45	-0.64	-2.20	0.39	0.62
							0.05	-0.11	-2.14	0.71	0.88
AR7								0.12	-2.17	1.42	1.66
								-0.20	-2.26	1.33	1.43
HARGARCH									-3.29	0.97	1.27
									-3.23	1.12	1.32
RHARGARCH										2.37	2.64
										2.54	2.66
VECM											0.81
											0.52

Table 10: Test statistics of Diebold Mariano West test (MSE and Q-Like) applied onto ranges forecasts of different models. Null hypothesis is of equal forecasting power and critical values corresponding to 95% confidence level are -1.96, 1.96. Insignificant values are printed in grey.

The observed data provide several conclusions:

- The first row confirms our findings in Table 10 as all models provide better than naive forecasts.
- SMA5 is significantly dominated by SMA5GARCH which is in turn significantly dominated by RSMA5GARCH. SMA22 is not dominated by SMA22GARCH, but both are (at least according to one measure) dominated by RSMA22GARCH. These rankings are in accordance with expectations, as in the first instance we allow for a more flexible model and in the second instance, we provide a more precise measure of volatility as a predictor.
- Better performance of HARGARCH compared to SMA5 is an expected result, however an impossibility to distinguish HARGARCH forecasts from other model's forecasts (except for RHARGARCH) is a surprising feature. As HARGARCH combines both SMA5GARCH and SMA22GARCH one would expect the combined model to perform better.
- The clearly best model is RHARGARCH, which is capable of dominating each model except for RSMA5GARCH. From this, we can conclude that for proper out of sample forecasting of daily ranges, using high quality volatility measures is critical.
- Lastly, turning to VECM models we observe a disappointing bad quality of forecasts. Neither VECM can beat any other model except for RW, moreover both are significantly worse in terms of forecast quality than RHARGARCH. The puzzle of a three-fold increase in $AdjR^2$ of VECMAUG over VECM in-sample not reflected in an increased forecasting performance is confirmed, as forecasts of both VECMs cannot be distinguished. The root of this puzzle can be investigated by analyzing in-sample range predictions of both VECMs. Apparently, both VECMs produce nearly identical in-sample range predictions as illustrated in Figure A.48. The increase in separate equations' $AdjR^2$ thus brings advantage when modeling daily highs and daily lows, however there is no guarantee that smaller errors in daily highs and daily lows equations in VECMAUG do not add up to produce larger errors in daily ranges forecast[32].

---

[32] Simply put, the difference of two large errors of the same sign can be small, however the difference of two small errors of opposing signs can be large.

To conclude, in this section we investigate out-of-sample forecasting performance of selected models. Models rooted in the co-integration of daily high and low prices dominate only the random-walk model. Moreover, no difference in the forecasts of base and augmented VECM is found. A general conclusion is that models incorporating realized ranges as opposed to daily ranges as predictors perform better out-of-sample. This confirms the results of our in-sample investigations. Surprisingly, a HAR model of daily ranges with a GARCH volatility-of-volatility component does not outperform models based solely on weekly or monthly average ranges with the same GARCH component. The clearly best model is an R-HAR model with a GARCH(1,1)-t volatility-of-volatility component.

Summary

In this thesis, we enrich the body of knowledge focused on daily ranges modeling by several new findings.

A general analysis of our dataset shows that market behavior significantly changes on intraday basis. The volume-volatility relationship does not hold in general. Based on surges in traded volume corresponding to a changing composition of traders by geographical location, we divide the trading day into separate sessions. For these sessions, we observe varying degrees of volume-volatility relationship, corresponding to different book thicknesses. Generally, however, volatility is significantly more correlated with average trade size and transaction count than with traded volume, contrary to popularly held belief. An extensive correlation analysis of trading variables (volatility, trading activity) reveals that these trading variables measured over different sessions convey unique information, which is reliably aggregated by daily volatility measures. In other words, considering daily volatilities and session variables separately does not yield any gain as daily volatilities already embody all information useful for next-day volatility forecasting.

Regressing daily ranges on range-based volatility estimates (Garman & Klass, Rogers & Satchell) is not expected to yield benefit, as our investigations show that all considered range-based estimators provide the same level of efficiency on real-world data. This is in sharp contrast with theoretical results, where Garman & Klass and Rogers &Satchell estimators show significant efficiency gains compared to daily ranges. Specifically, approximately 40% of information provided by herein considered range-based estimators as well as squared daily ranges is pure noise, while only 60% of information is related to the variance of the price generating process.

Using highly efficient realized ranges for the prediction of daily ranges shows small gains in terms of in-sample fit. Out-of-sample forecasting performance, however, shows advantages of regressing daily ranges on realized ranges. In accordance with intuition, realized ranges and daily ranges are found to contain the same useful information and the difference of these measures is pure noise.

While the information content of lagged sessions is fully reflected in lagged daily volatilities, information provided by current day's sessions can be used to improve end-of-day daily volatility forecasts. Specifically, if we utilize all up-to-date information provided by the market, traders active in the American sessions can gain at least 20-25

percentage points of $AdjR^2$ compared to their Asian counterparts. Variables most important for this gradual updating of volatility forecasts are model-dependent. In a HAR model of daily ranges and realized ranges, current-day session ranges and realized ranges are dominant in terms of predictive power. In a VECM of daily highs and daily lows, however, the most actual session's closing price is the most relevant predictor.

Comparison of models based on out-of-sample forecasting performance reveals several points. Firstly, even the simplest models based on average weekly or monthly volatility can beat the random-walk specification. Surprisingly, many models considered in volatility related literature provide forecasts of quality only comparable to forecasts of these simplest models. Namely, ARMA-GARCH and HAR-GARCH models of daily ranges can only beat a model which predicts the next day's volatility to be equal to average daily range of the last week. An R-HAR-GARCH model utilizing realized ranges for the prediction of daily ranges is the model of choice, as it can statistically beat nearly all models considered in this thesis. Models based on the co-integration of daily highs and daily lows are reported in literature to be of great usefulness in terms of trading strategy profitability. However, on our dataset, their high quality in-sample fits pertain only to daily high and daily low prices modeling. On out-of-sample daily ranges forecasting, VECM models are only capable of beating the random-walk specification.

As our main motivation for modeling daily ranges was to provide for a way of money management to intraday traders, a logical follow-up to our work would be to investigate out-of-sample forecasting properties of gradually updated models throughout the day. Next, drawing upon the results of Leitch & Tanner (1998), comparisons of models chosen by error-minimization as opposed to profit-maximization might bring interesting results. As the occurrence of news releases causes jumps in prices as well as sessions' ranges and realized ranges, including information on news releases might bring additional insights and improvements of cumulative volatility forecast updates. Possible methods of investigating these might be threshold models, whereby an occurrence of a news release is not modeled via a dummy variable representing a fundamental news being released. Instead, an above threshold session range/realized range could be taken as a proxy of a news event impact. Possible spillovers of news releases into increased or decreased volatilities of other sessions as well as whole days could be investigated. Lastly, a part of research suggests that order imbalance (a measure of whether buyers or sellers are more aggressive in the market at the moment) is a trading activity measure that needs to be taken into account, complementing herein discussed trading activity measures. Hence, investigations of the order imbalance might contribute to our understanding of volume-volatility relationship and might provide novel ways of volatility prediction.

Contents of CD

The attached compact-disc contains the whole dataset (*total_data.gdt*) in a format readable by gretl. All results can be replicated using this dataset. The *./carr* directory contains MLE scripts for the estimations of CARR with Exponentially as well as Weibull distributed disturbance terms. Lastly, the *./rolling_forecasts* directory contains all scripts necessary for automatic generation of out-of-sample forecasts presented in Section 6. All other models tested in-sample were created using gretl's built-in tools (OLS, VECM), hence no scripts are provided for these.

References

Ahoniemi, K. (2009): Modeling and Forecasting Implied Volatility, Helsinki School of Economics, ISBN 9524883031, 9789524883030

Alizadeth, S., Brandt, M., Diebold, F. (2001): Range-based Estimation of Stochastic Volatility Models of Exchange Rate Dynamics Are More Interesting than You Think, Journal of Finance 57, 1047-1092

Andersen, T. G., Bollerslev, T. (1998): Answering the Skeptics: Yes, Standard Volatility Models do Provide Accurate Forecasts, International Economic Review, Vol. 39, No. 4, Symposium on Forecasting and Empirical Methods in Macroeconomics and Finance (Nov., 1998), pp. 885-905

Andersen, T. G., Bollerslev, T., Diebold, F. X., Labys, P. (2003): Modeling and Forecasting Realized Volatility, Econometrica, Vol. 71, No. 2 (March, 2003), pp. 579-625

Baruník, J., Krištoufek, L. (2010): On Hurst Exponent Estimation Under Heavy-tailed Distributions , Physica. A: Statistical Mechanics and its Applications vol.389, 18 (2010), p. 3844-3855 (2010)

Bollerslev, T. (1986): Generalized Autoregressive Conditional Heteroskedasticity, Journal of Econometrics 31 (3): 307–327.

Chan, K., Fond, W.-M. (2000): Trade Size, Order Imbalance and the Volatility-Volume Relation, Journal of Financial Economics, Vol. 57 (2000), Issue, 2, pp. 247-523

Cheung, Y. W. (2007): An Empirical Model of Daily Highs and Lows, International Journal of Finance and Economics, Volume 12, Issue 1, pp. 1-20

Cheung, Y. L., Cheung, Y. W., He, A. W.W., Wan, A. T.K. (2010): A Trading Strategy Based on Callable Bull/Bear Contracts, Journal of Pacific Basic Finance, Vol. 18, Issue 2, pp. 186-198

Chou, R. Y. (2005): Forecasting Financial Volatilities With Extreme Values: The Conditional Autoregressive Range (CARR) Model, Journal of Money, Credit and Banking, Vol. 37, No. 3 (June 2005)

Christensen, K., Podolskij, M. (2007): Realized Range-based Estimation of Integrated Variance, Journal of Econometrics 141 (2007), pp. 323-349

Chu, C. F., Lam, K. P. (2008): Impact of Overnight Information on MEM Volatility Prediction, Statistics and Its Interface, Vol. 1 (2008), pp. 297-306

Corrado, Ch., Truong, C. (2007): Forecasting Stock Index Volatility: Comparing Implied Volatility and the Intraday High-Low Range, The Journal of Financial Research, Vol. 30, No. 2, pp. 201-215, 2007

Corsi, F., Reno, R. (2009): HAR Volatility Modeling with Heterogeneous Leverage and Jumps, Working paper

Corsi, F. (2007): A Simple Approximate Long-Memory Model of Realized Volatility, Journal of Financial Econometrics (2009), 7 (2), pp. 174-196

Dacorogna, M. M., Muller, U. A., Nagler, R. J., Olsen, R. B., Pictet, O. V. (1993): A Geographical Model For the Daily and Weekly Seasonal Volatility in the Foreign Exchange Market, Journal of International Money and Finance (1993), 12, pp. 413-438

Dacorogna, M. M., Muller, U. A., Davé, D. D., Olsen, R. B., Pictet, O. V., von Weizsacker, J. E. (1997): Volatilities of Different Time Resolutions - Analyzing the Dynamics of Market Components, Journal of Empirical Finance 4 (1997), pp. 219-239

Diebold, F. X., Mariano, R. S. (1995): Comparing Predictive Accuracy, Journal of Business and Economic Statistics, Vol. 20, Issue 1, pp. 134-144

Engle, R.F. & Russell, J.R. (1998): Autoregressive Conditional Duration: A New Model for Irregularly Spaced Transaction Data, Econometrica, 66, pp. 1127-1162.

Engle, R.F. (2002): New Frontiers for ARCH Models. Journal of Applied Econometrics, 17, pp. 425-446.

Engle, R. F., Gallo, G. M. (2006): A Multiple Indicators Model for Volatility Using Intra-daily Data, Journal of Econometrics 131 (2006), pp. 3 - 27

Garman, M. B., Klass, M. J. (1980): On the Estimation of Security Price Volatilities From Historical Data, Journal of Business, 1980, Vol. 53, No,. 1, pp. 67 - 78

Ghysels, E., Santa-Clara, P., Valkanov, R. (2006): Predicting Volatility: Getting the Most Out of Return Data Sampled at Different Frequencies, Journal of Econometrics 131 (2006), pp. 59-95

He, A. W.W, Kwok, J. T.K., Wan, A. T.K. (2010): An Empirical of Daily Highs and Lows of West Texas Intermediate Crude Oil Prices, Energy Economics, Volume 32, Issue 6, pp. 1499-1506

Jones, Ch., M., Kaul, G., Lipson, M. L. (1994): Transactions, Volume and Volatility, The Review of Financial Studies, Winter 1994, Vol. 7, No. 4, pp. 631-651

Jorion, P. (1995): Predicting Volatility in the Foreign Exchange Market, Journal of Finance, Vol. 50, No. 2

Leitch, G., Tanner, J. E. (1991): Economic Forecast Evaluation: Profits Versus the Conventional Error Measures, The American Economic Review (1991), Vol. 81, No. 3

Marten, M., van Dijk, D. (2007): Measuring Volatility with the Realized Range, Journal of Econometrics 138 (2007), pp. 181-207

Müller, U. A., Dacorogna, M. M., Dave, R. D., Pictet, O. V., Olsen, R. B. and Ward, J. R. R. (1993): Fractals and Intrinsic Time - A Challenge to Econometricians.
Presented in an opening lecture of the 39th International Conference of the Applied Econometrics Association (AEA), Real Time Econometrics — Sub monthly Time Series, 14-15 Oct 1993, in Luxembourg, and at the 4th International PASE Workshop, 22-26 Nov 1993, in Ascona (Switzerland)

Patton, A. J. (2011): Volatility Forecast Comparison Using Imperfect Volatility Proxies", Journal of Econometrics, Vol. 160, pp. 246-256

Pong, S., Shackleton, M. B., Taylor, S. J., Xu, X. (2004): Forecasting Currency Volatility: A Comparison of Implied Volatilities and AR(FI)MA Models, Journal of Banking and Finance, 2004, Vol. 28, Issue 10, pp. 2541-2563

Rogers, L. C. G., Satchell, S. E. (1991): Estimating Variance from High, Low and Closing Prices, The Annals of Applied Probability 1991, Vol. 1., No. 4, pp. 504-512

Yang, D., Zhang, Q. (2000): Drift-Independent Volatility Estimation Based on High, Low, Open and Close Prices, Journal of Business (2000), Vol. 73, No. 3

Sargan, J. (1964): Wages and Prices in the United Kingdom: A Study in Econometric Methodology, Econometric Analysis for National Economic Planning, London: Butterworths, 1964.

Starica, C. (2003): Is GARCH(1,1) as Good a Model as the Accolades of the Nobel Prize Would Imply? (November 2003). Available at SSRN: http://ssrn.com/abstract=637322 or http://dx.doi.org/10.2139/ssrn.637322

Zeng, T., Swansons, N. R. (1998): Predictive Evaluation of Econometric Forecasting Models in Commodity Futures Markets, Studies in Nonlinear Dynamics & Econometrics, Berkeley Electronic Press, Vol. 2

Appendix

Delivery	Start Date	End Date	Delivery	Start Date	End Date
12-Jul	09.11.2007	11.12.2007	3-Oct	11.12.2009	10.03.2010
3-Aug	12.12.2007	12.03.2008	6-Oct	11.03.2010	10.06.2010
6-Aug	13.03.2008	12.06.2008	9-Oct	11.06.2010	09.09.2010
9-Aug	13.06.2008	11.09.2008	12-Oct	10.09.2010	09.12.2010
12-Aug	12.09.2008	11.12.2008	3-Nov	10.12.2010	10.03.2011
3-Sep	12.12.2008	11.03.2009	6-Nov	11.03.2011	09.06.2011
6-Sep	12.03.2009	10.06.2009	9-Nov	10.06.2011	15.09.2011
9-Sep	11.06.2009	09.09.2009	12-Nov	16.09.2011	09.11.2011
12-Sep	10.09.2009	10.12.2009			

Table A.1: List of cut-off dates for the construction of a continuous contract (DD.MM.YYYY).

	coefficient	std. error	t-ratio	p-value
c	0.0010	0.00008	12.5812	<0.00001
R^{log}	1.30378	0.00594	219.2555	<0.00001

Log-Lik	5479.26	AdjR^2	0.98

Table A.2: Estimation results for model (17).

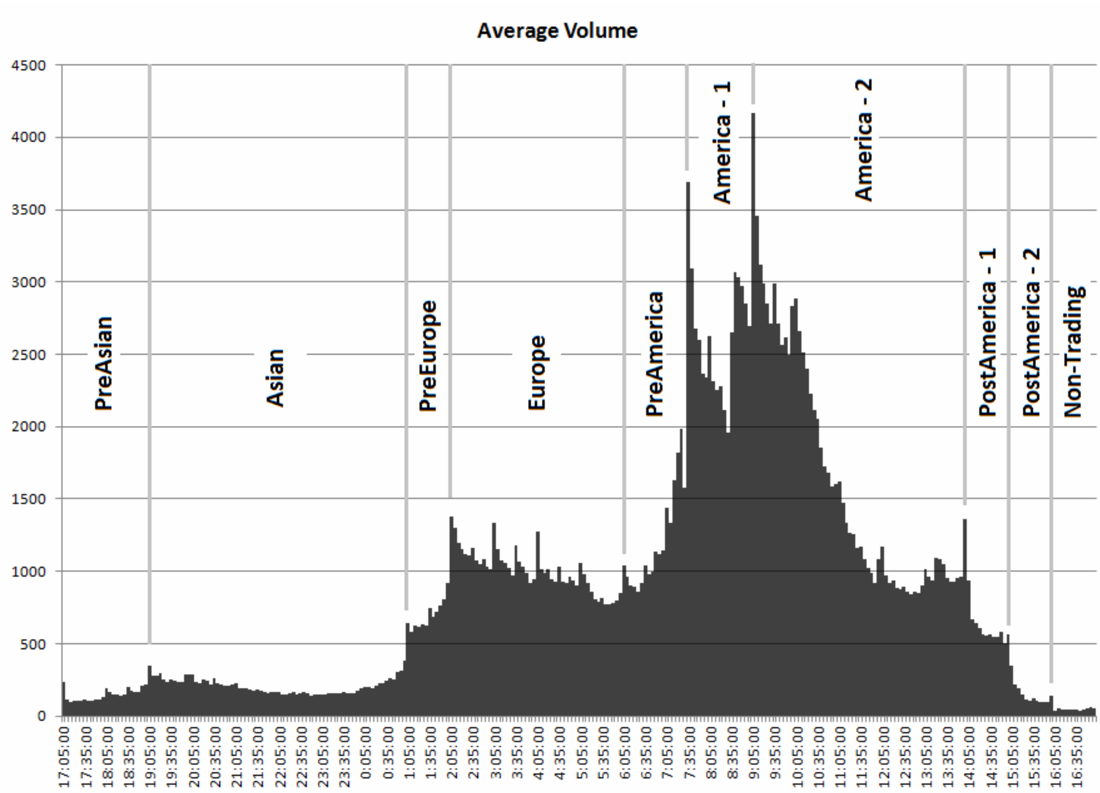


Figure A.3: Division of a trading day into trading sessions.

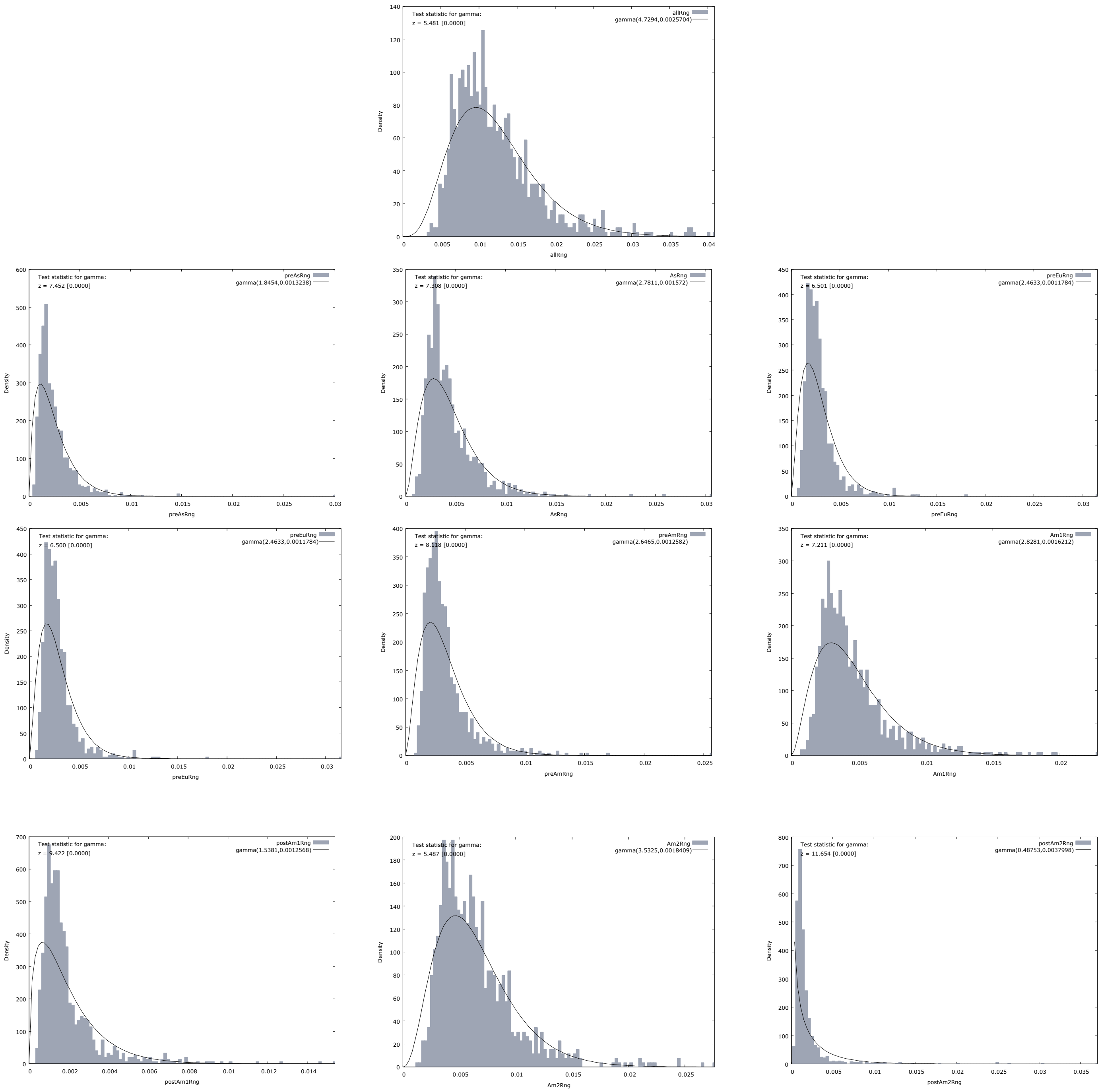


Figure A.4: Frequency distribution of daily ranges and session ranges with gamma p.d.f. giving the best fit.

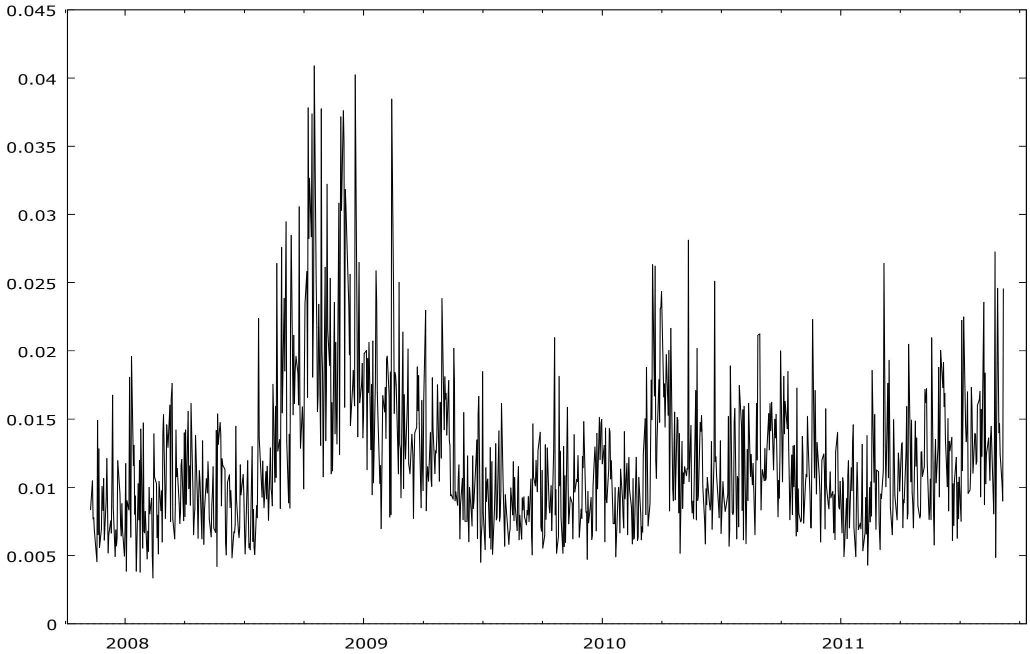


Figure A.5: Time plot of daily ranges (Nov 9th 2007 - Nov 9th 2011).

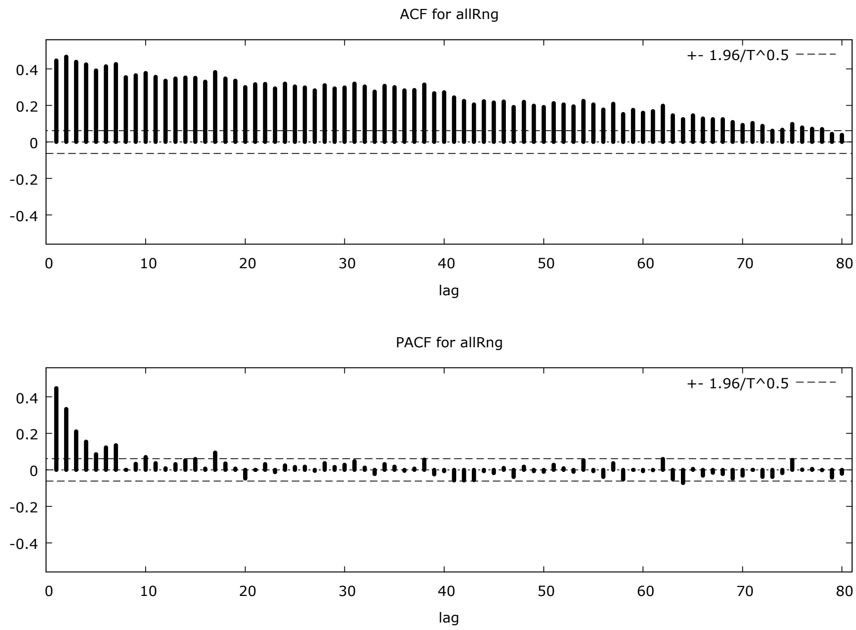


Figure A.6: ACF, PACF of daily ranges.

	Volume	Transactions	Trade Size
All day	0.2470	0.3936	-0.4298
preAsian	0.5312	0.5780	-0.0968
Asian	0.4853	0.5260	-0.1518
preEurope	0.3976	0.3362	-0.1979
Europe	0.4365	0.5285	-0.3057
preAmerica	0.4112	0.5425	-0.3739
America1	0.4930	0.5776	-0.3007
America2	0.3746	0.5127	-0.4041
postAmerica1	0.5265	0.6404	-0.3571
postAmerica2	0.1731	0.2106	-0.0688

Table A.7: Correlations of ranges with traded volume, number of transactions and average trade size for different sessions, including ETH session. For example, the correlation of preAsian range and preAsian volume is 0.5312.

	Volume	Transactions	Range	Trade Size
preAsian	0.1024	0.1723	0.4506	-0.2549
Asian	0.1524	0.2229	0.5422	-0.2976
preEuropean	0.1098	0.2178	0.4981	-0.3741
European	0.1489	0.2651	0.6140	-0.4251
preAmerican	0.1890	0.3221	0.5320	-0.4153
American1	0.1611	0.2625	0.5408	-0.3879
American2	0.2348	0.3825	0.7027	-0.4196
postAmerican1	0.3018	0.4046	0.5962	-0.3762
postAmerican2	0.1640	0.2561	0.4474	-0.2488

Table A.8: Correlations of daily ranges with volume, transaction count and average trade size of different sessions of the current day. For example, correlation of daily ranges with ranges of preAsian session is 0.4506.

Session	Variable	Lag							
		1	2	3	4	5	6	7	8
preAsian	**Volume**	-0.0128	-0.0081	0.0090	-0.0045	0.0013	0.0152	0.0080	0.0054
	Transactions	-0.3379	-0.3349	-0.3246	-0.3172	-0.3141	-0.3125	-0.3025	-0.3020
	Range	0.3185	0.2978	0.3041	0.3212	0.2946	0.2974	0.3069	0.2858
	Trade Size	-0.2572	-0.2447	-0.2305	-0.2551	-0.2512	-0.2109	-0.2320	-0.2257
Asian	**Volume**	0.0057	0.0161	0.0154	0.0191	0.0186	0.0046	0.0392	0.0086
	Transactions	0.0837	0.0879	0.0809	0.0889	0.0814	0.0634	0.0926	0.0656
	Range	0.3890	0.3796	0.4013	0.3740	0.3839	0.3496	0.3354	0.3526
	Trade Size	-0.3208	-0.3013	-0.2834	-0.3005	-0.2840	-0.2534	-0.2449	-0.2589
preEuropean	**Volume**	-0.0195	-0.0089	-0.0604	-0.0639	-0.0286	-0.0295	-0.0009	-0.0459
	Transactions	0.0763	0.0863	0.0321	0.0293	0.0590	0.0589	0.0742	0.0363
	Range	0.3090	0.3016	0.2921	0.3152	0.2964	0.2657	0.2959	0.2363
	Trade Size	-0.3745	-0.3847	-0.3837	-0.3663	-0.3720	-0.3465	-0.3201	-0.3430
European	**Volume**	-0.0248	0.0030	-0.0499	-0.0400	-0.0555	-0.0483	-0.0190	-0.0707
	Transactions	0.0984	0.1271	0.0709	0.0766	0.0535	0.0533	0.0822	0.0253
	Range	0.3295	0.3685	0.3203	0.3650	0.2992	0.2642	0.3197	0.2703
	Trade Size	-0.4701	-0.4609	-0.4542	-0.4429	-0.4409	-0.3946	-0.4026	-0.3967
preAmerican	**Volume**	0.0118	0.0140	-0.0379	-0.0046	-0.0637	-0.0347	0.0021	-0.0335
	Transactions	0.1358	0.1445	0.0805	0.1025	0.0553	0.0785	0.1082	0.0601
	Range	0.3227	0.3388	0.3004	0.3393	0.2743	0.2833	0.3027	0.2630
	Trade Size	-0.4205	-0.4442	-0.4077	-0.3926	-0.4043	-0.3787	-0.3617	-0.3565
American1	**Volume**	-0.0028	-0.0130	-0.0408	-0.0146	-0.0504	-0.0087	-0.0246	-0.0639
	Transactions	0.1018	0.0903	0.0547	0.0811	0.0384	0.0800	0.0562	0.0184
	Range	0.3195	0.3176	0.3022	0.3401	0.2626	0.3119	0.2543	0.2240
	Trade Size	-0.4073	-0.4016	-0.3829	-0.3826	-0.3686	-0.3514	-0.3362	-0.3421
American2	**Volume**	0.0185	0.0196	-0.0105	-0.0132	-0.0057	0.0065	0.0312	-0.0421
	Transactions	0.1819	0.1711	0.1319	0.1284	0.1254	0.1368	0.1499	0.0839
	Range	0.3942	0.3662	0.3571	0.3510	0.3534	0.3325	0.3407	0.3021
	Trade Size	-0.4495	-0.4305	-0.4231	-0.4205	-0.4034	-0.3909	-0.3625	-0.3848
postAmerican1	**Volume**	0.1310	0.1397	0.1015	0.0861	0.0830	0.1007	0.1620	0.0726
	Transactions	0.2363	0.2398	0.1920	0.1759	0.1730	0.1835	0.2306	0.1559
	Range	0.4389	0.4173	0.4172	0.4040	0.3825	0.3839	0.4234	0.3568
	Trade Size	-0.3700	-0.3421	-0.3387	-0.3292	-0.3347	-0.3311	-0.2674	-0.2832
postAmerican2	**Volume**	0.0192	0.0096	0.0129	0.0218	0.0158	0.0327	0.0550	0.0053
	Transactions	0.0983	0.0889	0.0943	0.0946	0.0818	0.0988	0.1242	0.0643
	Range	0.2977	0.2405	0.3014	0.2582	0.2616	0.2923	0.2725	0.2635
	Trade Size	-0.2340	-0.2503	-0.2506	-0.2513	-0.2339	-0.2359	-0.2105	-0.2123

Table A.9: Correlations of daily ranges with lagged trading variables of different sessions. For example, the correlation of today's range and yesterday's preAsian transaction count is -0.3379. All values significant at least on a 95% level are printed in black.

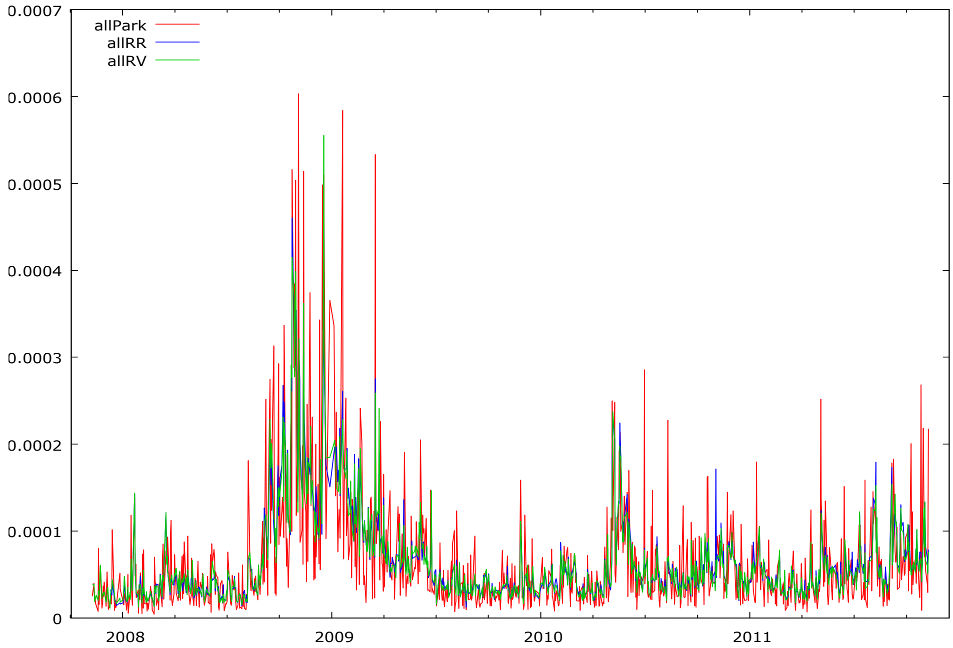


Figure A.10: Time plots of daily ranges, square root of daily realized ranges and daily realized volatilities constructed from 5-minute returns.

Session	Variable	Lag								
		0	1	2	3	4	5	6	7	8
ETH	Range	1.0000	0.4471	0.4673	0.4388	0.4254	0.3918	0.4142	0.4280	0.3545
	RV	0.7659	0.5889	0.5810	0.5486	0.5499	0.4989	0.5105	0.5024	0.4669
	RR	0.7584	0.5914	0.5873	0.5504	0.5539	0.5048	0.5176	0.5149	0.4684
preAsian	Range	0.4506	0.3185	0.2978	0.3041	0.3212	0.2946	0.2974	0.3069	0.2858
	RV	0.4817	0.4276	0.3959	0.4101	0.3885	0.3733	0.3922	0.3820	0.3638
	RR	0.5017	0.4152	0.3860	0.3993	0.3986	0.3910	0.3866	0.3858	0.3645
Asian	Range	0.5422	0.3890	0.3796	0.4013	0.3740	0.3839	0.3496	0.3354	0.3526
	RV	0.5217	0.4408	0.4900	0.4694	0.4553	0.4331	0.4254	0.4101	0.4044
	RR	0.5464	0.4557	0.4997	0.4711	0.4665	0.4330	0.4250	0.4162	0.4063
preEuropean	Range	0.4981	0.3090	0.3016	0.2921	0.3152	0.2964	0.2657	0.2959	0.2363
	RV	0.5979	0.4368	0.4197	0.4036	0.4224	0.3836	0.4020	0.4002	0.3492
	RR	0.6223	0.4762	0.4826	0.4433	0.4654	0.4350	0.4547	0.4406	0.4096
European	Range	0.6140	0.3295	0.3685	0.3203	0.3650	0.2992	0.2642	0.3197	0.2703
	RV	0.5999	0.4394	0.4541	0.4360	0.4358	0.3613	0.3623	0.3844	0.3525
	RR	0.6394	0.4790	0.5012	0.4731	0.4674	0.3969	0.4102	0.4365	0.3795
preAmerican	Range	0.5320	0.3227	0.3388	0.3004	0.3393	0.2743	0.2833	0.3027	0.2630
	RV	0.5174	0.3765	0.4199	0.3233	0.3372	0.3063	0.3457	0.3228	0.2920
	RR	0.5068	0.3539	0.4053	0.3271	0.3222	0.3367	0.3307	0.3220	0.2750
American1	Range	0.5408	0.3195	0.3176	0.3022	0.3401	0.2626	0.3119	0.2543	0.2240
	RV	0.4874	0.3402	0.3609	0.3181	0.3485	0.2817	0.3516	0.2604	0.2642
	RR	0.4981	0.3792	0.3796	0.3356	0.3905	0.3097	0.3920	0.3149	0.3030
American2	Range	0.7027	0.3942	0.3662	0.3571	0.3510	0.3534	0.3325	0.3407	0.3021
	RV	0.7059	0.5503	0.4835	0.4623	0.4664	0.4276	0.4201	0.4384	0.3964
	RR	0.6935	0.5492	0.4956	0.4759	0.4704	0.4335	0.4297	0.4467	0.3967
postAmerican1	Range	0.5962	0.4389	0.4173	0.4172	0.4040	0.3825	0.3839	0.4234	0.3568
	RV	0.5557	0.4230	0.3709	0.3792	0.3598	0.3549	0.3355	0.3587	0.3195
	RR	0.5910	0.4643	0.4237	0.3999	0.3952	0.3719	0.3633	0.3995	0.3595
postAmerican2	Range	0.4474	0.2977	0.2405	0.3014	0.2582	0.2616	0.2923	0.2725	0.2635
	RV	0.4883	0.3420	0.3423	0.3307	0.3186	0.3412	0.3372	0.3650	0.3009
	RR	0.4777	0.3475	0.3336	0.3506	0.3224	0.3365	0.3378	0.3494	0.2938

Table A.11: Correlations of ETH (daily) ranges with different sessions' lagged volatility measures. For example, the correlation of today's daily range and yesterday's preAsian realized range is 0.4152.

LHS Variable	α	β	$AdjR^2$	Dataset	Chow p-value
	0.000028	0.5681	0.58	whole	0.0000
RR_D	0.000029	0.5921	0.61	1st half	0.0000
	0.000032	0.4302	0.38	2nd half	0.0003
	0.000026	0.5888	0.60	whole	0.0000
RV_D	0.000028	0.6183	0.64	1st half	0.0007
	0.000031	0.4178	0.39	2nd half	0.0019

Table A.12: OLS results of $VolMeasure = \alpha + \beta \times \left(R_D^{\log}\right)^2 / 4\ln(2) + \varepsilon$ (model (19)). In Chow test, data sets were always halved, i.e. for example in the second row, the first half of dataset was halved and Chow test was carried out on first quarters of the whole dataset.

Regressor	Dependent variable					
	$(R^{log})^2/4\ln(2)$	$(R^{log})^2/4\ln(2)$	$(\sigma^{GK})^2$	$(\sigma^{GK})^2$	$(\sigma^{RS})^2$	$(\sigma^{RS})^2$
c	-0.0000009	-0.0000003	-0.0000036	**-0.0000035**	-0.0000041	-0.0000045
	0.0000027	0.0000020	0.0000024	0.0000021	0.0000037	0.0000037
RR_D	**1.0155**		**0.9837**		**1.0640**	
	0.0566		0.0517		0.0744	
RV_D		**1.0165**		**0.9918**		**1.0640**
		0.0399		0.0453		0.0744
$AdjR^2$	0.5765	0.5981	0.5453	0.5740	0.5991	0.5991

Regressor	Dependent variable					
	RR_D			*RV_D*		
c	**0.0000277**	**0.0000313**	0.0000281	**0.0000258**	**0.0000292**	0.0000256
	0.0000024	0.0000032	0.0000027	0.0000024	0.0000032	0.0000032
$(R^{log})^2/4\ln(2)$	**0.5680**			**0.5888**		
	0.0368			0.0391		
$(\sigma^{GK})^2$		**0.5547**			**0.5791**	
		0.0551			0.0588	
$(\sigma^{RS})^2$			**0.5634**			**0.5791**
			0.0387			0.0588
$AdjR^2$	0.5765	0.5453	0.5991	0.5981	0.5740	0.5740

Table A.13: Results of regressing range-based measures of variance on realized measures of variance and vice versa. HAC method for standard errors, bold estimates are significant on 99% level.

Mean Equation

Variable	coefficient	std. error	z-Statistic	p-value.
C	0.0100	0.0005	19.7114	0.0000
AR(1)	0.1308	0.0309	4.2265	0.0000
AR(2)	0.1520	0.0302	5.0327	0.0000
AR(3)	0.0679	0.0318	2.1329	0.0329
AR(4)	0.1136	0.0313	3.6305	0.0003
AR(5)	0.0699	0.0312	2.2372	0.0253
AR(6)	0.0858	0.0311	2.7556	0.0059
AR(7)	0.1243	0.0305	4.0771	0.0000

Variance Equation

Variable	coefficient	std. error	z-Statistic	p-value
C	0.0000	0.0000	1.8755	0.0607
RESID(-1)^2	0.0444	0.0148	2.9966	0.0027
GARCH(-1)	0.9378	0.0197	47.5484	0.0000
T-DIST. DOF	5.6714	0.9829	5.7700	0.0000
$AdjR^2$	0.3412		**Log-lik**	4049.6090

Table A.14: Estimation results of ARMA(7)-GARCH (1,1) -t applied to daily ranges

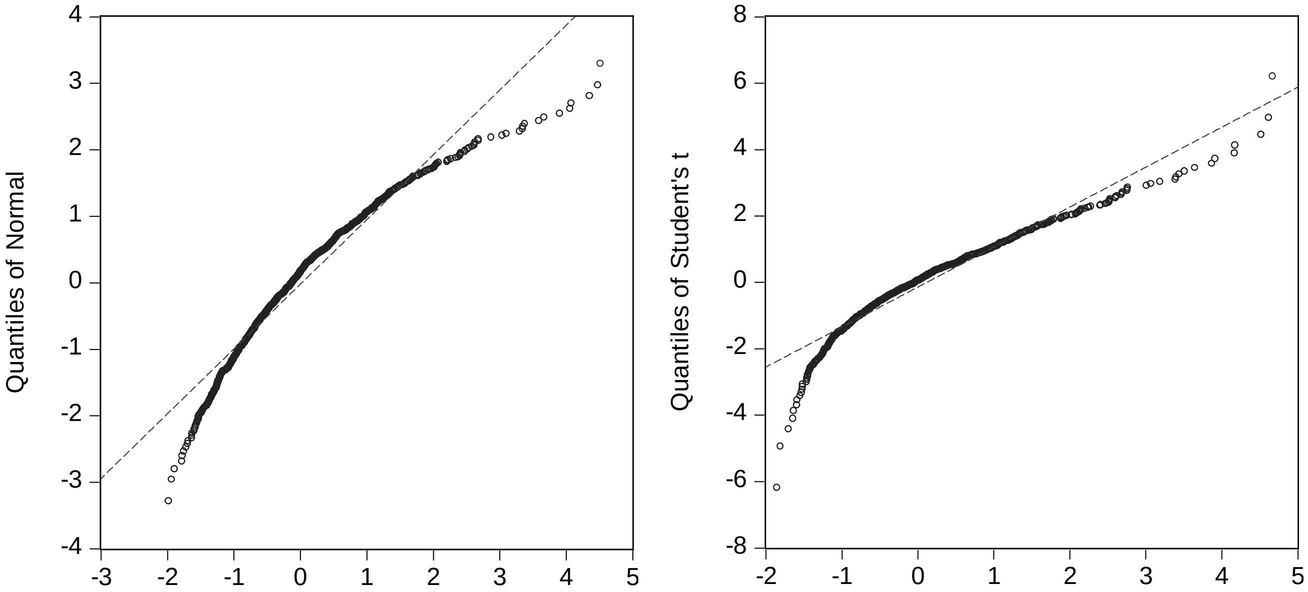


Table A.15: Q-Q plots of residuals from ARMA(7)-GARCH(1,1) with normally (left) and Student (right) distributed disturbances. Notice a better fit in the right figure.

Model	Range	I	RR	Trade Size	Trans. Count
	RHS variables				
HAR	×				
R-HAR			×		
LHAR	×	×			
LHAR-R	×	×	×		
LHAR-S	×	×		×	
LHAR-C	×	×			×
LHAR-SC	×	×		×	×
LHAR-RSC	×	×	×	×	×
R-LHAR		×	×		
R-LHAR-S		×	×	×	
R-LHAR-C		×	×		×
R-LHAR-SC		×	×	×	×

Table A.16: List of modified HAR models. For variance models, the explained variable is squared daily range and RR stands for realized range. For volatility models we explain daily ranges and RR stands for square root of realized ranges. I stands for the leverage effect.

	Variance Equation			Volatility Equation		
	HAR	**R-HAR**	**HAR-R**	**HAR**	**R-HAR**	**HAR-R**
c	**0.0000**	0.0000	0.0000	**0.0014**	0.0010	0.0011
$R^{(1)}(-1)$	0.0482		-0.0016	0.0466		-0.0023
$R^{(5)}(-1)$	**0.4356**		-0.0571	**0.3958**		-0.0450
$R^{(22)}(-1)$	**0.3884**		-0.1535	**0.4444**		-0.1942
$RR^{(1)}(-1)$		**0.7952**	0.7857		**0.4807**	**0.4957**
$RR^{(5)}(-1)$		**1.3463**	1.1959		**0.6917**	**0.7783**
$RR^{(22)}(-1)$		0.4086	0.8406		**0.2972**	0.5784
$AdjR^2$	0.3346	0.3684	0.3668	0.3560	0.3910	0.3900

Table A.17: Investigation of relationships between daily ranges and realized ranges from HAR perspective. Parameter estimates significant on 95% critical level are printed in bold.

	HAR	R-HAR	HAR-R	HAR-S	HAR-C	HAR-SC	HAR-RSC	LHAR	LHAR-R	LHAR-S	LHAR-C	LHAR-SC	LHAR-RSC	R-LHAR	R-LHAR-S	R-LHAR-C	R-LHAR-SC
c	0.0014**	0.0010	0.0011	0.0037	0.0035	0.0066	0.0138*	0.0017**	0.0011	0.0061**	0.0061	0.0131*	0.0142**	0.0011	0.0012	0.0107**	0.0122*
$R^{log,(1)}(-1)$	0.0466		-0.0023	0.0623	0.0188	0.0325	0.0028	0.0745	0.0247	0.0892*	0.0485	0.0616	0.0293				
$R^{log,(5)}(-1)$	0.3958***		-0.0450	0.3201***	0.3783***	0.2911**	-0.0368	0.3335***	-0.0505	0.2668**	0.3293**	0.2568*	-0.0422				
$R^{log,(22)}(-1)$	0.4444***		-0.1942	0.4495***	0.4959***	0.5103***	-0.3870	0.4410***	-0.2068	0.3970***	0.4803***	0.4313***	-0.4324				
$I^{(1)}(-1)$								0.0876**	0.0561	0.0878**	0.0855**	0.0853**	0.0548	0.0530	0.0546	0.0526	0.0533
$I^{(5)}(-1)$								-0.1362	-0.1145	-0.1476*	-0.1195	-0.1291	-0.1390*	-0.0973	-0.1063	-0.1068	-0.1087
$I^{(22)}(-1)$								-0.4054*	-0.0025	-0.4092*	-0.4293	-0.4454	0.0366	-0.0584	-0.0509	-0.0383	-0.0601
$RR^{(1)}(-1)$		0.4807***	0.4957***				0.5561**		0.4384**				0.4776**	0.4735***	0.4972***	0.4920**	0.5174**
$RR^{(5)}(-1)$		0.6917***	0.7783*				0.8932		0.7723*				0.9090*	0.6587***	0.5108**	0.9664***	0.8166*
$RR^{(22)}(-1)$		0.2972*	0.5784				0.6833		0.6268				0.7761	0.3272**	0.4368**	0.0308	0.1066
$size^{(1)}(-1)$				-0.0023		-0.0025	0.0002			-0.0019		-0.0022	0.0000		0.0007		0.0005
$size^{(5)}(-1)$				-0.0061		-0.0065	-0.0044			-0.0066		-0.0068	-0.0048		-0.0065		-0.0039
$size^{(22)}(-1)$				0.0071		0.0076	0.0036			0.0060*		0.0061	0.0040		0.0058*		0.0026
$count^{(1)}(-1)$					0.0010	0.0011	-0.0004				0.0009	0.0010	-0.0003			-0.0001	-0.0002
$count^{(5)}(-1)$					0.0001	0.0002	-0.0021				-0.0002	-0.0001	-0.0023			-0.0029	-0.0023
$count^{(22)}(-1)$					-0.0013	-0.0016	0.0014				-0.0012	-0.0014	0.0015			0.0022	0.0016
$AdjR^2$	0.3560	0.3910	0.3900	0.3581	0.3556	0.3586	0.3922	0.3648	0.3902	0.3676	0.3646	0.3680	0.3927	0.3913	0.3912	0.3939	0.3925

Table A.18: Estimation results of different HAR model specifications for daily ranges (volatility).

	HAR	R-HAR	HAR-R	HAR-S	HAR-C	HAR-SC	HAR-RSC	LHAR	LHAR-R	LHAR-S	LHAR-C	LHAR-SC	LHAR-RSC	R-LHAR	R-LHAR-S	R-LHAR-C	R-LHAR-SC
c	0.0000**	0.0000	0.0000	0.0001	0.0001	0.0003	0.0005**	0.0000**	0.0000*	0.0001*	0.0002	0.0005*	0.0005**	0.0000*	0.0000	0.0004**	0.0005**
$R^{log,(1)}$(-1)	0.0482		-0.0016	0.0598	0.0261	0.0349	-0.0027	0.0864	0.0395	0.1021*	0.0614	0.0684	0.0366				
$R^{log,(5)}$(-1)	0.4356***		-0.0571	0.3801***	0.4576***	0.4014**	0.0621	0.3866***	0.0465	0.3540**	0.4244***	0.3819**	0.0590				
$R^{log,(22)}$(-1)	0.3884***		-0.1535	0.3690***	0.3929***	0.3623**	-0.2991	0.3712***	-0.1734	0.3103**	0.3635***	0.2794*	-0.3309				
$I^{(1)}$(-1)								0.0046***	0.0037**	0.0047***	0.0047***	0.0046***	0.0037**	0.0033**	0.0034**	0.0034**	0.0035**
$I^{(5)}$(-1)								-0.0046	-0.0044	-0.0052	-0.0043	-0.0045	-0.0055*	-0.0037	-0.0039	-0.0045	-0.0044
$I^{(22)}$(-1)								-0.0161	0.0010	-0.0158	-0.0172	-0.0182*	0.0008	0.0004	0.0004	0.0002	-0.0010
$RR^{(1)}$(-1)		0.7952*	0.7857				0.7751		0.6911				0.6271	0.8202*	0.8595*	0.7280	0.7486
$RR^{(5)}$(-1)		1.3463**	1.1959				1.4415		1.2180				1.5039	1.3317**	1.1654*	1.8385***	1.7166**
$RR^{(22)}$(-1)		0.4086	0.8406				0.9069		0.9275				0.9978	0.4339	0.5105	0.0470	0.0151
$size^{(1)}$(-1)				-0.0001		-0.0001	0.0000			0.0002**		-0.0001	0.0000		0.0001		0.0000
$size^{(5)}$(-1)				-0.0002		-0.0002	-0.0001			-0.0004**		-0.0002	-0.0001		-0.0002		-0.0001
$size^{(22)}$(-1)				0.0002		0.0002	0.0001			0.0002		0.0001	0.0001		0.0001		0.0000
$count^{(1)}$(-1)					0.0000*	0.0000*	0.0000				0.0000*	0.0000**	0.0000			0.0000	0.0000
$count^{(5)}$(-1)					0.0000	0.0000	-0.0001				-0.0001	-0.0001	-0.0001*			-0.0001**	-0.0001*
$count^{(22)}$(-1)					0.0000	0.0000	0.0000				0.0000	0.0000	0.0000			0.0001	0.0001
$AdjR^2$	0.3346	0.3684	0.3668	0.3362	0.3340	0.3364	0.3687	0.3485	0.3705	0.3514	0.3488	0.3516	0.3728	0.3714	0.3705	0.3746	0.3732

Table A.19: Estimation results of different HAR model specifications for squared daily ranges (variance).

	HAR	R-HAR	HAR-R	HAR-S	HAR-C	HAR-SC	HAR-RSC	LHAR	LHAR-R	LHAR-S	LHAR-C	LHAR-SC	LHAR-RSC	R-LHAR	R-LHAR-S	R-LHAR-C	R-LHAR-SC
c	0.0004 **	0.0006 **	0.0004 **	-0.0002	0.0022	0.0016	0.0018	0.0005 ***	0.0005 ***	0.0009	0.0029	0.0038	0.0042 *	0.0008 ***	0.0033 ***	0.0003	0.0038
$R^{log,(1)}(-1)$	0.4586 ***		0.3600 ***	0.4691 ***	0.3921 ***	0.4149 ***	0.3390 ***	0.4072 ***	0.2978 ***	0.4203 ***	0.3355 ***	0.3582 ***	0.2634 ***				
$R^{log,(5)}(-1)$	0.3304 ***		0.2886 **	0.3148 ***	0.4671 ***	0.4360 ***	0.3292 **	0.2996 ***	0.2206 *	0.2750 ***	0.4337 ***	0.3893 ***	0.2344				
$R^{log,(22)}(-1)$	0.1605 ***		0.2542	0.1816 ***	0.0948	0.1146	0.2937	0.1999 ***	0.2755 *	0.1963 ***	0.1443 **	0.1424 *	0.3420 *				
$I^{(1)}(-1)$								-0.0032	0.0167	-0.0009	-0.0004	0.0006	0.0179	0.0335 **	0.0337 **	0.0314 **	0.0314 **
$I^{(5)}(-1)$								-0.1039 ***	-0.1174 ***	-0.1066 ***	-0.1086 ***	-0.1089 ***	-0.1231 ***	-0.1396 ***	-0.1436 ***	-0.1216 ***	-0.1252 ***
$I^{(22)}(-1)$								-0.0934	-0.1242 **	-0.0960	-0.0942	-0.1042	-0.1336 **	-0.2949 ***	-0.3077 ***	-0.3026 ***	-0.3173 ***
$RR^{(1)}(-1)$		0.0916 ***	0.0493 ***				0.0432 **		0.0566 ***				0.0516 **	0.0964 ***	0.1013 ***	0.0666 ***	0.0713 ***
$RR^{(5)}(-1)$		0.2503 ***	0.0362				0.0492		0.0520				0.0703	0.2029 ***	0.1866 ***	0.1955 ***	0.1754 ***
$RR^{(22)}(-1)$		0.2367 ***	-0.0554				-0.0924		-0.0372				-0.0973	0.2379 ***	0.1975 ***	0.2769 ***	0.2417 ***
$size^{(1)}(-1)$				0.0017 **		0.0012	0.0007			0.0015 **		0.0010	0.0004		-0.0007		-0.0009
$size^{(5)}(-1)$				-0.0025 **		-0.0014	-0.0015			-0.0029 ***		-0.0019	-0.0021		-0.0019		-0.0020
$size^{(22)}(-1)$				0.0011		0.0004	0.0009			0.0011		0.0005	0.0011		0.0010		0.0014
$count^{(1)}(-1)$					0.0005 *	0.0004	0.0003				0.0006 *	0.0004	0.0004			0.0011 ***	0.0011 ***
$count^{(5)}(-1)$					-0.0011 **	-0.0009	-0.0007				-0.0011 **	-0.0008	-0.0006			-0.0001	-0.0002
$count^{(22)}(-1)$					0.0004	0.0003	0.0002				0.0003	0.0002	0.0000			-0.0009	-0.0010
$AdjR^2$	0.7470	0.7040	0.7515	0.7476	0.7479	0.7476	0.7512	0.7564	0.7622	0.7568	0.7576	0.7573	0.7627	0.7346	0.7366	0.7415	0.7439

Table A.20: Estimation results of different HAR model specifications for square root of realized ranges (volatility).

	HAR	R-HAR	HAR-R	HAR-S	HAR-C	HAR-SC	HAR-RSC	LHAR	LHAR-R	LHAR-S	LHAR-C	LHAR-SC	LHAR-RSC	R-LHAR	R-LHAR-S	R-LHAR-C	R-LHAR-SC
c	0.0000**	0.0000***	0.0000**	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0001*	0.0001	0.0000**	0.0000**	0.0000	0.0001
$R^{log,(1)}$(-1)	0.4976***		0.3962***	0.5113***	0.4607***	0.4831***	0.4297***	0.4533***	0.3591***	0.4690***	0.4112***	0.4329***	0.3154***				
$R^{log,(5)}$(-1)	0.2755***		0.2489	0.2527**	0.3567***	0.3151***	0.2175	0.2492**	0.1959	0.2227**	0.3470***	0.3058***	0.1518				
$R^{log,(22)}$(-1)	0.1708***		0.2458	0.1851***	0.1297**	0.1487**	0.2315	0.1978***	0.2244	0.1862***	0.1465**	0.1291*	0.2660				
$I^{(1)}$(-1)								0.0003	0.0007	0.0003	0.0003	0.0004	0.0008*	0.0010***	0.0010***	0.0011***	0.0011***
$I^{(5)}$(-1)								-0.0023***	-0.0025***	-0.0024***	-0.0025***	-0.0025***	-0.0028***	-0.0030***	-0.0031***	-0.0028***	-0.0028***
$I^{(22)}$(-1)								-0.0022	-0.0029*	-0.0023	-0.0023	-0.0026	-0.0037**	-0.0076***	-0.0078***	-0.0078***	-0.0082***
$RR^{(1)}$(-1)		0.0629***	0.0334**				0.0008		0.0012**				0.0387**	0.0673***	0.0692***	0.0560***	0.0577***
$RR^{(5)}$(-1)		0.1237***	0.0100				0.0012		0.0006				0.0369	0.1024***	0.0948***	0.1085***	0.1003***
$RR^{(22)}$(-1)		0.1384***	-0.0237				-0.0011		-0.0003				-0.0326	0.1288***	0.1067***	0.1350***	0.1100***
$size^{(1)}$(-1)				0.0000**		0.0000*	0.0000			0.0000**		0.0000	0.0000		0.0000		0.0000
$size^{(5)}$(-1)				-0.0001**		0.0000	0.0000			-0.0001***		-0.0001**	-0.0001*		0.0000		0.0000
$size^{(22)}$(-1)				0.0000		0.0000	0.0000			0.0000		0.0000	0.0000		0.0000		0.0000
$count^{(1)}$(-1)					0.0000*	0.0000	0.0000				0.0000*	0.0000	0.0000			0.0000***	0.0000***
$count^{(5)}$(-1)					0.0000**	0.0000	0.0000				0.0000**	0.0000*	0.0000			0.0000	0.0000
$count^{(22)}$(-1)					0.0000	0.0000	0.0000				0.0000	0.0000	0.0000			0.0000	0.0000
$AdjR^2$	0.7234	0.6676	0.7295	0.7239	0.7238	0.7237	0.7272	0.7340	0.7401	0.7347	0.7352	0.7355	0.7451	0.7132	0.7154	0.7178	0.7210

Table A.21: Estimation results of different HAR model specifications for realized ranges (variance).

	HAR	HAR + R^{log}	HAR + RR	HAR + RV	HAR + size	HAR + count	HAR + vol
c	0.0014 **	0.0020 ***	0.0017 ***	0.0017 ***	0.0062 ***	-0.0015	-0.0007
$R^{log,(1)}(-1)$	0.0466	-0.0373	-0.0775 *	-0.0931 **	0.0627	0.0305	0.0313
$R^{log,(5)}(-1)$	0.3958 ***	0.3469 ***	0.1689	0.1441	0.3561 ***	0.4005 ***	0.4003 ***
$R^{log,(22)}(-1)$	0.4444 ***	0.3891 ***	0.2741 ***	0.2931 ***	0.3643 ***	0.4377 ***	0.4514 ***
preAsian(-1)		-0.0411	0.0534	0.3548	0.0005	-0.0002	-0.0001
Asian(-1)		0.1986 **	0.3623	0.2977	-0.0010	0.0002	0.0002
preEurope(-1)		-0.0675	0.1755	0.1834	0.0022	0.0000	0.0001
Europe(-1)		-0.0607	0.1043	0.0138	-0.0066 ***	-0.0004	-0.0007 *
preAmerican(-1)		-0.0624	-0.1394	0.0863	0.0031	-0.0001	0.0000
America1(-1)		0.0759	0.3148	0.2586	0.0024	0.0003	0.0003
America2(-1)		0.0780	0.6316 ***	0.7355 ***	-0.0049 **	0.0002	0.0000
postAm1(-1)		0.3169 **	0.6446 *	0.5111	0.0000	0.0005	0.0003
postAm2(-1)		0.1168	0.4086	0.2390	0.0018 **	-0.0002	0.0001
$AdjR^2$	0.3557	0.3638	0.3864	0.3888	0.3618	0.3527	0.3529

Table A.22: Estimation results of enriching a HAR of daily ranges (volatility) with yesterday's sessions variables.

	HAR	HAR + R^{log}	HAR + RR	HAR + RV	HAR + size	HAR + count	HAR + vol
c	0.0000 **	0.0000	0.0000 **	0.0000 ***	0.0002 **	0.0000	0.0000
$R^{log,(1)}(-1)$	0.0489	-0.0097	-0.0893 *	-0.1134 **	0.0604	0.0385	0.0389
$R^{log,(5)}(-1)$	0.4357 ***	0.4284	0.3034 *	0.2573 *	0.4033 ***	0.4483 ***	0.4467 ***
$R^{log,(22)}(-1)$	0.3885 ***	0.3578	0.2337 **	0.2364 **	0.3060 ***	0.3711 ***	0.3809 ***
preAsian(-1)		0.4144	3.3519	6.0777	0.0000	0.0000	0.0000
Asian(-1)		0.2719	0.6564	0.3899	0.0000	0.0000	0.0000
preEurope(-1)		-0.5601	0.6404	1.7077	0.0001	0.0000	0.0000
Europe(-1)		-0.1385	0.5418	0.1000	-0.0002 ***	0.0000	0.0000 *
preAmerican(-1)		-0.5208	-2.6053 **	-0.8287	0.0001	0.0000	0.0000
America1(-1)		0.1701	1.2421	0.8061	0.0001	0.0000	0.0000
America2(-1)		0.0956	2.1567 ***	2.6786	-0.0002 **	0.0000	0.0000
postAm1(-1)		1.0767	5.8685	4.4220 ***	0.0000	0.0000	0.0000
postAm2(-1)		0.1917	-1.3047	-2.5796	0.0001 **	0.0000	0.0000
$AdjR^2$	0.3346	0.3432	0.3713	0.3790	0.3402	0.3320	0.3317

Table A.23: Estimation results of enriching a HAR of daily ranges (variance) with yesterday's sessions variables.

	HAR	HAR + R^{log}	HAR + RR	HAR + RV	HAR + size	HAR + count	HAR + vol
c	0.0004	0.0006	0.0008	0.0007	-0.0006	0.0000	-0.0004
	**	***	***	***			
$RR^{(1)}$(-1)	0.4586	0.4230	**-1.1137**	**-0.1061**	0.4581	0.4492	0.4472
	***	***	***		***	***	***
$RR^{(5)}$(-1)	0.3304	0.3170	0.2191	0.2615	0.3318	0.3325	0.3369
	***	***	***	***	***	***	***
$RR^{(22)}$(-1)	0.1605	0.1488	0.1114	0.1597	0.1850	0.1519	0.1604
	***	***	**	***	***	***	***
preAsian(-1)		-0.0223	0.3710	0.2023	0.0000	0.0000	0.0000
			***	**			
Asian(-1)		0.0298	0.6418	0.2441	-0.0006	-0.0002	-0.0002
			***	***			
preEurope(-1)		0.0000	0.4778	0.2108	0.0002	0.0001	0.0001
			***	**			
Europe(-1)		-0.0179	0.5842	0.1276	-0.0010	-0.0001	-0.0002
			***	**			
preAmerican(-1)		0.0162	0.4367	0.1696	0.0013	0.0002	0.0003
			***	**	*		*
America1(-1)		-0.0219	0.6405	0.2005	0.0006	-0.0002	-0.0002
			***	***			
America2(-1)		-0.0077	0.8811	0.2822	-0.0005	0.0000	0.0000
			***	***			
postAm1(-1)		0.1508	0.7032	0.4030	0.0000	0.0002	0.0002
		***	***	***			
postAm2(-1)		0.0221	0.5867	0.3211	0.0005	0.0002	0.0002
			***	***			*
AdjR^2	0.7470	0.7511	0.7681	0.7642	0.7480	0.7489	0.7498

Table A.24: Estimation results of enriching a HAR of realized ranges (volatility) with yesterday's sessions variables.

	HAR	HAR + R^{log}	HAR + RR	HAR + RV	HAR + size	HAR + count	HAR + vol
c	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	**	***	***	***			
$RR^{(1)}$(-1)	0.2755	0.4925	-0.6123	-0.0242	0.4892	0.4971	0.4870
	***	***			***	***	***
$RR^{(5)}$(-1)	0.1708	0.2385	0.2377	0.2327	0.2797	0.2730	0.2836
	***	***	**	**	***	***	***
$RR^{(22)}$(-1)	0.4976	0.1728	0.1642	0.1853	0.1607	0.1910	0.1678
	***	***	***	***	***	***	***
preAsian(-1)		0.0081	1.3003	1.1314	0.0000	0.0000	0.0000
				**			
Asian(-1)		0.0953	1.1690	0.5978	0.0000	0.0000	0.0000
		**	*	**			
preEurope(-1)		-0.1028	2.2815	1.2675	0.0000	0.0000	0.0000
			***	**			
Europe(-1)		-0.0497	0.7993	0.0942	0.0000	0.0000	0.0000
						*	
preAmerican(-1)		0.0013	0.7431	0.4165	0.0000	0.0000	0.0000
						*	**
America1(-1)		-0.0137	0.9945	0.4091	0.0000	0.0000	0.0000
			*				
America2(-1)		-0.0194	0.9788	0.4479	0.0000	0.0000	0.0000
			*				
postAm1(-1)		0.2992	2.8740	1.6037	0.0000	0.0000	0.0000
		*	***	*			
postAm2(-1)		0.0337	2.1294	1.2603	0.0000	0.0000	0.0000
			*				
AdjR^2	0.7234	0.7331	0.7357	0.7423	0.7245	0.7244	0.7255

Table A.25: Estimation results of enriching a HAR of realized ranges (variance) with yesterday's sessions variables.

	HAR	+ preAs	+ As	+ preEu	+ Eu	+ preAm	+ Am1	+ Am2	+ postAm1	+ postAm2
c	0.0014 **	0.0070 **	0.0088 ***	0.0050 *	0.0099 ***	0.0042	0.0061	0.0111 ***	0.0092 ***	0.0041 *
$R^{log,(1)}(-1)$	0.0466	0.0513	0.0449	0.0443	0.0765 *	0.0507	0.0521	0.0090	-0.0080	0.0231
$R^{log,(5)}(-1)$	0.3958 ***	0.3269 ***	0.2648 **	0.2977 ***	0.2673 **	0.3639 ***	0.3428 ***	0.2432 ***	0.3391 ***	0.3520 ***
$R^{log,(22)}(-1)$	0.4444 ***	0.3678 ***	0.3507 ***	0.3963 ***	0.3093 ***	0.4033 ***	0.3810 ***	0.2487 ***	0.2763 ***	0.4025 ***
session_size(-1)		-0.0006	-0.0017	0.0000	-0.0025 **	-0.0006	-0.0007	-0.0012	-0.0001	0.0007
session_count(-1)		-0.0007 *	-0.0006 *	-0.0005	-0.0005	-0.0002	-0.0004	-0.0008 *	-0.0010 **	-0.0006
session_rr(-1)		1.5741 **	0.7989 **	1.5748 ***	1.1203 ***	0.5863	0.8348 **	1.3642 ***	3.5647 ***	1.4440 **
session_rng(-1)		-0.2103 *	-0.0045	-0.1602 *	-0.2492 ***	-0.1253	-0.1729	-0.1899 **	-0.7318 **	0.0314
$AdjR^2$	0.3557	0.3602	0.3636	0.3595	0.3669	0.3551	0.3605	0.3856	0.3802	0.3620

Table A.26: Estimation results of enriching a HAR of daily ranges (volatility) with all lagged variables of separate sessions.

	HAR	+ preAs	+ As	+ preEu	+ Eu	+ preAm	+ Am1	+ Am2	+ postAm1	+ postAm2
c	0.0000 **	0.0002 *	0.0001	0.0001	0.0003 **	0.0000	0.0001	0.0003 **	0.0002 *	0.0001
$R^{log,(1)}(-1)$	0.0489	0.0576	0.0636	0.0575	0.0855	0.0761	0.0657	0.0080	-0.0152	0.0332
$R^{log,(5)}(-1)$	0.4357 ***	0.3570 ***	0.3180 **	0.3325 **	0.3349 **	0.4181 ***	0.3993 ***	0.3086 **	0.4144 ***	0.4317 ***
$R^{log,(22)}(-1)$	0.3885 ***	0.3162 ***	0.3575 ***	0.3604 ***	0.2758 **	0.3977 ***	0.3397 ***	0.2060 *	0.2509 **	0.3761 ***
session_size(-1)		0.0000	0.0000	0.0000	-0.0001 **	0.0000	0.0000	0.0000	0.0000	0.0000
session_count(-1)		0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000 *	0.0000
session_rr(-1)		12.0724	2.6825	9.4780 *	3.6780 ***	-0.5068	2.7778 *	3.9588 ***	22.4499 ***	6.1127
session_rng(-1)		-0.4198 ***	-0.2372	-0.3946 ***	-0.5385 **	-0.4091 **	-0.4145	-0.3745 **	-2.3736	0.0406
$AdjR^2$	0.3346	0.3435	0.3401	0.3383	0.3486	0.3356	0.3384	0.3718	0.3583	0.3345

Table A.27: Estimation results of enriching a HAR of daily ranges (variance) with all lagged variables of separate sessions.

	HAR	+ preAs	+ As	+ preEu	+ Eu	+ preAm	+ Am1	+ Am2	+ postAm1	+ postAm2
c	0.0004 **	0.0011	0.0022 **	0.0010	-0.0002	-0.0021 *	-0.0013	-0.0006	0.0012	0.0000
$RR^{(1)}(-1)$	0.4586 ***	0.4541 ***	0.4323 ***	0.4247 ***	0.4704 ***	0.4536 ***	0.5046 ***	0.3919 ***	0.3496 ***	0.4129 ***
$RR^{(5)}(-1)$	0.3304 ***	0.3043 ***	0.3005 ***	0.2889 ***	0.3405 ***	0.3465 ***	0.3259 ***	0.3470 ***	0.3407 ***	0.3172 ***
$RR^{(22)}(-1)$	0.1605 ***	0.1543 ***	0.1446 ***	0.1694 ***	0.1675 ***	0.1963 ***	0.1910 ***	0.1695 ***	0.1311 **	0.1728 ***
session_size(-1)		0.0000	-0.0004	0.0002	0.0003	0.0008 **	0.0007 *	0.0005	0.0004	0.0004
session_count(-1)		-0.0001	-0.0002	-0.0001	0.0000	0.0001	0.0001	0.0000	-0.0002	0.0000
session_rr(-1)		0.3578 **	0.1375	0.5336 ***	0.0030	-0.0452	-0.0977	0.1498	0.8308 ***	0.4654 **
session_rng(-1)		-0.0836 **	0.0098	-0.0694 **	-0.0247	-0.0056	-0.0193	-0.0219	-0.0909	-0.0056
$AdjR^2$	0.7470	0.7476	0.7475	0.7490	0.7464	0.7474	0.7481	0.7474	0.7557	0.7511

Table A.28: Estimation results of enriching a HAR of realized ranges (volatility) with all lagged variables of separate sessions.

	HAR	+ preAs	+ As	+ preEu	+ Eu	+ preAm	+ Am1	+ Am2	+ postAm1	+ postAm2
c	0.0000 **	0.0000 *	0.0000	0.0000	0.0000	-0.0001 **	0.0000	0.0000	0.0000	0.0000
$RR^{(1)}(-1)$	0.2755 ***	0.4685 ***	0.4912 ***	0.4562 ***	0.5657 ***	0.5210 ***	0.5431 ***	0.4674 ***	0.4074 ***	0.4635 ***
$RR^{(5)}(-1)$	0.1708 ***	0.2455 **	0.2573 ***	0.2302 ***	0.2943 ***	0.2845 ***	0.2625 ***	0.2883 ***	0.2843 ***	0.2712 ***
$RR^{(22)}(-1)$	0.4976 ***	0.1638 ***	0.1715 ***	0.1911 ***	0.1584 ***	0.2102 ***	0.2050 ***	0.1834 ***	0.1579 ***	0.1815 ***
session_size(-1)		0.0000	0.0000	0.0000	0.0000	0.0000 *	0.0000	0.0000	0.0000	0.0000
session_count(-1)		0.0000	0.0000	0.0000	0.0000	0.0000 *	0.0000	0.0000	0.0000	0.0000
session_rr(-1)		0.1560	1.5039	2.4761 **	-0.3969	-0.6925 ***	-0.3587	0.1242	1.9402	2.0295
session_rng(-1)		0.0009	-0.0019 **	-0.0020 **	-0.0003	0.0002	-0.0001	-0.0004	0.0006	0.0000
$AdjR^2$	0.7234	0.7240	0.7241	0.7257	0.7245	0.7257	0.7247	0.7230	0.7317	0.7267

Table A.29: Estimation results of enriching a HAR of realized ranges (variance) with all lagged variables of separate sessions.

	HAR	+ preAs	+ As	+ preEu	+ Eu	+ preAm	+ Am1	+ Am2	+ postAm1	+ postAm2
c	0.0014 **	0.0069 **	0.0031	0.0052	0.0080 ***	-0.0078 **	-0.0081 *	-0.0051	0.0015	-0.0042 *
$R^{log,(1)}(-1)$	0.0466	0.0165	-0.0040	-0.0229	-0.0201	-0.0092	0.0267	0.0025	-0.0330	0.0013
$R^{log,(5)}(-1)$	0.3958 ***	0.3114 ***	0.2200 **	0.2204 **	0.1404	0.3036 ***	0.3312 ***	0.2029 **	0.2941 ***	0.2612 ***
$R^{log,(22)}(-1)$	0.4444 ***	0.2990 ***	0.3478 ***	0.2572 ***	0.2678 ***	0.3011 ***	0.3544 ***	0.1619 **	0.1544 *	0.3185 ***
session_size(-1)		-0.0008	-0.0006	-0.0010	0.0000	0.0010	0.0009	0.0021 *	-0.0003	-0.0012 *
session_count(-1)		-0.0006 *	-0.0001	-0.0004	-0.0009 **	0.0009 **	0.0008 *	0.0002	0.0003	0.0013 ***
session_rr(-1)		1.1895 **	0.5216	2.5973 ***	1.1779 ***	1.1137 **	-0.0237	0.8843 ***	3.1053 ***	1.9973 **
session_rng(-1)		0.4868 ***	0.5265 ***	0.4403 ***	0.6013 ***	0.4932 ***	0.6079 ***	0.6159 ***	0.0931	0.3552 ***
$AdjR^2$	0.3557	0.3926	0.4277	0.4380	0.4897	0.4449	0.4487	0.5803	0.4712	0.4570

Table A.30: Estimation results of enriching a HAR of daily ranges (volatility) with all non-lagged variables of separate sessions.

	HAR	+ preAs	+ As	+ preEu	+ Eu	+ preAm	+ Am1	+ Am2	+ postAm1	+ postAm2
c	0.0000 **	0.0002 **	0.0000	0.0002 *	0.0002 **	-0.0003 **	-0.0002	-0.0001	0.0000	-0.0002 **
$R^{log,(1)}(-1)$	0.0489	0.0154	-0.0171	-0.0411	-0.0305	0.0070	0.0450	0.0192	-0.0339	-0.0066
$R^{log,(5)}(-1)$	0.4357 ***	0.4279 ***	0.3389 **	0.2817 ***	0.2227 *	0.4447 ***	0.4245 ***	0.2915 ***	0.3900 ***	0.4089 ***
$R^{log,(22)}(-1)$	0.3885 ***	0.2057 *	0.2969 ***	0.1426	0.2424 **	0.2212 *	0.2493 **	0.0556	0.1248	0.1934 *
session_size(-1)		0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000 *
session_count(-1)		0.0000 *	0.0000	0.0000	0.0000 **	0.0000 ***	0.0000 *	0.0000	0.0000	0.0000 ***
session_rr(-1)		10.2267 ***	1.2881	26.1560 ***	4.0828 ***	1.5179	-1.2037	2.4316 **	18.7943 **	20.6278 *
session_rng(-1)		1.0867 ***	1.1967 ***	0.7474 ***	1.2093 ***	1.5482 ***	1.3817 ***	1.1150 ***	1.1391	0.7087 ***
$AdjR^2$	0.3346	0.3907	0.4320	0.4519	0.4969	0.4154	0.4140	0.5833	0.4689	0.4568

Table A.31: Estimation results of enriching a HAR of daily ranges (variance) with all non-lagged variables of separate sessions.

	HAR	+ preAs	+ As	+ preEu	+ Eu	+ preAm	+ Am1	+ Am2	+ postAm1	+ postAm2
c	0.0004 **	0.0016 **	0.0011	0.0008	0.0022 **	-0.0016	-0.0005	0.0015	0.0017 **	0.0002
$RR^{(1)}$(-1)	0.4586 ***	0.3660 ***	0.2940 ***	0.3179 ***	0.2281 ***	0.3454 ***	0.3375 ***	0.2683 ***	0.3454 ***	0.3908 ***
$RR^{(5)}$(-1)	0.3304 ***	0.2951 ***	0.1681 **	0.1971 ***	0.1816 ***	0.2614 ***	0.2864 ***	0.1598 ***	0.2554 ***	0.2558 ***
$RR^{(22)}$(-1)	0.1605 ***	0.1318 ***	0.1653 ***	0.1513 ***	0.1267 **	0.1020 **	0.1083 ***	-0.0245	0.0180	0.1380 ***
session_size(-1)		0.0000	-0.0002	-0.0002	0.0001	0.0001	-0.0002	-0.0004	-0.0004	-0.0004 *
session_count(-1)		-0.0002 *	0.0000	-0.0001	-0.0003 **	0.0002	0.0001	0.0000	0.0000	0.0001
session_rr(-1)		0.5574 ***	0.7572 ***	1.3075 ***	1.1860 ***	0.8643 ***	0.5935 ***	0.9763 ***	1.6599 ***	1.1601 ***
session_rng(-1)		0.1454 ***	0.0180	0.0639 *	-0.0270	0.0631	0.0609 *	0.0032	-0.1770 **	-0.0278
AdjR^2	0.7470	0.7720	0.8065	0.8009	0.8343	0.8083	0.8206	0.8967	0.8139	0.7784

Table A.32: Estimation results of enriching a HAR of realized ranges (volatility) with all non-lagged variables of separate sessions.

	HAR	+ preAs	+ As	+ preEu	+ Eu	+ preAm	+ Am1	+ Am2	+ postAm1	+ postAm2
const	0.0000 **	0.0000	0.0000	0.0000	0.0000	-0.0001 **	0.0000	0.0000	0.0000	0.0000
$RR^{(1)}$(-1)	0.2755 ***	0.4038 ***	0.3432 ***	0.3174 ***	0.2284 ***	0.4166 ***	0.3962 ***	0.3174 ***	0.4033 ***	0.4467 ***
$RR^{(5)}$(-1)	0.1708 ***	0.2615 ***	0.0654	0.1072	0.1779 *	0.2445 ***	0.2879 ***	0.1701 ***	0.2251 **	0.2459 **
$RR^{(22)}$(-1)	0.4976 ***	0.1320 **	0.2404 ***	0.1785 ***	0.1707 **	0.1126 **	0.0764	-0.0363	0.0466	0.1517 ***
session_size(-1)		0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000 *	0.0000 **	0.0000
session_count(-1)		0.0000	0.0000	0.0000	0.0000 **	0.0000 **	0.0000	0.0000	0.0000	0.0000 **
session_rr(-1)		3.2941 ***	1.5640 ***	7.6308 ***	2.5546 ***	1.8434 ***	1.2749 ***	1.5566 ***	6.3087 ***	4.2190 ***
session_rng(-1)		0.0866 **	0.0514	-0.0102	-0.0328	0.1180 ***	0.0968 *	0.0088	-0.3830 *	0.0034
AdjR^2	0.7234	0.7491	0.7928	0.7966	0.8314	0.7739	0.7880	0.8859	0.7892	0.7447

Table A.33: Estimation results of enriching a HAR of realized ranges (variance) with all non-lagged variables of separate sessions.

				+ preAs		+ As		+ preEu		+ Eu		+ preAm		+ Am1		+ Am2		+ postAm1		+ postAm2			
		HAR		**+ 1 session**		**+ 2 sessions**		**+ 3 sessions**		**+ 4 sessions**		**+ 5 sessions**		**+ 6 sessions**		**+ 7 sessions**		**+ 8 sessions**		**+ 9 sessions**			
HAR	**c**	0.0014	**	0.0069	**	0.0036		0.0032		0.0054		-0.0016		-0.0086	**	-0.0085	**	-0.0100	**	-0.0088	**	**c**	**HAR**
	$R^{log,(1)}(-1)$	0.0466		0.0165		-0.0074		-0.0369		-0.0575		-0.0729	*	-0.0717	**	-0.0504	*	-0.0598	**	-0.0561	*	**$R^{log,(1)}(-1)$**	
	$R^{log,(5)}(-1)$	0.3958	***	0.3114	***	0.2081	**	0.1456		0.0383		0.0275		0.0094		-0.0314		-0.0194		-0.0693		**$R^{log,(5)}(-1)$**	
	$R^{log,(22)}(-1)$	0.4444	***	0.2990	***	0.3163	***	0.2636	***	0.2335	***	0.2072	**	0.1948	**	0.0812		0.0682		0.0207		**$R^{log,(22)}(-1)$**	
preAs	**size**			-8.223E-04		0.0003		0.0001		-0.0003		-0.0005		-0.0008		-0.0006		-0.0005		-0.0005		**size**	**preAs**
	count			-0.0006	*	-0.0007		-0.0005		-0.0001		-0.0003		-0.0002		0.0001		0.0001		0.0003		**count**	
	rr			1.1895	**	0.9638		1.2342	**	0.9701	*	1.2331	**	1.2847	**	0.6371		0.3973		0.1999		**rr**	
	rng			0.4868	***	0.1751		-0.1475		-0.2888		-0.3788	**	-0.4299	**	-0.3536	**	-0.2640	*	-0.3549	**	**rng**	
As	**size**					-0.0007		-0.0003		-0.0009		-0.0003		0.0005		0.0000		0.0003		0.0004		**size**	**As**
	count					0.0004		0.0008		0.0003		0.0003		-0.0001		-0.0003		-0.0005		-0.0010	**	**count**	
	rr					0.2347		-0.0964		-0.0907		-0.1594		0.0365		-0.0060		-0.0028		0.3406		**rr**	
	rng					0.4616	***	0.2987	**	0.2862	**	0.3033	**	0.3182	***	0.3907	***	0.3819	***	0.3246	***	**rng**	
preEu	**size**							-0.0001		0.0006		0.0012		0.0004		0.0000		0.0000		0.0002		**size**	**preEu**
	count							-0.0007		-0.0009		-0.0011	*	-0.0009		-0.0006		-0.0004		-0.0005		**count**	
	rr							2.0927	***	2.1608	***	2.0065	***	1.8218	**	1.6384	***	1.4940	**	1.6822	***	**rr**	
	rng							0.3122	*	0.0090		0.0191		-0.0379		-0.2573	*	-0.2876	**	-0.3628	***	**rng**	
Eu	**size**									0.0010		0.0013		0.0017		-0.0003		-0.0007		-0.0006		**size**	**Eu**
	count									-0.0002		-0.0014	**	-0.0020	***	-0.0016	***	-0.0015	***	-0.0014	***	**count**	
	rr									0.3731		0.2776		0.2715		-0.3505		-0.2998		-0.2937		**rr**	
	rng									0.5704	***	0.5890	***	0.6073	***	0.7030	***	0.7007	***	0.6706	***	**rng**	
preAm	**size**											-0.0002		-0.0008		-0.0013		-0.0011		-0.0012		**size**	**preAm**
	count											0.0022	***	0.0020	***	0.0019	***	0.0018	***	0.0018	***	**count**	
	rr											0.6279		0.5540		0.4935		0.3884		0.4252		**rr**	
	rng											-0.0055		-0.1102		-0.2387	*	-0.2064		-0.2195	*	**rng**	
Am1	**size**													0.0014		0.0000		-0.0005		-0.0006		**size**	**Am1**
	count													0.0016	***	0.0017	***	0.0014	**	0.0012	**	**count**	
	rr													-0.2474		-0.3072		-0.1530		-0.0129		**rr**	
	rng													0.3560	***	0.2303	**	0.2327	**	0.2199	**	**rng**	
Am2	**size**															0.0047	***	0.0060	***	0.0051	***	**size**	**Am2**
	count															-0.0007		-0.0005		-0.0003		**count**	
	rr															0.5888	**	0.3675		0.2984		**rr**	
	rng															0.5762	***	0.5482	***	0.5478	***	**rng**	
postAm1	**size**																	-0.0012		-0.0016	**	**size**	**postAm1**
	count																	0.0005		-0.0001		**count**	
	rr																	-0.5446		-0.2208		**rr**	
	rng																	0.5942	**	0.4409	*	**rng**	
postAm2	**size**																			-0.0005		**size**	**postAm2**
	count																			0.0014	***	**count**	
	rr																			-0.6192		**rr**	
	rng																			0.3197	***	**rng**	
	AdjR^2	0.3557		0.4330		0.4330		0.4568		0.5213		0.5502		0.5903		0.7090		0.7189		0.7406		**AdjR^2**	

Table A.34: Estimation results of enriching a HAR of daily ranges (volatility) with all non-lagged variables of separate sessions, cumulatively.

			+ preAs	+ As	+ preEu	+ Eu	+ preAm	+ Am1	+ Am2	+ postAm1	+ postAm2		
		HAR	+ 1 session	+ 2 sessions	+ 3 sessions	+ 4 sessions	+ 5 sessions	+ 6 sessions	+ 7 sessions	+ 8 sessions	+ 9 sessions		
HAR	**c**	0.0000 **	0.0002 **	0.0001	0.0001	0.0002 *	0.0000	-0.0003 **	-0.0002	-0.0003 *	-0.0002	**c**	**HAR**
	$R^{log,(1)}(-1)$	0.0489	0.0154	-0.0224	-0.0526	-0.0779	-0.0890 *	-0.0835 *	-0.0537	-0.0716 *	-0.0692	**$R^{log,(1)}(-1)$**	
	$R^{log,(5)}(-1)$	0.4357 ***	0.4279 ***	0.3340 **	0.2474 **	0.1266	0.1219	0.1106	0.0628	0.0641	0.0152	**$R^{log,(5)}(-1)$**	
	$R^{log,(22)}(-1)$	0.3885 ***	0.2057 *	0.2513 **	0.1661	0.2164 *	0.2023 *	0.1907 *	0.0477	0.0474	-0.0366	**$R^{log,(22)}(-1)$**	
preAs	**size**		-3.046E-05	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	**size**	**preAs**
	count		0.0000 *	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	**count**	
	rr		10.2267 ***	7.7554 *	8.7800 **	5.3054	6.7751	7.0405	2.7240	1.4226	0.5757	**rr**	
	rng		1.0867 ***	0.0557	-0.8254	-1.0622	-1.2623 *	-1.1749	-0.8697	-0.6276	-1.0272 **	**rng**	
As	**size**			0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	**size**	**As**
	count			0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	**count**	
	rr			0.6409	-1.0789	-0.8673	-0.9632	-0.4845	-0.8453	-0.5588	0.6959	**rr**	
	rng			1.1062 ***	0.8034 **	0.7591 **	0.7823 ***	0.7597 ***	0.9195 ***	0.8725 ***	0.7124 ***	**rng**	
preEu	**size**				0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	**size**	**preEu**
	count				0.0000	0.0000	0.0000 *	0.0000 *	0.0000	0.0000	0.0000	**count**	
	rr				21.0380 ***	17.1055 ***	16.1247 ***	16.0165 ***	12.4311 **	11.3931 **	13.0675 ***	**rr**	
	rng				0.6858	-0.0017	0.1124	-0.0400	-0.5635	-0.6898	-0.8704 *	**rng**	
Eu	**size**					0.0000	0.0001	0.0001	0.0000	0.0000	0.0000	**size**	**Eu**
	count					0.0000	-0.0001 ***	-0.0001 ***	-0.0001 ***	-0.0001 ***	0.0000 ***	**count**	
	rr					1.5001	1.3675	1.0122	-1.9241	-1.4163	-1.5678	**rr**	
	rng					1.2097 ***	1.2493 ***	1.2791 ***	1.5272 ***	1.4925 ***	1.3603 ***	**rng**	
preAm	**size**						0.0000	-0.0001	-0.0001 *	-0.0001 *	-0.0001 **	**size**	**preAm**
	count						0.0001 ***	0.0001 ***	0.0001 ***	0.0001 ***	0.0001 ***	**count**	
	rr						1.5616	2.2873	2.1128	1.6732	2.0848 *	**rr**	
	rng						-0.0572	-0.4158	-0.5995	-0.4499	-0.5741	**rng**	
Am1	**size**							0.0000	0.0000	0.0000	0.0000	**size**	**Am1**
	count							0.0001 ***	0.0000 ***	0.0000 **	0.0000 **	**count**	
	rr							-1.3864	-0.4427	0.2105	0.4741	**rr**	
	rng							0.6569 **	0.2333	0.2367	0.3129	**rng**	
Am2	**size**								0.0001 ***	0.0002 ***	0.0001 ***	**size**	**Am2**
	count								0.0000	0.0000	0.0000	**count**	
	rr								2.2327 **	1.8245 *	1.6612 *	**rr**	
	rng								0.9364 ***	0.8471 ***	0.8823 ***	**rng**	
postAm1	**size**									0.0000	0.0000 *	**size**	**postAm1**
	count									0.0000 **	0.0000	**count**	
	rr									-9.9770 *	-8.2860 *	**rr**	
	rng									3.3281 ***	2.9421 ***	**rng**	
postAm2	**size**										0.0000	**size**	**postAm2**
	count										0.0000 ***	**count**	
	rr										0.0852	**rr**	
	rng										0.6358 ***	**rng**	
	$AdjR^2$	0.3346	0.3907	0.4352	0.4672	0.5350	0.5522	0.5738	0.7152	0.7339	0.7655	**$AdjR^2$**	

Table A.35: Estimation results of enriching a HAR of daily ranges (variance) with all non-lagged variables of separate sessions, cumulatively.

				+ preAs		+ As		+ preEu		+ Eu		+ preAm		+ Am1		+ Am2		+ postAm1		+ postAm2			
		HAR		+ 1 session		+ 2 sessions		+ 3 sessions		+ 4 sessions		+ 5 sessions		+ 6 sessions		+ 7 sessions		+ 8 sessions		+ 9 sessions			
HAR	c	0.0004	**	0.0016	**	0.0007		0.0004		0.0013		0.0004		-0.0006		0.0009	***	0.0010	***	0.0012	***	c	HAR
	$RR^{(1)}(-1)$	0.4586	***	0.3660	***	0.2741	***	0.2141	***	0.0991	***	0.0675	**	0.0040		-0.0108		-0.0116	*	-0.0168	***	$RR^{(1)}(-1)$	
	$RR^{(5)}(-1)$	0.3304	***	0.2951	***	0.1684	***	0.1242	**	0.0795	*	0.0666		0.0624	**	-0.0051		-0.0093		-0.0160		$RR^{(5)}(-1)$	
	$RR^{(22)}(-1)$	0.1605	***	0.1318	***	0.1699	***	0.1676	***	0.1444	***	0.1173	***	0.0951	***	-0.0055		-0.0156	*	-0.0189	**	$RR^{(22)}(-1)$	
preAs	size			4.569E-05		0.0006	**	0.0004	*	0.0003		0.0002		0.0000		0.0000		0.0000		0.0000		size	preAs
	count			-0.0002	*	-0.0002		-0.0002		-0.0002		-0.0002		-0.0001		0.0000		0.0000		0.0000		count	
	rr			0.5574	***	0.1655		0.3185	**	0.3654	***	0.4743	***	0.5128	***	0.3015	***	0.2589	***	0.2287	***	rr	
	rng			0.1454	***	0.1316	***	0.0248		-0.0001		-0.0451		-0.0664	*	-0.0233	**	-0.0057		-0.0038		rng	
As	size					-0.0005	*	-0.0004		-0.0005		-0.0004		0.0001		-0.0001		-0.0001		-0.0001		size	As
	count					0.0002		0.0003	*	0.0001		0.0002		0.0001		0.0001	**	0.0000		0.0000		count	
	rr					0.7003	***	0.5591	***	0.4729	***	0.4332	***	0.4868	***	0.3873	***	0.3886	***	0.4002	***	rr	
	rng					-0.0264		-0.0719	**	-0.0446	*	-0.0308		-0.0329		0.0063		0.0040		0.0031		rng	
preEu	size							-0.0001		0.0004		0.0006	*	0.0003		0.0001		0.0001		0.0001	*	size	preEu
	count							-0.0001		-0.0003	*	-0.0003	**	-0.0003	**	-0.0001		0.0000		0.0000		count	
	rr							0.8934	***	0.5975	***	0.4594	***	0.4039	***	0.2053	***	0.1764	***	0.1644	***	rr	
	rng							0.0600		0.0361		0.0531	*	0.0608	*	0.0133		0.0075		0.0034		rng	
Eu	size									0.0001		0.0001		0.0004		-0.0001		-0.0002		-0.0002		size	Eu
	count									0.0001		0.0000		0.0000		0.0001		0.0001		0.0001		count	
	rr									0.8521	***	0.6966	***	0.6217	***	0.3654	***	0.3757	***	0.3786	***	rr	
	rng									-0.0524	***	-0.0410	***	-0.0304	**	0.0058		0.0059		0.0049		rng	
preAm	size											0.0000		0.0003		0.0000		0.0000		0.0000		size	preAm
	count											0.0001		0.0002	*	-0.0001		-0.0001		-0.0001		count	
	rr											0.6816	***	0.3523	***	0.3312	***	0.3169	***	0.3087	***	rr	
	rng											-0.0328		-0.0143		-0.0056		-0.0029		-0.0016		rng	
Am1	size													-0.0005		0.0001		0.0000		-0.0001		size	Am1
	count													0.0000		0.0001	**	0.0001		0.0001		count	
	rr													0.5524	***	0.4197	***	0.4455	***	0.4502	***	rr	
	rng													0.0136		-0.0099		-0.0060		-0.0050		rng	
Am2	size															0.0002		0.0002		0.0001		size	Am2
	count															-0.0003	***	-0.0003	***	-0.0002	***	count	
	rr															0.7308	***	0.6696	***	0.6547	***	rr	
	rng															0.0063		0.0005		0.0016		rng	
postAm1	size																	0.0001		0.0000		size	postAm1
	count																	0.0000		0.0000		count	
	rr																	0.1258	***	0.1392	***	rr	
	rng																	0.0417	***	0.0102		rng	
postAm2	size																			0.0000		size	postAm2
	count																			0.0000		count	
	rr																			0.2074	***	rr	
	rng																			0.0025		rng	
	$AdjR^2$	0.7470		0.7720		0.8130		0.8299		0.8699		0.8899		0.9305		0.9928		0.9944		0.9952		$AdjR^2$	

Table A.36: Estimation results of enriching a HAR of realized ranges (volatility) with all non-lagged variables of separate sessions, cumulatively.

				+ preAs		+ As		+ preEu		+ Eu		+ preAm		+ Am1		+ Am2		+ postAm1		+ postAm2			
		HAR		+ 1 session		+ 2 sessions		+ 3 sessions		+ 4 sessions		+ 5 sessions		+ 6 sessions		+ 7 sessions		+ 8 sessions		+ 9 sessions			
HAR	c	0.0000	**	0.0000		0.0000		0.0000		0.0000		0.0000		0.0000		0.0000		0.0000		0.0000		c	HAR
	$RR^{(1)}(-1)$	0.2755	***	0.4038	***	0.3145	***	0.2166	***	0.0912	**	0.0638	**	0.0205		0.0102		0.0080		0.0004		$RR^{(1)}(-1)$	
	$RR^{(5)}(-1)$	0.1708	***	0.2615	***	0.0765		0.0195		0.0283		0.0205		0.0456		0.0021		0.0004		0.0010		$RR^{(5)}(-1)$	
	$RR^{(22)}(-1)$	0.4976	***	0.1320	**	0.2260	***	0.2145	***	0.2048	***	0.1828	***	0.1407	***	0.0173		0.0034		-0.0040	*	$RR^{(22)}(-1)$	
preAs	size			-3.258E-08		0.0000	*	0.0000		0.0000		0.0000		0.0000		0.0000		0.0000		0.0000		size	preAs
	count			0.0000		0.0000		0.0000		0.0000		0.0000		0.0000		0.0000		0.0000		0.0000	**	count	
	rr			3.2941	***	0.8838		1.1930	**	1.2554	***	1.7392	***	1.9460	***	1.3879	***	1.2123	***	1.1097	***	rr	
	rng			0.0866	**	0.1466	***	0.0695		0.0414		-0.0487		-0.0612		-0.0330	**	-0.0080		-0.0048		rng	
As	size					0.0000	*	0.0000		0.0000		0.0000		0.0000		0.0000		0.0000		0.0000		size	As
	count					0.0000		0.0000	**	0.0000		0.0000		0.0000		0.0000		0.0000		0.0000		count	
	rr					1.7171	***	1.2668	***	1.0578	***	1.0212	***	1.1396	***	0.9442	***	0.9694	***	0.9921	***	rr	
	rng					-0.0497		-0.0906	**	-0.0417		-0.0227		-0.0335		0.0152	*	0.0035		0.0012		rng	
preEu	size							0.0000		0.0000		0.0000		0.0000		0.0000		0.0000		0.0000		size	preEu
	count							0.0000		0.0000	*	0.0000	**	0.0000	**	0.0000		0.0000		0.0000		count	
	rr							5.8133	***	3.1851	***	2.4444	***	2.1868	***	1.2614	***	1.0782	***	0.9978	***	rr	
	rng							-0.0129		0.0160		0.0908		0.0859		0.0163		0.0112		0.0063		rng	
Eu	size									0.0000		0.0000		0.0000	*	0.0000		0.0000		0.0000		size	Eu
	count									0.0000		0.0000		0.0000		0.0000		0.0000		0.0000		count	
	rr									2.0119	***	1.8159	***	1.5508	***	0.9341	***	0.9883	***	0.9903	***	rr	
	rng									-0.0815	***	-0.0565	**	-0.0434	**	0.0052		0.0005		-0.0006		rng	
preAm	size											0.0000		0.0000		0.0000		0.0000		0.0000		size	preAm
	count											0.0000	**	0.0000	**	0.0000		0.0000		0.0000		count	
	rr											1.5430	***	1.1522	***	1.0516	***	1.0261	***	1.0212	***	rr	
	rng											-0.0690	*	-0.0691		-0.0219	**	-0.0080		-0.0088		rng	
Am1	size													0.0000	*	0.0000	***	0.0000		0.0000		size	Am1
	count													0.0000		0.0000	***	0.0000		0.0000		count	
	rr													1.0135	***	0.9510	***	0.9994	***	1.0039	***	rr	
	rng													0.0427		-0.0106	*	-0.0034		-0.0016		rng	
Am2	size															0.0000		0.0000		0.0000		size	Am2
	count															0.0000	**	0.0000		0.0000		count	
	rr															1.1284	***	1.0260	***	1.0101	***	rr	
	rng															0.0098	*	-0.0034		-0.0013		rng	
postAm1	size																	0.0000		0.0000		size	postAm1
	count																	0.0000		0.0000		count	
	rr																	0.8261	***	0.9564	***	rr	
	rng																	0.0715	***	0.0044		rng	
postAm2	size																			0.0000		size	postAm2
	count																			0.0000		count	
	rr																			1.0609	***	rr	
	rng																			0.0033	***	rng	
	$AdjR^2$	0.7234		0.7491		0.7994		0.8275		0.8773		0.8946		0.9247		0.9963		0.9988		0.9996		$AdjR^2$	

Table A.37: Estimation results of enriching a HAR of realized ranges (variance) with all non-lagged variables of separate sessions, cumulatively.

	coefficient	std. error	z-value	p-value	Signif
c	0.000237	0.000109	2.172	0.0299	**
$(\hat{\sigma}_D^{Park})^2$	0.125382	0.019581	6.403	0.0000	***
λ_t	0.855489	0.024893	34.37	0.0000	***
	Log-lik	3358.603	AIC	-6711.21	
	SchC	-6696.54	HQC	-6705.63	

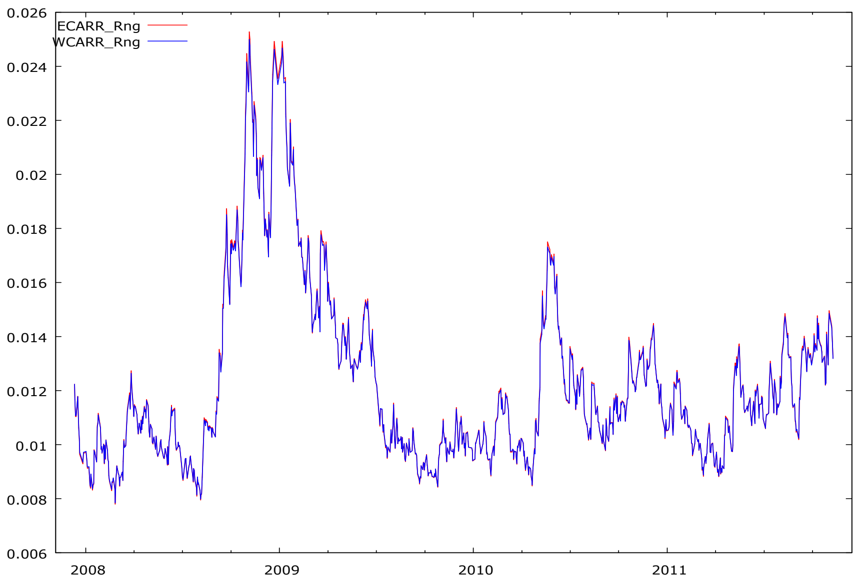
	coefficient	std. error	z-value	p-value	Signif
c	0.000239	0.000142	1.679	0.0931	*
$(\hat{\sigma}_D^{Park})^2$	0.120257	0.024412	4.926	0.0000	***
λ_t	0.860145	0.032996	26.07	0.0000	***
θ	2.91085	0.081051	35.91	1.86E-28	***
	Log-lik	3972.282	AIC	-7936.56	
	SchC	-7917.02	HQC	-7929.13	

Table A.38: Estimation results of a CARR(1,1) model applied to daily ranges with Exponentially (left) and Weibull (right) distributed error term

	coefficient	std. error	z-value	p-value	Signif
c	0.000318	0.000095	3.342	0.0008	***
$\sqrt{RR_D}$	0.432174	0.048672	8.879	0.0000	***
λ_t	0.526041	0.056391	9.328	0.0000	***
	Log-lik	3827.76	AIC	-7649.52	
	SchC	-7634.86	HQC	-7643.94	

	coefficient	std. error	z-value	p-value	Signif
c	0.000400	0.000187	2.138	0.0325	**
$\sqrt{RR_D}$	0.479382	0.079693	6.015	0.0000	***
λ_t	0.461144	0.098789	4.668	0.0000	***
θ	5.16716	0.248847	20.76	9.08E-96	***
	Log-lik	5021.462	AIC	-10034.9	
	SchC	-10015.3	HQC	-10027.5	

Table A.39: Estimation results of a CARR(1,1) model applied to square root of realized ranges with Exponentially (left) and Weibull (right) distributed error term.



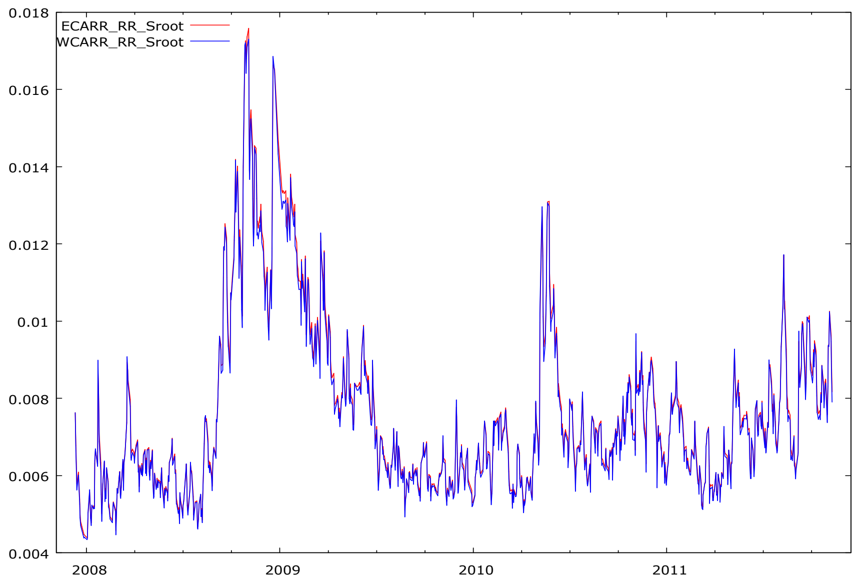
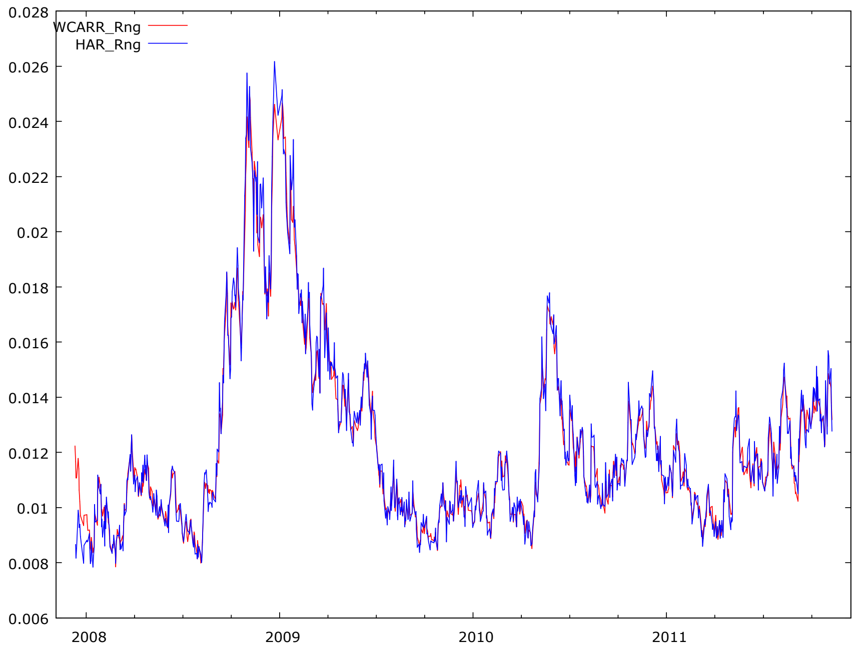


Figure A.40/A.41: In-sample comparison of E-CARR(1,1) and W-CARR(1,1) fitted values from Table A.38 (left) and Table A.39 (right).



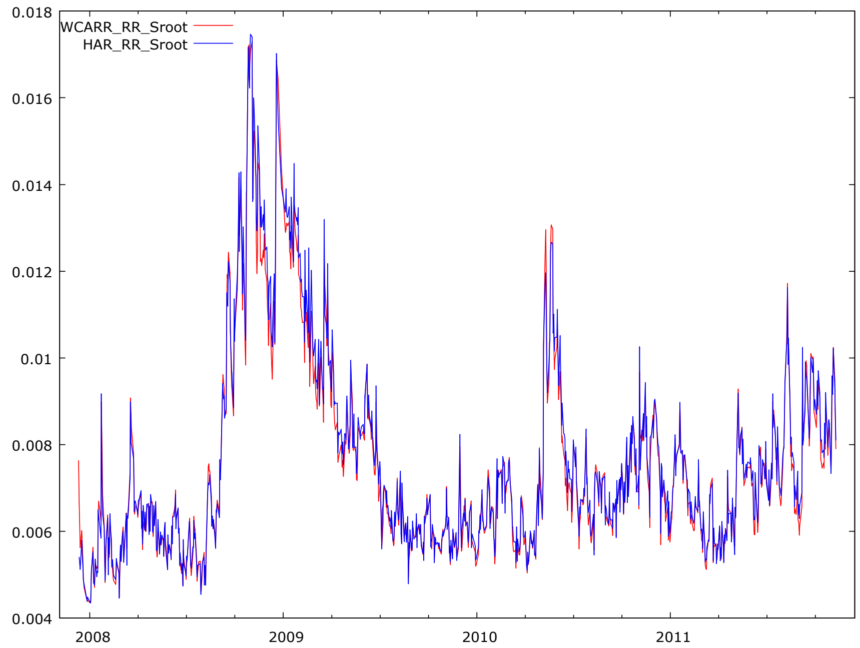


Figure A.42/A.43: In-sample comparison of W-CARR(1,1) and HAR applied to daily ranges and square roots of realized ranges.

	coefficient	std. error	t-ratio	p-value	
const	0.0229479	0.00085053	26.9807	<0.00001	***
AllLow	0.967126	0.00254251	380.3821	<0.00001	***

Log-Lik	3842.0	$AdjR^2$	0.99

Table A.44: Estimating the co-integrating vector or daily highs and lows. Dependent variable: AllHigh

A.45	Δ(h)		Δ(l)	
	coeff	**sign**	**coeff**	**sign**
c	0.0063	***	0.0013	
Δ(h(-1))	-0.0273		0.5955	***
Δ(h(-2))	-0.1253	*	0.3704	***
Δ(h(-3))	-0.0100		0.3332	***
Δ(h(-4))	-0.0061		0.2474	***
Δ(h(-5))	-0.0614		0.1550	**
Δ(h(-6))	-0.0375		0.0830	*
Δ(l(-1))	0.4229	***	-0.2431	***
Δ(l(-2))	-0.0748		-0.4959	***
Δ(l(-3))	0.0588		-0.2935	***
Δ(l(-4))	0.0434		-0.2068	***
Δ(l(-5))	0.0317		-0.1946	***
Δ(l(-6))	0.0395		-0.0988	*
EC(1)	-0.2668	***	-0.0577	
AdjR^2	0.2116		0.1465	
LB(20)	5.6029		12.3733	
p-value	0.9990		0.9030	

A.46	Δ(h)		Δ(l)	
	coeff	**sign**	**coeff**	**sign**
c	0.0068	*	0.0021	
Δ(h(-1))	0.0414		0.6070	***
Δ(h(-2))	-0.0655		0.3811	***
Δ(h(-3))	0.0377		0.3417	***
Δ(h(-4))	0.0277		0.2464	***
Δ(h(-5))	-0.0399		0.1576	**
Δ(h(-6))	-0.0281		0.0795	
Δ(l(-1))	0.3470	***	-0.2526	***
Δ(l(-2))	-0.1359	*	-0.5077	***
Δ(l(-3))	0.0094		-0.3011	***
Δ(l(-4))	0.0092		-0.2050	***
Δ(l(-5))	0.0095		-0.1991	***
Δ(l(-6))	0.0317		-0.0919	*
EC(1)	-0.3289	***	-0.0739	
Tue	0.0002		-0.0007	
Wed	0.0009		0.0006	
Thu	0.0009		-0.0001	
Fri	0.0002		-0.0005	
Vol(-1)	0.0000		0.0000	
Count(-1)	0.0000		0.0000	
Size(-1)	0.0000		-0.0001	
AdjR^2	0.2104		0.1451	
LB(20)	5.4636		12.0478	
p-value	0.9990		0.9140	

A.47	Δ(h)		Δ(l)	
	coeff	**sign**	**coeff**	**sign**
c	0.0042	***	-0.0005	
Δ(h(-1))	-0.7441	***	-0.0973	
Δ(h(-2))	-0.5372	***	-0.0283	
Δ(h(-3))	-0.3606	***	-0.0661	
Δ(h(-4))	-0.2617	***	-0.0591	
Δ(h(-5))	-0.2472	***	-0.0121	
Δ(h(-6))	-0.1103	***	0.0259	
Δ(l(-1))	-0.1451	**	-0.7871	***
Δ(l(-2))	-0.2639	***	-0.6608	***
Δ(l(-3))	-0.1234	*	-0.5160	***
Δ(l(-4))	-0.0941		-0.3833	***
Δ(l(-5))	-0.0745		-0.2833	***
Δ(l(-6))	0.0406		-0.0920	**
EC(1)	-0.1892	***	0.0200	
Δ(o)	0.7909	***	0.6558	***
Δ(o(-1))	1.1005	***	0.8642	***
Δ(o(-2))	0.6372	***	0.6485	***
Δ(o(-3))	0.4399	***	0.5545	***
Δ(o(-4))	0.3518	***	0.3754	***
Δ(o(-5))	0.2085	***	0.1681	***
Δ(c(-1))	-0.2976	*	-0.0914	
Δ(c(-2))	-0.2572	**	-0.0790	
ret(-1)	0.4838	**	0.4043	**
AdjR^2	0.5473		0.4947	
LB(20)	11.9264		19.6827	
p-value	0.9190		0.4780	

Tables A.45, A.46, A.47: Estimates of base VECM model, investigating the effect of additional variables.

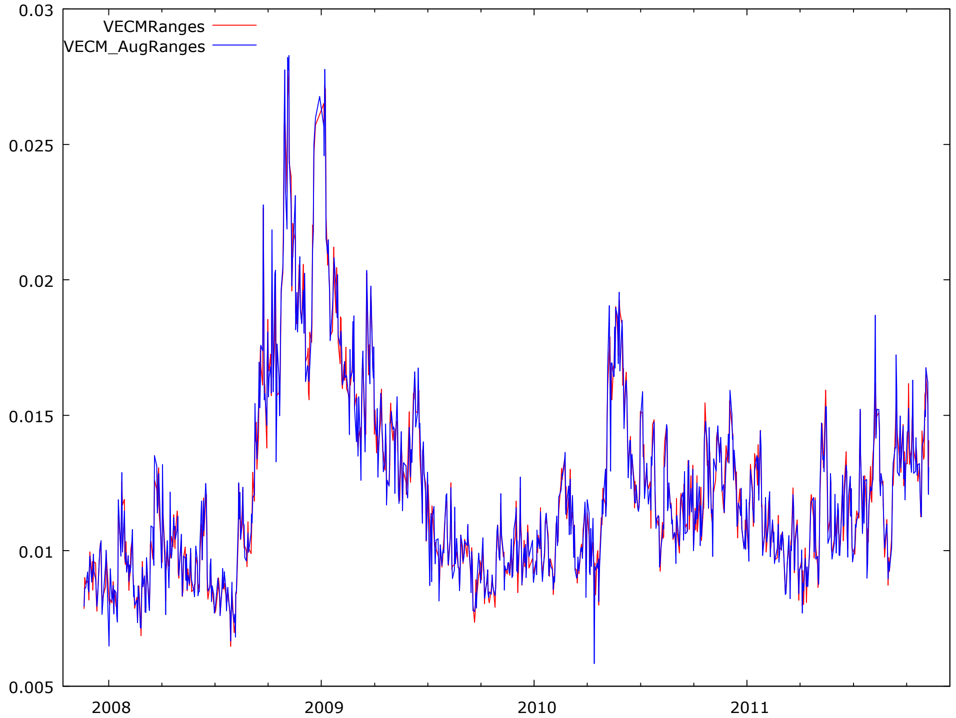


Table A.48: In-sample daily ranges predictions of both VECMs, namely models A.45 (red) and A.47 (blue), whole dataset.