

Multivariate volatility forecasts for stock market indices

Citation for published version (APA):

Wilms, I., Rombouts, J., & Croux, C. (2021). Multivariate volatility forecasts for stock market indices. *International Journal of Forecasting*, 37(2), 484-499. <https://doi.org/10.1016/j.ijforecast.2020.06.012>

Document status and date:

Published: 01/04/2021

DOI:

[10.1016/j.ijforecast.2020.06.012](https://doi.org/10.1016/j.ijforecast.2020.06.012)

Document Version:

Publisher's PDF, also known as Version of record

Document license:

Taverne

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain.
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.



Multivariate volatility forecasts for stock market indices

Ines Wilms ^{a,*}, Jeroen Rombouts ^b, Christophe Croux ^c



^a Department of Quantitative Economics, Maastricht University, The Netherlands

^b ESSEC Business School, France

^c EDHEC Business School, France

ARTICLE INFO

Keywords:

International stock markets

Lasso

Option-implied variance

Realized variance

Volatility spillover

ABSTRACT

Volatility forecasts aim to measure future risk and they are key inputs for financial analysis. In this study, we forecast the realized variance as an observable measure of volatility for several major international stock market indices and accounted for the different predictive information present in jump, continuous, and option-implied variance components. We allowed for volatility spillovers in different stock markets by using a *multivariate* modeling approach. We used heterogeneous autoregressive (HAR)-type models to obtain the forecasts. Based an out-of-sample forecast study, we show that: (i) including option-implied variances in the HAR model substantially improves the forecast accuracy, (ii) lasso-based lag selection methods do not outperform the parsimonious day-week-month lag structure of the HAR model, and (iii) cross-market spillover effects embedded in the multivariate HAR model have long-term forecasting power.

© 2020 International Institute of Forecasters. Published by Elsevier B.V. All rights reserved.

1. Introduction

Volatility is a measure of the degree of variation in the returns of a financial asset. Volatility measures risk and the forecasting volatility is very important in finance. For instance, risk management, asset allocation, option pricing, and trading in financial markets require reliable volatility forecasts (e.g., see Bauwens et al., 2012). In particular, asset managers, such as pension funds or hedge funds, can make better adjustments to the risk-return trade-off in their portfolios when accurate volatility forecasts are available. Industry standards such as the value-at-risk imposed by financial regulators for measuring and managing market risk are highly dependent on accurate volatility forecasts.

The inherent problem that affects the production of volatility forecasts is that volatility is unobservable. Therefore, latent variable models such as stochastic volatility

models or generalized autoregressive conditional heteroskedasticity (GARCH) models are generally employed (Bollerslev, 2001). However, accurate nonparametric estimates of volatility have become available since the emergence of high-frequency data (Andersen, Bollerslev, Diebold, & Labys, 2001; Barndorff-Nielsen & Shephard, 2002). Realized variances are computed as the sum of the squared intraday returns for a particular day, thereby making volatility “observable”. Volatility forecast models that rely on high-frequency data are now standard (e.g., see Haugom et al. (2014), Hua and Manzan (2013)).

Within this class of high-frequency-based models, the heterogeneous autoregressive (HAR) model (Corsi, 2009) has become very popular because of its simplicity and good forecast performance. The HAR model is based on the idea of investor heterogeneity where trading groups in financial markets make their decisions over daily, weekly, and monthly horizons. Recently, extensions of the HAR model have been introduced to allow for potentially different predictive information in jump versus continuous volatility components (Andersen et al., 2007; Hua & Manzan, 2013), or to allow for time-varying attenuation bias effects (see the HARQ model proposed by

* Corresponding author.

E-mail addresses: i.wilms@maastrichtuniversity.nl (I. Wilms), rombouts@essec.edu (J. Rombouts), christophe.croux@edhec.edu (C. Croux).

Bollerslev et al., 2016), which provide better forecasting accuracy than the simple HAR in most cases. Furthermore, option-implied variances may be considered as additional predictors, because (Bekaert & Hoerova, 2014) showed that they contain information about future stock market volatility.

Recently, Audrino and Knaus (2016) and Audrino et al. (2017) compared the standard HAR model to a lasso-based method (Tibshirani, 1996). Lasso-based methods achieve very good forecast performance in macroeconomics (e.g., De Mol et al., 2008; Li & Chen, 2014; Nicholson et al., 2018; Smeekes & Wijler, 2018; Wilms & Croux, 2016) and in finance, e.g., Nazemi and Fabozzi (2018) predicted US corporate bond recovery rates. Tian et al. (2015) considered corporate bankruptcy forecasting, and Audrino et al. (2020) analyzed the impacts of sentiment and attention variables on stock market volatility. Audrino and Knaus (2016) and Audrino et al. (2017) found that the more general lag structure allowed in the lasso-based approach does not improve the out-of-sample forecast performance of the parsimonious day-week month lag structure embedded in the HAR model. Their results support those obtained in previous studies regarding the difficulty of outperforming the forecast performance of the HAR model.

Audrino and Knaus (2016) and Audrino et al. (2017), and Bekaert and Hoerova (2014) only studied univariate models, but we also consider multivariate models in the present study. Common reactions of investors, policy makers, or central banks to news about macroeconomic or financial variables are likely to induce volatility interdependence across stocks, industries, or markets (e.g., Baele, 2005; Bekaert et al., 2005; Bollerslev et al., 2014; Cubadda et al., 2017). These volatility interdependencies can be captured in multivariate models. For instance, Bubák et al. (2011) detected significant intra-regional volatility spillovers among Central European foreign exchange markets. Moreover, Taylor (2015) considered a multivariate version of the HAR model proposed by Corsi (2009) for the realized variances of American, European, and Asian stock markets, and showed that it delivered improved out-of-sample forecasts. Diebold and Yilmaz (2015) observed that volatility connectedness tends to increase when stock markets pass through a period of high uncertainty, such as financial crises.

In this study, we forecast the realized variances for several international stock market indices using HAR-type models. By providing a comprehensive empirical analysis, we aim to guide practitioners and applied researchers regarding the forecast variables, lag structures, and estimation models that should be selected to obtain reliable volatility forecasts. We addressed the following three questions. (i) To what extent do jump and continuous realized variance and/or the option-implied variance *forecast variables* improve the forecast accuracy compared with the standard HAR model comprising only realized variance forecast variables? (ii) Do lasso-based *lag structures* achieve better forecast accuracy compared with the standard day-week-month (dwm) lag structure embedded in the HAR model? (iii) Are volatility forecasts more accurate when spillover effects between different stock

market indices are considered in a *multivariate* modeling framework?

We performed our empirical study based on the daily log-realized variances of several international stock market indices, four American stock market indices, and six European stock market indices. By measuring the out-of-sample forecast accuracy in terms of the absolute forecast errors in a rolling window setup, we show the following. (i) Augmenting standard HAR-type models with option-implied variances generally yields more accurate forecasts, which we confirmed with Diebold–Mariano (DM) tests. The gains obtained by disentangling the realized variances into jump and continuous components are generally highly limited. (ii) The parsimonious dwm lag structure of the HAR model is typically preferred to more sophisticated lasso-based lag structures, but also for multivariate models and when option-implied variances are included in the model. (iii) Univariate models can outperform multivariate models for short forecast horizons (i.e., daily or weekly) but multivariate models typically produce more accurate forecasts for longer forecast horizons (i.e., monthly or bimonthly), thereby demonstrating that spillover effects have long-term forecasting power. Furthermore, network analysis highlighted the time-varying nature of significant volatility spillovers among the international stock market indices. In fact, we found that the benefits of a multivariate model can increase during crisis periods such as the Global Financial Crisis and the European Sovereign Debt Crisis.

The remainder of this paper is structured as follows. In Section 2, we discuss the univariate and multivariate forecast models. In Section 3, we present the empirical set-up employed for the comparison of the models, and we describe the data in Section 4. The out-of-sample forecast analysis is presented in Section 5. Cross-market spillover effects are discussed in Section 6, as well as the effects of crisis periods. In Section 7, we give our conclusions.

2. Modeling realized variances

2.1. Univariate forecast model

First, we review the HAR model, which is the benchmark model for daily log-realized variances. We then introduce the extended HAR-model incorporating option-implied variances and/or jumps.

Standard HAR model. Let $RV_t^{(i)}$ be the realized variance of stock market index i at day t observed for $t = 1, \dots, T$. We write the log-transformed realized variance (LRV) as a function of its own past values:

$$LRV_t^{(i)} = \beta_1^{(i)} LRV_{t-1}^{(i)} + \beta_2^{(i)} LRV_{t-2}^{(i)} + \dots + \beta_p^{(i)} LRV_{t-p}^{(i)} + \epsilon_t^{(i)}, \quad (1)$$

for $t = p + 1, \dots, T$, where $p = 22$. Following standard practice, the realized variance is log-transformed to ensure that their values remain on the whole real line. Furthermore, although the variances typically have right-skewed distributions, the log-realized variances tend to have more Gaussian distributions, which might make it easier to forecast them with linear models (e.g., Bekaert

& Hoerova, 2014). It should be noted that the number of trading days in a month is approximately 22. Without loss of generality, we assume that all time series are centered such that no intercept is included and the error terms $\epsilon_t^{(i)}$ are uncorrelated with mean zero and constant variance.

The standard HAR model is an autoregressive model of order 22 but it includes the following restrictions:

$$\beta_\ell^{(i)} = \begin{cases} \beta_d^{(i)} + \beta_w^{(i)}/5 + \beta_m^{(i)}/22 & \text{for } \ell = 1 \\ \beta_w^{(i)}/5 + \beta_m^{(i)}/22 & \text{for } 2 \leq \ell \leq 5 \\ \beta_m^{(i)}/22 & \text{for } 6 \leq \ell \leq 22. \end{cases} \quad (2)$$

The HAR model can then be written more compactly as

$$\text{LRV}_t^{(i)} = \beta_d^{(i)} \text{LRV}_{t-1}^{(i)} + \beta_w^{(i)} \text{LRV}_{t-1}^{(i),(5)} + \beta_m^{(i)} \text{LRV}_{t-1}^{(i),(22)} + \epsilon_t^{(i)} \quad (3)$$

with

$$\text{LRV}_{t-1}^{(i),(n)} = (\text{LRV}_{t-1}^{(i)} + \text{LRV}_{t-2}^{(i)} + \cdots + \text{LRV}_{t-n}^{(i)})/n,$$

as the average log-realized variance over the last n days for $n = 5$ and 22. Thus, the HAR model explains the expected log-realized variance as a linear function of the log-realized variances of yesterday and the average over last week and last month. In the following, we add several state-of-the-art extensions to the standard HAR model.

Incorporating option-implied variances. Option-implied variances contain information about future stock market volatility (Bekaert & Hoerova, 2014), so the HAR models are augmented with option-implied variances (IV). Using $\text{LIV}_t^{(i)}$ as the log-transformed option-implied variance of stock market index i at day t , we set

$$\text{LRV}_t^{(i)} = \sum_{\ell=1}^p \left(\beta_\ell^{(i)} \text{LRV}_{t-\ell}^{(i)} + \beta_{\text{IV},\ell}^{(i)} \text{LIV}_{t-\ell}^{(i)} \right) + \epsilon_t^{(i)}. \quad (4)$$

We have no a priori information about the duration of the effect of a stock market's IV on its corresponding realized variance. Therefore, we added all lags to the autoregressive realized variance structure of model (1) and investigated whether lasso-based methods can help to select a shorter or more appropriate lag length (see Section 3.2). Note that if $\beta_{\text{IV},\ell}^{(i)} = 0$ for all lags ℓ , then model (4) reduces to the model in (1).

Incorporating jumps. We follow the methods employed in previous studies of realized variance forecasting (e.g., Andersen et al., 2007; Hua & Manzan, 2013) to account for the different predictive information in jump and continuous volatility components.

Let $\text{TBPV}_t^{(i)}$ be the threshold bipower variation of stock market index i at day t , as defined by Corsi et al. (2010). This is a measure of the realized variance when not considering jumps in the process. We then define the daily jump component $J_t^{(i)}$ of the realized variance of stock market index i and the daily continuous component $C_t^{(i)}$ as described by Bekaert and Hoerova (2014):

$$J_t^{(i)} = \max(\text{RV}_t^{(i)} - \text{TBPV}_t^{(i)}, 0) \quad \text{and} \quad C_t^{(i)} = \text{RV}_t^{(i)} - J_t^{(i)}. \quad (5)$$

The model for the log-realized variance of stock market index i with jumps, continuous realized variance compo-

nents, and option-implied variances is written as:

$$\text{LRV}_t^{(i)} = \sum_{\ell=1}^p \left(\beta_{J,\ell}^{(i)} \text{LJ}_{t-\ell}^{(i)} + \beta_{C,\ell}^{(i)} \text{LC}_{t-\ell}^{(i)} + \beta_{\text{IV},\ell}^{(i)} \text{LIV}_{t-\ell}^{(i)} \right) + \epsilon_t^{(i)}, \quad (6)$$

for $t = p+1, \dots, T$, where we use the notation $\text{LJ}_t^{(i)} := \log(1 + J_t^{(i)})$ and $\text{LC}_t^{(i)} := \log C_t^{(i)}$.

2.2. Multivariate forecast model

To allow for volatility spillovers between different stock market indices, we consider the multivariate extension of model (6). This model allows for the effects of variance components (jump, continuous, or option-implied) of one stock market index on future values of the realized variance of another stock market. The multivariate model for the log-realized variances of q stock market indices is then given by

$$\left\{ \begin{array}{l} \text{LRV}_t^{(1)} = \sum_{j=1}^q \sum_{\ell=1}^p \left(\beta_{J,\ell}^{(1,j)} \text{LJ}_{t-\ell}^{(j)} + \beta_{C,\ell}^{(1,j)} \text{LC}_{t-\ell}^{(j)} \right. \\ \quad \left. + \beta_{\text{IV},\ell}^{(1,j)} \text{LIV}_{t-\ell}^{(j)} \right) + \epsilon_t^{(1)} \\ \text{LRV}_t^{(2)} = \sum_{j=1}^q \sum_{\ell=1}^p \left(\beta_{J,\ell}^{(2,j)} \text{LJ}_{t-\ell}^{(j)} + \beta_{C,\ell}^{(2,j)} \text{LC}_{t-\ell}^{(j)} \right. \\ \quad \left. + \beta_{\text{IV},\ell}^{(2,j)} \text{LIV}_{t-\ell}^{(j)} \right) + \epsilon_t^{(2)} \\ \vdots = \vdots \\ \text{LRV}_t^{(q)} = \sum_{j=1}^q \sum_{\ell=1}^p \left(\beta_{J,\ell}^{(q,j)} \text{LJ}_{t-\ell}^{(j)} + \beta_{C,\ell}^{(q,j)} \text{LC}_{t-\ell}^{(j)} \right. \\ \quad \left. + \beta_{\text{IV},\ell}^{(q,j)} \text{LIV}_{t-\ell}^{(j)} \right) + \epsilon_t^{(q)}. \end{array} \right. \quad (7)$$

We assume that the multivariate error time series is uncorrelated over time, with mean zero and constant covariance matrix.

If the true model is multivariate, then each individual series can still be described by a univariate autoregressive moving average model. To link the multivariate model to these univariate marginalizations, the final equation representation of vector autoregressions (see Zellner & Palm, 1974, or Chevillon et al., 2018 for a more recent discussion) can be used. Hence, univariate models can still achieve good forecast accuracy even if the true model is multivariate. However, the multivariate model allows us to consider cross-market spillover effects (see Section 6).

3. Design of the forecast study

In this section, we discuss the different forecast models used in our empirical study.

3.1. Set of forecast variables

To investigate whether models with jumps and continuous components of realized variances and/or option-implied variances improve the forecast accuracy compared with standard models that only include past realized variances, we consider four sets of forecast variables, as shown in Table 1.

Table 1

Forecast variables (rows) included in each of the four forecast variable sets (columns).

	Forecast variable sets			
	RV	JC	RV+IV	JC+IV
Realized Variance	x		x	
Jump and Continuous Component		x		x
Option-implied Variance		x		x

By comparing the predictor sets “RV+IV” and “JC+IV” with “RV” and “JC”, respectively, we investigated whether adding option-implied variance information could improve the forecast accuracy. In addition, by comparing the predictor sets “JC” and “JC+IV” with “RV” and “RV+IV”, respectively, we investigated whether disentangling the realized variances into jump and continuous components could improve the forecast accuracy.

3.2. Lag structure

To investigate whether lasso-based lag selection can yield more accurate forecasts compared with the dwm lag structure of the standard HAR model, we considered the following three types of lag structures.

(1) *dwm*. The first lag structure imposed the dwm-based restrictions given in Eq. (2) on all of the forecast variables. Adding these restrictions reduced the 22 autoregressive parameters in Eq. (1) to the three parameters comprising β_d , β_w , and β_m in Eq. (3). In the univariate model (6) with jump, continuous, and option-implied variances, these restrictions led to a reduction from $3 \times 22 = 66$ autoregressive parameters to $3 \times 3 = 9$ parameters. In the multivariate model (7), there was a reduction from $3 \times 22 \times q = 660$ parameters to $3 \times 3 \times q = 90$ parameters per equation, where $q = 10$ is the number of stock market indices. This reduction in the number of parameters has been empirically demonstrated to obtain a good balance between model flexibility and parameter uncertainty. In terms of forecasting performance, the standard HAR model is a highly competitive benchmark.

(2) *lag selection*. We started with the fully unrestricted model (6) or (7) for multivariate models and used the ordered lasso (Tibshirani & Suo, 2016) to automatically select the lag length for each forecast variable (realized variance, jump, continuous, and/or option-implied variance). Estimation noise often degrades the out-of-sample prediction forecast performance (Bekaert & Hoerova, 2014), so simple models may outperform the full models. After selecting the lag lengths, we estimated the resulting model by least squares to reduce the bias introduced by lasso-type estimators (e.g., Belloni & Chernozhukov, 2013).

We used the ordered lasso for lag selection instead of the standard lasso because the former favors simple models with low maximal lag orders rather than more complex models with high maximal lag order. We review the ordered lasso and provide implementation details in Appendix A.

(3) *dwm + lag selection*. The third lag structure combined the previous two structure. Previous studies (e.g., Andersen et al., 2007; Hua & Manzan, 2013) have demonstrated the excellent performance of the dwm lag structure for past realized variances (possibly disentangled into jump and continuous components), so we only applied the ordered lasso to the option-implied variance (retaining dwm for the other forecast variables).

3.3. Univariate versus multivariate

To investigate whether more accurate volatility forecasts can be obtained by considering spillover effects between different stock market indices, we compared univariate models (Section 2.1) with the corresponding multivariate models (Section 2.2).

In contrast to univariate models, multivariate models can capture the spillover between different stock market indices. For example, if we consider model (7) with jump, continuous, and implied variances, then $\beta_{j,\ell}^{(i,j)}$, $\beta_{C,\ell}^{(i,j)}$, $\beta_{IV,\ell}^{(i,j)}$, capture the effects of the lagged jump, continuous realized variances, and option-implied variances of stock market index j , respectively, on the log-realized variance of stock market index i . If i and j are different, then refer to them as “spillover” effects. It should be noted that we cannot give a causal interpretation of these spillover effects, but instead these effects need to be interpreted as the co-movements of one time series with the future values of another.

4. Stock market data

In this section, we discuss the American and European stock market data used to perform our empirical study. We used realized variances data for the following four American and six European stock market indices: (1) Dow Jones Industrial Average “DJIA”, (2) Nasdaq stock index “NASDAQ”, (3) Russell 2000 “RUSSELL”, (4) Standard & Poor’s 500 market index “SP500”, (5) Amsterdam Exchange Index “AEX”, (6) French stock index Cotation Assistée en Continu “CAC”, (7) German stock index Deutscher Aktienindex “DAX”, (8) stock index of the Eurozone “EUROSTOXX”, (9) United Kingdom Financial Times Stock Exchange Index 100 “FTSE”, and (10) Swiss market index “SMI”.

Daily realized variances based on five-minute returns were taken from the Oxford-Man Institute of Quantitative Finance (publicly available for download at <http://realized.oxford-man.ox.ac.uk/data/>). Using these realized variances and the threshold bipower variation measures, we constructed the daily jump and continuous components according to Eq. (5). Daily option-implied variances were taken from Thomson Reuters Eikon and they were computed following the standard approach developed by Bakshi et al. (2003) and used by Bollerslev et al. (2009) and Carr and Wu (2009). It should be noted that the option-implied variance for the S&P 500 index is the well-known VIX index.

Our data set covered $T = 3686$ trading days between January 2, 2004 and April 30, 2019 after removing non-overlapping trading days. We used the same time span

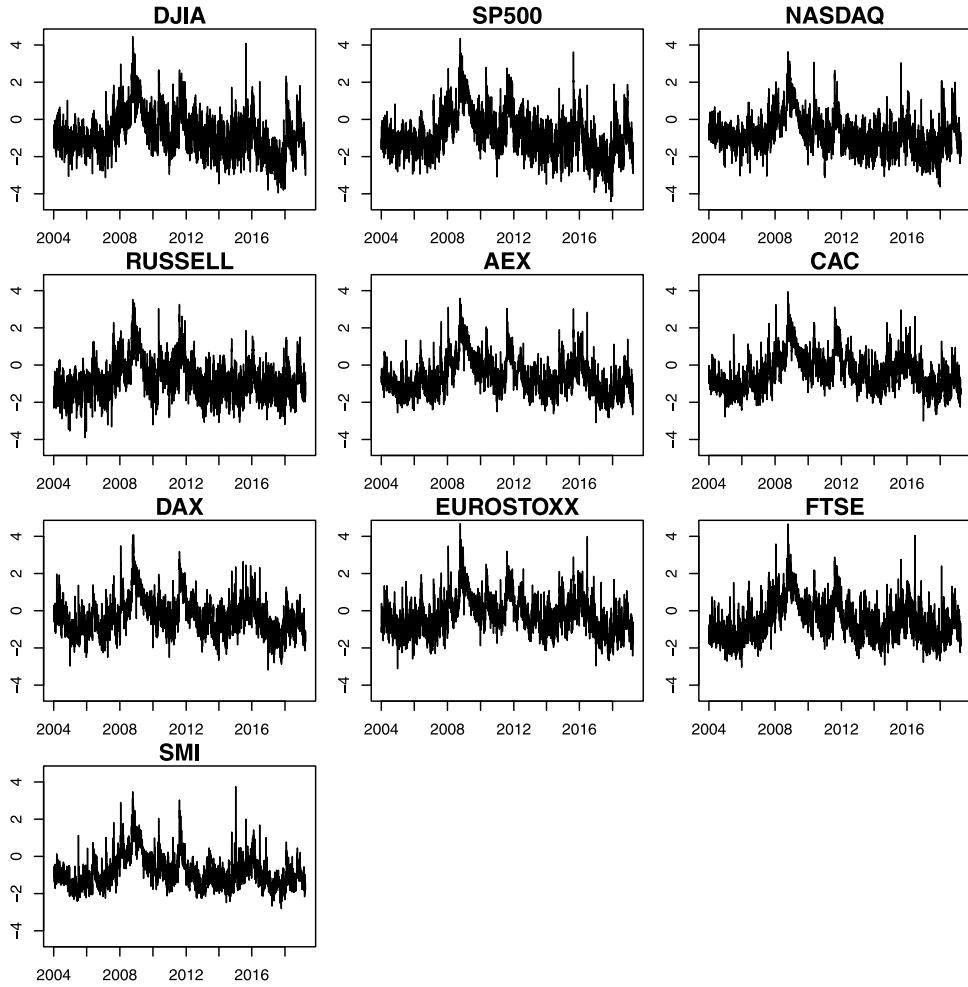


Fig. 1. Log-realized variances for the 10 American and European stock market indices from January 2004 to April 2019.

in our univariate and multivariate analyses to ensure the comparability of the results. We started in 2004 because the option-implied variance on the Russell 2000 is only available from that year on Thomson Reuters Eikon. Finally, the stock market indices considered are located in different time zones and this could potentially influence our empirical results, but (Demirer et al., 2018) noted that the use of volatilities rather than returns is likely to minimize this potential impact.

Fig. 1 plots the log-realized variances of the $q = 10$ stock market indices. Augmented Dickey–Fuller tests confirmed that the time series were stationary. Descriptive statistics are given in Table 2. All of the log-realized variances series were persistent according to the large values of their first order autocorrelations.

5. Forecast accuracy

We compared the forecast performance across the four forecast variable sets (Section 3.1), three lag structures (Section 3.2), and two modeling types, i.e., univariate or multivariate (Section 3.3), and thus we considered $4 \times 3 \times 2 = 24$ combinations in total.

Table 2

Descriptive statistics: average, standard deviation (St. Dev.), skewness, kurtosis, and first order autocorrelation for each log-realized variance time series.

Stock market	Average	St.Dev.	Skewness	Kurtosis	AutoCorr _{lag=1}
DJIA	-0.805	1.113	0.635	3.869	0.778
SP500	-0.830	1.144	0.617	3.729	0.818
NASDAQ	-0.748	0.957	0.657	3.938	0.812
RUSSELL	-0.792	0.968	0.701	3.918	0.755
AEX	-0.610	0.945	0.688	3.617	0.840
CAC	-0.385	0.930	0.602	3.524	0.830
DAX	-0.332	0.933	0.562	3.863	0.808
EUROSTOXX	-0.259	0.948	0.592	3.784	0.793
FTSE	-0.561	0.996	0.728	3.731	0.767
SMI	-0.813	0.845	1.059	4.602	0.846

To evaluate the out-of-sample forecast accuracy, we used a rolling window approach with a window size of $S = 252$ days (one trading year). As suggested by Corsi (2009), we considered both short-term (i.e., horizon $h = 1, 5$) and long-term (i.e., $h = 22, 44$) forecast horizons to recognize that the financial market comprises participants (such as dealers, market makers, and institutional

investors) with a large range of trading frequencies. Thus, we expected that the results obtained for $h = 1$ would mainly concern short-term traders with a daily frequency, the results for $h = 5$ would concern medium term investors who typically balance their positions weekly, and the results for $h = 22, 44$ would concern long-term agents (such as mutual funds and pension funds) with a typical time horizon of one or more months.

We considered particular combinations of the forecast variable set, lag structure, and the corresponding univariate or multivariate model to forecast the log-realized variances at each time point $t = S, \dots, T-h$. We obtained h -step ahead forecasts $\widehat{LRV}_{t+h}^{(i)}$ and the corresponding forecast errors $e_{t+h}^{(i)} = LRV_{t+h}^{(i)} - \widehat{LRV}_{t+h}^{(i)}$ for each stock market index $1 \leq i \leq 10$. The forecast performance of each stock market index i was then measured based on the mean absolute forecast error (MAFE):

$$\text{MAFE}_h^{(i)} = \frac{1}{T-h-S+1} \sum_{t=S}^{T-h} |e_{t+h}^{(i)}|.$$

The MAFEs were computed out-of-sample, so no penalty was needed for the model complexity. We used the MAFEs to avoid a few extreme errors driving the forecast comparison results (e.g., Andersen et al., 1999; Hansen & Lunde, 2005; Poon & Granger, 2003). Ordering the forecast models based on the MAFEs might have been biased relative to the actual performance because the unobserved volatility was substituted by a proxy (e.g., see Patton, 2011 for univariate volatility models and Laurent et al., 2013 for multivariate volatility models), but selecting the daily realized variances based on the five-minute returns eliminated most of the bias.

To forecast more than one day-ahead in time (i.e., forecast horizon $h > 1$), we used the direct forecast approach. For a given forecast horizon h , the direct approach re-estimated the univariate model (6) or multivariate model (7), but with $LRV_{t+h-1}^{(i)}$ as the dependent variable. By multiplying the newly obtained parameter estimates with the corresponding lagged predictors, we obtained the h -step ahead forecasts $\widehat{LRV}_{t+h}^{(i)}$ of the log-realized variances. We employed direct forecasts because they are more robust to model misspecification than iterated forecasts (Marcellino et al., 2006).

5.1. Results

The MAFEs averaged across the 10 stock market indices and per forecast variable set, lag structure, model, and forecast horizon are reported in Table 3. For each forecast horizon, the winning combination with the lowest MAFE is indicated by a shaded cell in Table 3. We conducted DM tests (Diebold & Mariano, 1995) to compare two competing methods. We implemented the modified DM test described by Harvey et al. (1997) because the standard test could be over-sized for larger forecast horizons. To compute the test statistic, Newey-West standard errors were used to account for serial correlation in the forecast errors.

(i) *Forecast variable set.* The winning combinations in Table 3 show that the forecast variable sets including option-implied variances performed best overall, thereby supporting the conclusion of Bekert and Hoerova (2014) that they contain information about future stock market volatility. DM tests confirmed the significant (at the 5% level) improvement in the overall forecast accuracy compared with the corresponding forecast variable set that did not include option-implied variances (all p -values < 0.01 except for $h = 5$: p -value = 0.07).

For short-term forecast horizons of $h = 1$ and $h = 5$, the winning forecast variable set included jump and continuous realized variance components to the option-implied variances rather than the realized variances themselves. DM-tests confirmed the statistical significance (at the 10% level: p -values=0.07 and 0.09 for $h = 1$ and $h = 5$, respectively) of disentangling the realized variances into jump and continuous components (with IV already in the model). However, for long-term forecasts with the multivariate model, disentangling did not yield further improvements if option-implied variances were included in the model.

(ii) *Lag structure.* The winning model combinations in Table 3 indicate the preference for the parsimonious dwm lag structure. However, the DM-tests indicated that the difference with the lasso-based lag structures might not have been significant (in particular, for “dwm+lag selection”: p -value = 0.22, 0.08, and 0.13 for $h = 5, 22, 44$, respectively). Nonetheless, given the simple implementation of the dwm lag structure, we advocate its usage in practice. Our results support the conclusions of Audrino and Knaus (2016) and Audrino et al. (2017) regarding the difficulty of outperforming HAR-based models when more general lag structures are selected with lasso-based estimators. We found that the ordered lasso's inability to select valuable information in the option-implied variances and/or higher order reduced its forecast performance with the dwm lag structure. The lag selection results obtained with ordered lasso are presented and discussed in Appendix B.

(iii) *Univariate versus multivariate.* It should be noted that it was more difficult to obtain accurate forecasts for longer horizons because MAFE increased with h . Table 3 shows that the preference for a univariate or multivariate model depended on the forecast horizon selected. For shorter forecast horizons (i.e., $h = 1, 5$), univariate models performed significantly better, whereas multivariate models were preferable for longer forecast horizons (i.e., $h = 22, 44$), as shown by the DM tests (all p -values < 0.01). As shown in Fig. 1, co-movement of the stock market volatility measures occurred. Incorporating spillover effects between the different stock market indices helped to improve the forecast accuracy for longer forecast horizons.

Conclusion. We suggest the following as recommendations for practitioners. (i) Option-implied variances should always be included in the forecasting variable set. Rather limited differences were observed for either the realized variances or their jump and continuous components, so

Table 3

Rolling window of 1 year. Mean absolute forecast errors (MAFEs) averaged across the 10 stock markets per forecast variable set, lag structure, model type, and forecast horizon. Winning combinations are indicated by a shaded cell for each forecast horizon. The model confidence set is indicated in bold for each forecast horizon. Rolling Window of one year. Mean Absolute Forecast Errors (MAFE) averaged across the ten stock markets per forecast variable set, lag structure, model type and forecast horizon. Winning combinations are indicated by a shaded cell for each forecast horizon. The Model Confidence Set (MCS) is indicated in bold for each forecast horizon.

Horizon	Model	Lag structure	Forecast variables			
			RV	JC	RV+IV	JC+IV
$h = 1$	Univariate	dwm	0.416	0.413	0.397	0.395
		lag selection	0.426	0.432	0.425	0.407
		dwm + lag selection	0.416	0.413	0.410	0.401
	Multivariate	dwm	0.440	0.497	0.489	0.543
		lag selection	0.432	0.447	0.431	0.422
		dwm + lag selection	0.440	0.497	0.488	0.552
$h = 5$	Univariate	dwm	0.510	0.506	0.505	0.503
		lag selection	0.523	0.533	0.525	0.515
		dwm + lag selection	0.510	0.506	0.508	0.506
	Multivariate	dwm	0.539	0.545	0.528	0.573
		lag selection	0.535	0.550	0.542	0.533
		dwm + lag selection	0.539	0.545	0.539	0.547
$h = 22$	Univariate	dwm	0.624	0.605	0.612	0.608
		lag selection	0.632	0.623	0.639	0.624
		dwm + lag selection	0.625	0.605	0.611	0.603
	Multivariate	dwm	0.578	0.556	0.540	0.589
		lag selection	0.636	0.638	0.633	0.634
		dwm + lag selection	0.578	0.556	0.560	0.554
$h = 44$	Univariate	dwm	0.677	0.641	0.645	0.634
		lag selection	0.681	0.666	0.682	0.664
		dwm + lag selection	0.677	0.641	0.647	0.637
	Multivariate	dwm	0.572	0.560	0.539	0.588
		lag selection	0.693	0.678	0.682	0.674
		dwm + lag selection	0.572	0.560	0.557	0.559

we propose adopting the simpler options of including only the former for more complex multivariate models and including the latter for univariate models. (ii) The dwm lag structure characteristic should be used for the standard HAR model. (iii) Use univariate models for $h = 1$ and $h = 5$ (short run), and multivariate models for $h = 22$ and $h = 44$ (long run). This is a simple recommendation based on the empirical results obtained and it is consistent with economic intuition.

To validate these recommendation, Fig. 2 presents $\text{MAFE}_h^{(i)}$ separately for each stock market index for the 24 possible combinations of the forecasting method, and for several values of the forecast horizon h . The colored dots in the dot plots corresponds to our recommendations (red for $h = 1$ and $h = 5$, and blue for $h = 22$ and $h = 44$). The associated MAFEs were not always the lowest but

they were close to the smallest value obtained, thereby confirming the validity of our recommendations. Furthermore, it should be noted that the MAFEs for forecasting variable sets JC+IV and RV+IV were very close in most cases.

Finally, we inspected the winning model combinations in more depth. Analyzing the signs of the forecast errors, which are important for risk management and option trading, did not reveal any systematic patterns in terms of over- or underestimation across time and/or stock markets. Next, we inspected the forecast performance across time. Fig. 3 shows the changes in the absolute forecast errors (averaged across the 10 stock markets) for the winning model combination at $h = 1$ as a function of time t until the end point of a 1-year rolling window. The 20 largest forecast errors are highlighted by red dots in Fig. 3. Similar findings were obtained for the

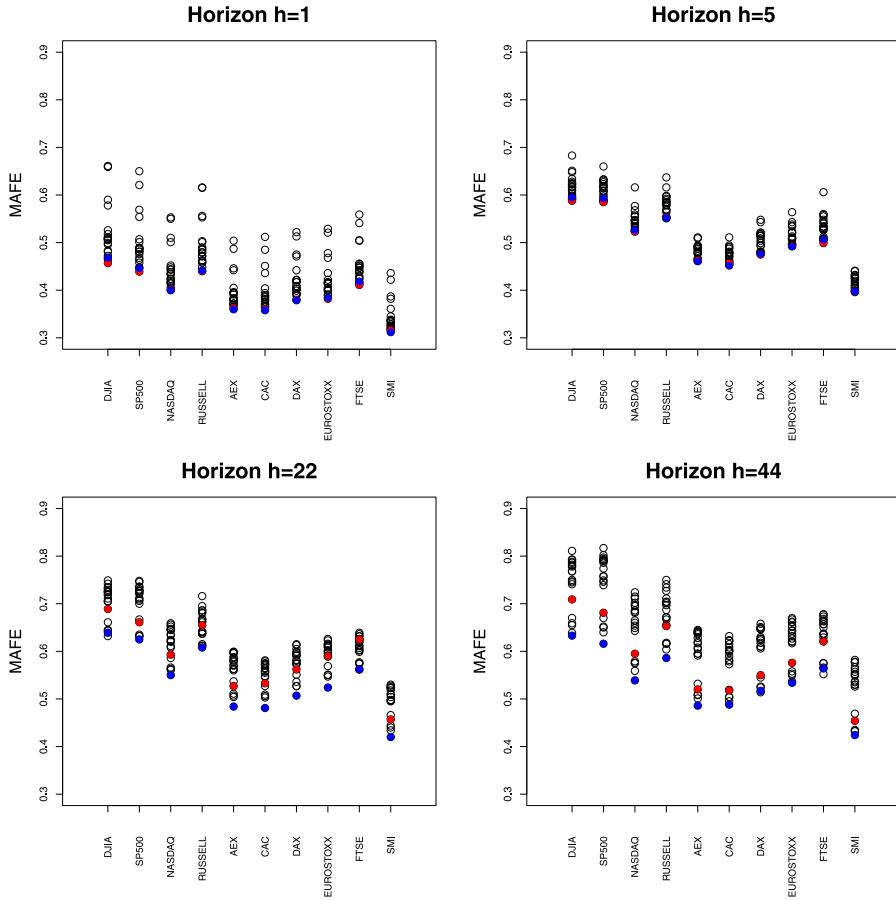


Fig. 2. Mean absolute forecast errors (MAFEs) for the 24 possible combinations (dots) of forecast methods for each stock market index on the horizontal axis based on a 1-year rolling window. MAFEs associated with our recommendation (dwm; univariate for $h = 1, 5$ and multivariate for $h = 22, 44$) are shown in blue for RV+IV and in red for JC+IV.

other forecast horizons, and thus they are not presented. The Global Financial Crisis (February 2008 to February 2009) and European Sovereign Debt Crisis (2011–2012) occurred during this period and they are indicated by the shaded areas. We observed that several of the largest errors occurred at the onset or during these two crisis periods. Other interesting findings include the fact that the largest forecast error was observed on June 24, 2016, which may be attributed to the reactions of the stock markets to the Brexit referendum on the previous day. The large forecast error on February 2, 2018 can be explained by the Dow Jones Industrial Average experiencing its biggest loss since the Brexit referendum.

5.2. Robustness checks

Model confidence set (MCS). To globally assess the significance of the differences in forecast performance among the 24 model combinations, we computed the MCS proposed by Hansen et al. (2011). MCS separates the set of best combinations from their competitors. There were no significant differences in the predictive ability of the combinations in the MCS. We obtained the MCS for each

forecast horizon.¹ The MCS is indicated in bold in Table 3 for each forecast horizon.

The MCS analysis results supported recommendations (i)–(iii) presented in the conclusion of Section 5.1. The recommended combinations appeared as the sole element in the MCS at horizons $h = 1, 22, 44$. At $h = 5$, the lag structure and forecast variables were unclear. However, the recommended combination (see shading in Table 3) was part of the MCS.

Choice of loss function. To assess the sensitivity of our results to the choice of loss function, we repeated our analysis using the mean squared forecast error (MSFE) instead of the MAFE. Overall, our main findings were unchanged after evaluating the forecast performance in terms of the MSFE values. The winning combinations were unchanged for horizons $h = 1, 5, 22$. For $h = 44$, the winning combination replaced “dwm” with the “dwm+lag selection” lag structure. These results are available from the authors upon request.

¹ We obtained the MCS using the MCSprocedure function in R (see Catania and Bernardi (2014)). We used the TR test statistic based on 5000 bootstrapped samples and with a 75% confidence level.

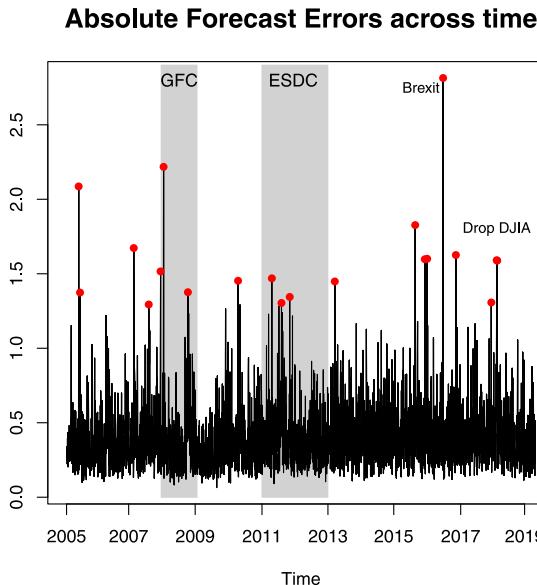


Fig. 3. Absolute forecast errors averaged across the 10 stock markets for the winning model combination at horizon $h = 1$. The horizontal axis represents the end date of a 1-year rolling window. The Global Financial Crisis (GFC) and European Sovereign Debt Crisis (ESDC) are indicated by the shaded areas.

Forecasting realized variances. As noted in Section 2, forecasting the log-realized variances with linear models typically obtains better statistical properties than forecasting the realized variances. Nevertheless, the variances are often used as inputs in many economic or financial decision-making processes, such as portfolio optimization, option pricing, value at risk modeling, and dynamic hedging (see Bauwens et al. (2012)). Therefore, we verified whether our main findings regarding forecasting based on log-realized variances could be extended to those based on realized variances.

To obtain the forecast $\widehat{RV}_t^{(i)}$ for the realized variance of stock market index i at time t for each forecast horizon, we used the same transformation described by Bekert and Hoerova (2014):

$$\widehat{RV}_t^{(i)} = \exp \left(\widehat{LRV}_t^{(i)} + \frac{1}{2} \text{Var} \left(\widehat{LRV}_t^{(i)} \right) \right), \quad (8)$$

where $\widehat{LRV}_t^{(i)}$ is the forecast of the log-realized variance at time t for stock market index i and the sample variance of the log-realized variance was used to compute the variance term. The second term in (8) is a transformation correction derived by assuming the log-normality of the realized variances. Graphs of the log-transformed realized variances did not indicate any strong deviation from normality, thereby agreeing with the findings obtained by Andersen, Bollerslev, Diebold, and Ebens (2001). We then computed the corresponding forecast errors $RV_{t+h}^{(i)} - \widehat{RV}_{t+h}^{(i)}$ and evaluated the forecast accuracy using the MAFEs averaged across the 10 stock market indices.

Overall, our main findings remained unchanged after evaluating the forecast performance in terms of the realized variance: (i) models with option-implied variances

tend to produce more accurate forecasts than those without, (ii) the HAR model's dwm lag structure is preferred over lasso-based lag structures, and (iii) multivariate models provide more accurate forecasts than univariate models for longer forecast horizons. These results are available from the authors upon request.

Rolling window size. To investigate the sensitivity of our results to the choice of rolling window size, we repeated our forecast study with a rolling window size of $S = 252 \times 5 = 1260$ (5 years). The results based on the MAFEs averaged across the 10 stock markets are shown in Table 6 in Appendix B. The results based on the MAFEs for the individual stock market indices are presented in Fig. 7 in Appendix B. Overall, our main findings remained unchanged.

Tables 3 and 6 demonstrate that there was no reduction in the magnitude of MAFE when using an estimation window of 5 years instead of 1 year. The amount of data in the shorter window was sufficiently large to obtain precise estimates and it avoided time series instabilities.² Figs. 2 to 7 show that our recommendations remained valid. The differences between JC+IV and RV+IV at horizons $h = 22$ and $h = 44$ for a 5-year rolling window were even smaller compared with the results obtained using a 1-year rolling window.

6. Cross-market spillover effects

The main aim of our study was to obtain appropriate information in order to provide practitioners and applied researchers with advice regarding modeling choices (i.e., forecast variables, lag structures, univariate versus multivariate modeling approach) when forecasting realized variances. In the following, we describe the cross-market feedback effects obtained with the multivariate models. We do not aim to explain all of the economic fundamentals related to our results but several economically interesting findings can be highlighted.

We investigated the cross-market feedback effects obtained with the multivariate model with a dwm lag structure based on the jump, continuous realized variances, and option-implied variances. The resulting model was estimated by least squares based on the full sample (January 2004–April 2016). Using the multivariate Portmanteau statistic (Tsay, 2014, p72), we found that the null hypothesis of multivariate white noise for the residuals was not rejected. Furthermore, inspecting the graph of the residuals also supported the validity of the model.

In order to detect significant spillovers from the option-implied variance, jump, and continuous components of one stock market index to the realized variance of another, we performed Granger causality tests. These tests allowed us to assess the predictive capacity of a group of time series for forecasting another, which is common in connectedness studies of stock returns or volatilities (e.g., see Billio et al. (2012) and Bubák et al. (2011)). In

² This finding supported our focus on $S = 1$ -year. Calibrating the time span to obtain equal forecasting periods for rolling windows of $S = 1$ and $S = 5$ did not alter this finding.

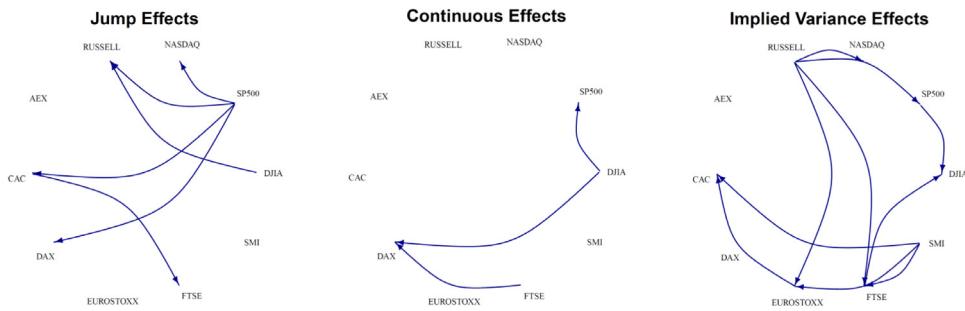


Fig. 4. Network of significant cross-market spillovers for the jump (left), continuous (middle), and option-implied variance (right) components of each stock market based on the log-realized variance of each other stock market index during the full sample (January 2004 to April 2016).

particular, we tested whether the day, week, and month jump components of a particular stock market index could jointly and significantly affect another stock market's realized variance. Similar test were performed to assess the effects of the continuous realized variance components and option-implied variance components. To correct for multiple testing, we used a significance level of 0.05/10 because a test was performed for each of the 10 stock markets. The Bonferroni correction is a conservative solution to the multiple testing problem, although it allowed us to detect fewer of the actual significant results. However, it had the advantage of focusing on the cross-market spillover effects with strong evidence.

We conducted network analysis to visualize the significant spillover effects. Fig. 4 shows three networks representing the significant cross-market effects of the jump (left panel), continuous (middle), and option-implied (right) variance components of each stock market index based on the log-realized variance of each other stock market index. The nodes shown in the network are the stock market indices. An edge was drawn from one stock market to another if Granger causality was detected.

Significant cross-market spillover effects were detected in all networks. For example, the left panel in Fig. 4 shows that significant spillover occurred from the jump components of SP500 to four other stock market indices (NASDAQ, RUSSELL, CAC, and DAX). We found the following asymmetries. (i) Among international stock market indices: American stock market indices typically had more predictive power for European stock market indices than the reverse. (ii) Among large and small stock market caps in the United States: Large caps such as SP500 and DJIA tended to influence small caps such as RUSSELL more than the reverse. Stock price volatility can be directly related to the rate of information flow into the stock market, so these asymmetries are consistent with a market where the prices of large stocks respond immediately to new information whereas the prices of small stocks exhibit a lagged response (e.g., see Harris & Pisedtasalasai, 2006). Furthermore, we observed that the transmission of volatility did not always occur through the largest stock market in a region, thereby agreeing with the results obtained by Koulakiotis et al. (2009). For instance, we found that the Swiss market index could transmit volatility to several other European stock markets (see the outgoing edges from SMI in Fig. 4, right panel) although it is not

the largest in terms of market capitalization. Similarly, the role of the Swiss market as a major exporter of volatility was reported in a study of cross-listed equities within European financial regions by Koulakiotis et al. (2009), and effects on volatility spillovers across emerging and developing stock markets were described by Yarovaya et al. (2016).

The cross-market spillover effects summarized in Fig. 4 were estimated based on the full sample. The Global Financial Crisis (February 2008 to February 2009) and European Sovereign Debt Crisis (2011–2012) occurred in this period. We also estimated the cross-market spillover effects individually for these sub-periods because crisis periods are expected to greatly affect the parameter estimates. Fig. 5 shows the significant cross-market spillovers during the Global Financial Crisis (top) and European Sovereign Debt Crisis (bottom), and these cross-market spillover effects differ considerably compared with the results in Fig. 4. It should be noted that compared with the full sample, we observed the opposite asymmetry among the international stock market indices during the considered crisis periods, as follows. (i) European stock markets tended to have greater predictive power for American stock markets, e.g., the steering role of FTSE during the Global Financial Crisis (Fig. 5, top middle panel) and of EUROSTOXX during the European Sovereign Debt Crisis (Fig. 5, bottom left panel). Similarly, Demirer et al. (2018) reported that according to their global bank network based on volatility connectedness, European banks played a more central role than American banks during the European Sovereign Debt Crisis. (ii) We found marginal evidence for volatility spillovers from small to large caps in the US (see the outgoing edge from RUSSELL to DJIA in Fig. 5, top left panel). Volatility spillovers from large to small caps are more robustly supported by previous studies, but Harris and Pisedtasalasai (2006) also reported volatility spillovers from small to large stocks in their study of the UK market and they varied with the time period considered. Similarly, BenSaïda et al. (2018) found frequent swings between a stock market's volatility transmission and reception during turbulent periods.

Table 4 compares the R^2 values per stock market index across the full sample, the Global Financial Crisis, and the European Sovereign Debt Crisis for the univariate and multivariate models. We found the following. (i) For the univariate models, the R^2 values were generally lower

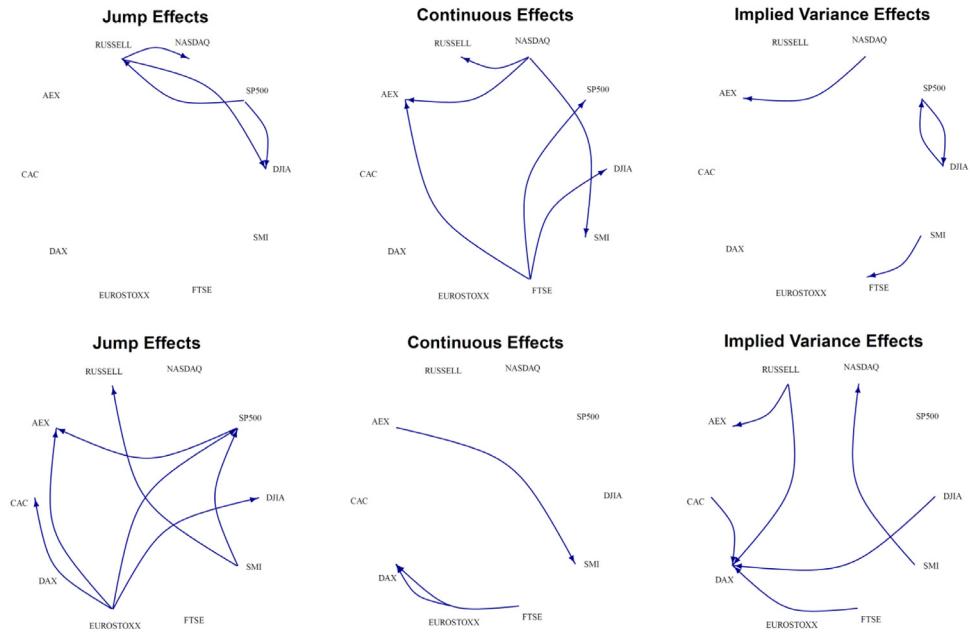


Fig. 5. Network of significant cross-market spillovers for the jump (left), continuous (middle), and option-implied variance (right) components of each stock market based on the log-realized variance of each other stock market index during the Global Financial Crisis (top) and European Sovereign Debt Crisis (bottom).

during crises, thereby agreeing with the findings obtained by Bekaert and Hoerova (2014). However, the opposite occurred for the multivariate models. (ii) Moreover, the added value increased for the multivariate models during crisis periods. Multivariate models based on the full sample obtained an average increase in R^2 of 2% compared with the univariate models. During the crisis periods, we found that the average increases were about 20% in the Global Financial Crisis and about 8% in the European Sovereign Debt Crisis. These results support previously reported empirical findings regarding increased volatility connectedness (e.g., Demirer et al., 2018) and return connectedness (e.g., Billio et al., 2012) during crisis periods.

7. Discussion

In this study, we used extended HAR models to obtain log-realized variance forecasts for several stock market indices. In our empirical study, we compared the forecast performance across 24 combinations of forecast variables, lag structures, and univariate versus multivariate models. Our out-of-sample results across stock market indices consistently indicated that: (i) augmenting the standard HAR model with option-implied variances yields more accurate forecasts; (ii) the standard HAR model's dwm lag structure is preferable compared with more general lasso-based lag structures; and (iii) multivariate HAR-type models that consider spillovers between stock market indices perform better than univariate models for longer forecast horizons. These spillovers can be visualized in networks, as described in Section 6.

Our out-of-sample results were confirmed by in-sample analysis in terms of R^2 , as shown in Table 5. We

Table 4

R^2 values for each stock market index with the univariate and multivariate models estimated based on the full sample and the Global Financial Crisis (GFC) and European Sovereign Debt Crisis (ESDC) sub-periods.

Model	Stock market index	Full sample	GFC	ESDC
Univariate	DJIA	0.728	0.654	0.578
	SP500	0.765	0.678	0.602
	NASDAQ	0.732	0.663	0.646
	RUSSELL	0.681	0.632	0.661
	AEX	0.772	0.628	0.766
	CAC	0.756	0.590	0.761
	DAX	0.733	0.589	0.787
	EUROSTOXX	0.729	0.574	0.756
	FTSE	0.725	0.647	0.680
	SMI	0.778	0.658	0.813
Multivariate	DJIA	0.748	0.810	0.667
	SP500	0.780	0.822	0.690
	NASDAQ	0.746	0.824	0.738
	RUSSELL	0.692	0.799	0.743
	AEX	0.798	0.851	0.851
	CAC	0.790	0.829	0.839
	DAX	0.758	0.810	0.861
	EUROSTOXX	0.760	0.806	0.833
	FTSE	0.756	0.854	0.785
	SMI	0.802	0.868	0.873

found that: (i) disentangling the realized variances into jump and continuous components led to minor improvements in the model fits, as shown by the increases in R^2 of less than 1% (compare the first two columns and the last two columns in Table 5); (ii) larger benefits were obtained after adding the option-implied variances to the realized variances or the disentangled version, i.e., an increase in R^2 of around 3%; and (iii) using a multivariate model instead of a univariate model led to an average increase in

Table 5

R^2 values for each stock market index with the univariate and multivariate models estimated based on the full sample with the dwm lag structure and the four sets of forecast variables.

Model	Stock market index	RV	JC	RV+IV	JC+IV
Univariate	DJIA	0.676	0.697	0.725	0.728
	SP500	0.721	0.727	0.766	0.765
	NASDAQ	0.701	0.699	0.733	0.732
	RUSSELL	0.632	0.637	0.677	0.681
	AEX	0.752	0.750	0.774	0.772
	CAC	0.741	0.741	0.756	0.756
	DAX	0.710	0.710	0.737	0.733
	EUROSTOXX	0.689	0.700	0.727	0.729
	FTSE	0.674	0.696	0.715	0.725
	SMI	0.763	0.759	0.780	0.778
Multivariate	DJIA	0.702	0.716	0.740	0.748
	SP500	0.738	0.748	0.776	0.780
	NASDAQ	0.706	0.713	0.742	0.746
	RUSSELL	0.647	0.659	0.686	0.692
	AEX	0.771	0.772	0.797	0.798
	CAC	0.760	0.764	0.788	0.790
	DAX	0.731	0.736	0.756	0.758
	EUROSTOXX	0.727	0.733	0.757	0.760
	FTSE	0.726	0.734	0.752	0.756
	SMI	0.778	0.779	0.801	0.802

R^2 of 2%. Clearly, part of this increase could be explained by the substantially higher number of parameters in the model and the average increase in the adjusted R^2 was 1%. Repeating the in-sample analysis for larger forecast horizons³ confirmed the larger gains with the multivariate model compared with the univariate model, i.e., for $h = 22$ the average increase in the adjusted R^2 was about 6%, whereas the increase was about 9% for $h = 44$. In addition, the findings described in points (i) and (ii) were quite stable across forecast horizons.

HAR-type models are typically estimated by least squares. Due to the imposed dwm lag restrictions, the number of parameters that needed to be estimated decreased substantially. However, when spillovers between stock markets were considered in the multivariate setting, the number of unknown parameters increased and it could even exceed the sample size for short estimation windows. In the latter case, the least squares estimator cannot be used and the ordered lasso procedure (or any other regularization method) is mandatory. The lasso procedure will implicitly select a subset of most predictable stock indices and yield a restricted model. If the number of stock markets is increased dramatically, the number of parameters could exceed the sample size and the least squares estimator cannot be used. In this case, the ordered lasso can be used to select a subset of the most predictive stock market indices to include in the multivariate model.

Network analysis clearly demonstrated the time-varying nature of the volatility spillovers among the international stock market indices. Furthermore, we found that the benefits of multivariate models increased during crisis periods, such as the Global Financial Crisis and European Sovereign Debt Crisis. Our models do not allow for time-varying parameters (e.g., see Dangl & Halling, 2012), but our forecast exercise was conducted with a short rolling

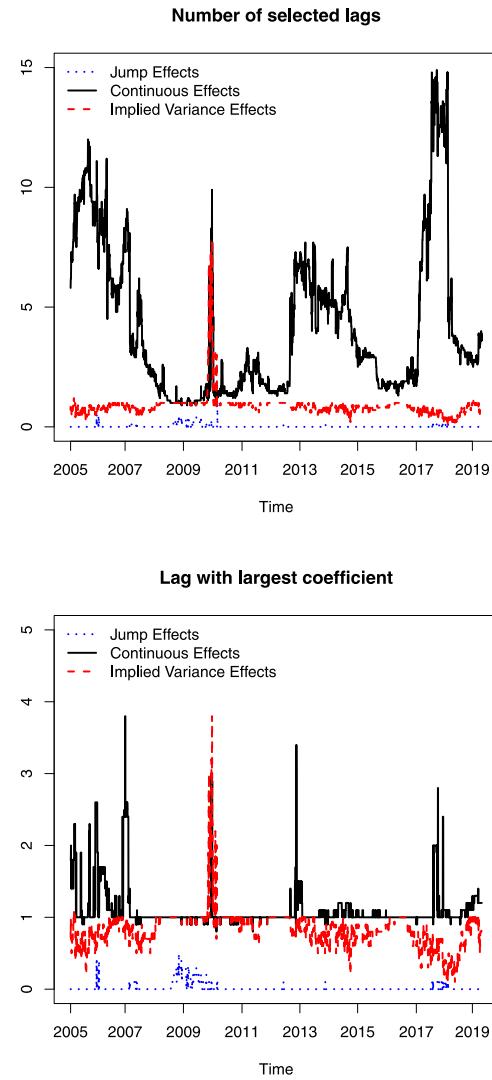


Fig. 6. Number of selected lags (top panel) and the lag with the largest coefficient value (bottom panel) averaged across the 10 stock market indices. The horizontal axis represents the end date of a 1-year rolling window.

window size (i.e., 1 year) to allow for some variation in the parameter estimates over time. This is a simple and practitioner-oriented approach, which is commonly used in studies of macroeconomic or financial forecasting.

In this study, we compared the forecast performance in terms of statistical loss functions based on absolute or squared forecast errors. In future research, it would be interesting to compare the forecast performance in terms of economic loss functions, such as by using pricing option-based strategies based on realized volatility similar to those described by Corsi et al. (2013) or by pricing synthetic variance swap contracts in the same manner as Rombouts et al. (2020).

Acknowledgments

We thank the reviewers for their thorough review and greatly appreciate their comments, which helped to

³ Results available from the authors upon request.

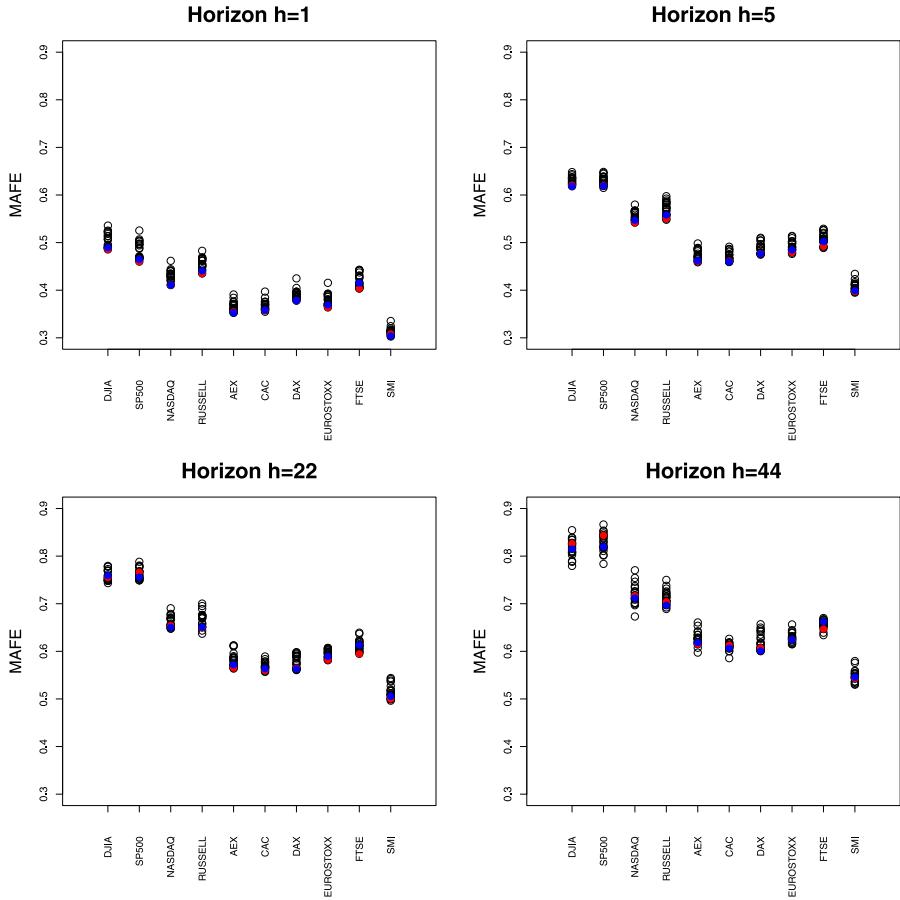


Fig. 7. Mean absolute forecast errors (MAFEs) for the 24 possible combinations (dots) of forecast methods for each stock market index on the horizontal axis over a 5-year rolling window. MAFEs associated with our recommendation (dwm; univariate for $h = 1, 5$ and multivariate for $h = 22, 44$) are shown in blue for RV+IV and in red for JC+IV.

substantially improve the quality of the manuscript. The authors gratefully acknowledge financial support from the European Union's Horizon 2020 research and innovation programme under Marie Skłodowska-Curie grant agreement No. 832671.

Appendix A. Ordered lasso

We present the ordered lasso for a univariate model. For the multivariate model (7), the ordered lasso is applied equation by equation.

Review and implementation of the ordered lasso. Let $\mathbf{y} = (y_{p+1}, \dots, y_T)'$ denote the response time series and $\mathbf{X} \in \mathbb{R}^{(T-p) \times d}$ is the predictor matrix. The response contains the log-realized variance of a stock market index and the predictor matrix contains the lagged values of the k forecast variables. Let $\boldsymbol{\beta}$ be the $d = p \times k$ -dimensional regression parameter vector. We denote the elements of $\boldsymbol{\beta}$ by a double sub-index $\beta_{j,\ell}$, where the index j denotes the forecast variable type and the index ℓ is the lag.

Write each $\beta_{j,\ell} = \beta_{j,\ell}^+ - \beta_{j,\ell}^-$, with $\beta_{j,\ell}^+, \beta_{j,\ell}^- \geq 0$, for all $1 \leq j \leq k$ and $1 \leq \ell \leq p$. The ordered lasso estimator with sparsity parameter $\lambda > 0$ is defined as (Tibshirani &

Suo, 2016)

$$\widehat{\boldsymbol{\beta}}_\lambda = \underset{\boldsymbol{\beta}^+, \boldsymbol{\beta}^-}{\operatorname{argmin}} \frac{1}{2(T-p)} \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2^2 + \lambda \sum_{i=1}^q \sum_{j=1}^q \sum_{\ell=1}^p (\beta_{j,\ell}^+ + \beta_{j,\ell}^-), \quad (9)$$

subject to the constraints $\beta_{j,1}^+ \geq \beta_{j,2}^+ \geq \dots \geq \beta_{j,p}^+ \geq 0$, and $\beta_{j,1}^- \geq \beta_{j,2}^- \geq \dots \geq \beta_{j,p}^- \geq 0$. Each component $\beta_{j,\ell}$ of the parameter vector is split into a positive part $\beta_{j,\ell}^+$ and a negative part $\beta_{j,\ell}^-$ to ensure that the resulting optimization problem in (9) is convex. Typically, either $\beta_{j,\ell}^+$ or $\beta_{j,\ell}^-$ is exactly zero, which ensures that the absolute values of the coefficients $\widehat{\beta}_{j,\ell}$ decrease in the lag order.

The first term in Eq. (9) is a least squares criterion. The second term in Eq. (9) is a penalty term on the parameters, with sparsity parameter λ . The selected maximal lag is smaller when the sparsity parameter λ is larger. Similar to the lasso penalty, this penalty encourages sparsity in $\widehat{\beta}_{j,\ell}^+$ and $\widehat{\beta}_{j,\ell}^-$, and thus some (or many) elements of $\boldsymbol{\beta}$ will be estimated as exactly zero. In addition, the side constraints ensure that a higher order lag can only be included in the model if its lower order lags are already

Table 6

Rolling window of 5 years. Mean absolute forecast errors (MAFEs) averaged across the 10 stock markets per forecast variable set, lag structure, model type, and forecast horizon. Winning combinations are indicated by a shaded cell for each forecast horizon. The model confidence set is indicated in bold for each forecast horizon.

Horizon	Model	Lag structure	Forecast variables				
			RV	JC	RV+IV	JC+IV	
<i>h</i> = 1	Univariate	dwm	0.420	0.418	0.398	0.396	
		lag selection	0.421	0.427	0.421	0.405	
		dwm + lag selection	0.420	0.418	0.414	0.401	
	Multivariate	dwm	0.412	0.415	0.400	0.405	
		lag selection	0.420	0.441	0.418	0.404	
		dwm + lag selection	0.412	0.415	0.402	0.404	
	<i>h</i> = 5	Univariate	Forecast variables				
			RV	JC	RV+IV	JC+IV	
			0.519	0.515	0.513	0.510	
		lag selection	0.523	0.530	0.523	0.514	
		dwm + lag selection	0.519	0.515	0.514	0.508	
		Multivariate	dwm	0.526	0.530	0.535	0.543
			lag selection	0.534	0.540	0.535	0.523
			dwm + lag selection	0.526	0.530	0.519	0.528
	<i>h</i> = 22	Univariate	Forecast variables				
			RV	JC	RV+IV	JC+IV	
			0.635	0.616	0.622	0.619	
		lag selection	0.634	0.630	0.634	0.622	
		dwm + lag selection	0.635	0.616	0.620	0.618	
		Multivariate	dwm	0.632	0.629	0.632	0.636
			lags	0.652	0.646	0.650	0.639
			dwm + lag selection	0.632	0.629	0.626	0.629
	<i>h</i> = 44	Univariate	Forecast variables				
			RV	JC	RV+IV	JC+IV	
			0.696	0.660	0.670	0.674	
		lag selection	0.685	0.678	0.681	0.679	
		dwm + lag selection	0.696	0.660	0.668	0.670	
		Multivariate	dwm	0.690	0.673	0.679	0.651
			lag selection	0.698	0.679	0.689	0.675
			dwm + lag selection	0.690	0.673	0.681	0.672

included. It should be noted that each forecast variable is allowed to have its own maximal lag structure.

Thus, for a given sparsity parameter, the ordered lasso selects the lags. The model is the re-estimated by least squares to yield the log-realized variance forecast $\widehat{LRV}_{t,\lambda}$ at time t for a given stock market index, and we also keep track of the associated Bayesian information criterion BIC_λ .

Algorithm. According to standard regularization studies of lasso-based estimation, the variables are standardized before computing the ordered lasso. The optimization problem (9) can be solved efficiently using a proximal gradient method. For further details, we refer the reader to the studies by [Jenatton et al. \(2011\)](#) and [Tibshirani and Suo \(2016\)](#), and the references therein. All computations were conducted in R version 3.2.1. The codes for the algorithms are based on the R-package `orderedLasso`

([Friedman et al., 2015](#)) and they are available from the authors upon request.

Selection of sparsity parameter. The selection of the sparsity parameter λ is important. Previous studies typically used information criteria as the BIC to select the “optimal” sparsity parameter. Forecasts are then obtained from the lasso-based estimator with the selected sparsity parameter. By contrast, we did not select an “optimal” sparsity parameter, but instead we combined forecasts with several sparsity parameters. These forecast combinations are typically more robust to an incorrectly selected sparsity parameter. Our empirical analysis results showed that the forecast combinations yielded significantly better or similar forecast errors to an information criteria-based approach.

We consider a logarithmically spaced grid of sparsity parameters of length $M = 50$ in the interval $[\lambda_1, \lambda_M]$,

where $\lambda_1 = 0$ and λ_M is a value for the sparsity parameter that shrinks all parameters to zero (e.g., see Friedman et al. (2010)). The combined forecast is then

$$\widehat{LRV}_t = \sum_{m=1}^M w_m \cdot \widehat{LRV}_{t,\lambda_m},$$

with forecast combination weights w_m

$$w_m = \frac{\exp(-0.5\text{BIC}_{\lambda_m})}{\sum_{m=1}^M \exp(-0.5\text{BIC}_{\lambda_m})},$$

as used for standard Bayesian model averaging (e.g., see Koop, 2003, Chapter 11).

Lag selection results. The lag selection results for the ordered lasso are shown in Fig. 6, where the top panel presents the number of selected jump, continuous, and option-implied variance lags (out of 22), and the bottom panel presents the jump/continuous/implied variance lag with the largest absolute coefficient value. The results were averaged across the 10 stock markets for brevity but the findings also hold for the individual stock markets.

We observed the following. (i) The ordered lasso mainly selected lags of continuous RV components (about four on average), but very few of the jump RV components and implied variances (about one on average). The selected lag structure varied considerably across time, as also found by Audrino et al. (2017). (ii) The largest coefficient value always corresponded to one of the first four lags and usually the first lag.

Appendix B. Additional results for a 5-year rolling window

See Fig. 7 and Table 6.

References

- Andersen, T. G., Bollerslev, T., & Diebold, F. X. (2007). Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility. *The Review of Economics and Statistics*, 89(4), 701–720.
- Andersen, T. G., Bollerslev, T., Diebold, F. X., & Ebens, H. (2001). The distribution of realized stock return volatility. *Journal of Financial Economic*, 61(1), 43–76.
- Andersen, T. G., Bollerslev, T., Diebold, F. X., & Labys, P. (2001). The distribution of realized exchange rate volatility. *Journal of the American Statistical Association*, 96(453), 42–55.
- Andersen, T. G., Bollerslev, T., & Lange, S. (1999). Forecasting financial market volatility: Sample frequency vis-a-vis forecast horizon. *Journal of Empirical Finance*, 6(5), 457–477.
- Audrino, F., Camponovo, L., & Roth, C. (2017). Testing the lag structure of assets' realized volatility dynamics. *Quantitative Finance and Economics*, 1(4), 363–387.
- Audrino, F., & Knaus, S. D. (2016). Lassoing the HAR model: A model selection perspective on realized volatility dynamics. *Econometric Reviews*, 35(8–10), 1485–1521.
- Audrino, F., Sigris, F., & Ballinari, D. (2020). The impact of sentiment and attention measures on stock market volatility. *International Journal of Forecasting*, 36(2), 334–357.
- Baele, L. (2005). Volatility spillover effects in European equity markets. *Journal of Financial and Quantitative Analysis*, 40(2), 373–401.
- Bakshi, G., Kapadia, N., & Madan, D. (2003). Stock return characteristics, Skew laws, and the differential pricing of individual equity options. *Review of Financial Studies*, 16(1), 101–143.
- Barndorff-Nielsen, O. E., & Shephard, N. (2002). Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society. Series B*, 64(2), 253–280.
- Bauwens, L., Hafner, C. M., & Laurent, S. (2012). *Handbook of volatility models and their applications*. Wiley.
- Bekaert, G., Harvey, C. R., & Ng, A. (2005). Market integration and contagion. *Journal of Business*, 78(1), 39–69.
- Bekaert, G., & Hoerova, M. (2014). The VIX, the variance premium and stock market volatility. *Journal of Econometrics*, 183(2), 181–192.
- Belloni, A., & Chernozhukov, V. (2013). Least squares after model selection in high-dimensional sparse models. *Bernoulli*, 19(2), 521–547.
- BenSaid, A., Litimi, H., & Abdallah, O. (2018). Volatility spillover shifts in global financial markets. *Economic Modelling*, 73, 343–353.
- Billio, M., Getmansky, M., Lo, A. W., & Pelizzon, L. (2012). Econometric measures of connectedness and systemic risk in the finance and insurance sectors. *Journal of Financial Economic*, 104(3), 535–559.
- Bollerslev, T. (2001). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 100(1), A307–A327.
- Bollerslev, T., Marrone, J., Xu, L., & Zhou, H. (2014). Stock return predictability and variance risk premia: Statistical inference and international evidence. *Journal of Financial and Quantitative Analysis*, 49(3), 633–661.
- Bollerslev, T., Patton, A. J., & Quaedvlieg, R. (2016). Exploiting the errors: A simple approach for improved volatility forecasting. *Journal of Econometrics*, 192(1), 1–18.
- Bollerslev, T., Tauchen, G., & Zhou, H. (2009). Expected stock returns and variance risk premia. *Review of Financial Studies*, 22(11), 4463–4492.
- Bubák, V., Kočenda, E., & Žikeš, F. (2011). Volatility transmission in emerging European foreign exchange markets. *Journal of Banking & Finance*, 35(11), 2829–2841.
- Carr, P., & Wu, L. (2009). Variance risk premia. *Review of Financial Studies*, 22(3), 1311–1341.
- Catania, L., & Bernardi, M. (2014). MCS: Model confidence set procedure. R package version 0.1.1. <https://cran.r-project.org/web/packages/MCS/index.html>.
- Chevillon, G., Hecq, A., & Laurent, S. (2018). Generating univariate fractional integration within a large VAR(1). *Journal of Econometrics*, 204(1), 54–65.
- Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Economic*, 7(2), 174–196.
- Corsi, F., Fusari, N., & La Vecchia, D. (2013). Realizing smiles: Options pricing with realized volatility. *Journal of Financial Economic*, 107(2), 284–304.
- Corsi, F., Pirino, D., & Renò, R. (2010). Threshold bipower variation and the impact of jumps on volatility forecasting. *Journal of Econometrics*, 159(2), 276–288.
- Cubadda, G., Guardabascio, B., & Hecq, A. (2017). A vector heterogeneous autoregressive index model for realized volatility measures. *International Journal of Forecasting*, 33(2), 337–344.
- Dangl, T., & Halling, M. (2012). Predictive regressions with time-varying coefficients. *Journal of Financial Economic*, 106(1), 157–181.
- De Mol, C., Giannone, D., & Reichlin, L. (2008). Forecasting using a large number of predictors: Is Bayesian shrinkage a valid alternative to principal components? *Journal of Econometrics*, 146(2), 318–328.
- Demirer, M., Diebold, F. X., Liu, L., & Yilmaz, K. (2018). Estimating global bank network connectedness. *Journal of Applied Econometrics*, 33(1), 1–15.
- Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3), 253–263.
- Diebold, F. X., & Yilmaz, K. (2015). *Financial and macroeconomics connectedness: A network approach to measurement and monitoring*. Oxford University Press, New York, US.
- Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, 33(1), 1–22.
- Friedman, J., Suo, X., & Tibshirani, R. (2015). Package: orderedLasso. R package version 1.7. <https://cran.r-project.org/web/packages/orderedLasso/index.html>.
- Hansen, P. R., & Lunde, A. (2005). A forecast comparison of volatility models: Does anything beat a GARCH(1, 1)? *Journal of Applied Econometrics*, 20(7), 873–889.

- Hansen, P. R., Lunde, A., & Nason, J. M. (2011). The model confidence set. *Econometrica*, 79(2), 453–497.
- Harris, R. D. F., & Pisetsorasasai, A. (2006). Return and volatility spillovers between large and small stocks in the UK. *Journal of Business Finance & Accounting*, 33(9–10), 1556–1571.
- Harvey, D., Leybourne, S., & Newbold, P. (1997). Testing the equality of prediction mean squared errors. *International Journal of Forecasting*, 13(2), 281–291.
- Haugom, E., Langeland, H., Molnar, P., & Westgaard, S. (2014). Forecasting volatility of the US oil market. *Journal of Banking & Finance*, 47, 1–14.
- Hua, J., & Manzan, S. (2013). Forecasting the return distribution using high-frequency volatility measures. *Journal of Banking & Finance*, 37(11), 4381–4403.
- Jenatton, R., Audibert, J. Y., & Bach, F. (2011). Structured variable selection with sparsity-inducing norms. *Journal of Machine Learning Research*, 12, 2777–2824.
- Koop, G. (2003). *Bayesian econometrics*. Wiley-Interscience.
- Koulakiotis, A., Dasilas, A., & Papasyriopoulos, N. (2009). Volatility and error transmission spillover effects: Evidence from three European financial regions. *The Quarterly Review of Economics and Finance*, 49(3), 858–869.
- Laurent, S., Rombouts, J. V. K., & Violante, F. (2013). On loss functions and ranking forecasting performances of multivariate volatility models. *Journal of Econometrics*, 173(1), 1–10.
- Li, J., & Chen, W. (2014). Forecasting macroeconomic time series: LASSO-based approaches and their forecast combinations with dynamic factor models. *International Journal of Forecasting*, 30(4), 996–1015.
- Marcellino, M., Stock, J. H., & Watson, M. W. (2006). A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series. *Journal of Econometrics*, 135(1–2), 499–526.
- Nazemi, A., & Fabozzi, F. J. (2018). Macroeconomic variable selection for creditor recovery rates. *Journal of Banking & Finance*, 89, 14–25.
- Nicholson, W., Wilms, I., Bien, J., & Matteson, D. S. (2018). High dimensional forecasting via interpretable vector autoregression. arXiv:1412.5250v3.
- Patton, A. J. (2011). Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics*, 160(1), 246–256.
- Poon, S.-H., & Granger, C. (2003). Forecasting volatility in financial markets: A review. *Journal of Economic Literature*, 41(2), 478–539.
- Rombouts, J. V. K., Stentoft, L., & Violante, F. (2020). Variance swap payoffs, risk premia and extreme market conditions. *Econometrics and Statistics*, 13, 106–124.
- Smeekes, S., & Wijler, E. (2018). Macroeconomic forecasting using penalized regression methods. *International Journal of Forecasting*, 34(3), 408–430.
- Taylor, N. (2015). Realized volatility forecasting in an international context. *Applied Economics Letters*, 22(6), 503–509.
- Tian, S., Yu, Y., & Guo, H. (2015). Variable selection and corporate bankruptcy forecasts. *Journal of Banking & Finance*, 52, 89–100.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B*, 58(1), 267–288.
- Tibshirani, R., & Suo, X. T. (2016). An ordered lasso and sparse time-lagged regression. *Technometrics*, 58(4), 415–423.
- Tsay, R. (2014). *Multivariate time series analysis. With R and financial applications*. John Wiley & Sons, Inc.
- Wilms, I., & Croux, C. (2016). Forecasting using sparse cointegration. *International Journal of Forecasting*, 32(4), 1256–1267.
- Yarovaya, L., Brzeszczyński, J., & Lau, C. K. M. (2016). Intra-and inter-regional return and volatility spillovers across emerging and developed markets: Evidence from stock indices and stock index futures. *International Review of Financial Analysis*, 43, 96–114.
- Zellner, A., & Palm, F. (1974). Time series analysis and simultaneous equation econometric models. *Journal of Econometrics*, 2(1), 17–54.