

Benčík, Daniel

Working Paper

Range-based volatility estimation and forecasting

IES Working Paper, No. 34/2014

Provided in Cooperation with:

Charles University, Institute of Economic Studies (IES)

Suggested Citation: Benčík, Daniel (2014) : Range-based volatility estimation and forecasting, IES Working Paper, No. 34/2014, Charles University in Prague, Institute of Economic Studies (IES), Prague

This Version is available at:

<https://hdl.handle.net/10419/120419>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

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

Range-based Volatility Estimation and Forecasting

Daniel Benčík

IES Working Paper: 34/2014



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

[UK FSV – IES]

Opletalova 26
CZ-110 00, Prague
E-mail : ies@fsv.cuni.cz
<http://ies.fsv.cuni.cz>

Institut ekonomických studií
Fakulta sociálních věd
Univerzita Karlova v Praze

Opletalova 26
110 00 Praha 1

E-mail : ies@fsv.cuni.cz
<http://ies.fsv.cuni.cz>

Disclaimer: The IES Working Papers is an online paper series for works by the faculty and students of the Institute of Economic Studies, Faculty of Social Sciences, Charles University in Prague, Czech Republic. The papers are peer reviewed, but they are *not* edited or formatted by the editors. The views expressed in documents served by this site do not reflect the views of the IES or any other Charles University Department. They are the sole property of the respective authors. Additional info at: ies@fsv.cuni.cz

Copyright Notice: Although all documents published by the IES are provided without charge, they are licensed for personal, academic or educational use. All rights are reserved by the authors.

Citations: All references to documents served by this site must be appropriately cited.

Bibliographic information:

Benčík D. (2014). “Range-based Volatility Estimation and Forecasting” IES Working Paper 34/2014. IES FSV. Charles University.

This paper can be downloaded at: <http://ies.fsv.cuni.cz>

Range-based Volatility Estimation and Forecasting

Daniel Benčík^a

^aInstitute of Economic Studies, Faculty of Social Sciences, Charles University in Prague, Smetanovo nábřeží 6, 111 01 Prague 1, Czech Republic

December 2014

Abstract:

In this paper, we analyze new possibilities in predicting daily ranges, i.e. differences between daily high and low prices. We empirically assess efficiency gains in volatility estimation when using range-based estimators as opposed to simple daily ranges and explore the use of these more efficient volatility measures as predictors of daily ranges. The array of models used in this paper include the heterogeneous autoregressive model, conditional autoregressive ranges model and a vector error-correction model of daily highs and lows. Contrary to intuition, models based on co-integration of daily highs and lows fail to produce good quality out of sample forecasts of daily ranges. The best one-day-ahead daily ranges forecasts are produced by a realized range based HAR model with a GARCH volatility-of-volatility component.

Keywords: volatility, returns, futures contracts, cointegration, prediction

JEL: C52, C53, C58

1. 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 (e.g. the Nobel Prize awarded ARCH type models and their generalizations (Engle 1982, Bollerslev 1986)).

The work presented in this paper views volatility modeling from the standpoint of a day-trader. A crucial question for such an investor is how large an intraday move is to be expected once a position is open. For this purpose, a good quality prediction of the day's range (difference of highest and lowest daily price) is very useful, while predictions of other volatility measures (e.g. standard deviation of returns) provide only limited help.

Contributions of this paper to the existing body of volatility related literature are numerous. Firstly, we investigate the efficiency of several range-based volatility estimators proposed in the second half of the 20th century (Parkinson 1980, Garman & Klass 1980, Rogers & Satchell 1990). In theory, these range-based volatility measures claim significant efficiency gains compared to e.g. standard deviation of returns. We test whether efficiency gains reported in theory are observable in practice.

Next, we focus on predicting daily ranges. We use daily ranges themselves (Chou 2005) as well as range-based measures of volatility as predictors, driven by the idea that more precise measurement of past volatility should lead to superior forecasts. For this purpose, already published models on linkages between different volatility measures (Engle & Gallo 2003) are adapted specifically to daily ranges' prediction.

Lastly, we combine the results of our investigations described above with several existing models used for daily volatility modeling/forecasting. Additionally, we investigate possible improvements of daily ranges modeling by using lagged trading activity variables. The work finishes by picking the best model for daily ranges prediction based on out-of-sample forecasting performance.

The work is organized as follows. In Section 1, we provide motivation for daily ranges modeling. Section 2 describes our dataset. In Section 3, we compare various daily variance estimators in terms of efficiency and usefulness for daily ranges prediction. In Section 4, 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 5 and conclusion follows.

2. Motivation

For clarity of explanations, let $P_{t,D}$ denote the price of a financial asset measured at time $0 \leq t \leq T$ on day D and let us assume that log-price $p_{t,D} = \log_e(P_{t,D})$ evolves according to a diffusion process

$$dp_{t,D} = \mu_{t,D}dt + \sigma_D dW_{t,D} + c_D dJ_{t,D} \quad (1)$$

where $\mu_{t,D}, \sigma_D, c_D$ correspond to drift, volatility and jump terms and $W_{t,D}, J_{t,D}$ are Wiener and constant-intensity Poisson processes. The daily (log)range is then defined as

$$R_D = \sup_{0 \leq t \leq T} P_{t,D} - \inf_{0 \leq t \leq T} P_{t,D} \quad R_D^{\log} = \sup_{0 \leq t \leq T} p_{t,D} - \inf_{0 \leq t \leq T} p_{t,D} \quad (2),(3)$$

and the daily log-return, daily high, low, open and close prices are denoted

$$r_D = P_{T,D} - P_{T,D-1} \quad (4)$$

$$h_D = \sup_{0 \leq t \leq T} p_{t,D} \quad l_D = \inf_{0 \leq t \leq T} p_{t,D} \quad o_D = p_{f,D} \quad c_D = p_{T,D} \quad (5),(6)$$

where f is a portion of a trading day during which trading activity is minimal.

The most popular measures used for the inference of recent (one-day) variance, i.e. of the squared diffusion coefficient σ_D , are the squared daily return and the absolute daily return

$$r_D^S = r_D^2 \quad r_D^A = |r_D| \quad (7), (8)$$

Parkinson (1980) showed that assuming $\mu_{t,D} = 0$, σ_D^2 can be estimated by

$$(\hat{\sigma}_D^{Park})^2 = [4 \ln 2]^{-1} (R_D^{\log})^2 \quad (9)$$

along with a five times higher efficiency of variance estimation compared to (7). 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. Keeping the assumption of $\mu_{t,D} = 0$, Garman & Klass (1980) suggest an estimator

$$(\hat{\sigma}_D^{GK})^2 = 0.12 \frac{[o_D - c_{D-1}]}{f} + 0.78 \frac{0.5(h_D - l_D)^2 + (2 \ln(2) - 1)[c_D - o_D]}{1 - f} \quad (10)$$

and claim that the efficiency gain compared to (7) is approximately 7.4 regardless of f .

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

$$(\hat{\sigma}_D^{RS})^2 = (l_D - o_D)(l_D - c_D) + (h_D - o_D)(h_D - c_D) \quad (11)$$

which has only slightly lower efficiency compared to $\hat{\sigma}_D^{GK}$. 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 (9) - (11) do not require tick by tick data needed for the construction of finely spaced intraday returns.

When intraday data are available, volatility of the price generating process can be estimated using finely spaced intraday returns. Andersen et al (2001) introduce the concept of realized variance for the estimation of daily variance as a sum of squared intraday returns. Martens & van Dijk (2007) and Christensen & Podolskij (2007) in independent studies build upon the work Parkinson (1980) and apply it to the topic of measuring daily variance on intraday data. Specifically, they propose the replacement of squared intraday returns in realized variance by squared intraday ranges to create the so called realized range. According to empirical studies, realized range provides efficiency gains over realized variance. Compared to range-based estimators of volatility, both realized variance and realized range offer significant efficiency gains and are sometimes considered as measurements as opposed to estimations of volatility.

In spite of high efficiency of aforementioned range-based/realized 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, day-traders are more likely to benefit from a precise prediction of R_D^{\log} as the range is directly related to profit-target and stop-loss settings. Other volatility estimates, including the previously defined range-based ones and realized ones, cannot be used in such a manner.

To understand this, firstly notice that volatility measures built using close-close or open-close returns (i.e. absolute/squared returns, standard deviation of returns) provide help only to investors who close their positions at the end of day. Investors exiting their positions in any part of the day are much rather interested in whole day volatility, hence realized measures of volatility as well as range-based measures of volatility can be considered. Realized measures of volatility, however, do not discriminate between days where price moves without direction (sideways) and days where the price trends – two days with similar realized variance can have strikingly different daily ranges (high for a directional day, low for a sideways day). As intraday traders care about the extent or directional price move, realized measures of volatility offer limited help for this purpose. Range-based estimators, on the other hand, offer information directly related to the extent of directional price movement. The inclusion of open/close prices/returns in some of aforementioned range-based estimators makes the connection between a volatility reading and the extent of directional move hard to assess. The daily range is thus left as the most useful measure of volatility for an intraday investor as the relation of daily range to the extent of directional intraday movement is clear.

For this purpose, we focus solely on daily ranges prediction. Using the above defined range-based and realized estimators of volatility, we will investigate whether it is possible to benefit from higher precision of past volatility measurement for creating better daily range forecasts.

3. Initial data analysis

The data used in this paper relate to the EUR/USD Forex futures contract traded on CME from Nov, 9 2007 to Nov, 9 2011. In this time span, different delivery months were traded from which a continuous contract was created based on the maximum volume rule. Trade prices were used, hence the effect market microstructure was not removed. Timing

conventions used in this paper follow Electronic Trading Hours (ETH) used by CME. Lastly, so as not to contaminate our dataset with outlying observations, we omitted all days for which the total trading volume was below 80,000 contracts (mostly U.S. national holidays).

Basic statistical analysis of data at hand coupled with persistence of volatility is summarized in Table 1.

| | Mean | St. Dev | Skew | Kurt | GHE |
|-----------------|-----------|-----------|--------|--------|-------|
| R_{1D}^{\log} | 0.0121566 | 0.0055872 | 1.6842 | 4.1943 | 0.990 |

Table 1: Descriptive statistics of daily ranges sampled at daily frequencies.

Positive skew coupled with high kurtosis are typical for volatility readings known for their heavy-tailed behavior. The Hurst Exponent measured using Generalized Exponent method is high indicating very strong persistence. Turning to distributional properties of daily ranges, Locke's nonparametric test rejects the null hypothesis of gamma distributed daily ranges at 99.9% confidence level (a frequency distribution of daily ranges, coupled with the best fitting gamma distribution p.d.f. is shown in Figure 2).

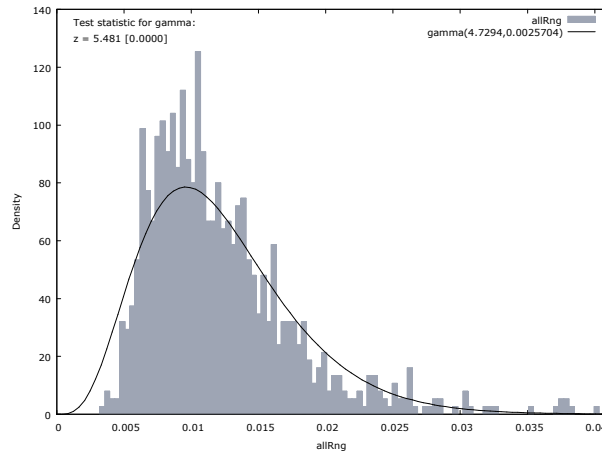


Figure 2: Distribution of daily ranges including the best gamma distribution fit

Alizadeh, Brandt & Diebold (2001) argue that logs of ranges are approximately normally distributed. Even though visually daily log-ranges correspond to a normal distribution (Figure 3), the Jarque-Bera test of normality rejects the null hypothesis of daily ranges' normal distribution

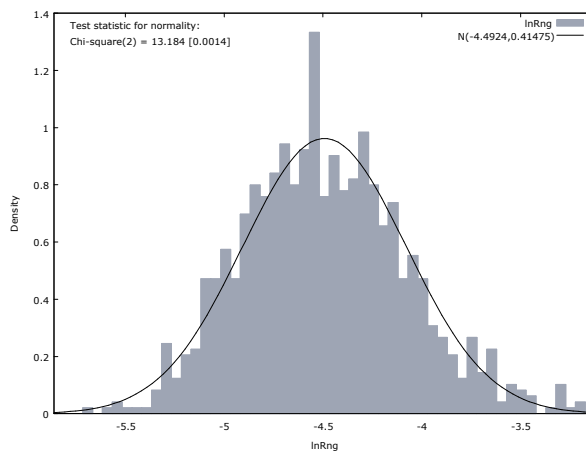


Figure 3: Distribution of daily log-ranges including the best normal distribution fit

Hence, the distributional properties of daily ranges on our dataset remain unknown.

When testing for stationarity of our main time series, we find the following. Both ADF and ADF-GLS testing rejects the null hypothesis of unit-root in daily ranges at 99.9% confidence level. On the other hand KPSS test reject the null hypothesis of stationarity. We conclude that daily ranges are borderline-stationary, which is a common finding in volatility related literature.

4. Comparing volatility measures

As previously mentioned, range-based volatility estimators (9) - (11) differ in their efficiency. In this section, we investigate whether this feature holds on real-world data or whether the differing efficiency of range-based volatility estimators is confined to simulated processes. 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.

Our reasons for such a comparison are several. Firstly, practitioners working in illiquid markets cannot reliably estimate volatility using realized measures of volatility (infrequent trading poses an obstacle). However, as OHLC data are readily available even for illiquid instruments, range-based estimators can be used for volatility assessment. Thus, we want to generally infer how precisely volatility can be estimated when data is restricted to OHLC readings. Secondly, moving to liquid markets where realized volatility measures can be calculated, we are interested in the forecasting accuracy gains stemming from using more precise lagged volatility measurements as predictors of daily ranges. Instead of estimating each daily range model in numerous variants (each with different measure of volatility as predictor of daily range), we firstly assess the empirical existence of efficiency gains of various volatility estimators. In case range-based measures of volatility are found to be of equal efficiency, regressing daily ranges on Garmann & Klass and Rogers & Satchell measures does not yield anything. Similarly, based on superior empirical efficiency, we can decide whether to use realized variance or realized range in the forecasting part of the paper.

A time plot mutually comparing range-based measures of volatility is presented in Figure 3, a comparison of Parkinson's measure with realized measures of volatility is presented in Figure 4¹. When comparing range-based measures of volatility (Figure 3), we observe some differences. Most notably, Parkinson's and Rogers & Satchell's estimators sometimes report a high volatility reading not reported by the remaining two volatility estimators. When comparing realized measures of volatility to the Parkinson's measure (Figure 4), we observe a significant difference of Parkinson's measure from the realizes ones. Owing to similarity of realized range/realized variance construction, we observe a near identical development of these measures.

¹ In our work, we chose to sample returns at 5-minute intervals for the construction of realized variance and realized range, as this sampling frequency was firstly proposed by Andersen & Bollerslev (1998). Also this sampling frequency was originally used in Andersen et al (2001)

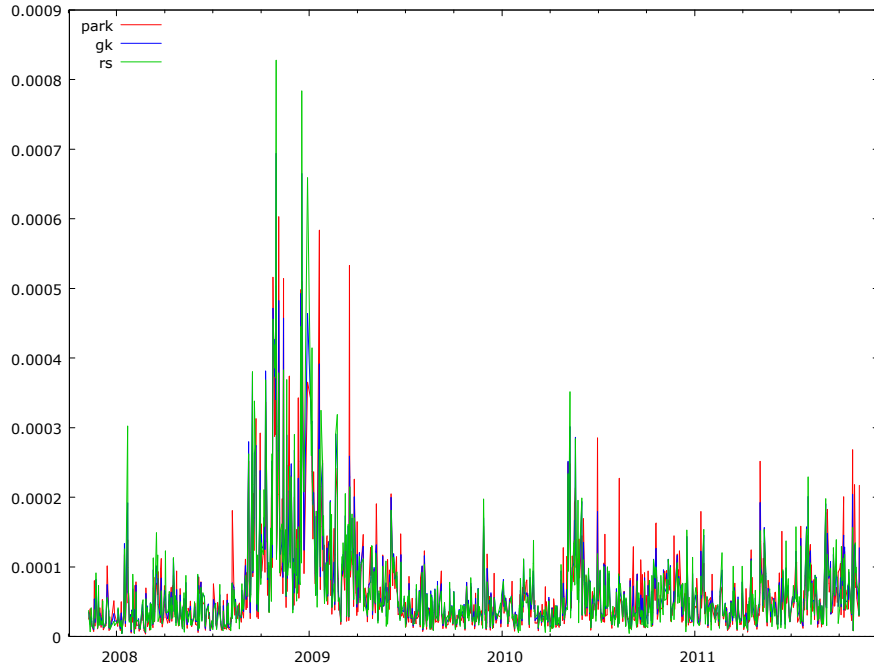


Figure 3: Time plots of Parkinson's, Rogers & Satchell's and Garmann & Klass' measure of variance

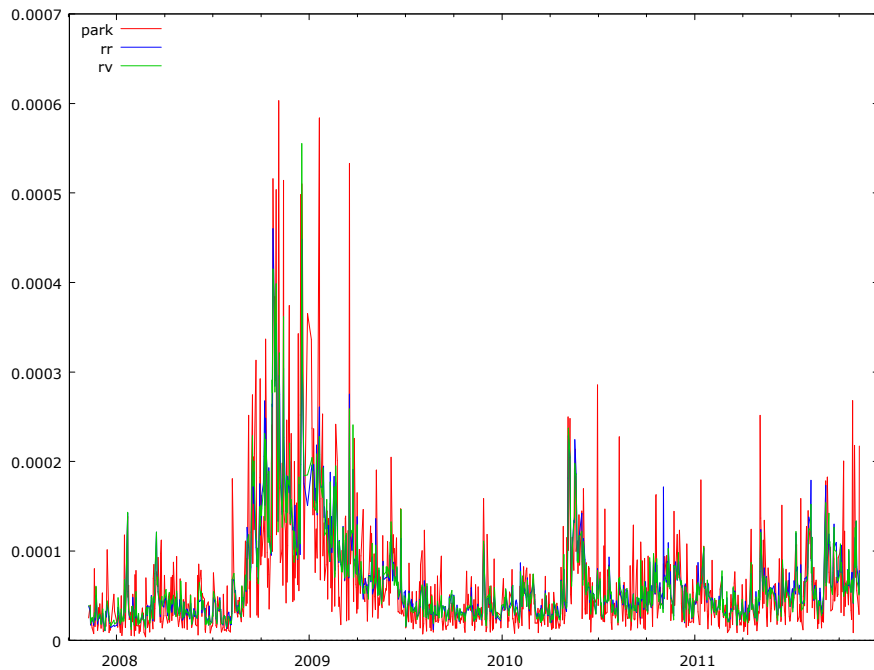


Figure 4: Time plots of Parkinson's measure of variance, realized variance and realized range

Next, we firstly investigate the general relationship between daily ranges and realized variance/realized ranges. We are interested in assessing the inefficiency of daily ranges over realized estimators. Secondly, we empirically assess the efficiencies of different range-based estimators presented in Section 1. Theoretically, range-based estimators herein considered differ significantly in efficiency. We assess whether different efficiencies are observable on empirical data.

For comparative quality assessment of daily ranges vis-à-vis realized ranges/realized variance, i.e. for information on the level of noise contained in daily ranges, we use Mincer-Zarnowitz regressions². Specifically, we estimate

$$RR_D = \alpha + \beta(\hat{\sigma}_D^{Park})^2 + \varepsilon_E \quad RV_D = \alpha + \beta(\hat{\sigma}_D^{Park})^2 + \varepsilon_E \quad (12)$$

with estimation results of (12) shown in Table 5. If squared daily ranges and realized ranges/variance were comparable in terms of estimation 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. Intuitively, this can be attributed to the fact that daily ranges neglect a large part of intraday information.

| LHS Variable | α | β | AdjR ² |
|-----------------|----------|---------|-------------------|
| RR _D | 0.000028 | 0.5681 | 0.58 |
| RV _D | 0.000026 | 0.5888 | 0.60 |

Table 5: OLS results of $VolMeasure = \alpha + \beta \times (R_D^{log})^2 / 4 \ln(2) + \varepsilon$.

To assess the efficiency of Parkinson's, Garman & Klass' and Rogers & Satchell's range-based volatility measures, equations resembling (12) were estimated with HAC standard errors ($f = 1/24$ was used to represent the non-trading session in G&K measure of volatility). Specifically, we estimated

$$\begin{aligned} (\hat{\sigma}_D^{Park})^2 &= \alpha + \beta RV_D + \varepsilon_D & (\hat{\sigma}_D^{Park})^2 &= \alpha + \beta RR_D + \varepsilon_D \\ (\hat{\sigma}_D^{GK})^2 &= \alpha + \beta RV_D + \varepsilon_D & (\hat{\sigma}_D^{GK})^2 &= \alpha + \beta RR_D + \varepsilon_D \\ (\hat{\sigma}_D^{RS})^2 &= \alpha + \beta RV_D + \varepsilon_D & (\hat{\sigma}_D^{RS})^2 &= \alpha + \beta RR_D + \varepsilon_D \end{aligned} \quad (13)$$

$$\begin{aligned} RV_D &= \alpha + \beta(\hat{\sigma}_D^{Park})^2 + \varepsilon_D & RR_D &= \alpha + \beta(\hat{\sigma}_D^{Park})^2 + \varepsilon_D \\ RV_D &= \alpha + \beta(\hat{\sigma}_D^{GK})^2 + \varepsilon_D & RR_D &= \alpha + \beta(\hat{\sigma}_D^{GK})^2 + \varepsilon_D \\ RV_D &= \alpha + \beta(\hat{\sigma}_D^{RS})^2 + \varepsilon_D & RR_D &= \alpha + \beta(\hat{\sigma}_D^{RS})^2 + \varepsilon_D \end{aligned} \quad (14)$$

with results presented in Table 6 and Table 7. As all five variables are measures of the same variance and as 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 Table 6 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.

² Mincer, J. A., Zarnowitz, V. (1969) The Evaluation of Economic Forecasts: Analysis of Forecasting Behavior and Performance, NBER Books, National Bureau of Economic Research

| Regressor | Dependent variable | | | | | |
|-------------------|------------------------|------------------------|--------------------------|--------------------------|--------------------------|--------------------------|
| | $(R^{\log})^2/4\ln(2)$ | $(R^{\log})^2/4\ln(2)$ | $(\sigma^{\text{GK}})^2$ | $(\sigma^{\text{GK}})^2$ | $(\sigma^{\text{RS}})^2$ | $(\sigma^{\text{RS}})^2$ |
| c | -0.000009 | -0.000003 | -0.000036 | -0.000035 | -0.000041 | -0.000045 |
| | 0.000027 | 0.000020 | 0.000024 | 0.000021 | 0.000037 | 0.000037 |
| 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 |

Table 6: Regressing range-based measures on realized measures of variance. Bold estimates are significant on 99% level.

| Regressor | Dependent variable | | | | | |
|--------------------------|--------------------|-----------------|---------------|-----------------|-----------------|---------------|
| | RR _D | | | RV _D | | |
| C | 0.000277 | 0.000313 | 0.000281 | 0.000258 | 0.000292 | 0.000256 |
| | 0.000024 | 0.000032 | 0.000027 | 0.000024 | 0.000032 | 0.000032 |
| $(R^{\log})^2/4\ln(2)$ | 0.5680 | | | 0.5888 | | |
| | 0.0368 | | | 0.0391 | | |
| $(\sigma^{\text{GK}})^2$ | | 0.5547 | | | 0.5791 | |
| | | 0.0551 | | | 0.0588 | |
| $(\sigma^{\text{RS}})^2$ | | | 0.5634 | | | 0.5791 |
| | | | 0.0387 | | | 0.0588 |
| AdjR ² | 0.5765 | 0.5453 | 0.5991 | 0.5981 | 0.5740 | 0.5740 |

Table 7: Regressing realized measures of variance on range-based measures. Bold estimates are significant on 99% level.

On the other hand, β estimates in Table 7 indicate a decomposition of range-based volatility measures into information on variance and noise. These β estimates as well as AdjR² 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. While in theory, Garman & Klass as well as Rogers & Satchell estimators promise efficiency increases over Parkinson's measure, our results do not confirm this. We thus do not have any reason to use the Garman & Klass or Rogers & Satchell measure as predictors in daily ranges modeling instead of daily ranges themselves.

5. Modeling daily ranges

Unlike many variables in economics with usually one predominant approach to modeling, volatility can be modeled using various approaches. Focusing on models that do not assume long-memory, simple AR (ARCH type, see Engle 1982) or more refined GARCH models can be used for volatility modeling. Recent developments in volatility modeling applicable to daily ranges are, for example, mixtures of long, medium and short-term volatilities (HAR of Corsi 2004), vector error-correction models (Cheung et al 2007) or models incorporating data sampled at different frequencies (Ghysels 2003).

In this paper, we specifically focus on models that do not incorporate long memory (e.g. ARFIMA, FIVECM). This choice is motivated by several drawbacks of long-memory models (estimation issues, dubious interpretation) coupled with questionable improvement of

fit/forecasting accuracy. For example, Corsi (2004) compares the fit/forecasting performance of a pseudo long-memory HAR model and a long-memory ARFIMA model applied to realized volatility. He finds that both in-sample and out-of-sample ARFIMA is dominated by HAR. Moreover, with ARFIMA there is no possibility to forecast daily ranges using more precise measures of volatility as predictors, which is one of the main ideas presented in this paper. Except for the last argument, similar logics apply to the choice of VECM over fractionally integrated VECM, strengthened by the facts that VECM is merely a restricted FIVECM.

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

5.1 ARMA-GARCH

In order to have a benchmark in daily ranges modeling, we chose an ARMA-GARCH model (see e.g. Pong et al (2003), Ahoniemi (2009) for ARMA modeled volatility). For brevity purposes, we limit ourselves to stating that an ARMA(7,0)-GARCH(1,1) model with Student-distributed residuals was the model offering best in-sample fit.

5.2 Heterogeneous autoregressive model

Dacorogna et al (1997) propose, as an extension of the GARCH model, to combine volatility views of market participants with differing investment horizons. Corsi (2003) follows up to this idea by proposing his own model and argues that short-term, medium-term and long-term volatilities should be used for modeling. Specifically, Corsi proposes a model of the following specification

$$RV_D = \alpha_0 + \alpha_1 RV_{D-1} + \alpha_2 RV_{D-1}^{(5)} + \alpha_3 RV_{D-1}^{(22)} + \varepsilon_t \quad (15)$$

where $RV_D^{(a)}$ is a simple average of realized variances over days $(D-a; D]$. Corsi & Reno (2009) assess the leverage effect by estimating

$$RV_D = \alpha_0 + \alpha_1 RV_{D-1} + \alpha_2 RV_{D-1}^{(5)} + \alpha_3 RV_{D-1}^{(22)} + \beta_1 I_{D-1}^{(1)} r_{D-1}^{(1)} + \beta_2 I_{D-1}^{(5)} r_{D-1}^{(5)} + \beta_3 I_{D-1}^{(22)} r_{D-1}^{(22)} + \varepsilon_D \quad (16)$$

where $I_D^{(a)}$ is an indicator variable equal to one in case the average daily return $r_D^{(a)}$ measured over days $(D-a; D]$ is negative. According to Corsi & Reno this new (LHAR - Leveraged HAR) model performs better than the original specification (15).

Here, we estimate a battery of HAR models with daily ranges as the explained variable. All models discussed here were estimated using HAC standard errors. Lastly, as realized variance and realized ranges provide nearly identical estimates of variance, we chose to work with realized ranges only from here on.

In light of borderline stationarity of daily ranges on our dataset, consistency of OLS used in HAR estimation might be questionable. In order to satisfy OLS assumptions, the cointegration of daily ranges and their weekly/monthly averages was investigated. Using the Engle-Granger test, daily ranges were found to be cointegrated with their weekly and monthly averages, which makes OLS estimates in HAR super-consistent³.

The base HAR specification for daily ranges modeling is

³ The same exercise was performed for averages of realized ranges in R-HAR modeling with the same results. For brevity reasons, we do not report details of statistical testing, but we kindly provide them upon request.

$$R_D^{\log} = \alpha_0 + \alpha_1 R_{D-1}^{\log} + \alpha_2 R_{D-1}^{\log,(5)} + \alpha_3 R_{D-1}^{\log,(22)} + \varepsilon_D \quad (17)$$

Including the leverage effect leads to the LHAR specification given by

$$R_D^{\log} = \alpha_0 + \alpha_1 R_{D-1}^{\log} + \alpha_2 R_{D-1}^{\log,(5)} + \alpha_3 R_{D-1}^{\log,(22)} + \beta_1 I_{D-1}^{(1)} r_{D-1} + \beta_2 I_{D-1}^{(5)} r_{D-1}^{(5)} + \beta_3 I_{D-1}^{(22)} r_{D-1}^{(22)} + \varepsilon_D \quad (18)$$

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.

$$R_D^{\log} = \alpha_0 + \alpha_1 RR_{D-1} + \alpha_2 RR_{D-1}^{(5)} + \alpha_3 RR_{D-1}^{(22)} + \varepsilon_D \quad (19)$$

To investigate the added benefit of using variables representing trading activity (average trade size/transaction count⁴) and their possible long-term influence on volatility, we define models with -S/-C suffixes as

$$R_D^{\log} = \alpha_0 + \alpha_1 R_{D-1}^{\log} + \alpha_2 R_{D-1}^{\log,(5)} + \alpha_3 R_{D-1}^{\log,(22)} + \gamma_1 TS_{D-1} + \gamma_2 TS_{D-1}^{(5)} + \gamma_3 TS_{D-1}^{(22)} + \varepsilon_D \quad (20)$$

$$R_D^{\log} = \alpha_0 + \alpha_1 R_{D-1}^{\log} + \alpha_2 R_{D-1}^{\log,(5)} + \alpha_3 R_{D-1}^{\log,(22)} + \delta_1 TC_{D-1} + \delta_2 TC_{D-1}^{(5)} + \delta_3 TC_{D-1}^{(22)} + \varepsilon_D \quad (21)$$

For example, R-LHAR-SC model is a HAR model using realized ranges as regressors. On top of that, the specification is enriched by information on leverage effect, average trade size and transaction count as described in (18), (19), (20) and (21).

Firstly, we focus on HAR and R-HAR specifications. The estimated results are presented in Table 8 and indicate that neither daily ranges nor realized ranges are strictly superior in terms of goodness-of-fit when modeling daily ranges. For this reason, all further models are evaluated with either daily or realized ranges being the RHS measure of (average) volatility.

| | HAR | R-HAR |
|------------------------------|---------------|---------------|
| c | 0.0000 | 0.0000 |
| R⁽¹⁾(-1) | 0.0482 | |
| R⁽⁵⁾(-1) | 0.4356 | |
| R⁽²²⁾(-1) | 0.3884 | |
| RR⁽¹⁾(-1) | | 0.7952 |
| RR⁽⁵⁾(-1) | | 1.3463 |
| RR⁽²²⁾(-1) | | 0.4086 |
| AdjR² | 0.3346 | 0.3684 |

Table 8: Investigation of relationships between daily ranges and realized ranges from HAR perspective.

⁴ Logarithms (not levels) of average trade size and transaction count were used in all regressions contained in this thesis.

Parameter estimates significant on 95% critical level are printed in bold.

In-sample estimation results of all specifications are shown in Table A.1 in the Appendix. In these results, we do not observe any significant increase in AdjR^2 for either model. Contrary to Corsi, the "bad-news" effect is not 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.

5.3 CARR

For modeling daily ranges, Chou (2005) proposes an adoption of a MEM model originally developed in Engle & Russell (1998). Chou's conditional autoregressive range (CARR) model of order p, q is specified by

$$R_D^{\log} = \lambda_D \varepsilon_D \quad \lambda_D = \omega + \sum_{i=1}^q \alpha_i R_{D-i}^{\log} + \sum_{i=1}^p \beta_i \lambda_{D-i} \quad (22)$$

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 or Exponentially distributed ε_D . Investigations carried out on the S&P 500 futures contract reveal superior volatility forecasts of CARR models as compared to GARCH models, presence of a strong leverage 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. The assumption of Weibull distributed ε_D seems plausible, as estimated residuals $\hat{\varepsilon}_D = R_D^{\log} / \hat{\lambda}_D$ are near-Weibull distributed.

Estimations on our dataset were carried out using robust standard error estimation techniques (QML covariance matrix). The results for daily ranges prediction are presented in Table 9 for $\varepsilon_D \approx \text{Exp}(\bullet)$ and Table 10 for $\varepsilon_D \approx \text{Weibull}(\theta, \bullet)$ respectively. Optimal values of lags $p = q = 1$ were determined based on parameter significance and we find strong evidence for $\varepsilon_D \approx \text{Weibull}(\theta, \bullet)$ ⁵.

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

Table 9: Estimation results of a CARR model with Exponentially distributed residuals.

⁵ The estimated θ is significantly different from 1.0 in which case Weibull distribution collapses into an Exponential.

| | 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_i | 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 10: Estimation results of a CARR model with Weibull distributed residuals.

Before delving into augmentations of the CARR(1,1) model, we compared the in-sample performance of CARR(1,1) model with $\varepsilon_D \approx Weibull(\theta, \bullet)$ to the performance of a base HAR specification (17). When neglecting the buildup period of MLE estimation, we observed a striking similarity of fitted values indicating a near identity of both models. For this reason, 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.

5.4 Cointegration of high and low prices

Cheung (2007) investigates the usefulness of vector error-correction-models for daily ranges modeling on several stock indices. Cheung's intuition is that daily highs/lows are integrated of order one while the daily range is stationary. Cheung's tests confirm these hypotheses and a VECM model is adopted for $\Delta h_D, \Delta l_D$:

$$X_D = (\Delta h_D, \Delta l_D)^T \quad X_D = \alpha + \sum_{i=1}^p \beta_i X_{D-i} + \gamma \mathcal{R}_{D-1}^{\log} + \xi D + \varepsilon_D \quad (23)$$

where the error correcting term (daily range) turns out significant. Generally the predictive power of the model ranges from 8% to 17% in terms of AdjR^2 depending on instrument.

In order to improve the model's predictive power, Cheung includes several exogenous variables. Cheung adds changes of daily open and daily close prices as well as daily returns into (23). 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 this section, we formally investigate the assumptions of a VECM model, assess the predictive power and proper specification of (23) on our dataset, including Cheung's augmentations. 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 upon the results of which we conclude that daily highs and lows are co-integrated.

Estimation results of Chou's basic model presented in Table 11 have signs of significant parameters in both equations according to expectations and speak in favor of a mean-reverting process in daily ranges. 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. The model specification thus changes to

$$X_D = \alpha + \sum_{i=1}^p \beta_i X_{D-i} + \sum_{j=0}^q \gamma_j \Delta o_{D-j} + \sum_{k=1}^r \delta_k \Delta c_{D-k} + \sum_{m=1}^s \phi_m co_{D-m} + \varphi EC_{D-1} + \varepsilon_D \quad (24)$$

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 12 for $q=4, r=2, s=1$. Several points deserve mentioning:

- In both daily high/daily low equations, significance of parameters changes strongly and AdjR^2 rises noticeably.
- 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.

| | $\Delta(\mathbf{h})$ | | $\Delta(\mathbf{l})$ | | $\Delta(\mathbf{h})$ | | $\Delta(\mathbf{l})$ | | |
|--------------------------|----------------------|-----|----------------------|-----|--------------------------|---------|----------------------|---------|-----|
| | coeff | | coeff | | coeff | | coeff | | |
| c | 0.0063 | *** | 0.0013 | | c | 0.0042 | *** | -0.0005 | |
| $\Delta(\mathbf{h}(-1))$ | -0.0273 | | 0.5955 | *** | $\Delta(\mathbf{h}(-1))$ | -0.7441 | *** | -0.0973 | |
| $\Delta(\mathbf{h}(-2))$ | -0.1253 | * | 0.3704 | *** | $\Delta(\mathbf{h}(-2))$ | -0.5372 | *** | -0.0283 | |
| $\Delta(\mathbf{h}(-3))$ | -0.0100 | | 0.3332 | *** | $\Delta(\mathbf{h}(-3))$ | -0.3606 | *** | -0.0661 | |
| $\Delta(\mathbf{h}(-4))$ | -0.0061 | | 0.2474 | *** | $\Delta(\mathbf{h}(-4))$ | -0.2617 | *** | -0.0591 | |
| $\Delta(\mathbf{h}(-5))$ | -0.0614 | | 0.1550 | ** | $\Delta(\mathbf{h}(-5))$ | -0.2472 | *** | -0.0121 | |
| $\Delta(\mathbf{h}(-6))$ | -0.0375 | | 0.0830 | * | $\Delta(\mathbf{h}(-6))$ | -0.1103 | *** | 0.0259 | |
| $\Delta(\mathbf{l}(-1))$ | 0.4229 | *** | -0.2431 | *** | $\Delta(\mathbf{l}(-1))$ | -0.1451 | ** | -0.7871 | *** |
| $\Delta(\mathbf{l}(-2))$ | -0.0748 | | -0.4959 | *** | $\Delta(\mathbf{l}(-2))$ | -0.2639 | *** | -0.6608 | *** |
| $\Delta(\mathbf{l}(-3))$ | 0.0588 | | -0.2935 | *** | $\Delta(\mathbf{l}(-3))$ | -0.1234 | * | -0.5160 | *** |
| $\Delta(\mathbf{l}(-4))$ | 0.0434 | | -0.2068 | *** | $\Delta(\mathbf{l}(-4))$ | -0.0941 | | -0.3833 | *** |
| $\Delta(\mathbf{l}(-5))$ | 0.0317 | | -0.1946 | *** | $\Delta(\mathbf{l}(-5))$ | -0.0745 | | -0.2833 | *** |
| $\Delta(\mathbf{l}(-6))$ | 0.0395 | | -0.0988 | * | $\Delta(\mathbf{l}(-6))$ | 0.0406 | | -0.0920 | ** |
| EC(1) | -0.2668 | *** | -0.0577 | | EC(1) | -0.1892 | *** | 0.0200 | |
| | | | | | $\Delta(\mathbf{o})$ | 0.7909 | *** | 0.6558 | *** |
| | | | | | $\Delta(\mathbf{o}(-1))$ | 1.1005 | *** | 0.8642 | *** |
| | | | | | $\Delta(\mathbf{o}(-2))$ | 0.6372 | *** | 0.6485 | *** |
| | | | | | $\Delta(\mathbf{o}(-3))$ | 0.4399 | *** | 0.5545 | *** |
| | | | | | $\Delta(\mathbf{o}(-4))$ | 0.3518 | *** | 0.3754 | *** |
| | | | | | $\Delta(\mathbf{o}(-5))$ | 0.2085 | *** | 0.1681 | *** |
| | | | | | $\Delta(\mathbf{c}(-1))$ | -0.2976 | * | -0.0914 | |
| | | | | | $\Delta(\mathbf{c}(-2))$ | -0.2572 | ** | -0.0790 | |
| | | | | | ret(-1) | 0.4838 | ** | 0.4043 | ** |
| AdjR² | 0.2116 | | 0.1465 | | AdjR² | 0.5473 | | 0.4947 | |
| LB(20) | 5.6029 | | 12.3733 | | LB(20) | 11.9264 | | 19.6827 | |
| p-value | 0.9990 | | 0.9030 | | p-value | 0.9190 | | 0.4780 | |

Tables 11, 12: Estimates of base VECM model, investigating the effect of additional variables.

6. Out-of-sample Forecasting Exercise

In this section, we assess the out of sample forecasting performance of the models so far discussed. We focus solely on one-day-ahead forecasts of daily ranges using a rolling window for estimation. Window length was set to the minimum number of observations allowing for a MLE estimation of all models (400 observations, yielding 599 one-step-ahead forecasts). For assessing differences in forecast quality RMSE and Q-LIKE loss functions were used. 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

$$MSE = n^{-1} \sum_{i=1}^n e_i^2 \quad QLIKE = n^{-1} \sum_{i=1}^n (\ln h_i + \sigma_i / h_i) \quad (25)$$

In order to statistically compare the performance of different models a volatility forecast comparison test based on Diebold & Mariano (1995) and West (1996) was used.

Following literature on volatility forecasting, we decided to compare the forecasting power of models discussed in Sections 4.2 and 4.4 to several benchmark models - specifically the random-walk model, lagged average weekly, lagged average monthly volatility and a more sophisticated AR(7)-GARCH(1,1)-t discussed in Section 4.1. 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.

As the number of HAR models estimated in this paper is large, only base specifications and specifications with a clear increase in explanatory power stemming from additional variables were used for forecasting. The list of models chosen for out-of-sample forecasting is listed in Table 13 with an overview of forecasting performance as measured by RMSE and Q-LIKE following in Table 14.

| Model | Description |
|-----------|---|
| 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 |
| AR7 | AR(7)-GARCH(1,1)-t model derived forecasts |
| HARGARCH | Forecasts of a HAR model with leverage effect from the previous trading day, GARCH modeling of residual volatility included |
| RHARGARCH | Forecasts of a R-HAR model without any leverage effect, GARCH modeling of residual volatility included |
| VECM | Forecasts of VECM of Highs and Lows with 6 lags |
| VECMAUG | Forecasts of VECM of Highs and Lows with 6 lags enriched by information on changes of closing and opening prices |

Table 13: List of models considered for out-of-sample forecasting evaluation.

| Model | RMSE | Q-LIKE | Model | RMSE | Q-LIKE |
|-------|----------|-----------|-----------|----------|-----------|
| RW | 0.000027 | -3.452526 | HARGARCH | 0.000015 | -3.500247 |
| SMA5 | 0.000017 | -3.496000 | RHARGARCH | 0.000015 | -3.502932 |
| SMA22 | 0.000016 | -3.499393 | VECM | 0.000016 | -3.498304 |
| AR7 | 0.000015 | -3.500018 | VECMAUG | 0.000016 | -3.497844 |

Table 14: Average RMSE and Q-LIKE of one-step-ahead rolling-window forecasts.

The most striking observation is that despite a marked increase of in-sample AdjR^2 (in modeling changes of high and low) the augmented VECM model does not generate smaller errors (in modeling daily ranges) compared to the base VECM model.

To assess differences in forecasting accuracy statistically, we performed Diebold-Mariano-West test for both RMSE and Q-LIKE loss functions. The resulting matrix is listed below as Table 15. A negative test statistic in row *A* and column *B* indicates that model *B* provides better forecasts than model *A*.

| | SMA5 | SMA22 | AR7 | HARGARCH | RHARGARCH | VECM | VECMAUG |
|-------------|-------|-------|-------|----------|-----------|-------|---------|
| RW | -6.41 | -6.02 | -7.21 | -6.79 | -7.31 | -6.82 | -6.51 |
| | -6.92 | -6.79 | -7.68 | -7.31 | -7.76 | -7.43 | -7.15 |
| SMA5 | | -1.07 | -3.49 | -2.99 | -4.49 | -1.47 | -1.00 |
| | | -1.55 | -3.17 | -3.07 | -4.29 | -1.36 | -0.95 |
| SMA22 | | | -1.05 | -1.41 | -2.68 | -0.06 | 0.20 |
| | | | -0.32 | -0.65 | -2.33 | 0.47 | 0.65 |
| AR7 | | | 0.05 | -0.11 | -2.14 | 0.71 | 0.88 |
| | | | | 0.12 | -2.17 | 1.42 | 1.66 |
| HAR-GARCH | | | | -0.20 | -2.26 | 1.33 | 1.43 |
| | | | | | -3.29 | 0.97 | 1.27 |
| R-HAR-GARCH | | | | | -3.23 | 1.12 | 1.32 |
| | | | | | | 2.37 | 2.64 |
| VECM | | | | | | 2.54 | 2.66 |
| | | | | | | | 0.81 |
| | | | | | | | 0.52 |

Table 15: 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. All models provide better than naive forecasts. The clearly best model is RHARGARCH. From this, we can conclude that for proper out of sample forecasting of daily ranges, using high quality volatility measures is critical. Turning to VECM models we observe a disappointing bad quality of forecasts. Both VECM models can beat only beat the random-walk model, 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 4. 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 forecast⁶.

⁶ 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.

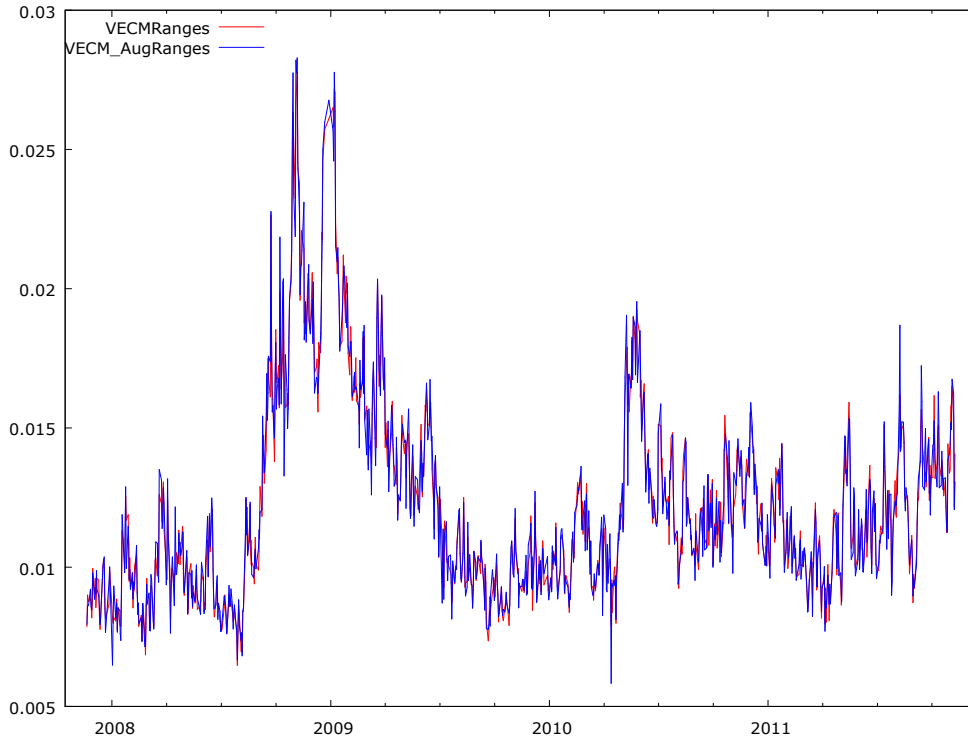


Figure 4: In-sample daily ranges predictions of VECM and VECM_AUG models

To conclude, models rooted in the co-integration of daily high and low prices dominate only the random-walk model. 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. The best model is an R-HAR model with a GARCH(1,1) volatility-of-volatility component.

Summary

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

Regressing daily ranges on range-based volatility estimates (Garman & Klass, Rogers & Satchell) is not expected to yield benefit, as we 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 noise, while only 60% of information is related to the variance of the price generating process. Using 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.

Comparison of models based on out-of-sample forecasting performance reveals several points. A model utilizing realized ranges for the prediction of daily ranges is the model of choice, as it can statistically beat all models considered in this paper. Models based on the co-integration show good in-sample fits pertaining 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 feasible follow-up to our work would be to investigate the possibility to update current day's volatility using intraday data. Using such an approach, traders active in the later parts of the trading day might obtain higher forecasting accuracy compared to traders active early in 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 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.

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., Krištofuk, 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

| | HAR | R-HAR | HAR-S | HAR-C | HAR-SC | LHAR | LHAR-S | LHAR-C | LHAR-SC | R-LHAR | R-LHAR-S | R-LHAR-C | R-LHAR-SC |
|---------------------------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|--------------|
| c | 0.0014 ** | 0.0010 | 0.0037 | 0.0035 | 0.0066 | 0.0017 ** | 0.0061 ** | 0.0061 | 0.0131 * | 0.0011 | 0.0012 | 0.0107 ** | 0.0122 * |
| R^{log,(1)}(-1) | 0.0466 | | 0.0623 | 0.0188 | 0.0325 | 0.0745 | 0.0892 * | 0.0485 | 0.0616 | | | | |
| R^{log,(5)}(-1) | 0.3958 *** | | 0.3201 *** | 0.3783 *** | 0.2911 ** | 0.3335 *** | 0.2668 ** | 0.3293 ** | 0.2568 * | | | | |
| R^{log,(22)}(-1) | 0.4444 *** | | 0.4495 *** | 0.4959 *** | 0.5103 *** | 0.4410 *** | 0.3970 *** | 0.4803 *** | 0.4313 *** | | | | |
| I⁽¹⁾(-1) | | | | | | 0.0876 ** | 0.0878 ** | 0.0855 ** | 0.0853 ** | 0.0530 | 0.0546 | 0.0526 | 0.0533 |
| I⁽⁵⁾(-1) | | | | | | -0.1362 | -0.1476 * | -0.1195 | -0.1291 | -0.0973 | -0.1063 | -0.1068 | -0.1087 |
| I⁽²²⁾(-1) | | | | | | -0.4054 * | -0.4092 * | -0.4293 | -0.4454 | -0.0584 | -0.0509 | -0.0383 | -0.0601 |
| RR⁽¹⁾(-1) | | 0.4807 *** | | | | | | | | 0.4735 *** | 0.4972 *** | 0.4920 ** | 0.5174 ** |
| RR⁽⁵⁾(-1) | | 0.6917 *** | | | | | | | | 0.6587 *** | 0.5108 ** | 0.9664 *** | 0.8166 * |
| RR⁽²²⁾(-1) | | 0.2972 * | | | | | | | | 0.3272 ** | 0.4368 ** | 0.0308 | 0.1066 |
| size⁽¹⁾(-1) | | | -0.0023 | | -0.0025 | | -0.0019 | | -0.0022 | | 0.0007 | | 0.0005 |
| size⁽⁵⁾(-1) | | | -0.0061 | | -0.0065 | | -0.0066 | | -0.0068 | | -0.0065 | | -0.0039 |
| size⁽²²⁾(-1) | | | 0.0071 | | 0.0076 | | 0.0060 * | | 0.0061 | | 0.0058 * | | 0.0026 |
| count⁽¹⁾(-1) | | | | 0.0010 | 0.0011 | | | 0.0009 | 0.0010 | | | -0.0001 | -0.0002 |
| count⁽⁵⁾(-1) | | | | 0.0001 | 0.0002 | | | -0.0002 | -0.0001 | | | -0.0029 | -0.0023 |
| count⁽²²⁾(-1) | | | | -0.0013 | -0.0016 | | | -0.0012 | -0.0014 | | | 0.0022 | 0.0016 |
| AdjR² | 0.3560 | 0.3910 | 0.3581 | 0.3556 | 0.3586 | 0.3648 | 0.3676 | 0.3646 | 0.3680 | 0.3913 | 0.3912 | 0.3939 | 0.3925 |

Table A.1: Estimation results of different HAR model specifications for daily ranges (volatility).

IES Working Paper Series

2014

1. Diana Zigraiova, Petr Jakubik: *Systemic Event Prediction by Early Warning System*
2. Pavla Blahova, Karel Janda, Ladislav Kristoufek: *The Perspectives for Genetically Modified Cellulosic Ethanol in the Czech Republic*
3. Oxana Babecká Kucharčuková: *The impact of exchange rate volatility on trade: Evidence for the Czech Republic*
4. Oliver Polyák: *The Impact of Euro Adoption on Export Performance: Comparison of the Czech Republic and Slovakia*
5. Lucie Davidová, Vladimír Benáček: *Determinants of Austrian International Trade: Analysis Based on the Gravity Model*
6. Julie Chytilová, Václav Korbel: *Individual and Group Cheating Behavior: A Field Experiment with Adolescents*
7. Petr Janský: *Consumer Demand System Estimation and Value Added Tax Reforms in the Czech Republic*
8. Pavel Hait, Petr Janský: *Inflation Differentials among Czech Households*
9. Hana Džmuráňová, Petr Teplý: *Risk management of savings accounts*
10. Tomáš Fiala, Tomáš Havránek: *Ailing Mothers, Healthy Daughters? Contagion in the Central European Banking Sector*
11. Petra Andrlíková: *Is Barrier version of Merton model more realistic? Evidence from Europe*
12. Martin Pospíšil, Jiří Schwarz: *Bankruptcy, Investment, and Financial Constraints: Evidence from a Post-Transition Economy*
13. Matt Hunt, Goran Serdarević, Tom Ovington, Clive Kenny: *Evidence for a ladder of investment in Central and Eastern European countries*
14. Petra Andrlíková: *Bayesian default probability models*
15. Petra Buzková: *Has the Relationship Between Market and Model CDS Price Changed during the EMU Debt Crisis?*
16. Jakub Černý, Jiří Witzany: *Interest Rate Swap Credit Valuation Adjustment*
17. Martin Gregor: *Receiver's access fee for a single sender*
18. Pavel Ryska: *Deflation and Economic Growth in Long-Term Perspective*
19. Michal Paulus, Eva Michalíková, Vladimír Benáček: *German International Trade: Interpreting Export Flows According to the Gravity Model*
20. Michal Bauer, Nathan Fiala, Ian Lively: *Trusting Former Rebels: An Experimental Approach to Understanding Reintegration after Civil War*
21. Michal Skořepa: *Family Matters: Concurrent Capital Buffers in a Banking Group*
22. Martin Gregor: *Access Fees for Competing Lobbies*
23. Jozef Baruník, František Čech: *On the Modelling and Forecasting Multivariate Realized Volatility: Generalized Heterogeneous Autoregressive (GHAR) Model*
24. Roman Horváth, Ayaz Zeynalov: *The Natural Resource Curse and Institutions in Post-Soviet Countries*
25. Petr Janský: *Impact of Excise Duties on Households in the Czech Republic*
26. Petr Jakubík, Thomas Reininger: *What Are the Key Determinants of Nonperforming Loans in CESEE?*

27. Jaromír Baxa, Tomáš Olešňaník: *Current Accounts in the Eurozone Countries: The Role of Euro, Fiscal Policies and Financial Developments*
28. Petr Polák: *The Productivity Paradox: A Meta-analysis*
29. Tomáš Havránek, Zuzana Iršová, Karel Janda, David Zilberman: *Selective Reporting and the Social Cost of Carbon*
30. Jozef Baruník, Tomáš Křehlík: *Coupling High-Frequency Data with Nonlinear Models in Multiple-Step-ahead Forecasting of Energy Markets' Volatility*
31. Lenka Štašná, Jana Votápková: *Efficiency of Hospitals in the Czech Republic: Conditional Efficiency Approach*
32. Eva Michalíková, Michal Paulus: *Gravity Model Analysis: Robust Evidence from the Czech Republic and Corruption Matching*
33. Jozef Baruník, Lucie Kraicová: *Estimation of Long Memory in Volatility Using Wavelets*
34. Daniel Benčík: *Range-based Volatility Estimation and Forecasting*

All papers can be downloaded at: <http://ies.fsv.cuni.cz>



Univerzita Karlova v Praze, Fakulta sociálních věd
Institut ekonomických studií [UK FSV – IES] Praha 1, Opletalova 26
E-mail : ies@fsv.cuni.cz <http://ies.fsv.cuni.cz>