Separating Positive and Negative Jumps in HAR Volatility Forecasting Models

Martin Becker*
Saarland University, Saarbrücken, Germany

February 17, 2009

Preliminary Version

Abstract

We propose new variations of the HAR volatility forecasting model class originally introduced by Corsi (2004). Our models differ from the HAR models of Andersen et al. (2007) and Corsi et al. (2008) in a separate treatment of positive and negative jumps based on estimators and tests of Barndorff-Nielsen et al. (2008) and Klößner (2008). An extensive empirical study is carried out to compare the (out-of-sample) forecasting performance of the different HAR models.

Keywords: HAR models, realized variance, integrated volatility, volatility forecasting, OHLC data

*Tel.: +49 681 302 3571, Fax.: +49 681 302 3551, Address: Saarland University, Campus C3.1, Room 206, 66123 Saarbrücken, email: `martin.becker@mx.uni-saarland.de`

1 Introduction

As mentioned in Andersen et al. (2007), volatility is central to asset pricing, asset allocation, and risk management. A decade ago, volatility was usually measured on the basis of daily returns, or sometimes daily open, high, low, and close prices[1]. The main vehicles for volatility forecasting were the popular GARCH and (discrete time) stochastic volatility models. With the increased availability of high frequency data, other concepts have emerged for measuring and forecasting volatility, including realized variance[2] (see, e.g., Andersen and Bollerslev (1998) and Andersen et al. (2001)) or, more generally, realized (power) variation (see, e.g., Barndorff-Nielsen and Shephard (2002a) and Barndorff-Nielsen and Shephard (2002b)).

The empirical results[3] in Andersen et al. (2003) show that out-of-sample forecasting models of realized variance (RV) outperform GARCH and stochastic volatility models.[4] Corsi (2004) adopt the Heterogeneous Market Hypothesis[5], which was already employed in Müller et al. (1997) for the construction of the HARCH model class, to develop a simple heterogeneous autoregressive model for realized variance, the so-called HAR-RV model. In the HAR-RV model, realized volatility is affected by different aggregations of past realized volatility, which corresponds to appropriate parameter restrictions in the underlying autoregressive model. Despite their simplicity, HAR-RV models are able to capture most of the empirically observable stylized facts of realized variance.

Andersen et al. (2007) and Corsi et al. (2008) extend the original HAR-RV model by separately measuring the continuous sample path variation[6] and the discontinuous jump part of the quadratic variation, i.e., the sum of squared jumps (SSJ), using realized power variations. Using newly introduced tests for the presence of jumps on particular (trading) days based on asymptotic results of Barndorff-Nielsen and Shephard (2006), their empirical results indicate better forecasting performance for models with separate regressors for the (aggregated) continuous and jump parts (where positively detected by the jump test) of the process' variation. Recent works of Barndorff-Nielsen et al. (2008) and Klößner (2008) introduce methods for individual estimation of the sum of squared positive (SSpJ) and the sum of squared negative jumps (SSnJ) based on high-frequency data, the latter making use of high-frequency open, high, low, and close prices. We address the question whether incorporating these estimators in HAR-RV models leads to an improved (out-of-sample) forecast performance for realized

[1] for estimators based on open, high, low, and close prices, see, e.g., Rogers et al. (1994) and the references therein.

[2] also called realized *volatility*

[3] supported by analytical results in Andersen et al. (2004)

[4] This is not surprising, because of the different meanings of *volatility*: daily (squared) returns bare only very noisy information of the volatility, i.e. the conditional variance of the daily returns, in GARCH and SV models, whereas realized variance is a very good proxy for quadratic variation, i.e. the volatility in RV models, see the discussion in Andersen and Bollerslev (1998).

[5] According to the Heterogeneous Market Hypothesis, different types of agents behave according to different time resolutions.

[6] i.e., integrated volatility

variance. For this purpose we propose some new HAR-RV model classes as competitors to the HAR-RV-type models of Corsi (2004), Andersen et al. (2007), and Corsi et al. (2008). An extensive empirical study is carried out to compare the out-of-sample forecasting performance in the spirit of Corsi et al. (2008) by considering (out-of-sample) Mincer-Zarnowitz-R^2 and RMSE.

The remainder is organized as follows: Section 2 introduces the theoretical framework and some basic notation. In section 3 we collect various estimators for quadratic variation, integrated volatility, and the sum of squared (positive and negative) jumps for individual trading days. Section 4 summarizes the tests for the presence of jumps, which are used to construct the HAR-RV forecasting model classes in section 5. In section 6, the empirical results are presented in aggregated form, while section 7 concludes.

2 Preliminaries and Notation

We adopt the setting of Klößner (2008), where the log-price process $(p_{\tau,t})_{\tau\in[0,1]}$ on each day[7] $t \in \{1,\dots,T\}$, $T \in \mathbb{N}$, is given as the sum of a Brownian semimartingale $p_{0,t} + \int_0^\tau \mu_{s,t} ds + \int_0^\tau \sigma_{s,t} dW_{s,t}$ and a finite activity pure jump process $J^{(p)}_{\tau,t}$,

$$p_{\tau,t} = p_{0,t} + \int_0^\tau \mu_{s,t} ds + \int_0^\tau \sigma_{s,t} dW_{s,t} + J^{(p)}_{\tau,t} \ ,$$

with locally bounded and predictable drift rate $\mu_{\tau,t}$, Brownian motion $W_{\tau,t}$ and never vanishing spot volatility $\sigma_{\tau,t}$ given by

$$\sigma_{\tau,t} = \sigma_{0,t} + \int_0^\tau \widetilde{\mu}_{s,t} ds + \int_0^\tau \widetilde{\sigma}_{s,t} dW_{s,t} + \int_0^\tau v_{s,t} dB_{s,t} + J^{(\sigma)}_{\tau,t} \ ,$$

where $\widetilde{\mu}_{\tau,t}$, $\widetilde{\sigma}_{\tau,t}$, $v_{\tau,t}$ are càdlàg, $\widetilde{\mu}_{\tau,t}$ is locally bounded as well as predictable, $B_{\tau,t}$ is a Brownian motion independent of $W_{\tau,t}$, and $J^{(\sigma)}_{\tau,t}$ is a finite-activity pure-jump process accounting for the possibility of rare jumps in volatility.

It is well known that, for each day $t \in \{1,\dots,T\}$, the quadratic variation of $p_{\cdot,t}$ on $[0,1]$,

$$\mathrm{QV}_t = \lim_{N\to\infty} \sum_{i=1}^{N} \left(p_{\frac{i}{N},t} - p_{\frac{i-1}{N},t} \right)^2 \ ,$$

can be decomposed into integrated volatility, $\mathrm{IV}_t := \int_0^1 \sigma^2_{\tau,t} d\tau$, and the sum of squared jumps, $\mathrm{SSJ}_t := \sum_{\tau\in[0,1]} (\Delta p_{\tau,t})^2$, with $\Delta p_{\tau,t} := p_{\tau,t} - p_{\tau-,t}$ and $p_{\tau-,t} := \lim_{h\to 0^+} p_{\tau-h,t}$.

[7] the (trading) time horizon is normalized to $[0,1]$ for each single day

Obviously, SSJ_t can be split up into the contributions due to positive and negative jumps,

$$\mathrm{SSpJ}_t := \sum_{\substack{\Delta p_{\tau,t}>0 \\ \tau\in[0,1]}} (\Delta p_{\tau,t})^2 \qquad \text{and} \qquad \mathrm{SSnJ}_t := \sum_{\substack{\Delta p_{\tau,t}<0 \\ \tau\in[0,1]}} (\Delta p_{\tau,t})^2 \ ,$$

leading to the representations $\mathrm{QV}_t = \mathrm{IV}_t + \mathrm{SSJ}_t$ and $\mathrm{QV}_t = \mathrm{IV}_t + \mathrm{SSpJ}_t + \mathrm{SSnJ}_t$.
To facilitate the exposition of estimators and tests concerning these quantities based on (equally spaced) intradaily high-frequency log-price data[8], we define for fixed $N \in \mathbb{N}$ the intradaily open, high, low, and close (OHLC) log-prices

$$\mathsf{o}_{i,t} := p_{\frac{i-1}{N},t}, \quad \mathsf{h}_{i,t} := \sup_{\tau\in[\frac{i-1}{N},\frac{i}{N}]} p_{\tau,t}, \quad \mathsf{l}_{i,t} := \inf_{\tau\in[\frac{i-1}{N},\frac{i}{N}]} p_{\tau,t}, \quad \mathsf{c}_{i,t} := p_{\frac{i}{N},t},$$

as well as the intradaily log-returns $r_{i,t} := \mathsf{c}_{i,t} - \mathsf{o}_{i,t}$, for $i \in \{1, \ldots, N\}$ and $t \in \{1, \ldots, T\}$.

3 Estimating QV, IV and Sum of Squared Jumps

We now collect various estimators for QV_t, IV_t, SSJ_t, SSpJ_t, and SSnJ_t from the literature. We start with estimators relying only on the discrete log-price quotes $p_{\frac{i}{N},t}$, for $i \in \{0, \ldots, N\}$ and $t \in \{0, \ldots, T\}$.

3.1 Estimators without Highs/Lows

3.1.1 Estimating QV

It is well known that QV_t can be consistently[9] estimated with the realized variance, $\mathrm{RV}_t = \sum_{i=1}^N r_{i,t}^2$. In our setting, RV_t can be regarded as the 'standard' estimator for QV_t, see e.g. Andersen et al. (2001).

3.1.2 Estimating IV

A large class of estimators for quantities of the form

$$\int_0^1 |\sigma_{s,t}|^p ds$$

[8] including high and low quotes, in parts

[9] throughout the paper, we consider consistent estimation w.r.t. $N \to \infty$, in this case: $\operatorname{plim}_{N\to\infty} \mathrm{RV}_t = \mathrm{QV}_t$

for $p = \sum_{k=1}^{M} \gamma_k > 0$, $M \in \mathbb{N}$, $\gamma_k \geq 0$ for $k = 1, \ldots, M$, is given by the scaled normalized[10] realized multipower variation

$$\mathrm{MPV}_t^{(\gamma_1,\ldots,\gamma_M)} := \frac{1}{1-\frac{M-1}{N}} \frac{1}{N^{1-\frac{1}{2}\sum_k \gamma_k}} \left(\prod_{k=1}^{M} \mu_{\gamma_k}^{-1} \right) \sum_{i=M}^{N} \prod_{k=1}^{M} |r_{i-k+1,t}|^{\gamma_k}$$

with $\mu_\gamma := \mathrm{E}\left(|u|^\gamma\right) = 2^{\gamma/2} \frac{\Gamma(\frac{\gamma+1}{2})}{\Gamma(1/2)}$ for $u \sim \mathcal{N}(0,1)$.[11]

In our setting, $\mathrm{MPV}_t^{(\gamma_1,\ldots,\gamma_M)}$ provides a consistent estimator of $\int_0^1 |\sigma_{s,t}|^p ds$, if the condition $\max\{\gamma_1, \ldots, \gamma_M\} < 2$ holds, cf. Barndorff-Nielsen et al. (2006).

A special case is the well-known (scaled normalized) realized bipower variation,

$$\mathrm{BPV}_t^{\mathrm{IV}} := \mathrm{MPV}_t^{(1,1)} = \frac{1}{1-\frac{1}{N}} \frac{\pi}{2} \sum_{i=2}^{N} |r_{i-1,t}| \cdot |r_{i,t}| \ ,$$

which has been proposed in Barndorff-Nielsen and Shephard (2004) as a consistent estimator for IV_t.

To overcome some of the problems induced by microstructure noise, Andersen et al. (2007) suggest the use of a staggered form of multipower variation[12], where $l \in \mathbb{N}$ returns are skipped in the products instead of using adjacent returns. A special case of the resulting staggered realized multipower variations,

$$\text{s-MPV}_t^{(\gamma_1,\ldots,\gamma_M;l)} := \frac{1}{1-\frac{(l+1)(M-1)}{N}} \frac{1}{N^{1-\frac{1}{2}\sum_k \gamma_k}} \left(\prod_{k=1}^{M} \mu_{\gamma_k}^{-1} \right) \sum_{i=1+(l+1)(M-1)}^{N} \prod_{k=1}^{M} |r_{i-(l+1)(k-1),t}|^{\gamma_k} \ ,$$

is the staggered realized bipower variation

$$\text{s-BPV}_t^{\mathrm{IV}} := \text{s-MPV}_t^{(1,1;1)} = \frac{1}{1-\frac{2}{N}} \frac{\pi}{2} \sum_{i=3}^{N} |r_{i-2,t}| \cdot |r_{i,t}| \ .$$

Corsi et al. (2008) introduce the concept of realized threshold multipower variation to mitigate the well-known positive finite-sample[13] bias of realized multipower measures due to jumps in the log-price process.

For given threshold functions $\vartheta_t : [0,1] \to \mathbb{R}^+$ and $\vartheta_{i,t} := \vartheta_t(\frac{i}{N})$, they consider (realized)

---

[10]realized multipower variation is often considered without scaling factor $\prod_{k=1}^{M} \mu_{\gamma_k}^{-1}$, normalization factor $\frac{1}{N^{1-\frac{1}{2}\sum_k \gamma_k}}$, and/or (finite-sample) bias correction factor $\frac{1}{1-\frac{M-1}{N}}$

[11]particularly, $\mu_1 = \sqrt{\frac{2}{\pi}}$. $\mathcal{N}(0,1)$ denotes the standard normal distribution.

[12]see also Huang and Tauchen (2005)

[13]i.e. for fixed $N \in \mathbb{N}$

scaled[14] threshold multipower variations

$$\mathrm{TMPV}_t^{(\gamma_1,\dots,\gamma_M)} := \frac{1}{N^{1-\frac{1}{2}\sum_k \gamma_k}} \left(\prod_{k=1}^{M} \mu_{\gamma_k}^{-1}\right) \sum_{i=M}^{N} \prod_{k=1}^{M} |r_{i-k+1,t}|^{\gamma_k} \mathbb{1}_{|r_{i-k+1,t}|^2 \le \vartheta_{i-k+1,t}} ,$$

where increments $r_{i-k+1,t}$ with $r_{i-k+1,t}^2 > \vartheta_{i-k+1,t}$ are treated as jumps and are thus eliminated. Corsi et al. (2008) show that the particular choice of the threshold function ϑ_t is almost immaterial for the considered applications. They chose a multiple $c_\vartheta^2 \cdot \widehat{V}_{i,t}^Z$ of an iteratively determined[15] estimator for the local variance $\sigma^2_{\frac{i}{N},t}$,

$$\widehat{V}_{i,t}^Z := \frac{\sum\limits_{\substack{j=-L \\ j\neq -1,0,1 \\ i+j\in\{1,\dots,N\}}}^{L} K\left(\frac{j}{L}\right) r_{i+j,t}^2 \mathbb{1}_{r_{i+j,t}^2 \le c_V^2 \cdot \widehat{V}_{j,t}^{Z-1}}}{\sum\limits_{\substack{j=-L \\ j\neq -1,0,1 \\ i+j\in\{1,\dots,N\}}}^{L} K\left(\frac{j}{L}\right) \mathbb{1}_{r_{i+j,t}^2 \le c_V^2 \cdot \widehat{V}_{j,t}^{Z-1}}} , \quad Z = 1,2,\dots ,$$

with

$$K(y) := \frac{1}{\sqrt{2\pi}} \exp(-\frac{y^2}{2}) ,$$

$L = 25$, $c_V = 3$, and $\widehat{V}_{i,t}^0 := +\infty$ (thus using all returns in the first iteration).
The robustness of the choice of the threshold is further investigated by considering different values of c_ϑ in the family of threshold functions $\vartheta_t^{c_\vartheta}$ with $\vartheta_{i,t}^{c_\vartheta} := c_\vartheta^2 \cdot \widehat{V}_{i,t}^Z$ for $i \in \{1,\dots,N\}$, $t \in \{1,\dots,T\}$. As in Corsi et al. (2008), we chose $c_\vartheta = 3$ for the following applications.

3.1.3 Estimating SSJ, SSpJ, and SSnJ

According to the decomposition of quadratic variation into integrated volatility and the sum of squared jumps, consistent estimators of SSJ_t are obviously provided by the difference of consistent estimators for QV_t and IV_t.
Building on the respective estimators for IV_t, in Andersen et al. (2007), the estimator

$$\widehat{\mathrm{SSJ}}_t^{\mathrm{ABD}} = \mathrm{RV}_t - \mathrm{BPV}_t^{\mathrm{IV}}$$

for SSJ_t is considered in a first approach, whereas Corsi et al. (2008) make use of their threshold estimator for IV_t and obtain

$$\widehat{\mathrm{SSJ}}_t^{\mathrm{CPR}} = \mathrm{RV}_t - \mathrm{TBPV}_t^{\mathrm{IV}}$$

for estimating SSJ_t.

[14] In Corsi et al. (2008), scaling is not incorporated in the definition of $\mathrm{TMPV}_t^{(\gamma_1,\dots,\gamma_M)}$.
[15] the iteration stops if $\widehat{V}_{i,t}^Z \equiv \widehat{V}_{i,t}^{Z-1}$ or if $Z = 100$ (which has never occured in our empirical applications)

estimator	calculation
RUVp_t	$2\sum_{i=1}^N 2\,(\mathsf{h}_{i,t}-\mathsf{o}_{i,t})\,(\mathsf{h}_{i,t}-\mathsf{c}_{i,t})\,\mathbb{1}_{r_{i,t}>0}$
RDVp_t	$2\sum_{i=1}^N 2\,(\mathsf{o}_{i,t}-\mathsf{l}_{i,t})\,(\mathsf{c}_{i,t}-\mathsf{l}_{i,t})\,\mathbb{1}_{r_{i,t}>0}$
RGRVp_t	$2\sum_{i=1}^N \frac{1}{2\ln 2-1}\,(\mathsf{h}_{i,t}-\mathsf{o}_{i,t})\,(\mathsf{o}_{i,t}-\mathsf{l}_{i,t})\,\mathbb{1}_{r_{i,t}>0}$
RTrGRVp_t	$2\sum_{i=1}^N \frac{1}{2\ln 2-1}\,(\mathsf{h}_{i,t}-\mathsf{c}_{i,t})\,(\mathsf{c}_{i,t}-\mathsf{l}_{i,t})\,\mathbb{1}_{r_{i,t}>0}$
RpJVp_t	$\frac{8}{7}\sum_{i=1}^N \frac{1}{2}\left((\mathsf{h}_{i,t}-\mathsf{o}_{i,t})^2+(\mathsf{c}_{i,t}-\mathsf{l}_{i,t})^2\right)\mathbb{1}_{r_{i,t}>0}$
RnJVp_t	$8\sum_{i=1}^N \frac{1}{2}\left((\mathsf{h}_{i,t}-\mathsf{c}_{i,t})^2+(\mathsf{o}_{i,t}-\mathsf{l}_{i,t})^2\right)\mathbb{1}_{r_{i,t}>0}$
RDVn_t	$2\sum_{i=1}^N 2\,(\mathsf{o}_{i,t}-\mathsf{l}_{i,t})\,(\mathsf{c}_{i,t}-\mathsf{l}_{i,t})\,\mathbb{1}_{r_{i,t}<0}$
RUVn_t	$2\sum_{i=1}^N 2\,(\mathsf{h}_{i,t}-\mathsf{o}_{i,t})\,(\mathsf{h}_{i,t}-\mathsf{c}_{i,t})\,\mathbb{1}_{r_{i,t}<0}$
RGRVn_t	$2\sum_{i=1}^N \frac{1}{2\ln 2-1}\,(\mathsf{h}_{i,t}-\mathsf{o}_{i,t})\,(\mathsf{o}_{i,t}-\mathsf{l}_{i,t})\,\mathbb{1}_{r_{i,t}<0}$
RTrGRVn_t	$2\sum_{i=1}^N \frac{1}{2\ln 2-1}\,(\mathsf{h}_{i,t}-\mathsf{c}_{i,t})\,(\mathsf{c}_{i,t}-\mathsf{l}_{i,t})\,\mathbb{1}_{r_{i,t}<0}$
RnJVn_t	$8\sum_{i=1}^N \frac{1}{2}\left((\mathsf{h}_{i,t}-\mathsf{c}_{i,t})^2+(\mathsf{o}_{i,t}-\mathsf{l}_{i,t})^2\right)\mathbb{1}_{r_{i,t}<0}$
RpJVn_t	$\frac{8}{7}\sum_{i=1}^N \frac{1}{2}\left((\mathsf{h}_{i,t}-\mathsf{o}_{i,t})^2+(\mathsf{c}_{i,t}-\mathsf{l}_{i,t})^2\right)\mathbb{1}_{r_{i,t}<0}$
RUVz_t	$2\sum_{i=1}^N (\mathsf{h}_{i,t}-\mathsf{o}_{i,t})^2\,\mathbb{1}_{r_{i,t}=0}$
RDVz_t	$2\sum_{i=1}^N (\mathsf{o}_{i,t}-\mathsf{l}_{i,t})^2\,\mathbb{1}_{r_{i,t}=0}$
RGVz_t	$\frac{12}{\pi^2-6}\sum_{i=1}^N (\mathsf{h}_{i,t}-\mathsf{o}_{i,t})\,(\mathsf{o}_{i,t}-\mathsf{l}_{i,t})\,\mathbb{1}_{r_{i,t}=0}$

Table 1: 12+3 basic estimators of Klößner (2008)

Disentangling SSJ_t in SSpJ_t and SSnJ_t for estimation purposes was first achieved by Barndorff-Nielsen et al. (2008), who proposed realized semi-variances, defined as

$$\text{RS}_t^+ := \sum_{i=1}^N r_{i,t}^2 \mathbb{1}_{r_{i,t}>0} \qquad \text{and} \qquad \text{RS}_t^- := \sum_{i=1}^N r_{i,t}^2 \mathbb{1}_{r_{i,t}<0} \ ,$$

to obtain estimators for the sum of squared positive jumps and the sum of squared negative jumps. As RS_t^+ is a consistent estimator for $\frac{1}{2}\text{IV}_t + \text{SSpJ}_t$, and RS_t^- consistently estimates $\frac{1}{2}\text{IV}_t + \text{SSnJ}_t$, consistent estimators for SSpJ_t and SSnJ_t are easily obtained by

$$\widehat{\text{SSpJ}}_t^{\text{RS}} := \text{BPUV}_t := \text{RS}_t^+ - \frac{1}{2}\text{BPV}_t^{\text{IV}} \qquad \text{and} \qquad \widehat{\text{SSnJ}}_t^{\text{RS}} := \text{BPDV}_t := \text{RS}_t^- - \frac{1}{2}\text{BPV}_t^{\text{IV}} \ .$$

3.2 Estimators using Highs/Lows

In Klößner (2008), estimators for QV_t, IV_t, SSJ_t, SSpJ_t, and SSnJ_t based on intradaily open, high, low, and close quotes $\mathsf{o}_{i,t}$, $\mathsf{h}_{i,t}$, $\mathsf{l}_{i,t}$, and $\mathsf{c}_{i,t}$ for $i \in \{1,\dots,N\}$, $t \in \{1,\dots,T\}$, are developed. The estimators are constructed as (in different senses) optimal linear combinations of a set of 12 basic estimators, supplemented by 3 (theoretically irrelevant) estimators for the (empirically observable) case of zero increments ($r_{i,t} = \mathsf{c}_{i,t} - \mathsf{o}_{i,t} = 0$). The definitions of these 12 (+3) basic estimators are summarized in Table 1.

By solving quadratic programs with linear constraints, Klößner (2008) designes (efficient) es-

basic est. \ est.	$\widehat{\mathrm{QV}}_t^{(c)}$	$\widehat{\mathrm{IV}}_t^{(c)}$	$\widehat{\mathrm{SSJ}}_t^{(c)}$	$\widehat{\mathrm{SSpJ}}_t^{(c)}$	$\widehat{\mathrm{SSnJ}}_t^{(c)}$
RUVp_t	-0.89954	0.12208	-1.02162	-1.13861	0.11699
RDVp_t	-0.89954	0.12208	-1.02162	-1.13861	0.11699
RGRVp_t	0.18738	0.13376	0.05362	0.07141	-0.01779
$\mathrm{RTrGRVp}_t$	0.18738	0.13376	0.05362	0.07141	-0.01779
RpJVp_t	0.87500	0.00000	0.87500	0.87500	0.00000
RnJVp_t	1.04933	-0.01168	1.06101	1.20625	-0.14524
RDVn_t	-0.89954	0.12208	-1.02162	0.11699	-1.13861
RUVn_t	-0.89954	0.12208	-1.02162	0.11699	-1.13861
RGRVn_t	0.18738	0.13376	0.05362	-0.01779	0.07141
$\mathrm{RTrGRVn}_t$	0.18738	0.13376	0.05362	-0.01779	0.07141
RnJVn_t	0.87500	0.00000	0.87500	0.00000	0.87500
RpJVn_t	1.04933	-0.01168	1.06101	-0.14524	1.20625
RUVz_t	0.25537	0.25537	0.00000	0.00000	0.00000
RDVz_t	0.25537	0.25537	0.00000	0.00000	0.00000
RGVz_t	0.48926	0.48926	0.00000	0.00000	0.00000

Table 2: weights of consistent estimators in Klößner (2008)

timators for special purposes, including consistent estimators for QV_t, IV_t, SSJ_t, SSpJ_t, and SSnJ_t with feasible central limit theorems. Table 2 contains the weights of the basic estimators for these consistent estimators.

4 Testing for the Presence of Jumps

In this section, we collect tests for the presence of jumps from the recent literature. Based on distributional results developed in Barndorff-Nielsen and Shephard (2006), Corsi et al. (2008), and Klößner (2008), the asymptotic distribution of certain test statistics under the null hypothesis of an absent jump component $J_{\tau,t}^{(p)}$ is known[16], allowing for a simple construction of test recipes.

All of the following tests for the presence of jumps need a consistent estimator for integrated quarticity (IQ) to be feasible. A summary of these estimators follows.

4.1 Estimators for IQ

Applying the aforementioned results for realized (threshold) multipower variation, an extensive family of estimators for IQ_t based on $p_{\frac{i}{N},t}$, $i \in \{0,\ldots,N\}$, $t \in \{1,\ldots,T\}$, can be constructed. We revert to the (integrated) quarticity measures used in Andersen et al. (2007), Corsi et al. (2008) and Huang and Tauchen (2005), more precisely:

[16] in particular, the resulting asymptotic distribution is $\mathcal{N}(0,1)$ for all following test statistics

- realized tripower variation quarticity measure:

$$\mathrm{TriPV}_t^{\mathrm{IQ}} := \mathrm{MPV}_t^{(\frac{4}{3},\frac{4}{3},\frac{4}{3})} = \frac{N\,\Gamma(\frac{1}{2})^3}{4\,\Gamma(\frac{7}{6})^3} \sum_{i=3}^{N} |r_{i-2,t}|^{\frac{4}{3}} \cdot |r_{i-1,t}|^{\frac{4}{3}} \cdot |r_{i,t}|^{\frac{4}{3}}$$

- realized quadpower variation quarticity measure:

$$\mathrm{QPV}_t^{\mathrm{IQ}} := \mathrm{MPV}_t^{(1,1,1,1)} = \frac{N\,\pi^2}{4} \sum_{i=4}^{N} |r_{i-3,t}| \cdot |r_{i-2,t}| \cdot |r_{i-1,t}| \cdot |r_{i,t}|$$

- staggered realized tripower variation quarticity measure:

$$\text{s-TriPV}_t^{\mathrm{IQ}} := \text{s-MPV}_t^{(\frac{4}{3},\frac{4}{3},\frac{4}{3};1)} = \frac{N\,\Gamma(\frac{1}{2})^3}{4\,\Gamma(\frac{7}{6})^3} \frac{1}{1-\frac{4}{N}} \sum_{i=5}^{N} |r_{i-4,t}|^{\frac{4}{3}} \cdot |r_{i-2,t}|^{\frac{4}{3}} \cdot |r_{i,t}|^{\frac{4}{3}}$$

- staggered realized quadpower variation quarticity measure:

$$\text{s-QPV}_t^{\mathrm{IQ}} := \text{s-MPV}_t^{(1,1,1,1;1)} = \frac{N\,\pi^2}{4} \frac{1}{1-\frac{6}{N}} \sum_{i=7}^{N} |r_{i-6,t}| \cdot |r_{i-4,t}| \cdot |r_{i-2,t}| \cdot |r_{i,t}|$$

Klößner (2008) presents the following new estimator for IQ_t, which makes additionally use of intradaily high and low (log-)prices $\mathsf{h}_{i,t}$ and $\mathsf{l}_{i,t}$, $i \in \{1,\dots,N\}$, $t \in \{1,\dots,T\}$:

$$\begin{aligned}
\widehat{\mathrm{IQ}}_t^{\mathrm{hl}} &= \frac{N}{2} \sum_{i=1}^{N} \frac{16}{3} \left((\mathsf{h}_{i,t} - \mathsf{c}_{i,t})^4 + (\mathsf{o}_{i,t} - \mathsf{l}_{i,t})^4 \right) \mathbb{1}_{r_{i,t}>0} \\
&+ \frac{N}{2} \sum_{i=1}^{N} \frac{16}{3} \left((\mathsf{h}_{i,t} - \mathsf{o}_{i,t})^4 + (\mathsf{c}_{i,t} - \mathsf{l}_{i,t})^4 \right) \mathbb{1}_{r_{i,t}<0} \\
&+ N \sum_{i=1}^{N} \left((\mathsf{h}_{i,t} - \mathsf{o}_{i,t})^4 + (\mathsf{o}_{i,t} - \mathsf{l}_{i,t})^4 \right) \mathbb{1}_{r_{i,t}=0}
\end{aligned}$$

$\widehat{\mathrm{IQ}}_t^{\mathrm{hl}}$ will be used for the tests based on intradaily open, high, low, and close data.

4.2 Tests based on (staggered) MPV and threshold MPV

In Huang and Tauchen (2005), many different tests for the presence of jumps based on RV and (staggered) MPV are constructed and compared. As a result of their comprehensive Monte Carlo studies they suggest employing a ratio form test statistics, which has the nice interpretation as a Hausman-type test.

Andersen et al. (2007) make direct use of this asymptotically $\mathcal{N}(0,1)$-distributed ratio test statistics,

$$Z_t := \sqrt{N}\frac{(\mathrm{RV}_t - \text{s-BPV}_t^{\mathrm{IV}})\cdot \mathrm{RV}_t^{-1}}{\sqrt{(\frac{\pi^2}{4}+\pi-5)\max\left\{1, \frac{\text{s-TriPV}_t^{\mathrm{IQ}}}{(\text{s-BPV}_t^{\mathrm{IV}})^2}\right\}}} ,$$

while Corsi et al. (2008) build upon a similar structure, but use corrected versions of their threshold estimators in place of s-BPV$_t^{\mathrm{IV}}$ and s-TriPV$_t^{\mathrm{IQ}}$.

Under the null hypothesis of no jumps, the threshold estimators $\mathrm{TMPV}_t^{(\gamma_1,\ldots,\gamma_M)}$ are biased downwards, since the contributions $|r_{i-k+1,t}|^{\gamma_k}$ for $r_{i-k+1,t}$ are excluded if $r_{i-k+1,t}$ is too big (i.e. $r_{i-k+1,t}^2 > \vartheta_{i-k+1,t}$). To eliminate this bias, Corsi et al. (2008) replace $|r_{i-k+1,t}|^{\gamma_k}$ by (an estimator of) its (conditional) expectation under the null[17] (instead of zero) in these cases. They arrive at the corrected versions of (realized) threshold MPV,

$$\text{C-TMPV}_t^{(\gamma_1,\ldots,\gamma_M)} := \frac{1}{N^{1-\frac{1}{2}\sum_k \gamma_k}} \sum_{i=M}^{N} \prod_{k=1}^{M} Z_{\gamma_k}(r_{i-k+1,t}, \vartheta_{i-k+1,t}) ,$$

with

$$Z_\gamma(x,y) := \begin{cases} |x|^\gamma & \text{if } x^2 \le y \\ \frac{1}{2\Phi(-c_\vartheta)\sqrt{\pi}} \left(\frac{2}{c_\vartheta^2} y\right)^{\frac{\gamma}{2}} \Gamma\left(\frac{\gamma+1}{2}, \frac{c_\vartheta^2}{2}\right) & \text{if } x^2 > y \end{cases} .$$

The resulting corrected test statistics

$$\text{C-Z}_t := \sqrt{N}\frac{(\mathrm{RV}_t - \text{C-TBPV}_t^{\mathrm{IV}})\cdot \mathrm{RV}_t^{-1}}{\sqrt{(\frac{\pi^2}{4}+\pi-5)\max\left\{1, \frac{\text{C-TTriPV}_t^{\mathrm{IQ}}}{(\text{C-TBPV}_t^{\mathrm{IV}})^2}\right\}}}$$

is again asymptotically $\mathcal{N}(0,1)$-distributed.

4.3 Tests based on OHLC data

In Klößner (2008), special purpose estimators for tests on the presence of jumps are also developed. In particular, separate test for the presence of arbitrary jumps, positive jumps, and negative jumps are constructed. The test statistics rely on certain estimators for SSJ_t, SSpJ_t, and SSnJ_t, again constructed as linear combinations of the twelve aforementioned basic estimators, with the weights given in Table 3. Table 3 also contains variance factors in the last

[17] given by

$$\mathrm{E}\left(|r_{i-k+1,t}|^{\gamma_k}\middle| r_{i-k+1,t}^2 > \vartheta_{i-k+1,t}\right) = \frac{\Gamma\left(\frac{\gamma_k+1}{2}, \frac{c_\vartheta^2}{2}\right)}{2\Phi\left(-\frac{\sqrt{\vartheta_{i-k+1,t}}}{\sigma_{\frac{i}{N},t}}\right)\sqrt{\pi}} (2\sigma_{\frac{i}{N},t})^{\frac{1}{2}\gamma_k} ,$$

where $\Gamma(\cdot,\cdot)$ denotes the upper incomplete Gamma function, i.e. $\Gamma(\alpha,x) = \int_x^{+\infty} s^{\alpha-1}e^{-s}ds$, and $\Phi(\cdot)$ denotes the cumulative distribution function of the standard normal distribution.

basic est. \ est.	$\widehat{\text{SSJ}}_t^{(t)}$	$\widehat{\text{SSpJ}}_t^{(t)}$	$\widehat{\text{SSnJ}}_t^{(t)}$
RUVp_t	-1.02162	-1.50326	0.00000
RDVp_t	-1.02162	-1.50326	0.00000
RGRVp_t	0.05362	0.20185	0.00000
RTrGRVp_t	0.05362	0.20185	0.00000
RpJVp_t	0.87500	0.87500	0.00000
RnJVp_t	1.06101	1.72782	0.00000
RDVn_t	-1.02162	0.00000	-1.50326
RUVn_t	-1.02162	0.00000	-1.50326
RGRVn_t	0.05362	0.00000	0.20185
RTrGRVn_t	0.05362	0.00000	0.20185
RnJVn_t	0.87500	0.00000	0.87500
RpJVn_t	1.06101	0.00000	1.72782
variance factors	1.30144	0.86020	0.86020

Table 3: weights of estimators for tests in Klößner (2008)

row, which will be needed for normalization of the test statistics.
The resulting test statistics

$$\text{TJ}_t^{\text{hl}} := \sqrt{N}\frac{\widehat{\text{SSJ}}_t^{(t)}}{\sqrt{1.30144\,\widehat{\text{IQ}}_t^{\text{hl}}}},\ \text{TJp}_t^{\text{hl}} := \sqrt{N}\frac{\widehat{\text{SSpJ}}_t^{(t)}}{\sqrt{0.86020\,\widehat{\text{IQ}}_t^{\text{hl}}}},\ \text{TJn}_t^{\text{hl}} := \sqrt{N}\frac{\widehat{\text{SSnJ}}_t^{(t)}}{\sqrt{0.86020\,\widehat{\text{IQ}}_t^{\text{hl}}}}$$

are asymptotically $\mathcal{N}(0,1)$-distributed under the null hypotheses of no jumps at all, no positive jumps, and no negative jumps, resp.

5 Forecasting Models for Realized Volatility

5.1 HAR models based on RV and SSJ

In this section we present various forecasting models for (transformations of) the volatility of the log-price process $p_{\tau,t}$. We build upon the stochastic additive cascade of three realized variance components corresponding to daily, weekly, and monthly time horizons of market activity, the so-called HAR model, which was originaly introduced by Corsi (2004).
To simplify notation, the arithmetic mean of a subseries $(X_t)_{t=t_1+1}^{t_2}$ of a time series $(X_t)_{t=1}^{T}$ will be denoted by

$$X_{t_1:t_2} := \frac{1}{t_2 - t_1}\sum_{t=t_1+1}^{t_2} X_t$$

for $t_1,\, t_2 \in \mathbb{N}$, $1 \leq t_1 < t_2 \leq T$.
The first HAR model of Corsi (2004), which we further refer to as HAR-RV model, can then

be written out as[18]

$$\mathrm{RV}_{t:t+h} = \beta_0 + \beta_d \mathrm{RV}_t + \beta_w \mathrm{RV}_{t-5:t} + \beta_m \mathrm{RV}_{t-22:t} + \varepsilon_t \ ,$$

where $h \in \{1, 5, 22\}$ is the aggregation frequency of the forecast[19] and ε_t is assumed to be IID. Corsi (2004) also proposed square-root and logarithmic versions[20] of the HAR-RV model, i.e.

$$\mathrm{RV}^{\frac{1}{2}}_{t:t+h} = \beta_0 + \beta_d \mathrm{RV}^{\frac{1}{2}}_t + \beta_w \mathrm{RV}^{\frac{1}{2}}_{t-5:t} + \beta_m \mathrm{RV}^{\frac{1}{2}}_{t-22:t} + \varepsilon_t$$

and

$$\log \mathrm{RV}_{t:t+h} = \beta_0 + \beta_d \log \mathrm{RV}_t + \beta_w \log \mathrm{RV}_{t-5:t} + \beta_m \log \mathrm{RV}_{t-22:t} + \varepsilon_t \ ,$$

resp.

Andersen et al. (2007) extend the HAR-RV model first by adding a jump component, namely an estimator J_t for the sum of squared jumps, to the model, i.e.

$$\mathrm{RV}_{t:t+h} = \beta_0 + \beta_d \mathrm{RV}_t + \beta_w \mathrm{RV}_{t-5:t} + \beta_m \mathrm{RV}_{t-22:t} + \beta_j \mathrm{J}_t + \varepsilon_t \ ,$$

where $\mathrm{J}_t := \max\{\widehat{\mathrm{SSJ}}^{\mathrm{ABD}}_t, 0\} = \max\{\mathrm{RV}_t - \mathrm{BPV}^{\mathrm{IV}}_t, 0\}$, resulting in the so-called HAR-RV-J model with corresponding square-root and logarithmic[21] versions.

They refine their original model in three ways: first, they make use of their test for the presence of jumps on day t by replacing J_t with

$$\mathrm{J}_{t,\alpha} := \left(\mathrm{RV}_t - \mathrm{BPV}^{\mathrm{IV}}_t\right) \cdot \mathbb{1}_{Z_t > \Phi^{-1}(\alpha)} \ ,$$

obviously replicating J_t for $\alpha = 0.5$.

Second, they replace RV_t with the difference of RV_t and $\mathrm{J}_{t,\alpha}$, i.e. with

$$C_{t,\alpha} := \mathrm{RV}_t - \mathrm{J}_{t,\alpha} = \mathrm{RV}_t \cdot \mathbb{1}_{Z_t \leq \Phi^{-1}(\alpha)} + \mathrm{BPV}^{\mathrm{IV}}_t \cdot \mathbb{1}_{Z_t > \Phi^{-1}(\alpha)},$$

and third, Andersen et al. (2007) include weekly and monthly aggregated jump components as well, resulting in their so-called HAR-RV-CJ model

$$\mathrm{RV}_{t:t+h} = \beta_0 + \beta_{cd} C_{t,\alpha} + \beta_{cw} C_{t-5:t,\alpha} + \beta_{cm} C_{t-22:t,\alpha} + \beta_{jd} \mathrm{J}_{t,\alpha} + \beta_{jw} \mathrm{J}_{t-5:t,\alpha} + \beta_{jm} \mathrm{J}_{t-22:t,\alpha} + \varepsilon_t$$

with adequately defined square-root and logarithmic versions.

The HAR-RV-TCJ model of Corsi et al. (2008) emerges from the HAR-RV-CJ model by

[18] For the coefficients, we use the abbreviations d for daily, w for weekly $(t-5:t)$, and m for monthly $(t-22:t)$ aggregation levels, resp.

[19] $h = 1$ corresponds to daily volatility forecasts, $h = 5$ to weekly, and $h = 22$ to monthly forecasts.

[20] the transformations are applied to the aggregated quantities, e.g. $\mathrm{RV}^{\frac{1}{2}}_{t-5:t} := (\mathrm{RV}_{t-5:t})^{\frac{1}{2}}$.

[21] for the logarithmic transformation of J_t, $\log(\mathrm{J}_t + 1)$ is considered instead of $\log \mathrm{J}_t$.

replacing $\mathrm{BPV}_t^{\mathrm{IV}}$ with $\mathrm{TBPV}_t^{\mathrm{IV}}$, s-$\mathrm{BPV}_t^{\mathrm{IV}}$ with C-$\mathrm{TBPV}_t^{\mathrm{IV}}$, and s-$\mathrm{TriPV}_t^{\mathrm{IQ}}$ with C-$\mathrm{TTriPV}_t^{\mathrm{IQ}}$, resp., resulting in[22]

$$\mathrm{RV}_{t:t+h} = \beta_0+\beta_{cd}\mathrm{TC}_{t,\alpha}+\beta_{cw}\mathrm{TC}_{t-5:t,\alpha}+\beta_{cm}\mathrm{TC}_{t-22:t,\alpha}+\beta_{jd}\mathrm{TJ}_{t,\alpha}+\beta_{jw}\mathrm{TJ}_{t-5:t,\alpha}+\beta_{jm}\mathrm{TJ}_{t-22:t,\alpha}+\varepsilon_t \ ,$$

where $\mathrm{TJ}_{t,\alpha} := \left(\mathrm{RV}_t - \mathrm{TBPV}_t^{\mathrm{IV}}\right) \cdot \mathbb{1}_{\text{C-Z}_t > \Phi^{-1}(\alpha)}$,

$$\mathrm{TC}_{t,\alpha} := \mathrm{RV}_t - \mathrm{TJ}_{t,\alpha} = \mathrm{RV}_t \cdot \mathbb{1}_{\text{C-Z}_t \leq \Phi^{-1}(\alpha)} + \mathrm{TBPV}_t^{\mathrm{IV}} \cdot \mathbb{1}_{\text{C-Z}_t > \Phi^{-1}(\alpha)} \ ,$$

and $\mathrm{TC}_{t-5:t,\alpha}$, $\mathrm{TC}_{t-22:t,\alpha}$, $\mathrm{TJ}_{t-5:t,\alpha}$ as well as $\mathrm{TJ}_{t-22:t,\alpha}$ defined as usual.

5.2 HAR models based on RV, SSpJ and SSnJ

All previous HAR models[23] can be regarded as regression models of the form

$$\widehat{\mathrm{QV}}_{t+1} = \beta_0 + \beta_d\widehat{\mathrm{QV}}_t + \beta_w\widehat{\mathrm{QV}}_{t-5:t} + \beta_m\widehat{\mathrm{QV}}_{t-22:t} + \varepsilon_t$$

or

$$\widehat{\mathrm{QV}}_{t+1} = \beta_0 + \beta_{cd}\widehat{\mathrm{IV}}_t + \beta_{cw}\widehat{\mathrm{IV}}_{t-5:t} + \beta_{cm}\widehat{\mathrm{IV}}_{t-22:t} + \beta_{jd}\widehat{\mathrm{SSJ}}_t + \beta_{jw}\widehat{\mathrm{SSJ}}_{t-5:t} + \beta_{jm}\widehat{\mathrm{SSJ}}_{t-22:t} + \varepsilon_t \ .$$

With other estimators for QV_t, IV_t, SSJ_t or potentially SSpJ_t and SSnJ_t at hand, further HAR-type models are easily constructed. While (daily) tests for the presence of jumps are used in most specifications, there is apparently no implicit need for such tests. Using the square-root or logarithmic transformation requires non-negative/positive explanatory variables, which is guaranteed by using one of the aforementioned tests for the presence of jumps (for $\alpha \leq 0.5$), but using only the positive part (as J_t in the HAR-RV-J model) or using other transformations, e.g. the *symmetric square root transformation*

$$\mathrm{ssqrt}: \mathbb{R} \to \mathbb{R}; \mathrm{ssqrt}(x) = \begin{cases} +\sqrt{x} & : x \geq 0 \\ -\sqrt{-x} & : x < 0 \end{cases}$$

or the *symmetric logarithm transformation*

$$\mathrm{slog}: \mathbb{R} \to \mathbb{R}; \mathrm{slog}(x) = \begin{cases} +\log(1+x) & : x \geq 0 \\ -\log(1-x) & : x < 0 \end{cases} ,$$

is also feasible.

We propose some new HAR-type models, which are based on the work of Barndorff-Nielsen

[22]Corsi et al. (2008) finally leave the weekly and monthly jump aggregates out of their model

[23]apart from the (intermediate) HAR-RV-J model

et al. (2008) and Klößner (2008):

- the HAR-RV-RS model:

$$\begin{aligned}\mathrm{RV}_{t:t+h} = \beta_0 &+ \beta_d \mathrm{BPV}_t^{\mathrm{IV}} + \beta_w \mathrm{BPV}_{t-5:t}^{\mathrm{IV}} + \beta_m \mathrm{BPV}_{t-22:t}^{\mathrm{IV}} \\ &+ \beta_{jd^+} \widehat{\mathrm{SSpJ}}_t^{\mathrm{RS}} + \beta_{jd^-} \widehat{\mathrm{SSnJ}}_t^{\mathrm{RS}} + \beta_{jw^+} \widehat{\mathrm{SSpJ}}_{t-5:t}^{\mathrm{RS}} + \beta_{jw^-} \widehat{\mathrm{SSnJ}}_{t-5:t}^{\mathrm{RS}} \\ &+ \beta_{jm^+} \widehat{\mathrm{SSpJ}}_{t-22:t}^{\mathrm{RS}} + \beta_{jm^-} \widehat{\mathrm{SSnJ}}_{t-22:t}^{\mathrm{RS}} + \varepsilon_t \end{aligned} \tag{1}$$

- the HAR-RV-HL model:

$$\begin{aligned}\mathrm{RV}_{t:t+h} = \beta_0 &+ \beta_d \widehat{\mathrm{IV}}_t^{(c)} + \beta_w \widehat{\mathrm{IV}}_{t-5:t}^{(c)} + \beta_m \widehat{\mathrm{IV}}_{t-22:t}^{(c)} \\ &+ \beta_{jd^+} \widehat{\mathrm{SSpJ}}_t^{(c)} + \beta_{jd^-} \widehat{\mathrm{SSnJ}}_t^{(c)} + \beta_{jw^+} \widehat{\mathrm{SSpJ}}_{t-5:t}^{(c)} + \beta_{jw^-} \widehat{\mathrm{SSnJ}}_{t-5:t}^{(c)} \\ &+ \beta_{jm^+} \widehat{\mathrm{SSpJ}}_{t-22:t}^{(c)} + \beta_{jm^-} \widehat{\mathrm{SSnJ}}_{t-22:t}^{(c)} + \varepsilon_t \end{aligned} \tag{2}$$

- the HAR-RV-HL-T model:

$$\begin{aligned}\mathrm{RV}_{t:t+h} = \beta_0 &+ \beta_d \widehat{C}_t^{\mathrm{hl}} + \beta_w \widehat{C}_{t-5:t}^{\mathrm{hl}} + \beta_m \widehat{C}_{t-22:t}^{\mathrm{hl}} \\ &+ \beta_{jd^+} \widehat{\mathrm{Jp}}_t^{\mathrm{hl}} + \beta_{jd^-} \widehat{\mathrm{Jn}}_t^{\mathrm{hl}} + \beta_{jw^+} \widehat{\mathrm{Jp}}_{t-5:t}^{\mathrm{hl}} + \beta_{jw^-} \widehat{\mathrm{Jn}}_{t-5:t}^{\mathrm{hl}} \\ &+ \beta_{jm^+} \widehat{\mathrm{Jp}}_{t-22:t}^{\mathrm{hl}} + \beta_{jm^-} \widehat{\mathrm{Jn}}_{t-22:t}^{\mathrm{hl}} + \varepsilon_t \ , \end{aligned} \tag{3}$$

  where $\widehat{\mathrm{Jp}}_t^{\mathrm{hl}} := \widehat{\mathrm{SSpJ}}_t^{(c)} \cdot \mathbb{1}_{\mathrm{TJp}_t^{\mathrm{hl}} > \Phi^{-1}(\alpha)}$, $\widehat{\mathrm{Jn}}_t^{\mathrm{hl}} := \widehat{\mathrm{SSnJ}}_t^{(c)} \cdot \mathbb{1}_{\mathrm{TJn}_t^{\mathrm{hl}} > \Phi^{-1}(\alpha)}$,

  $$\widehat{C}_t^{\mathrm{hl}} := \widehat{\mathrm{QV}}_t^{(c)} - \widehat{\mathrm{Jp}}_t^{\mathrm{hl}} - \widehat{\mathrm{Jn}}_t^{\mathrm{hl}} = \widehat{\mathrm{IV}}_t^{(c)} + \widehat{\mathrm{SSpJ}}_t^{(c)} \cdot \mathbb{1}_{\mathrm{TJp}_t^{\mathrm{hl}} \le \Phi^{-1}(\alpha)} + \widehat{\mathrm{SSnJ}}_t^{(c)} \cdot \mathbb{1}_{\mathrm{TJn}_t^{\mathrm{hl}} \le \Phi^{-1}(\alpha)} \ ,$$

  and $\widehat{C}_{t-5:t}^{\mathrm{hl}}$, $\widehat{C}_{t-22:t}^{\mathrm{hl}}$, $\widehat{\mathrm{Jp}}_{t-5:t}^{\mathrm{hl}}$, $\widehat{\mathrm{Jn}}_{t-5:t}^{\mathrm{hl}}$, $\widehat{\mathrm{Jp}}_{t-22:t}^{\mathrm{hl}}$ as well as $\widehat{\mathrm{Jn}}_{t-22:t}^{\mathrm{hl}}$ defined as usual.

It is important to note that OHLC data is required for the estimation of HAR-RV-HL and HAR-RV-HL-T, whereas HAR-RV-RS does not rely on high and low quotes.
Of course, many more HAR-type model specifications are imaginable. In the next section we compare the aforementioned HAR models in the spirit of Corsi et al. (2008), i.e., we estimate all models (with Newey-West robust standard errors) and calculate[24] the (out-of-sample) Mincer-Zarnowitz-R^2 and the out-of-sample relative RMSE of the prediction $\widehat{\mathrm{RV}}_t^{1/2}$ of $\mathrm{RV}_t^{1/2}$ (based

[24]The out-of-sample Mincer-Zarnowitz-R^2 of a forecasting model is the R^2 of the regression $y_t = \alpha + \beta \widehat{y}_t + \varepsilon_t$, $t = t_0, \ldots, T$, where $\widehat{y}_t$ is the forecast of y_t based on the information up to time $t-1$.

on information up to time $t-1$), defined by

$$\text{RMSE} = \frac{1}{T-t_0+1}\left(\sum_{t=t_0}^{T}\left(\frac{\text{RV}_t^{\frac{1}{2}} - \widehat{\text{RV}}_t^{\frac{1}{2}}}{\text{RV}_t^{\frac{1}{2}}}\right)^2\right)^{\frac{1}{2}} .$$

We chose $t_0 = 10$ for all models.

6 Empirical Results

We apply the HAR models of the previous section to various intradaily OHLC data with interval length of 5 minutes.[25] The data set consists of

- stock indices: Dow Jones Industrial Average, NASDAQ Composite, NASDAQ 100, S&P 100, S&P 500,
- foreign exchange markets: AUD/USD, CHF/USD, EUR/USD, GBP/USD, JPY/USD,
- futures: Crude Oil (01/2009), E-Mini S&P 500 (12/2008), Gold (02/2009), E-Mini NASDAQ 100, S&P 500 (12/2008)

for the time range from 2000-01-01 up to 2007-12-31.[26]
We varied the HAR model parameters as follows:

- The significance level of all tests is set to $\alpha = 0.001$.
- Forecast horizons of $h = 1$ day and $h = 5$ days.
- Single jump component (1 day) and three jump components (aggregation of 1, 5, 22 days)
- Different transformations: none, sqrt, ssqrt, log, slog

The individual estimation results are available in a separate appendix. We present (highly) aggregated results here, consisting of

1. mean values of Mincer-Zarnowitz-R^2, RMSE, and mean percentage of days with significant jumps (only for models featuring tests for the presence of jumps),
2. mean ranks (over the different HAR models) of Mincer-Zarnowitz-R^2 and RMSE[27],

[25] The log- and slog-transformations are of course not scale-independent. Following Andersen et al. (2007) and Corsi et al. (2008), the data has been annualized, i.e. all log-prices are multiplied by 250.
[26] Data source: disktrading.com
[27] for RMSE, lower ranks are 'better'

3. mean percentages from maximum (over the different HAR models) of Mincer-Zarnowitz-R^2 and RMSE[28],

in Tables 4 – 6 (overall) and Tables 7 – 18 (for different values of h and different jump components). The values in Tables 4 – 18 are printed in greyscales to improve the illustration of the main results: the darker the numbers, the better the result, i.e. for R^2, bigger values and ranks are darker, for RMSE, smaller values and ranks are darker, resp.

In parts, our results for HAR-RV, HAR-RV-J, HAR-RV-CJ, and HAR-RV-TCJ are in line with the results of Andersen et al. (2007) and Corsi et al. (2008). Particularly, the forecasts for $h = 5$ outperform[29] the forecasts for $h = 1$. Furthermore, adding the aggregated jump components for 5 and 22 days as explanatory variables does not result in a general improvement of the predictions. As observed in Andersen et al. (2007), the forecasts of the HAR-RV-CJ model are often superior to the HAR-RV and HAR-RV-J, especially w.r.t. the (out-of-sample) Mincer-Zarnowitz-R^2 for the (s)sqrt and (s)log transformations. In contrast to Corsi et al. (2008), the advantages of the HAR-RV-TCJ model forecasts seem to be less pronounced. While the aggregated results indicate superiority of the HAR-RV-TCJ model w.r.t. R^2 in the untransformed and (s)sqrt tranformed models, HAR-RV-CJ often performs better w.r.t. to RMSE and other transformations.

Unlike the HAR-RV-J, HAR-RV-CJ, and HAR-RV-TCJ model, the newly proposed HAR-RV-RS model discriminates between (the daily sums of) positive and negative (squared) jumps. The question whether this individual treatment of the positive and negative jump component leads to improved forecasts of realized variance can be answered positively. The aggregated results indicate that the HAR-RV-RS model outperforms all other models w.r.t. RMSE for all transformations. HAR-RV, HAR-RV-J, and HAR-RV-CJ are inferior to HAR-RV-RS for all transformations w.r.t. to R^2, only for the untransformed and the (s)sqrt transformed models, HAR-RV-TCJ outperforms HAR-RV-RS w.r.t. R^2.

It is remarkable, that the HAR-RV-RS model specification does not include tests for the presence of jumps and thus does not rely on choosing significance levels or other tuning parameters. In contrast to the estimation of the HAR-RV-TCJ model, which involves iterative estimation of the local volatility, the computing costs for the HAR-RV-RS model estimation are comparable to the HAR-RV, HAR-RV-J and HAR-RV-CJ models. Aside from the improved volatility predictions, the HAR-RV-RS model estimations[30] reveal interesting differences between the effect of positive and negative jumps. We omit a detailed evaluation of the individual results, but in summary, for futures and indices the effect of positive jumps on future realized variances if often (significantly) negative, whereas negative jumps often have a (significantly) positive effect on future volatility. Furthermore, the estimated individual effects of positive and negative jumps are sometimes (both) significant although the (combined) effect of arbitrary jumps

[28] for RMSE, lower percentages are 'better'

[29] concerning Mincer-Zarnowitz-R^2 and RMSE

[30] see the separate appendix

(in other models) is not.
A comparison of the aforementioned models with the HAR-RV-HL and the HAR-RV-HL-T models is difficult, because the regressors in these models are estimated with OHLC data whereas the explained realized variance is estimated without high and low price data. Hence, the bad forecasting performance is not surprising. It is nevertheless remarkable that the test for the presence of jumps indicates jumps far more often than the test in the HAR-RV-CJ and HAR-RV-TCJ model specifications.

trafo	measure	RV	RV-J	RV-CJ	RV-TCJ	RV-RS	RV-HL	RV-HL-T
none	R^2	0.529989	0.525758	0.519488	0.534523	0.525985	0.495167	0.533772
	RMSE	0.006142	0.006074	0.006056	0.005967	0.005818	0.006347	0.006025
sqrt	R^2	0.652589	0.651210	0.654986	0.657465	0.656328	0.616291	0.654051
	RMSE	0.005129	0.005106	0.005055	0.005042	0.005016	0.005509	0.005224
ssqrt	R^2	0.652589	0.651210	0.654986	0.657465	0.655897	0.617097	0.654051
	RMSE	0.005129	0.005106	0.005055	0.005042	0.005013	0.005476	0.005224
log	R^2	0.686785	0.686954	0.688914	0.687309	0.692535	0.650302	0.685440
	RMSE	0.004811	0.004803	0.004934	0.004821	0.004739	0.005283	0.004981
slog	R^2	0.687756	0.688307	0.690208	0.690392	0.692443	0.649366	0.687347
	RMSE	0.004878	0.004859	0.004945	0.004853	0.004794	0.005289	0.005039
	sign. (%)			18.158130	21.543148			27.217662

Table 4: Overall mean values, $\alpha = 0.001$

trafo	measure	RV	RV-J	RV-CJ	RV-TCJ	RV-RS	RV-HL	RV-HL-T
none	R^2	4.350000	3.733333	4.250000	4.433333	3.633333	2.516667	5.083333
	RMSE	4.366667	4.033333	4.150000	3.583333	3.033333	5.100000	3.733333
sqrt	R^2	3.816667	3.283333	4.366667	5.166667	4.900000	2.050000	4.416667
	RMSE	4.133333	4.366667	3.366667	2.750000	2.800000	6.016667	4.566667
ssqrt	R^2	3.833333	3.283333	4.383333	5.216667	4.783333	2.033333	4.466667
	RMSE	4.150000	4.416667	3.366667	2.800000	2.733333	5.916667	4.616667
log	R^2	4.216667	4.133333	4.750000	3.800000	5.500000	2.016667	3.583333
	RMSE	3.333333	3.700000	3.350000	3.666667	2.616667	6.383333	4.950000
slog	R^2	3.933333	3.833333	4.716667	4.566667	5.133333	2.016667	3.800000
	RMSE	3.633333	3.750000	3.283333	3.433333	2.816667	6.200000	4.883333

Table 5: Overall mean ranks, $\alpha = 0.001$

trafo	measure	RV	RV-J	RV-CJ	RV-TCJ	RV-RS	RV-HL	RV-HL-T
none	R^2	96.045454	95.626866	94.478417	97.434294	95.211900	88.565801	96.698926
	RMSE	91.946696	91.015770	90.789974	89.708710	87.877302	96.786887	90.459391
sqrt	R^2	98.189841	97.977013	98.576553	98.984977	98.553082	91.365133	98.442905
	RMSE	91.108112	90.787405	89.925249	89.673120	89.297185	98.864271	92.603315
ssqrt	R^2	98.208220	97.995336	98.595003	99.003486	98.522165	91.610646	98.461377
	RMSE	91.575073	91.252490	90.389478	90.134360	89.700385	98.781680	93.070171
log	R^2	98.683927	98.697202	98.983713	98.775195	99.405854	92.340304	98.531473
	RMSE	87.964573	87.868364	88.737187	88.172564	86.723161	97.411926	90.716289
slog	R^2	98.685168	98.756465	99.041307	99.097364	99.236768	91.904095	98.673849
	RMSE	89.673364	89.402869	90.000231	89.394006	88.287916	97.906254	92.273564

Table 6: Overall mean percentages of row-wise maxima, $\alpha = 0.001$

trafo	measure	RV	RV-J	RV-CJ	RV-TCJ	RV-RS	RV-HL	RV-HL-T
none	R^2	0.466302	0.470205	0.451151	0.481667	0.477007	0.440519	0.476296
	RMSE	0.007020	0.006959	0.006894	0.006798	0.006753	0.007381	0.006937
sqrt	R^2	0.605783	0.607945	0.608266	0.614308	0.614824	0.576847	0.608086
	RMSE	0.006047	0.005998	0.005948	0.005929	0.005927	0.006566	0.006199
ssqrt	R^2	0.605783	0.607945	0.608266	0.614308	0.614426	0.576265	0.608086
	RMSE	0.006047	0.005998	0.005948	0.005929	0.005927	0.006509	0.006199
log	R^2	0.636229	0.638015	0.638685	0.638471	0.643639	0.604106	0.634123
	RMSE	0.005687	0.005647	0.005846	0.005633	0.005602	0.006271	0.005935
slog	R^2	0.642382	0.644867	0.645196	0.647332	0.649806	0.608433	0.641590
	RMSE	0.005781	0.005730	0.005870	0.005708	0.005687	0.006341	0.006024
	sign. (%)			18.158130	21.543148			27.217662

Table 7: Mean values, $h = 1$, jump comp.: 1, $\alpha = 0.001$

trafo	measure	RV	RV-J	RV-CJ	RV-TCJ	RV-RS	RV-HL	RV-HL-T
none	R^2	3.866667	4.066667	4.066667	4.533333	3.933333	2.200000	5.333333
	RMSE	4.333333	3.933333	3.666667	3.133333	3.066667	5.733333	4.133333
sqrt	R^2	3.200000	3.266667	4.266667	5.333333	5.600000	2.000000	4.333333
	RMSE	4.466667	4.200000	3.000000	2.333333	2.733333	6.400000	4.866667
ssqrt	R^2	3.200000	3.266667	4.266667	5.400000	5.533333	2.000000	4.333333
	RMSE	4.533333	4.200000	3.000000	2.333333	2.733333	6.266667	4.933333
log	R^2	4.200000	4.666667	4.800000	3.733333	5.800000	1.666667	3.133333
	RMSE	3.733333	3.133333	3.533333	3.533333	2.400000	6.666667	5.000000
slog	R^2	3.733333	4.400000	4.733333	4.466667	5.466667	1.666667	3.533333
	RMSE	3.933333	3.133333	3.266667	3.400000	2.600000	6.666667	5.000000

Table 8: Mean ranks, $h = 1$, jump comp.: 1, $\alpha = 0.001$

trafo	measure	RV	RV-J	RV-CJ	RV-TCJ	RV-RS	RV-HL	RV-HL-T
none	R^2	93.852213	95.122449	91.818835	97.760073	96.277012	87.835491	95.693124
	RMSE	91.610151	90.783829	90.035955	88.919526	88.559696	97.871086	90.774669
sqrt	R^2	97.574262	97.906973	98.037781	99.035511	98.829004	91.549237	97.990989
	RMSE	90.483290	89.760839	89.065248	88.680024	88.749774	99.119946	92.611495
ssqrt	R^2	97.632421	97.965075	98.096155	99.094019	98.846879	91.525877	98.049550
	RMSE	91.370260	90.643266	89.946073	89.553171	89.632341	99.157345	93.497234
log	R^2	98.398739	98.657960	98.790574	98.771346	99.438485	92.245223	98.175103
	RMSE	87.577762	87.004381	89.148465	86.763738	86.352341	97.334971	91.026672
slog	R^2	98.298810	98.666982	98.755943	99.110756	99.319519	91.745879	98.280633
	RMSE	88.982045	88.244475	89.870667	87.884887	87.646569	98.174464	92.361293

Table 9: Mean percentages of row-wise maxima, $h = 1$, jump comp.: 1, $\alpha = 0.001$

trafo	measure	RV	RV-J	RV-CJ	RV-TCJ	RV-RS	RV-HL	RV-HL-T
none	R^2	0.466302	0.452846	0.441808	0.473351	0.455166	0.427561	0.465497
	RMSE	0.007020	0.006943	0.006956	0.006810	0.006581	0.006877	0.006722
sqrt	R^2	0.605783	0.598345	0.604246	0.610592	0.604477	0.565282	0.608128
	RMSE	0.006047	0.006031	0.005993	0.005944	0.005917	0.006207	0.006006
ssqrt	R^2	0.605783	0.598345	0.604246	0.610592	0.603747	0.573585	0.608128
	RMSE	0.006047	0.006031	0.005993	0.005944	0.005912	0.006178	0.006006
log	R^2	0.636229	0.633434	0.635094	0.636724	0.639563	0.599369	0.636276
	RMSE	0.005687	0.005689	0.006113	0.005653	0.005638	0.006076	0.005781
slog	R^2	0.642382	0.639631	0.641302	0.644997	0.644300	0.609292	0.643610
	RMSE	0.005781	0.005771	0.006074	0.005725	0.005702	0.006033	0.005852
	sign. (%)			18.158130	21.543148			27.217662

Table 10: Mean values, $h = 1$, jump comp.: 1, 5, 22, $\alpha = 0.001$

trafo	measure	RV	RV-J	RV-CJ	RV-TCJ	RV-RS	RV-HL	RV-HL-T
none	R^2	4.933333	3.600000	4.066667	5.066667	2.933333	2.333333	5.066667
	RMSE	4.666667	4.266667	5.066667	3.800000	3.133333	3.800000	3.266667
sqrt	R^2	4.266667	2.800000	3.866667	5.600000	4.200000	2.200000	5.066667
	RMSE	4.333333	5.000000	4.266667	2.200000	3.000000	5.133333	4.066667
ssqrt	R^2	4.266667	2.800000	3.933333	5.666667	4.066667	2.133333	5.133333
	RMSE	4.200000	5.066667	4.200000	2.133333	3.000000	5.333333	4.066667
log	R^2	4.666667	3.466667	3.800000	4.266667	5.000000	2.466667	4.333333
	RMSE	3.000000	3.800000	4.200000	3.266667	3.333333	6.000000	4.400000
slog	R^2	4.333333	3.133333	3.866667	4.800000	4.866667	2.400000	4.600000
	RMSE	3.600000	4.200000	4.133333	2.600000	3.333333	5.600000	4.533333

Table 11: Mean ranks, $h = 1$, jump comp.: 1, 5, 22, $\alpha = 0.001$

trafo	measure	RV	RV-J	RV-CJ	RV-TCJ	RV-RS	RV-HL	RV-HL-T
none	R^2	95.611331	93.578045	91.366226	97.751403	92.862201	86.164886	95.429420
	RMSE	95.825862	94.893542	95.056696	93.290687	90.952981	95.275634	92.420344
sqrt	R^2	98.394479	97.143252	98.116676	99.153534	97.713205	89.800114	98.730444
	RMSE	95.189636	94.983920	94.437025	93.606302	93.421900	98.415623	94.594044
ssqrt	R^2	98.402791	97.151399	98.124852	99.161741	97.635778	91.496679	98.738744
	RMSE	95.516617	95.310495	94.764541	93.929913	93.659437	98.287875	94.920801
log	R^2	98.811088	98.326926	98.566346	98.869632	99.073356	91.490750	98.794188
	RMSE	89.566629	89.653355	92.954102	89.121855	88.965575	96.376134	90.904869
slog	R^2	98.905265	98.419911	98.673751	99.292551	98.901242	92.137237	99.070843
	RMSE	92.296370	92.203857	94.881338	91.507869	91.286998	96.764946	93.321610

Table 12: Mean percentages of row-wise maxima, $h = 1$, jump comp.: 1, 5, 22, $\alpha = 0.001$

trafo	measure	RV	RV-J	RV-CJ	RV-TCJ	RV-RS	RV-HL	RV-HL-T
none	R^2	0.593676	0.592219	0.594548	0.597223	0.585125	0.554292	0.594876
	RMSE	0.005265	0.005234	0.005168	0.005133	0.005144	0.005870	0.005438
sqrt	R^2	0.699396	0.700193	0.703862	0.704083	0.704106	0.657525	0.697268
	RMSE	0.004210	0.004191	0.004125	0.004129	0.004110	0.004853	0.004480
ssqrt	R^2	0.699396	0.700193	0.703862	0.704083	0.703716	0.656168	0.697268
	RMSE	0.004210	0.004191	0.004125	0.004129	0.004111	0.004806	0.004480
log	R^2	0.737341	0.738029	0.741014	0.737294	0.742895	0.693856	0.732779
	RMSE	0.003936	0.003921	0.003879	0.003896	0.003844	0.004537	0.004214
slog	R^2	0.733129	0.734314	0.737211	0.735551	0.737878	0.685796	0.729332
	RMSE	0.003975	0.003954	0.003909	0.003921	0.003886	0.004568	0.004256
	sign. (%)			18.158130	21.543148			27.217662

Table 13: Mean values, $h = 5$, jump comp.: 1, $\alpha = 0.001$

trafo	measure	RV	RV-J	RV-CJ	RV-TCJ	RV-RS	RV-HL	RV-HL-T
none	R^2	4.400000	3.666667	4.600000	4.266667	3.600000	2.600000	4.866667
	RMSE	3.866667	3.666667	3.333333	3.533333	3.266667	6.066667	4.266667
sqrt	R^2	3.733333	3.533333	5.066667	4.933333	5.533333	1.400000	3.800000
	RMSE	3.733333	3.933333	2.733333	3.400000	2.800000	6.666667	4.733333
ssqrt	R^2	3.800000	3.466667	5.066667	5.000000	5.266667	1.466667	3.933333
	RMSE	3.733333	3.933333	2.733333	3.400000	2.800000	6.600000	4.800000
log	R^2	4.266667	4.066667	5.400000	3.733333	5.866667	1.400000	3.266667
	RMSE	3.533333	3.666667	2.666667	3.733333	2.266667	6.666667	5.466667
slog	R^2	4.000000	3.800000	5.400000	4.533333	5.466667	1.400000	3.400000
	RMSE	3.533333	3.533333	2.600000	3.733333	2.666667	6.733333	5.200000

Table 14: Mean ranks, $h = 5$, jump comp.: 1, $\alpha = 0.001$

trafo	measure	RV	RV-J	RV-CJ	RV-TCJ	RV-RS	RV-HL	RV-HL-T
none	R^2	97.898199	97.766342	98.157788	98.676780	96.404826	90.523942	98.136914
	RMSE	86.506547	85.998497	85.021699	84.694268	84.769031	97.984173	88.457947
sqrt	R^2	98.589916	98.722308	99.275642	99.311989	99.208294	91.725704	98.391467
	RMSE	85.246330	84.915081	83.710159	83.746617	83.399258	98.926439	89.619487
ssqrt	R^2	98.621445	98.753756	99.307192	99.343539	99.193497	91.524713	98.422783
	RMSE	86.225032	85.893482	84.684289	84.717544	84.400225	99.006650	90.599328
log	R^2	98.998022	99.098012	99.529973	98.987975	99.725167	92.372712	98.478339
	RMSE	84.901445	84.616922	83.781083	84.122281	83.070629	98.533025	89.865020
slog	R^2	98.961966	99.138082	99.557865	99.306530	99.584145	91.579767	98.553132
	RMSE	85.576710	85.180383	84.331435	84.549524	83.854842	98.850322	90.526258

Table 15: Mean percentages of row-wise maxima, $h = 5$, jump comp.: 1, $\alpha = 0.001$

trafo	measure	RV	RV-J	RV-CJ	RV-TCJ	RV-RS	RV-HL	RV-HL-T
none	R^2	0.593676	0.587760	0.590446	0.585853	0.586643	0.558297	0.598421
	RMSE	0.005265	0.005160	0.005207	0.005127	0.004794	0.005260	0.005005
sqrt	R^2	0.699396	0.698359	0.703571	0.700877	0.701906	0.665510	0.702723
	RMSE	0.004210	0.004205	0.004155	0.004165	0.004109	0.004410	0.004212
ssqrt	R^2	0.699396	0.698359	0.703571	0.700877	0.701700	0.662372	0.702723
	RMSE	0.004210	0.004205	0.004155	0.004165	0.004100	0.004410	0.004212
log	R^2	0.737341	0.738338	0.740863	0.736746	0.744043	0.703878	0.738580
	RMSE	0.003936	0.003955	0.003898	0.004104	0.003874	0.004247	0.003996
slog	R^2	0.733129	0.734416	0.737124	0.733690	0.737786	0.693942	0.734857
	RMSE	0.003975	0.003980	0.003928	0.004060	0.003901	0.004211	0.004025
	sign. (%)			18.158130	21.543148			27.217662

Table 16: Mean values, $h = 5$, jump comp.: 1, 5, 22, $\alpha = 0.001$

trafo	measure	RV	RV-J	RV-CJ	RV-TCJ	RV-RS	RV-HL	RV-HL-T
none	R^2	4.200000	3.600000	4.266667	3.866667	4.066667	2.933333	5.066667
	RMSE	4.600000	4.266667	4.533333	3.866667	2.666667	4.800000	3.266667
sqrt	R^2	4.066667	3.533333	4.266667	4.800000	4.266667	2.600000	4.466667
	RMSE	4.000000	4.333333	3.466667	3.066667	2.666667	5.866667	4.600000
ssqrt	R^2	4.066667	3.600000	4.266667	4.800000	4.266667	2.533333	4.466667
	RMSE	4.133333	4.466667	3.533333	3.333333	2.400000	5.466667	4.666667
log	R^2	3.733333	4.333333	5.000000	3.466667	5.333333	2.533333	3.600000
	RMSE	3.066667	4.200000	3.000000	4.133333	2.466667	6.200000	4.933333
slog	R^2	3.666667	4.000000	4.866667	4.466667	4.733333	2.600000	3.666667
	RMSE	3.466667	4.133333	3.133333	4.000000	2.666667	5.800000	4.800000

Table 17: Mean ranks, $h = 5$, jump comp.: 1, 5, 22, $\alpha = 0.001$

trafo	measure	RV	RV-J	RV-CJ	RV-TCJ	RV-RS	RV-HL	RV-HL-T
none	R^2	96.820071	96.040628	96.570819	95.548919	95.303561	89.738883	97.536248
	RMSE	93.844226	92.387210	93.045547	91.930360	87.227498	96.016657	90.184605
sqrt	R^2	98.200708	98.135519	98.876115	98.438874	98.461824	92.385478	98.658720
	RMSE	93.513194	93.489781	92.488565	92.659536	91.617810	98.995075	93.588236
ssqrt	R^2	98.176224	98.111115	98.851813	98.414646	98.412508	91.895316	98.634433
	RMSE	93.188381	93.162716	92.163010	92.336813	91.109535	98.674852	93.263320
log	R^2	98.527858	98.705910	99.047959	98.471825	99.386408	93.252530	98.678263
	RMSE	89.812458	90.198799	89.065098	92.682381	88.504099	97.403575	91.068596
slog	R^2	98.574631	98.800885	99.177670	98.679620	99.142167	92.153496	98.790790
	RMSE	91.838330	91.982763	90.917483	93.633742	90.363257	97.835283	92.885093

Table 18: Mean percentages of row-wise maxima, $h = 5$, jump comp.: 1, 5, 22, $\alpha = 0.001$

7 Conclusion

Andersen et al. (2007) and Corsi et al. (2008) already illustrate that dividing quadratic variation into integrated volatility and the sum of squared jumps yields a substantial improvement in volatility forecasting. This paper shows that distiguishing positive and negative jumps by using realized semi-variances leads to further improvements for the prediction of realized variance without increasing computational costs or data requirements.

Extending the data source to intradaily open, high, low, and close quotes and using OHLC-based estimators for integrated volatility and the sum of squared positive and negative jumps of Klößner (2008) for the prediction of realized variance does not lead to better forecasts. However, the comparison seems quite unfair, because in these models, the regressors are very different from the explained quantity, realized variance.

Instead of the prediction of realized variance, it would be interesting to build univariate forecasting models for OHLC-based estimators of quadratic variation, or even better, multivariate models for integrated volatility and the sum of squared positive and negative jumps. Forecasting models for related quantities, such as the relative contribution of positive and negative jumps to the total price variation (similar to the investigations in Huang and Tauchen (2005)) based on the estimators of Klößner (2008) provide another interesting branch for future work.

References

Andersen, T. G. and T. Bollerslev (1998) Answering the Skeptics: Yes, Standard Volatility Models Do Provide Accurate Forecasts, *International Economic Review*, 39(4), pp. 885–905.

Andersen, T. G., T. Bollerslev, and F. X. Diebold (2007) Roughing It Up: Including Jump Components in the Measurement, Modeling and Forecasting of Return Volatility, *The Review of Economics and Statistics*, 89(4), pp. 701–720.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys (2001) The Distribution of Realized Exchange Rate Volatility, *Journal of the American Statistical Asscociation*, 96, pp. 42–55.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys (2003) Modeling and Forecasting Realized Volatility, *Econometrica*, 71(2), pp. 579–625.

Andersen, T. G., T. Bollerslev, and N. Meddahi (2004) Analytic Evaluation of Volatility Forecasts, *International Economic Review*, 45(4), pp. 1079–1110.

Barndorff-Nielsen, O. E., S. Kinnebrock, and N. Shephard (2008) Measuring downside risk — realised semivariance, Working paper.

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

Barndorff-Nielsen, O. E. and N. Shephard (2002b) Estimating Quadratic Variation Using Realized Variance, *Journal of Applied Econometrics*, 17(5), pp. 457–477.

Barndorff-Nielsen, O. E. and N. Shephard (2004) Power and Bipower Variation with Stochastic Volatility and Jumps, *Journal of Financial Econometrics*, 2(1), pp. 1–37.

Barndorff-Nielsen, O. E. and N. Shephard (2006) Econometrics of Testing for Jumps in Financial Econometrics Using Bipower Variation, *Journal of Financial Econometrics*, 4(1), pp. 1–30.

Barndorff-Nielsen, O. E., N. Shephard, and M. Winkel (2006) Limit theorems for multipower variation in the presence of jumps, *Stochastic Processes and their Applications*, 116(5), pp. 796–806.

Corsi, F. (2004) A Simple Long Memory Model of Realized Volatility, Manuscript, University of Southern Switzerland.

Corsi, F., D. Pirino, and R. Renò (2008) Volatility forecasting: the jumps do matter, Working paper, University of Siena.

Huang, X. and G. Tauchen (2005) The Relative Contribution of Jumps to Total Price Variance, *Journal of Financial Econometrics*, 3(4), pp. 456–499.

Klößner, S. (2008) Separating Risk due to Diffusion,Positive Jumps, and Negative Jumps, Working Paper.

Müller, U. A., M. M. Dacorogna, R. M. Davé, R. B. Olsen, O. V. Pictet, and J. E. von Weizsäcker (1997) Volatilities of different time resolutions - Analyzing the dynamics of market components, *Journal of Empirical Finance*, 4(2-3), pp. 213–239.

Rogers, L. C. G., S. E. Satchell, and Y. Yoon (1994) Estimating the volatility of stock prices: A comparison of methods that use high and low prices, *Applied Financial Economics*, 4(3), pp. 241–247.