International Review of Economics and Finance 101 (2025) 104171

Contents lists available at ScienceDirect

International Review of Economics and Finance

journal homepage: www.elsevier.com/locate/iref

Forecasting realised volatility using regime-switching models

Yi Ding [a,*], Dimos Kambouroudis [b], David G. McMillan [b]

[a] *Department of Banking and Finance, University of Southampton, UK*
[b] *Division of Accounting and Finance, University of Stirling, UK*

A R T I C L E I N F O

JEL classification:
C22
C53
C58
G32

Keywords:
Realised volatility
Non-linearity
Regime switching
Value at risk
Expected shortfall

A B S T R A C T

This paper extends standard AR and HAR models for realised volatility (RV) forecasting to include nonlinearity through two broad regime-switching approaches, the smooth transition and Markov-switching methods. Using daily data for eight international stock markets over the period 2007–2021, a comprehensive comparison is provided using a range of forecast tests that includes statistical and economic (risk management) based metrics. The results show that regime-switching models provide a better in-sample fit and out-of-sample forecasting, although this latter result is less clear-cut at the daily horizon. In comparing the two nonlinear approaches, we find that the abrupt transition technique of the Markov-switching model is preferred to the smooth transition one. It is believed that our results will be of interest to those especially engaged in risk management practice as well as for those modelling market behaviour.

1. Introduction

High-frequency volatility in financial asset prices and risk management has become increasingly important, creating new theoretical and computational challenges for volatility estimation and prediction. One prominent representative in this sphere is the employment of Realised Volatility (RV), which began with the work of Andersen and Bollerslev (1998) and provides observable proxies of daily volatility. In recent decades, the development of forecasting RV has witnessed debate move in two distinct directions. First, RV is noted for its strong tendency to persist over time and its ability to retain information over the long term (i.e., high persistence; see, for example, Andersen et al., 2003; Lieberman & Phillips, 2008). This is evident in how the lags in data decay gradually and exponentially within the autocorrelation function. Consequently, the widespread application of the RV with high persistence excites new generation of linear volatility models, especially the HAR model and its extensions (see e.g., Andersen et al., 2007; Barndorff-Nielsen et al., 2008; Corsi, 2009; Patton & Sheppard, 2015). Second, several papers provide evidence supporting the view that volatility prediction patterns may take a nonlinear structure, perhaps where the persistence of volatility can vary in a way that is associated with extreme events in the market, including financial crises and sudden policy changes (McAleer & Medeiros, 2008a, 2008b). This research, conducted across a range of forecasting models, demonstrates that nonlinear persistence can enhance RV predictive accuracy, for which previous examples including equity returns (Mcmillan, 2007; Raggi & Bordignon, 2012), interest rates (Bohl et al., 2011), and cryptocurrencies (Caporale & Zekokh, 2019; Cheikh et al., 2020).

The view from the body of literature is that both nonlinearity and high persistence can exist within RV. Thus, to obtain more accurate forecasts, such models should consider alternative approaches to allow for potential state transition of volatility, extending linear models to incorporate nonlinear and regime-switching frameworks. Indeed, recent literature suggests that predictive models

* Corresponding author.
E-mail address: Y.Ding@soton.ac.uk (Y. Ding).

https://doi.org/10.1016/j.iref.2025.104171
Received 8 February 2023; Received in revised form 5 June 2024; Accepted 11 May 2025
Available online 13 May 2025

with regime-switching can provide a richer understanding of both in-sample dynamics and out-of-sample forecasts (see, for example, Cheikh et al., 2020; Gallo & Otranto, 2015; Li et al., 2022; Liu et al., 2020; Luo et al., 2022). However, there is less certainty as to which nonlinear model or which transition function is more suitable for predictive accuracy. A reason why this is important is due to the different potential sources of the nonlinear volatility. One interpretation of nonlinear volatility is that it arises from business cycle dynamics (Hamilton, 1989), which is related to economic expansions and contractions. An alternative explanation, which is more widely applied, suggests that nonlinearity is caused by financial breaks usually associated with major events such as financial crises or sudden changes in policy. Therefore, high regimes are short-lived, whereas low regimes have greater persistence (McAleer & Medeiros, 2008a, 2008b).

To deal with predictive information switching between different states, one common approach is to use the smooth transition method. Following the work of González-Rivera (1998), who proposes the smooth transition GARCH (ST-GARCH) model, Taylor (2004) implements this model with time-varying parameters to capture regime dependency. Khemiri (2011) and Cheikh et al. (2020) utilize the ST-GARCH for international stock index and cryptocurrency markets, respectively. Additionally, the smooth transition model has been combined with the HAR model. For instance, Qu et al. (2016) employ logistic smooth transition in the HAR model and find that it can improve in-sample fit and out-of-sample forecasting performance. Izzeldin et al. (2021) use the exponential smooth transition HAR model to assess the impact of Covid-19 at both aggregate and sectoral levels. Moreover, although two regimes are commonly used, smooth transition models can accommodate multiple regimes. McAleer and Medeiros (2008a, 2008b) introduce a multiple-regime smooth transition HAR model (HARST). However, in their forecasting results, the HARST model performs worse than the linear HAR model and forecasting ability only improves when the HARST and linear HAR models are combined.

An alternative approach to the smooth transition model is the family of nonlinear regime-switching models with unobserved transition variables, i.e., the Markov-switching model. Raggi and Bordignon (2012) provide evidence of such nonlinearity and indicate that accounting for it significantly improves the description of RV. Notably, they demonstrate that nonlinearities improve forecasting ability over several horizons. Subsequently, Ma et al. (2017) report that HAR-type models augmented with a Markov-switching approach yield more accurate forecasts. Ma et al. (2019) also note that adding Markov-switching and the volatility of volatility to HAR-type models can lead to improved forecasting accuracy. The literature notes that the Markov-switching approach combined with HAR models can significantly improve forecasting accuracy.

Following the reported good performance of Markov-switching models, two extensions are considered. First, a drawback of the Markov-switching model is that the transition probability between regimes is fixed. As a result, Bazzi et al. (2017) propose a new Markov-switching model with time-varying transition probability (MS-TVTP) where the transition probability is driven by the mean of observations over time. Subsequently, Wang et al. (2019) and Wang et al. (2022) incorporate the MS-TVTP model with a linear HAR model and find that a MS-HAR model with TVTP can obtain superior forecasting performance compared to both the HAR model and the MS-HAR model with constant transition probability. Second, the variances in the Markov-switching model are state-independent. This means that the variance of each regime is the same; the switching only involves the predictive regression parameters. Kim and Nelson (1999) consider an alternative extension of the MS model that involves variance shifts in the Markov-switching model. Subsequently, Guidolin and Timmermann (2006) extend the Markov-switching model and incorporate heteroscedasticity in the Markov-switching dynamics. In Guidolin et al. (2009, 2014), the Markov-switching model with heteroscedasticity exhibits strong forecasting performance for stock and bond returns.

In this paper, our aim is to perform a systematic forecast performance assessment of the alternate nonlinear frameworks compared to the linear models. Specifically, this paper seeks to improve upon the literature in three ways. First, we explore the question of which transition function can improve RV forecasting performance over linear alternatives, by seeking to identify the preferred nonlinear model. We achieve this by modelling and forecasting the RV of eight international stock indices, including four developed and four emerging markets. We begin with the linear AR and HAR models, and then consider a wide range of smooth transition and Markov-switching frameworks to compare forecast performance. Recent work (e.g., Li et al., 2022; Liu et al., 2020; Wang et al., 2022) only focuses on one or two forms of regime-switching approaches to forecast volatility. Thus, we consider a more comprehensive set of regime-switching models.

In terms of empirical methodology, the second important contribution of this paper is to provide two different extensions of the Markov-switching model. The first extension incorporates time-varying transition probability into the Markov-switching framework. As noted, Markov-switching models are generally specified with constant probabilities, whereas the work of Wang et al. (2019, 2022) and suggest a potential gain by allowing time-varying probability. This paper, therefore, treats the regime probability values as parameters to be estimated. This allows the fitted volatility to dynamically adapt to recent market conditions. The second extension allows variance shifts between regimes, implying heteroscedasticity in the Markov-switching dynamics. The variance is independent of the regime in a simple Markov-switching model, with only the predictive regression parameters switching. We employ a model that allows switching of not only predicting regression parameters, but also variance. Our paper, thus, builds on Guidolin et al. (2009, 2014) and we consider not only the AR model but also the HAR model when estimating heteroscedasticity dynamics.

Third, this paper evaluates the forecasting performance not only from a statistical perspective, similar to existing work, but also considers economic-based forecast evaluation by calculating both Value-at-Risk (VaR) and Expected Shortfall (ES). It is worth noting that the usefulness of a forecasting model is not only in describing data but also in informing market agents or policymakers of its economic application. This we achieve by employing the forecasts to generate VaR and ES across different risk levels. Expanding on previous work, the contribution we seek here is to provide an answer as to whether regime-switching approaches improve forecasts and what is the most appropriate regime-switching function for risk management application.

The main findings are summarized as follows. First, for in-sample results, the regime-switching models are preferred over linear models. Additionally, the Markov-switching models have a better goodness-of-fit than the smooth transition approaches. Second, for out-of-sample results, there is no clear evidence that the regime-switching models improve forecast performance at the daily horizon, where the linear HAR model also performs well. Third, and in contrast, the Markov-switching HAR model with time-varying transition probability dominates forecasting performance for weekly and monthly horizons. Fourth, in regard of risk management applications, our results suggest that regime-switching models, again, provide greater improvements at the weekly and monthly levels, while no model is preferred at the daily level.

The paper is structured as follows: Section 2 introduces the smooth transition and Markov-switching models considered in this paper, as well as the forecasting evaluation methodology. Section 3 describes the data. Section 4 provides the empirical data and findings of in-sample estimation and out-of-sample exercise. Section 5 presents the results of risk management applications, while Section 6 summarises and concludes.

2. Forecasting models and evaluation methodologies

2.1. AR model

According to the RV method of Andersen and Bollerslev (1998), assume that $\mathrm{r}_{t,i} = \mathrm{p}_{t,i} - \mathrm{p}_{t,i-1}, (t = 1, \cdots, T;\ i = 1, \cdots, N)$, where T is the number of trading days, $\mathrm{p}_{t,I}$ is the log asset price at day t and *i* th intraday interval, and $\mathrm{r}_{t,I}$ is the log assets return at day *t* and *i* th intraday interval. Therefore, the RV on trading day *t* (RV_t) can be calculated as:

$$RV_t = \sum_{i=1}^{N} (r_{t,i})^2 \tag{1}$$

The first linear model used in this paper is the autoregressive (AR) model. The AR model is a regression of the variable against itself. In this paper, the AR model is restricted to be an autoregressive order to be one. Thus, the AR(1) model is given as:

$$logRV_t = \theta_0 + \theta_i \log RV_{t-1} + u_t \tag{2}$$

where $u_t \sim iid(0, \sigma^2)$.

2.2. HAR model

Corsi (2009) proposes the HAR model using daily RV, as well as the weekly and monthly averages of RV, the standard HAR-RV model is given as:

$$logRV_t = \beta_0 + \beta_d \log RV_{t-1} + \beta_w logRV_{t-1:t-5} + \beta_m logRV_{t-1:t-22} + u_t \tag{3}$$

where weekly and monthly averages of RV are calculated as:

$$RV_{t-1:t-5} = \frac{1}{5} \sum_{i=1}^{5} RV_{t-i} \tag{4}$$

$$RV_{t-1:t-22} = \frac{1}{22} \sum_{i=1}^{22} RV_{t-i} \tag{5}$$

2.3. Smooth transition models

It is argued that financial volatility can present as two regimes: low volatility and high volatility regimes. Consequently, this paper allows the AR and HAR models to depend on the smooth transition autoregressive (STAR) model in which the transition variable is observed. Different from the abrupt transition of the threshold model, the smooth transition model allows for a gradual transition between regimes. A two-regime smooth transition model is given as:

$$\log RV_t = X_t \alpha + G(s_t; \gamma, \psi) Z_t' \beta + (1 - G(s_t; \gamma, \psi)) Z_t' \delta + \varepsilon_t \tag{6}$$

where X_t is a vector containing regime invariant variables, Z_t is a vector containing regime transition variables, G denotes the continuous transition function with a value between 0 and 1, and ε_t is the stochastic error team. The smooth transition models allow different types of market behaviour according to the nature of transition function G.

First, the logistic smooth transition (LSTAR) model depends on whether the transition variable is above or below the transition value, and the logistic transition function is shown as:

$$G(s_t; c, \gamma) = \frac{1}{1 + exp(-\gamma(s - c))}, \gamma > 0 \tag{7}$$

where s is the smoothing parameter, c is the transition parameter to determine the point at which regimes are weighted equally, and γ is the slope value to control the speed and smoothness of the transition process.

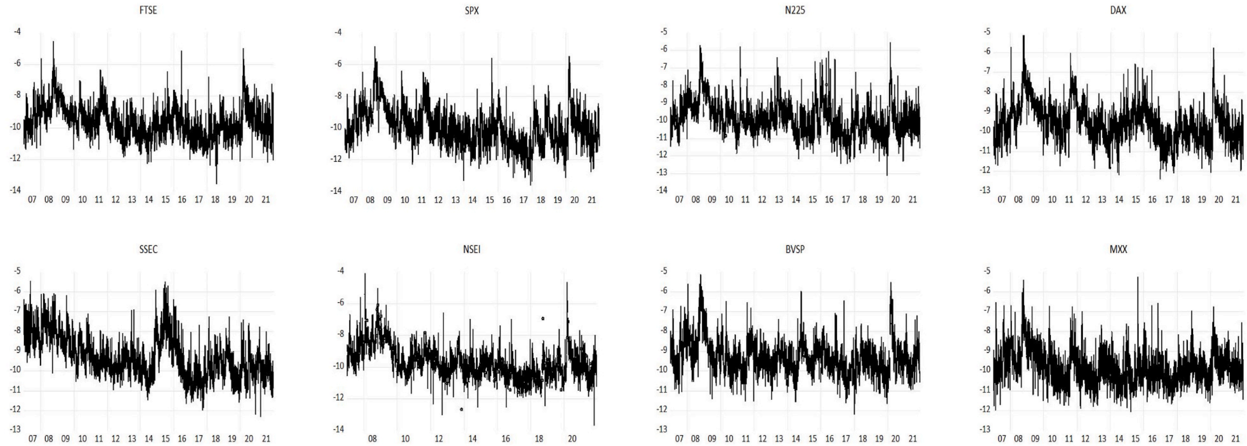

Fig. 1. The plots of the whole period of log-RV of eight market index.

Table 1
Statistic descriptive of Log-RV.

	FTSE	SPX	N225	DAX	SSEC	NSEI	BSVP	MXX
Mean	−9.6370	−9.9985	−9.8629	−9.5424	−9.2883	−9.6866	−9.3072	−9.8401
St.Dev.	1.0198	1.2276	0.9614	0.9809	1.0697	1.0127	0.8794	0.8783
Kurtosis	3.8138	3.3843	4.0353	3.7011	2.9309	4.0384	4.8559	4.1222
Skewness	0.6413	0.5003	0.6564	0.5032	0.5102	0.7194	0.8669	0.7916
Median	−9.7620	−10.1098	−9.9518	−9.6096	−9.4283	−9.8207	−9.3988	−9.9686
25 %-quantile	−10.563	−11.124	−10.679	−10.3924	−10.295	−10.581	−9.949	−10.553
75 %-quantile	−9.434	−9.717	−9.543	−9.2234	−9.168	−9.615	−9.036	−9.599
AutoCorr	0.6350	0.7660	0.7090	0.719	0.7920	0.6450	0.6880	0.5020
Jarque-Bera	364.33***	180.28***	426.28***	238.00***	158.84***	485.41***	991.72***	590.63***
Obs.	3790	3766	3660	3797	3645	3700	3690	3764

Note: This table reports the summary statistics of log-RV of eight stock index for the whole period from 1st January 2007 to 31st December 2021. *** indicate the significant level at 1 %.

Second, the exponential smooth transition (ESTAR) function depends on the distance between the transition value and threshold value, and is given as:

$$G(s_t; c, \gamma) = 1 - exp\left(-\gamma(s-c)^2\right), \gamma > 0 \tag{8}$$

where s is the smoothing parameter, c is the transition parameter, and γ is the slope value.

Following common practice (e.g., Guidolin et al., 2009, 2014), the LSTAR and ESTAR models restrict the lag of the delay parameter to be one. Moreover, this paper also uses the smooth transition HAR model, which allows for a smooth transition between two regimes governed by both logistic and exponential functions, namely the LST-HAR model and EST-HAR model. Due to there being multiple parameters in the HAR model, to determine the regime transition variable, Z_t' in equation (6), the model selected is the one that minimises the sum-of-squared residuals.

2.4. Markov-switching models

An alternative approach to smooth transition models is nonlinear regime-switching models in which the transition variable is unobserved, i.e., the state-dependent process of the Markov-switching model. Specifically, this paper combines the Markov-switching method with the linear HAR model, namely the Markov-switching HAR (MS-HAR) model, given as:

$$logRV_t = \beta_{0,S_t} + \beta_{d,S_t} logRV_{t-1} + \beta_{w,S_t} \log RV_{t-1:t-5} + \beta_{m,S_t} \log RV_{t-1:t-22} + u_{t,S_t} \tag{9}$$

where $u_{t,S_t} \sim N\left(0, v_{t,S_t}\right)$. The variance v_{t,S_t}, constant β_{0,S_t}, HAR model coefficients $\beta_{d,S_t}, \beta_{w,S_t} and \beta_{m,S_t}$ all depend on the unobservable states. From an economic point of view, this paper imposes and estimates a two-state predictive regression ($S_t = 1$ or $S_t = 2$), which represents low and high volatility regimes, respectively. The unobserved regime variable S_t is assumed to follow the first order Markov chain process with a constant transition probability matrix with the generic element P_{ji}, which is defined as:

$$Pr(S_t = i | S_{t-1} = j) = P_{ji} \quad for\ i, j = 1, 2 \tag{10}$$

This is the probability of transition from regime j to regime i between $t-1$ and t. In matrix notation, is shown as:

Table 2
In-sample comparison.

	FTSE	SPX	N225	DAX	SSEC	NSEI	BSVP	MXX
				AIC				
AR	1.8630	2.0324	1.6505	1.6636	1.7278	1.7457	1.8484	2.1919
HAR	1.7055	1.8730	1.5088	1.5304	1.5682	1.5933	1.7129	2.0135
LSTAR	1.8663	2.0309	1.6420	1.6388	1.7179	1.7357	1.8315	2.1893
ESTAR	1.8637	2.0334	1.6449	1.6387	1.7179	1.7358	1.8312	2.1905
LST-HAR	1.7049	1.8685	1.5063	1.5275	1.5673	1.5930	1.7134	2.0114
EST-HAR	1.7017	1.8598	1.6449	1.5241	1.5645	1.5935	1.7080	2.0131
MS-AR	1.8177	1.9925	1.6159	1.6458	1.7019	1.7183	1.8306	2.1418
MS-TVTP-AR	1.8345	1.9924	1.6156	1.6482	1.6974	1.7156	1.8296	2.1441
MSH-AR	1.8194	1.9743	1.6102	1.6175	1.6453	1.6577	1.7895	2.1400
MS-HAR	1.6624	1.8580	1.4465	1.4991	1.5291	1.5394	1.6843	1.9933
MS-TVTP-HAR	1.6655	1.8612	1.4474	1.5114	1.5288	1.5392	1.6894	1.9906
MSH-HAR	**1.6575**	**1.8280**	**1.4417**	**1.4669**	**1.4897**	**1.5024**	**1.6778**	**1.9822**
				BIC				
AR	1.8711	2.0406	1.6589	1.6716	1.7362	1.7540	1.8567	2.2000
HAR	1.7220	1.8896	1.5257	1.5468	1.5852	1.6101	1.7297	2.0301
LSTAR	1.8907	2.0554	1.6671	1.6631	1.7430	1.7606	1.8564	2.2138
ESTAR	1.8881	2.0579	1.6700	1.6630	1.7431	1.7607	1.8561	2.2150
LST-HAR	1.7462	1.9099	1.5487	1.5685	1.6099	1.6350	1.7554	2.0528
EST-HAR	1.7430	1.9011	1.6700	1.5651	1.6070	1.6355	1.7500	2.0545
MS-AR	1.8462	2.0211	1.6452	1.6741	1.7312	1.7473	1.8596	2.1704
MS-TVTP-AR	1.8712	2.0292	1.6532	1.6847	1.7352	1.7529	1.8669	2.1808
MSH-AR	1.8520	2.0070	1.6437	1.6499	1.6789	1.6909	1.8226	2.1726
MS-HAR	1.7078	1.9035	1.4931	1.5442	1.5758	1.5856	1.7306	2.0389
MS-TVTP-HAR	1.7191	1.9150	1.5025	1.5647	1.5841	1.5938	1.7441	2.0444
MSH-HAR	**1.7070**	**1.8776**	**1.4925**	**1.5161**	**1.5407**	**1.5528**	**1.7282**	**2.0319**
				LL				
AR	−1174.5	−1277.4	−1006.5	−1056.0	−1049.3	−1076.0	−1139.4	−1375.6
HAR	−1055.1	−1155.4	−902.03	−953.23	−933.78	−963.12	−1035.8	−1240.4
LSTAR	−1172.6	−1272.5	−997.28	−1036.3	−1039.3	−1065.8	−1125.0	−1370.0
ESTAR	−1170.9	−1274.0	−999.02	−1036.2	−1039.4	−1065.9	−1124.8	−1370.7
LST-HAR	−1048.8	−1146.6	−894.52	−945.45	−927.27	−956.93	−1030.0	−1233.0
EST-HAR	−1046.8	−1141.2	−999.02	−943.30	−925.57	−957.25	−1026.8	−1234.1
MS-AR	−1140.9	−1247.3	−980.34	−1039.7	−1028.6	−1054.0	−1123.4	−1339.1
MS-TVTP-AR	−1149.5	−1245.2	−978.11	−1039.3	−1023.9	−1050.4	−1120.8	−1338.5
MSH-AR	−1140.9	−1234.8	−975.85	−1020.7	−993.19	−1015.6	−1097.0	−1337.0
MS-HAR	−1021.4	−1139.1	−857.63	−926.69	−903.37	−923.41	−1011.4	−1220.9
MS-TVTP-HAR	−1021.3	−1139.1	−856.15	−932.37	−901.23	−921.30	−1012.5	−1217.2
MSH-HAR	**−1017.3**	**−1119.5**	**−853.73**	**−905.54**	**−878.82**	**−899.94**	**−1006.4**	**−1213.0**

Note: This table reports the AIC, BIC and Log Likelihood (LL) of eight RV indices for all forecasting models considered for the in-sample period from 1st January 2007 to 31st December 2011. The forecasting model with the best performance is highlighted with bold fonts.

$$P = \begin{bmatrix} P^{11} & P^{21} \\ P^{12} & P^{22} \end{bmatrix} = \begin{bmatrix} P^{11} & 1 - P^{11} \\ 1 - P^{22} & P^{22} \end{bmatrix} \tag{11}$$

In order to gain more flexibility, the transition probability of the Markov-switching model can follow a time-varying transition probability (MS-TVTP). Following Diebold et al. (1994) and Filardo (1994), we let the transition probability follow an independent regime-switching process, which depends on the exogenous variable, d_i. Equation (10) is changed as:

$$P(S_t = i | S_{t-1} = i) = p_{ii} = \frac{exp(c_i + d_i \delta_{t-1})}{1 + exp(c_i + d_i \delta_{t-1})} \tag{12}$$

where c_i is a constant, d_i is the transition probabilities of MS-TVTP model and if constant would coincide with MS model, and δ_{t-1} is the coefficient value.

In addition, while the variance of the simple MS model is independent of the state $(v_{t,S_1} = v_{t,S_2})$, i.e., the homoscedastic case, this can also be relaxed. Hence, this paper follows Guidolin et al. (2009, 2014) to consider the heteroskedastic case, in which the variance is also regime-specific $(v_{t,S_1} \neq v_{t,S_2})$. This Markov-switching dynamics with heteroscedasticity (MSH) is also combined with AR and HAR models.

2.5. Loss functions

To evaluate and compare the accuracy of all forecasting models, the four loss functions used in this paper are given by:

Table 3
Maximum Likelihood Estimates of MSH-HAR models in in-sample period.

	FTSE		SPX		N225		DAX	
	Regime 1	Regime 2	Regime 1	Regime 2	Regime 1	Regime 2	Regime 1	Regime 2
β_{0,S_t}	−0.8037***	−0.7605	−0.7771***	−0.4207	−0.6785***	−1.2491	−1.0055***	−0.1138
	(0.2242)	(0.8077)	(0.2199)	(0.3679)	(0.1898)	(1.0033)	(0.3025)	(0.2126)
β_{d,S_t}	0.3436***	0.5151***	0.3947***	0.0838	0.3802***	0.8035***	0.4842***	0.3418***
	(0.0358)	(0.1742)	(0.0423)	(0.1511)	(0.0400)	(0.2455)	(0.0527)	(0.0573)
β_{w,S_t}	0.2764***	0.6661**	0.4012***	0.6971***	0.2759***	0.6056	0.3403***	0.4255***
	(0.0585)	(0.2643)	(0.0606)	(0.1735)	(0.0557)	(0.4019)	(0.0809)	(0.0797)
β_{m,S_t}	0.3101***	−0.3017	0.1289***	0.1803*	0.2867***	−0.5955**	0.0697	0.2276***
	(0.0502)	(0.2065)	(0.0479)	(0.1013)	(0.0452)	(0.2768)	(0.0687)	(0.0628)
$\log(\sigma_t)$	−0.8133***	−0.2866***	−0.3980***	−0.9761***	−0.8540***	−0.3999***	−0.4332***	−1.0320***
	(0.0392)	(0.0682)	(0.0273)	(0.0743)	(0.0385)	(0.0994)	(0.0445)	(0.0527)
p_{11}	0.7393		0.9917		0.8698		0.9437	
AIC	1.6575		1.8280		1.4417		1.4669	
BIC	1.7070		1.8776		1.4925		1.5161	
LL	−1017.3		−1119.5		−853.73		−905.54	

	SSEC		**NSEI**		**BSVP**		**MXX**	
	Regime 1	Regime 2	Regime 1	Regime 2	Regime 1	Regime 2	Regime 1	Regime 2
β_{0,S_t}	−1.5537**	−1.0091***	−0.7837	−0.9722***	−0.7358***	−0.8785*	−1.3492**	−1.2961***
	(0.6399)	(0.2472)	(0.6960)	(0.1864)	(0.2279)	(0.5071)	(0.6555)	(0.3056)
β_{d,S_t}	0.3097***	0.3518***	0.3284***	0.3616***	0.3094***	0.5250**	0.3104***	0.1767**
	(0.0752)	(0.0494)	(0.0994)	(0.0481)	(0.0575)	(0.0887)	(0.1062)	(0.0718)
β_{w,S_t}	0.1806	0.1372*	0.2837	0.3534***	0.4019***	0.1976	0.3171	0.4359***
	(0.1290)	(0.0763)	(0.1924)	(0.0693)	(0.0860)	(0.1494)	(0.2146)	(0.1083)
β_{m,S_t}	0.3066**	0.4163***	0.2801*	0.1912***	0.2209***	0.1727	0.2228	0.2705***
	(0.1174)	(0.0679)	(0.1587)	(0.0458)	(0.0613)	(0.1214)	(0.1898)	(0.0823)
$\log(\sigma_t)$	−0.3790***	−0.9323***	−0.2388***	−0.9246***	−0.7806***	−0.2719***	−0.1390**	−0.6683***
	(0.0555)	(0.0509)	(0.0668)	(0.0424)	(0.0444)	(0.0645)	(0.0665)	(0.0813)
p_{11}	0.8893		0.7780		0.9532		0.8214	
p_{22}	0.9491		0.9295		0.8818		0.9145	
AIC	1.4897		1.5024		1.6778		1.9822	
BIC	1.5407		1.5528		1.7282		2.0319	
LL	−878.82		−899.94		−1006.4		−1213.0	

Note: This table reports two-regime MSH-HAR model of each RV index considered for the in-sample period from 1st January 2007 to 31st December 2011. σ is the standard deviation of error in each regime. p_{11} and p_{22} are the transition probability of being in the low and high regime, respectively. Standard errors are in parentheses. ***, ** and * refer the significant at 99 %, 95 % and 90 % confidence level, respectively.

$$QLIKE = \frac{1}{n}\sum_{t=1}^{n}\left(\log(\widehat{RV}_t) + \frac{RV_t}{\widehat{RV}_t}\right) \tag{13}$$

$$MSE = \frac{1}{n}\sum_{t=1}^{n}(RV_t - \widehat{RV}_t)^2 \tag{14}$$

$$MAE = \frac{1}{n}\sum_{t=1}^{n}|RV_t - \widehat{RV}_t| \tag{15}$$

$$MZ\ Regression : RV_t = a_0 + a_1\widehat{RV}_t + \epsilon_t \tag{16}$$

where RV_t and $\widehat{RV}_t$ denote the actual volatility and forecast volatility, respectively.

The loss functions are used to determine the best model with the smallest forecasting error overall time (largest MZ regression R-squared value). However, the forecasting performance of models may change over time, especially during unstable periods. To evaluate whether the forecasting performance changes over time, this paper also considers the cumulative forecast error of the MSE to display the sum of the forecasting error over time.

2.6. Model Confidence Set (MCS) test

When considering the best model to adequately describe the data generating process from several alternative models, this leads to issues about whether (often small) differences in, for example, MSE values are significantly different. To address this, we consider the Model Confidence Set (MCS) test of Hansen et al. (2011). The MCS test utilizes bootstrap implementation and removes the worst performing model sequentially according to the rejection of the null hypothesis of equal predictive ability (EPA). The specific process

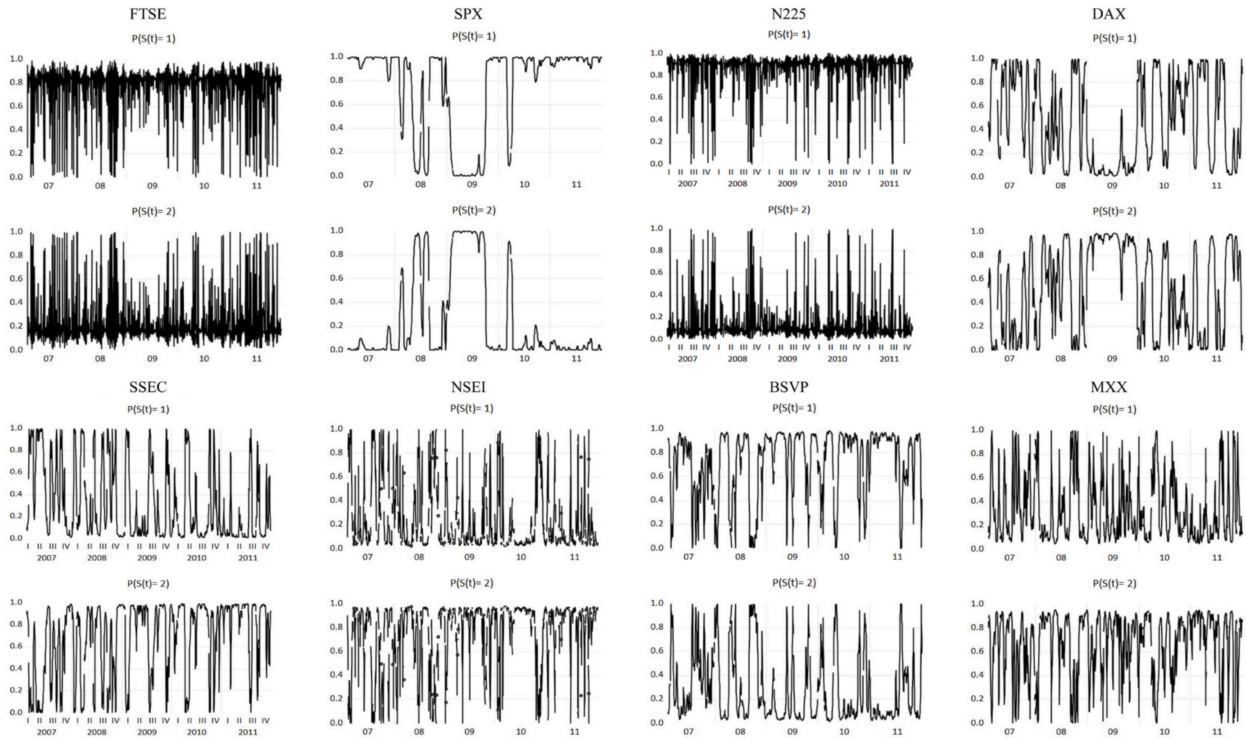

Fig. 2. The smoothed transition probability of MSH-HAR model.

of MCS is as follows. First, assume there are m_0 alternative forecasting models to be tested, so $M_0 = \{1, 2, \cdots, m_0\}$. Let $d_{ij,t}$ demote the loss function differential between any two models at time t:

$$d_{ij,t} = l_{i,t} - l_{j,t} \ (i, j \in M_0) \tag{17}$$

Second, the MCS test is a process that sequentially removes the worst forecasting model from M_0. Thus, in each step, the null hypothesis is set as any two models that have EPA:

$$H_{0,M} : E(d_{ij,t} = 0), \textit{for all } \ i,j \in M_0 \tag{18}$$

$$H_{A,M} : E(d_{ij,t} \neq 0), \textit{for some } \ i,j \in M_0$$

Third, in each step of the MCS test, if the null hypothesis of EPA is rejected at a certain significant level, the worst forecasting model is removed sequentially until the null hypothesis of EPA is not rejected. However, in application of this test and the prediction ability of any two forecast models, requires the test statistic to be recalculate at every step of the process. In order to overcome this limitation, Hansen et al. (2011) construct the Range Statistics and Semi-Quadratic Statistics to test the hypotheses above, and the two tests are shown as follows:

$$T_R = \max_{i,j \in M_0} \left| \frac{\overline{d}_{i,j}}{\sqrt{\widehat{var}(\overline{d}_{i,j})}} \right| \ \textit{and} \ T_{SQ} = \sum_{i,j \in M_0} \frac{(\overline{d}_{i,j})^2}{\widehat{var}(\overline{d}_{i,j})} \tag{19}$$

where $\overline{d}_{i,j}$ is the mean value of the loss functions difference, calculated as $\overline{d}_{i,j} = \frac{1}{M} \sum d_{ij,t}$. Through sequentially removing the worst model, $\widehat{M}_0$ is a subset of models which contains the surviving models from M_0.

3. Data description

The data is obtained from the Oxford-Man Institute of Quantitative Finance. To offset the influence of microstructure noise, our empirical study follows Andersen et al. (2011) and Liu et al. (2015) in employing 5 min RV for eight international stock indices. As most papers focus on developed markets, and as we wish our results to be applicable for a wider range of market participants, our data sample is selected from developed and emerging markets, notably, stock indices for the UK (FTSE), Japan (N225), USA (SPX), Germany (DAX), China (SSEC), Brazil (BVSP), India (NSEI), and Mexico (MXX). The 5 min RVs are used in logarithmic form (log-RV) to produce results with a more normal distribution. We focus on the fifteen-year data sample, ranging from 1st January 2007 to 31st December

Table 4
Out-of-sample forecasting performances of MSE.

	FTSE	SPX	N225	DAX	SSEC	NSEI	BSVP	MXX
				H = 1				
AR	0.5118	0.4953	0.4284	0.3691	0.3535	0.4020	0.2974	0.4367
HAR	0.4157	0.4397	**0.3760**	**0.3032**	**0.2870**	0.3267	0.2583	0.3704
ESTAR	0.5043	0.4938	0.4271	0.3641	0.3512	0.3874	0.2932	0.4351
LSTAR	0.5043	0.4938	0.4285	0.3641	0.3514	0.3864	0.2933	0.4342
EST-HAR	0.4167	0.4403	0.3764	0.3044	0.2886	0.3230	**0.2582**	**0.3701**
LST-HAR	0.4172	0.4405	0.3752	0.3061	0.2876	**0.3226**	0.2591	0.3831
MS-AR	0.4878	0.4742	0.4129	0.3502	0.3375	0.3938	0.2890	0.4141
MS-TVTP-AR	0.4859	0.4735	0.4194	0.3498	0.3382	0.3931	0.2865	0.4057
MSH-AR	0.4902	0.4806	0.4285	0.3434	0.3538	0.3922	0.2840	0.4056
MS-HAR	**0.4151**	0.4392	0.3765	0.3055	0.2904	0.3308	0.2618	0.3699
MS-TVTP-HAR	0.4153	**0.4329**	0.3772	0.3064	0.2877	0.3271	0.2602	0.3685
MSH-HAR	0.4155	0.4392	0.3787	0.3036	0.2897	0.3244	0.2590	0.3755
				H = 5				
AR	0.2643	0.2753	0.2349	0.1789	0.1960	0.2184	0.1539	0.2197
HAR	0.1693	0.2001	0.1683	0.1124	0.1271	0.1367	0.1065	0.1561
ESTAR	0.2555	0.2717	0.2325	0.1764	0.1923	0.2018	0.1467	0.2187
LSTAR	0.2571	0.2720	0.2334	0.1764	0.1924	0.2023	0.1467	0.2181
EST-HAR	0.1685	0.2018	0.1690	0.1129	0.1282	0.1333	0.1052	0.1568
LST-HAR	0.1732	0.2357	0.1693	0.1132	0.1320	0.1335	0.1064	0.1598
MS-AR	0.2015	0.2419	0.1938	0.1530	0.1628	0.1793	0.1341	0.1581
MS-TVTP-AR	0.2026	0.2254	0.1891	0.1400	0.1528	0.1659	0.1322	0.1524
MSH-AR	0.2075	0.2368	0.1894	0.1382	0.1620	0.1760	0.1276	0.1581
MS-HAR	0.1327	0.1707	0.1398	0.0900	0.0977	0.1071	0.0917	0.1258
MS-TVTP-HAR	**0.1273**	**0.1690**	**0.1370**	**0.0877**	**0.0975**	**0.1052**	**0.0911**	**0.1203**
MSH-HAR	0.1627	0.1901	0.1566	0.1067	0.1127	0.1293	0.0980	0.1434
				H = 22				
AR	0.2897	0.3734	0.2829	0.2159	0.2505	0.2527	0.1786	0.2298
HAR	0.1377	0.2011	0.1584	0.0960	0.0986	0.1107	0.0827	0.1189
ESTAR	0.2818	0.3653	0.2829	0.2064	0.2405	0.2385	0.1702	0.2258
LSTAR	0.2816	0.3652	0.2837	0.2060	0.2406	0.2384	0.1701	0.2251
EST-HAR	0.1369	0.1980	0.1588	0.0931	0.0979	0.1086	0.0804	0.1182
LST-HAR	0.1346	0.1996	0.1623	0.1029	0.0979	0.1223	0.0805	0.1265
MS-AR	0.1934	0.2511	0.1756	0.1204	0.1641	0.1461	0.1115	0.1172
MS-TVTP-AR	0.1737	0.2085	0.1756	0.1198	0.1457	0.1459	0.1122	0.1062
MSH-AR	0.2020	0.2485	0.1669	0.1371	0.1422	0.1598	0.1117	0.1218
MS-HAR	0.0932	**0.1225**	0.0900	**0.0605**	0.0602	0.0742	0.0526	0.0674
MS-TVTP-HAR	**0.0855**	0.1264	**0.0885**	0.0682	**0.0574**	**0.0731**	**0.0512**	**0.0642**
MSH-HAR	0.1159	0.1669	0.1410	0.0804	0.0625	0.0875	0.0691	0.0908

Note: This table reports the forecasting evaluation (MSE) of eight RV indices for all forecasting models considered using recursive method over daily, weekly and monthly horizons (h = 1, 5 and 22) and the out-of-sample period from 1st January 2012 to 31st December 2021. The forecasting model with the best performance is highlighted with bold fonts.

2021. The first five years are set as the in-sample period (i.e., 1st January 2007 to 31st December 2011), and the last ten years (i.e., 1st January 2012 to 31st December 2021) are used in the out-of-sample forecast generation.

Fig. 1 shows the time plots of all log-RVs for the full sample period. Those plots reveal several periods of high volatility. For all indices, the financial crisis of 2007–2009 and the global pandemic in 2020 lead to the most extreme levels of volatility. Particular high volatility is related different major events such the European sovereign debt crisis in 2015 and Chinese stock market turbulence around the same time. Table 1 presents the summary statistics of the daily 5 min log-RV of eight stock indices. All series of log-RV exhibit a non-normal distribution with excess kurtosis and is right-skewed. Further evidence of non-normality is given by the Jarque-Bera test statistic. The first-order autocorrelation values indicate that all log-RV series are highly persistent and allow for further modelling analysis.

4. Empirical results

In-sample estimation results.

We start with two linear models, the AR and basic HAR models, as the benchmark. We then extend these models using the smooth

Table 5
Out-of-sample forecasting performances of MAE.

	FTSE	SPX	N225	DAX	SSEC	NSEI	BSVP	MXX
				H = 1				
AR	0.5510	0.5589	0.5016	0.4765	0.4633	0.4771	0.4214	0.5138
HAR	0.4883	0.5241	0.4627	**0.4337**	**0.4071**	0.4259	0.3859	0.4639
ESTAR	0.5463	0.5604	0.5016	0.4754	0.4617	0.4694	0.4166	0.5123
LSTAR	0.5464	0.5603	0.5012	0.4753	0.4618	0.4694	0.4165	0.5114
EST-HAR	0.4904	0.5264	0.4634	0.4351	0.4087	0.4241	**0.3845**	0.4635
LST-HAR	0.4883	0.5256	**0.4615**	0.4366	0.4079	**0.4238**	0.3854	0.4767
MS-AR	0.5347	0.5468	0.4916	0.4636	0.4514	0.4718	0.4155	0.5015
MS-TVTP-AR	0.5343	0.5475	0.4961	0.4636	0.4505	0.4718	0.4108	0.4944
MSH-AR	0.5376	0.5505	0.5015	0.4608	0.4623	0.4742	0.4097	0.4912
MS-HAR	**0.4880**	0.5235	0.4629	0.4354	0.4107	0.4273	0.3867	0.4635
MS-TVTP-HAR	0.4881	**0.5194**	0.4633	0.4360	0.4074	0.4261	0.3862	**0.4625**
MSH-HAR	0.4883	0.5239	0.4642	0.4339	0.4096	0.4250	0.3858	0.4665
				H = 5				
AR	0.3915	0.4082	0.3670	0.3304	0.3446	0.3557	0.3024	0.3581
HAR	0.2953	0.3244	0.2871	0.2495	0.2632	0.2586	0.2308	0.2776
ESTAR	0.3860	0.4074	0.3659	0.3281	0.3410	0.3488	0.2947	0.3562
LSTAR	0.3866	0.4079	0.3657	0.3281	0.3409	0.3496	0.2944	0.3557
EST-HAR	0.2956	0.3305	0.2892	0.2511	0.2648	0.2586	0.2312	0.2783
LST-HAR	0.3034	0.3603	0.2880	0.2507	0.2773	0.2589	0.2313	0.2825
MS-AR	0.3386	0.3828	0.3367	0.3033	0.3136	0.3233	0.2811	0.2986
MS-TVTP-AR	0.3396	0.3601	0.3315	0.2863	0.2989	0.3045	0.2764	0.2941
MSH-AR	0.3403	0.3747	0.3308	0.2838	0.3111	0.3186	0.2692	0.2925
MS-HAR	0.2616	0.2933	0.2673	0.2175	0.2257	0.2265	0.2150	0.2523
MS-TVTP-HAR	**0.2568**	**0.2925**	**0.2648**	**0.2104**	**0.2254**	**0.2285**	**0.2150**	**0.2512**
MSH-HAR	0.2837	0.3086	0.2724	0.2388	0.2410	0.2454	0.2184	0.2616
				H = 22				
AR	0.4202	0.4826	0.4191	0.3665	0.3947	0.3957	0.3307	0.3724
HAR	0.2615	0.3168	0.2857	0.2267	0.2293	0.2288	0.2070	0.2371
ESTAR	0.4160	0.4797	0.4179	0.3586	0.3863	0.3883	0.3223	0.3678
LSTAR	0.4160	0.4793	0.4182	0.3582	0.3860	0.3886	0.3223	0.3672
EST-HAR	0.2614	0.3168	0.2849	0.2248	0.2279	0.2238	0.2045	0.2357
LST-HAR	0.2584	0.3181	0.2917	0.2439	0.2299	0.2589	0.2049	0.2527
MS-AR	0.3529	0.3955	0.3293	0.2735	0.3208	0.2874	0.2586	0.2640
MS-TVTP-AR	0.3293	0.3536	0.3288	0.2732	0.2966	0.2867	0.2599	0.2505
MSH-AR	0.3538	0.3824	0.3146	0.2856	0.2927	0.3052	0.2571	0.2648
MS-HAR	0.2224	**0.2563**	**0.2197**	**0.1808**	0.1794	0.1913	0.1679	0.1825
MS-TVTP-HAR	**0.2141**	0.2570	0.2208	0.1922	**0.1749**	0.1913	**0.1664**	**0.1796**
MSH-HAR	0.2222	0.2684	0.2495	0.1926	0.1781	**0.1904**	0.1739	0.1940

Note: This table reports the forecasting evaluation (MAE) of eight RV indices for all forecasting models considered using recursive method over daily, weekly and monthly horizons (h = 1, 5 and 22) and the out-of-sample period from 1st January 2012 to 31st December 2021. The forecasting model with the best performance is highlighted with bold fonts.

transition and Markov-switching frameworks. As noted above, this paper considers the two-regime predictive regression in the smooth transition and Markov-switching frameworks, which represents the low and high-regime volatility. The first five years of data samples are set as the in-sample period.[1]

Table 2 presents three model selection criteria for each forecasting model, namely the Akaike information criteria (AIC), Bayesian (or Schwarz) information criterion (BIC), and log-likelihood (LL), respectively. Overall, all three criteria support the MSH-HAR model as providing the best in-sample fit among all forecast models. Comparing the linear AR and HAR models, the nonlinear regime-switching models of the smooth transition and Markov-switching are generally preferred in terms of these criteria value. This in-sample result is consistent with the work of Raggi and Bordignon (2012), which suggests accounting for high persistence and non-linearities leads to an improved description of the data. In addition, we find the Markov-switching approach provides a better goodness-of-fit, compared to the smooth transition models.

As the MSH-HAR model is preferred based on the model selection criteria, we provide some further estimation details for this model. Table 3 reports the parameter estimates of the MSH-HAR for each index. In this model, all parameters, including the variance, are allowed to switch between two regimes ($S_t = 1$ and $S_t = 2$) that represent low and high volatility. Table 3 notably shows that the MSH-HAR models do have different standard deviations of the error, σ_t, across each regime, and are statistically significant at the 1 %

[1] As the recursive method are used to obtain forecasts in the paper, the in-sample estimates will change each period continuously, so the final in-sample estimates are quite different from the initial ones. The in-sample results presented here are the initial in-sample estimates, ranging from 1st January 2007 to 31st December 2011, it is included largely for illustrative purpose.

Table 6
Out-of-sample forecasting performances of QLIKE.

	FTSE	SPX	N225	DAX	SSEC	NSEI	BSVP	MXX
				H = 1				
AR	0.002793	0.002461	0.002284	0.001998	0.002089	0.002115	0.001746	0.002322
HAR	0.002253	0.002168	**0.002005**	**0.001634**	**0.001689**	**0.001714**	**0.001510**	0.001964
ESTAR	0.002762	0.002446	0.002284	0.001979	0.002076	0.002079	0.001723	0.002319
LSTAR	0.002763	0.002447	0.002295	0.001979	0.002079	0.002068	0.001723	0.002313
EST-HAR	0.002256	0.002170	0.002008	0.001643	0.001698	0.001697	0.001509	0.001965
LST-HAR	0.002266	0.002175	0.002006	0.001648	0.001694	0.001693	0.001514	0.002030
MS-AR	0.002653	0.002345	0.002199	0.001892	0.001987	0.002066	0.001694	0.002195
MS-TVTP-AR	0.002643	0.002343	0.002237	0.001894	0.001994	0.002070	0.001681	0.002158
MSH-AR	0.002670	0.002366	0.002281	0.001849	0.002091	0.002070	0.001671	0.002149
MS-HAR	0.002250	0.002166	0.002010	0.001648	0.001711	0.001752	0.001538	0.001960
MS-TVTP-HAR	**0.002249**	**0.002135**	0.002018	0.001658	0.001695	0.001719	0.001525	**0.001953**
MSH-HAR	0.002251	0.002165	0.002022	0.001637	0.001704	0.001703	0.001514	0.001995
				H = 5				
AR	0.001483	0.001412	0.001278	0.000996	0.001179	0.001184	0.000925	0.001208
HAR	0.000950	0.001026	0.000918	0.000623	0.000758	0.000754	0.000648	0.000858
ESTAR	0.001447	0.001387	0.001271	0.000988	0.001159	0.001115	0.000886	0.001207
LSTAR	0.001458	0.001389	0.001277	0.000988	0.001160	0.001115	0.000885	0.001201
EST-HAR	0.000945	0.001031	0.000921	0.000627	0.000763	0.000736	0.000641	0.000864
LST-HAR	0.000993	0.001188	0.000927	0.000627	0.000786	0.000738	0.000648	0.000880
MS-AR	0.001144	0.001231	0.001050	0.000851	0.000974	0.000986	0.000803	0.000867
MS-TVTP-AR	0.001150	0.001157	0.001018	0.000783	0.000920	0.000911	0.000790	0.000829
MSH-AR	0.001189	0.001211	0.001030	0.000776	0.000974	0.000971	0.000769	0.000876
MS-HAR	0.000746	0.000878	0.000769	0.000496	0.000579	0.000599	0.000558	0.000695
MS-TVTP-HAR	**0.000715**	**0.000871**	**0.000749**	**0.000487**	**0.000579**	**0.000588**	**0.000553**	**0.000657**
MSH-HAR	0.000917	0.000976	0.000859	0.000594	0.000678	0.000720	0.000598	0.000793
				H = 22				
AR	0.001661	0.001963	0.001561	0.001226	0.001537	0.001413	0.001102	0.001256
HAR	0.000801	0.001070	0.000885	0.000555	0.000587	0.000650	0.000517	0.000645
ESTAR	0.001625	0.001911	0.001572	0.001180	0.001483	0.001362	0.001049	0.001240
LSTAR	0.001625	0.001908	0.001582	0.001178	0.001483	0.001360	0.001049	0.001235
EST-HAR	0.000794	0.001042	0.000888	0.000537	0.000584	0.000635	0.000499	0.000642
LST-HAR	0.000779	0.001058	0.000907	0.000592	0.000584	0.000708	0.000501	0.000691
MS-AR	0.001089	0.001308	0.000949	0.000685	0.000972	0.000823	0.000678	0.000633
MS-TVTP-AR	0.000988	0.001105	0.000950	0.000682	0.000870	0.000823	0.000682	0.000574
MSH-AR	0.001148	0.001305	0.000924	0.000785	0.000861	0.000908	0.000690	0.000682
MS-HAR	0.000542	**0.000653**	0.000506	**0.000353**	0.000354	0.000446	0.000330	0.000372
MS-TVTP-HAR	**0.000494**	0.000680	**0.000494**	0.000399	**0.000340**	**0.000434**	**0.000320**	**0.000354**
MSH-HAR	0.000673	0.000888	0.000797	0.000464	0.000370	0.000523	0.000432	0.000500

Note: This table reports the forecasting evaluation (QLIKE) of eight RV indices for all forecasting models considered using recursive method over daily, weekly and monthly horizons (h = 1, 5 and 22) and the out-of-sample period from 1st January 2012 to 31st December 2021. The forecasting model with the best performance is highlighted with bold fonts.

level. Of note, the estimated p_{22} of FTSE and N225 is much smaller than p_{11}, which means the volatility will switch between regime 1 and 2 more frequently. Here, the switching structure is more abrupt in comparison to other countries, suggesting less persistence in regime 2.

The smoothed transition probability series, shown as $P(S_t = 1)$ and $P(S_t = 2)$, estimated from MSH-HAR models are plotted in Fig. 2. In addition to the observation regarding the FTSE and N225, we can see that the transition probabilities of the four emerging markets are more volatile than for the developed markets. Volatility generally lies in regime 1 with short switches to regime 2 during 2007–2008, the volatility series tend to switch back to regime 2 and remain there in 2009, the volatility series mainly in regime 1 and after 2010. It is roughly consistent with the Global Financial Crisis that affected the macro-economy in 2008.

4.1. Out-of-sample forecasting results

The out-of-sample forecasts, ranging from 1st January 2012 to 31st December 2021, are recursively generated from all forecasting models. In addition, we also report the results for 5- and 22-day-ahead forecasts, approximately corresponding to one trading week and month. For long-term forecasting, we replace the data frequency of the volatility model to generate multi-step-ahead forecasts. Moreover, forecasting performance is measured by the loss functions of MSE, MAE, QLIKE, and the adjusted R^2 of the MZ regression, as well as the MCS tests of MSE and QLIKE criterion to select the optimal models with EPA.

The out-of-sample results for the four loss functions are reported over daily, weekly, and monthly horizons in Tables 4–7. Table 4 presents the MSE loss functions over the daily, weekly, and monthly horizons (h = 1, 5, and 22). The results for the one-day-ahead forecasts are mixed with no single model consistently outperforming. The HAR model performs the best for three series, whilst the EST-HAR model offers good performance for two series, with three other models preferred for a single series. For the one-week-ahead

Table 7
Out-of-sample forecasting performances of MZ regression adjusted R2.

	FTSE	SPX	N225	DAX	SSEC	NSEI	BSVP	MXX
				H = 1				
AR	0.4000	0.5861	0.5029	0.5184	0.6283	0.4186	0.4712	0.2453
HAR	0.4939	0.6210	**0.5566**	**0.5918**	**0.6869**	0.4926	**0.5299**	0.3319
ESTAR	0.4053	0.5885	0.5042	0.5198	0.6304	0.4294	0.4724	0.2416
LSTAR	0.4055	0.5882	0.5030	0.5198	0.6302	0.4292	0.4723	0.2436
EST-HAR	0.4926	0.6215	0.5559	0.5897	0.6854	**0.4968**	0.5291	0.3298
LST-HAR	0.4918	0.6214	0.5570	0.5874	0.6863	0.4966	0.5277	0.3180
MS-AR	0.4218	0.6010	0.5185	0.5390	0.6429	0.4261	0.4835	0.2745
MS-TVTP-AR	0.4232	0.6025	0.5117	0.5386	0.6415	0.4233	0.4852	0.2769
MSH-AR	0.4144	0.5963	0.5026	0.5404	0.6278	0.4302	0.4854	0.2827
MS-HAR	**0.4942**	0.6214	0.5562	0.5892	0.6841	0.4872	0.5226	0.3306
MS-TVTP-HAR	0.4941	**0.6255**	0.5559	0.5874	0.6863	0.4932	0.5258	**0.3339**
MSH-HAR	**0.4942**	0.6214	0.5543	0.5908	0.6848	0.4960	0.5283	0.3249
				H = 5				
AR	0.5992	0.7215	0.6616	0.7046	0.7678	0.5978	0.6460	0.4800
HAR	0.7309	0.7869	0.7528	0.8047	0.8407	0.7189	0.7435	0.6058
ESTAR	0.6079	0.7269	0.6643	0.7041	0.7713	0.6158	0.6537	0.4761
LSTAR	0.6059	0.7265	0.6635	0.7040	0.7713	0.6144	0.6538	0.4774
EST-HAR	0.7314	0.7859	0.7517	0.8033	0.8392	0.7238	0.7447	0.6027
LST-HAR	0.7303	0.7559	0.7510	0.8028	0.8398	0.7234	0.7437	0.6020
MS-AR	0.6803	0.7544	0.7203	0.7395	0.8074	0.6515	0.6817	0.6003
MS-TVTP-AR	0.6786	0.7594	0.7224	0.7575	0.8088	0.6560	0.6797	0.6114
MSH-AR	0.6765	0.7588	0.7227	0.7616	0.8083	0.6623	0.6919	0.6079
MS-HAR	0.7898	0.8188	0.7942	0.8429	0.8772	0.7771	0.7777	0.6789
MS-TVTP-HAR	**0.7962**	**0.8190**	**0.7981**	**0.8465**	**0.8774**	**0.7796**	**0.7781**	**0.6925**
MSH-HAR	0.7421	0.7984	0.7697	0.8151	0.8596	0.7351	0.7644	0.6366
				H = 22				
AR	0.4776	0.5525	0.5199	0.5719	0.6851	0.5002	0.4974	0.3147
HAR	0.7299	0.7365	0.7221	0.7898	0.8625	0.7339	0.7443	0.5945
ESTAR	0.4832	0.5651	0.5185	0.5791	0.6925	0.5133	0.5043	0.3137
LSTAR	0.4834	0.5651	0.5190	0.5803	0.6925	0.5118	0.5043	0.3141
EST-HAR	0.7309	0.7414	0.7214	0.7951	0.8624	0.7372	0.7492	0.5949
LST-HAR	0.7354	0.7388	0.7157	0.7847	0.8627	0.7335	0.7482	0.5951
MS-AR	0.6410	0.7705	0.7087	0.7339	0.7890	0.6399	0.6494	0.5989
MS-TVTP-AR	0.6601	0.7243	0.7081	0.7356	0.7959	0.6406	0.6474	0.6305
MSH-AR	0.6212	0.6978	0.7210	0.7115	0.8008	0.6489	0.6560	0.6129
MS-HAR	0.8175	0.8373	0.8424	**0.8661**	0.9177	0.8162	0.8354	0.7642
MS-TVTP-HAR	**0.8318**	**0.8374**	**0.8447**	0.8497	**0.9191**	**0.8190**	**0.8389**	**0.7735**
MSH-HAR	0.7717	0.7803	0.7542	0.8238	0.9146	0.7893	0.7858	0.6872

Note: This table reports the Mincer-Zarnowitz regression adjusted R^2 of eight RV indices for all forecasting models considered using recursive method over daily, weekly and monthly horizons (h = 1, 5 and 22) and the out-of-sample period from 1st January 2012 to 31st December 2021. The forecasting model with the highest adjusted R^2 is highlighted with bold fonts.

forecasts, the MS-TVTP-HAR model provides the best forecast performance across all indices. Additionally, while the Markov-switching models provide more accurate forecasts than smooth transition models, the set of nonlinear models outperform the linear AR and HAR models. The MS-TVTP-HAR model also generally has the best forecasting ability at the monthly horizon, while the MSH-HAR model is preferred for the SPX and DAX.

Table 5 shows the MAE forecast results, consistent with Table 4, the AR and its smooth transition extensions do not present strong forecasts. At the one-day horizon, the best model is, again, mixed with evidence for the HAR model and its nonlinear extensions. The MS-TVTP-HAR model dominates all stock indices at the one-week-ahead horizon, while varinats of the MS-HAR model are preferred for one-month-ahead forecasts. For both the QLIKE results in Table 6 and the adjusted R^2 of MZ regression in Table 7, they present similar results to those in Tables 4 and 5 The results are mixed at the one-day horizon with some evidence for the HAR model but also its nonlinear extensions. At the weekly and (to a lesser extent) monthly horizons, the MS-TVTP-HAR model dominates.

To further strengthen the out-of-sample results and observe the forecasting error for all models over time, the cumulative forecasting error of the MSE loss function over the daily, weekly, and monthly horizons are plotted in Figs. 3–5. The daily forecasting error in Fig. 3 shows that the AR and regime-switching AR models perform worse, while there is no observable difference in daily forecasting error among the HAR model, smooth transition HAR models, and Markov-switching HAR models for all indices. In Fig. 4, it can be observed that the MS-TVTP-HAR model (purple line) generates the lowest cumulative MSE all the time. This reveals that the MS-TVTP-HAR model consistently maintains the most accurate weekly forecasting performance over time. The three types of Markov-switching models can also be seen to perform better than smooth transition models. The cumulative monthly forecasting error in Fig. 5 presents similar results to Fig. 4. The MS-TVTP-HAR model produces the lowest forecasting error, except for DAX and SPX, for which the MS-

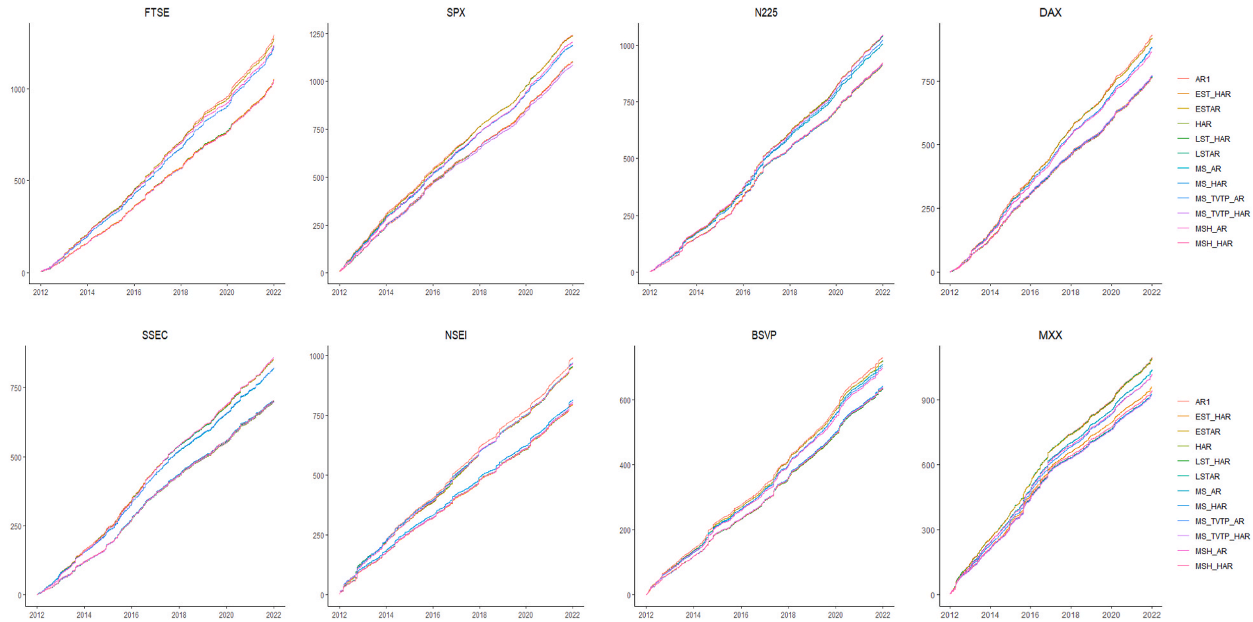

Fig. 3. The sum of cumulative MSE forecasting error for one-day-ahead forecast.

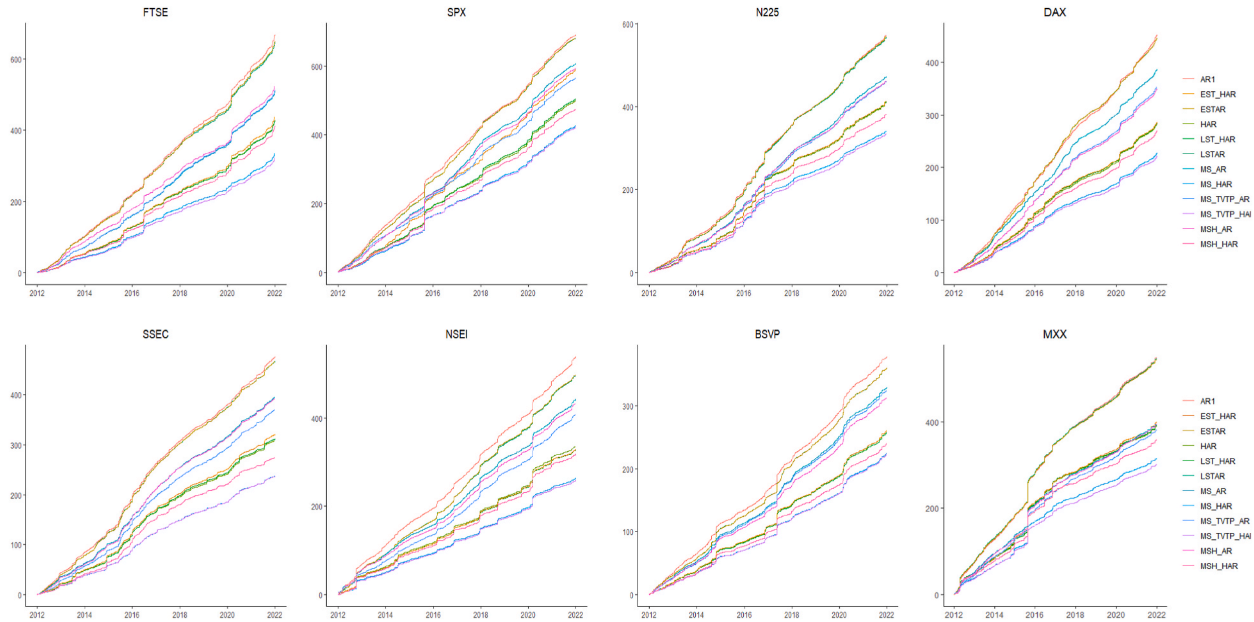

Fig. 4. The sum of cumulative MSE forecasting error for one-week-ahead forecast.

HAR model performs well.

The Range statistics of the MCS test results for the MSE and QLIKE criteria are provided in Tables 8 and 9.[2] The MCS test chooses a subset of models with EPA at the 90 % confidence level, the value 1 in the table means that the optimal model is chosen, and the value 0 means the model is eliminated. For the MCS test of the MSE criterion in Table 8, the AR model and its regime-switching models perform worse, with all eliminated from EPA selection. For daily forecasts, the results of the best model are mixed, the HAR model, the LST-HAR model and the MS-TVTP-HAR model perform equally well. For weekly and monthly forecasting horizons, there is overwhelming evidence in favour of the MS-TVTP-HAR model across all stock indices.

[2] We do not present the Semi-Quadratic statistics, as it gives practically the same results as the Range statistics.

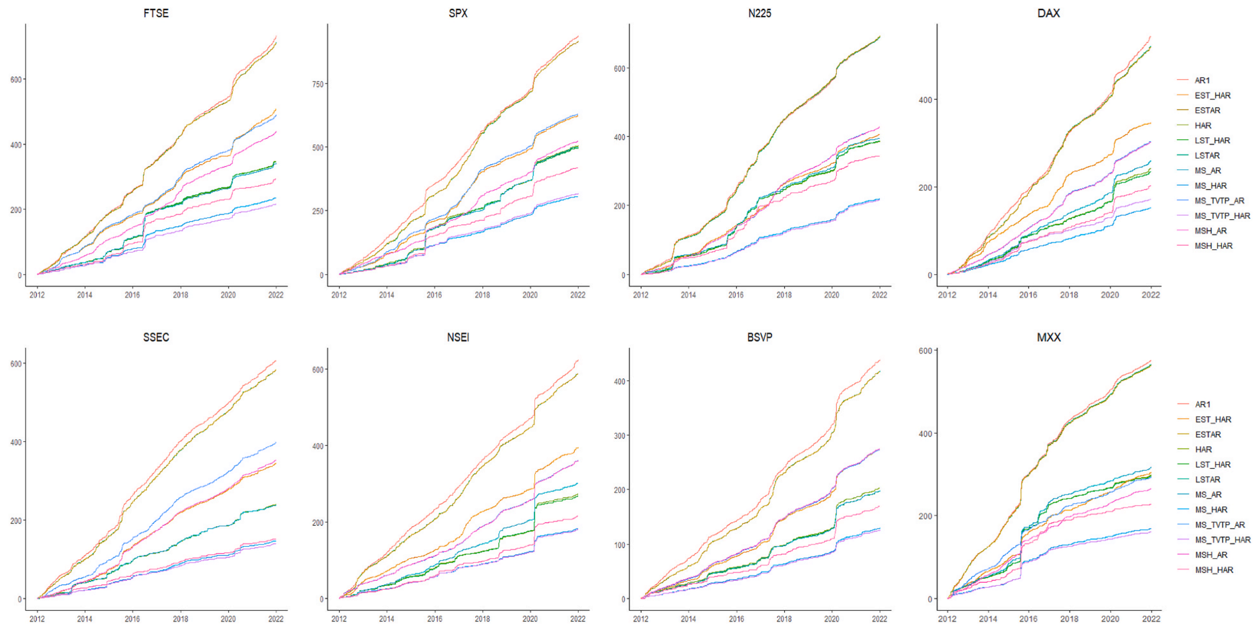

Fig. 5. The sum of cumulative MSE forecasting error for one-month-ahead forecast.

Table 9 presents the MCS test of the QLIKE criterion. Compared to Table 8, similar results are obtained. The linear HAR model and the MS-TVTP-HAR model outperform for daily forecasting. Specifically, the linear HAR model provides the best forecasting performance for N225, DAX and SSEC, and the MS-TVTP-HAR model is the best model for FTSE, SPX, and MXX. Again, the weekly and monthly results indicate that the MS-TVTP-HAR model is the best performing forecast model, whilst the MS-HAR also has good forecasting performance on SPX and DAX at the monthly horizon.

To sum, according to the different forecast metrics and the MCS test, the linear HAR model has broadly equivalent (and possibly better) predictive ability to the regime-switching models at the daily forecasting horizon, with only weak evidence in support of the nonlinear models. In contrast, the MS-TVTP-HAR model achieves a superior forecasting performance across all stock indices at the weekly horizon and for most at the monthly. In addition, we also find the Markov-switching models provide more accurate forecasts than smooth transition models. Compared with existing work, the results here, differ slightly from Wang et al. (2019) and Zhang et al. (2020). We show that the need for nonlinear regime-switching is weak at the one-day-ahead forecast horizon. We emphasise, however, that the Markov-switching model with time-varying transition probability has better performance for the long-term forecasting horizon. Of further interest, there are no clearly observed differences between the developed and emerging countries in the out-of-sample forecast comparison.

5. Risk management application

Since a key application of financial volatility forecasting concerns risk management, we further consider evaluation based upon Value-at-Risk (VaR) and Expected Shortfall (ES). VaR is calculated to measure the maximum amount of loss for financial assets under a certain confidence level, and ES is designed to measure the expected loss value when a VaR violation has occurred. The VaR of an asset is calculated as:

$$VaR = \mu_t + \sigma_t N(\alpha) \tag{21}$$

where μ_t is the mean of asset's log-return, σ_t is the predicted volatility, and $N(\alpha)$ defines the left α th quantile of the normal distribution. The calculation of ES given as:

$$ES = \mu_t + \sigma_t \frac{f(N(\alpha))}{1-\alpha} \tag{22}$$

where μ_t and σ_t are defined as above and $f(N(\alpha))$ is the density function of a standard normal distribution at the left α th quantile.

To examine the accuracy of VaR forecasts, we compute the violation rate (VR, or exceedances) for daily return, which is the number of daily returns exceed the forecasted VaR divided by the total forecasted observations. VaR accuracy is also examined using both the Kupiec (Kupiec, 1995) and Christoffersen (Christoffersen, 1998) tests, which are the unconditional and conditional coverage tests for the correct number of exceedances. Specifically, the Kupiec test is an unconditional coverage test with the null hypothesis that the observed violation rate is statistically equal to the expected violation rate. Christoffersen's conditional coverage test extends this analysis and examines whether exceedances are independent. Here, the null hypothesis that exceedances occur independently at every

Table 8
The Model confidence set test of MSE criterion.

	FTSE	SPX	N225	DAX	SSEC	NSEI	BSVP	MXX
				H = 1				
AR	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
HAR	0.0000	0.0000	0.0000	**1.0000**	**1.0000**	0.0000	0.9336	0.0000
ESTAR	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
LSTAR	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
EST-HAR	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	**1.0000**	0.0000
LST-HAR	0.0000	0.0000	**1.0000**	0.0000	0.0000	**1.0000**	0.0000	0.0000
MS-AR	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
MS-TVTP-AR	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
MSH-AR	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
MS-HAR	**1.0000**	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
MS-TVTP-HAR	0.0000	**1.0000**	0.0000	0.0000	0.0000	0.0000	0.0000	**1.0000**
MSH-HAR	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
				H = 5				
AR	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
HAR	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
ESTAR	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
LSTAR	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
EST-HAR	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
LST-HAR	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
MS-AR	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
MS-TVTP-AR	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
MSH-AR	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
MS-HAR	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
MS-TVTP-HAR	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**
MSH-HAR	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
				H = 22				
AR	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
HAR	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
ESTAR	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
LSTAR	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
EST-HAR	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
LST-HAR	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
MS-AR	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
MS-TVTP-AR	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
MSH-AR	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
MS-HAR	0.0000	**1.0000**	0.0000	**1.0000**	0.0000	0.0000	0.0000	0.0000
MS-TVTP-HAR	**1.0000**	0.0000	**1.0000**	0.0000	**1.0000**	**1.0000**	**1.0000**	**1.0000**
MSH-HAR	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Note: This table reports the MCS test in term of MSE criterion for eight RV indices over daily, weekly and monthly horizons (h = 1, 5 and 22). The forecasting models with EPA at 90 % confidence level are highlighted in table. The value 1 in the table means that the optimal model is chosen, the value 0 means the model is eliminated.

point in time and against the alternative hypothesis that the failure rate exhibits clustering. Both tests are carried out in the likelihood ratio (LR) framework. The LR for each test is given as:

1. LR statistic for the test of correct unconditional coverage:

$$LR_{UC} = 2\,log\left((1-\pi_0)^{T-N}\pi_0^N\right) - 2\,\log\left((1-\alpha_0)^{T-N}\alpha_0^N\right) \sim \chi_1^2 \tag{23}$$

where π_0 is the observed violation rate and calculated by the number of the days N when violations occurred divide by forecasting period T.

2. LR statistic for the test of correct conditional coverage:

$$LR_{CC} = 2\,log\left((1-\pi_{01})^{n_{00}}\pi_{01}^{n_{01}}(1-\pi_{11})^{n_{10}}\pi_{11}^{n_{11}}\right) - 2\,\log\left((1-\pi_0)^{n_{00}+n_{10}}\pi_0^{n_{01}+n_{11}}\right) \sim \chi_1^2 \tag{24}$$

where the $n_{i,j}$ refer to the number of i values followed by j (for $i,j = 0,1$), $\pi_{i,j}$ is the probability that i occurs at time t followed by the j occurs at time $t-1$.

We further consider the Dynamic Quantile (DQ) test of Engle and Manganelli (2004) to examine if present VaR violations are not correlated with past violations. The DQ test defines the hit sequence as follows:

$$Hit_t = I(r_t < -VaR_t) - a \tag{25}$$

This sequence assumes that value $(1-a)$ whenever the actual returns are less than the VaR quantile and the value $(-a)$ otherwise.

Table 9
The Model confidence set test of QLIKE criterion.

	FTSE	SPX	N225	DAX	SSEC	NSEI	BSVP	MXX
H = 1								
AR	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
HAR	0.0000	0.0000	**1.0000**	**1.0000**	**1.0000**	0.0000	0.9542	0.0000
ESTAR	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
LSTAR	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
EST-HAR	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	**1.0000**	0.0000
LST-HAR	0.0000	0.0000	0.8936	0.0000	0.0000	**1.0000**	0.0000	0.0000
MS-AR	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
MS-TVTP-AR	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
MSH-AR	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
MS-HAR	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
MS-TVTP-HAR	**1.0000**	**1.0000**	0.0000	0.0000	0.0000	0.0000	0.0000	**1.0000**
MSH-HAR	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
H = 5								
AR	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
HAR	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
ESTAR	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
LSTAR	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
EST-HAR	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
LST-HAR	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
MS-AR	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
MS-TVTP-AR	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
MSH-AR	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
MS-HAR	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
MS-TVTP-HAR	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**
MSH-HAR	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
H = 22								
AR	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
HAR	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
ESTAR	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
LSTAR	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
EST-HAR	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
LST-HAR	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
MS-AR	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
MS-TVTP-AR	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
MSH-AR	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
MS-HAR	0.0000	**1.0000**	0.0000	**1.0000**	0.0000	0.0000	0.0000	0.0000
MS-TVTP-HAR	**1.0000**	0.0000	**1.0000**	0.0000	**1.0000**	**1.0000**	**1.0000**	**1.0000***
MSH-HAR	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Note: This table reports the MCS test in term of QLIKE criterion for eight RV indices over daily, weekly and monthly horizons (h = 1, 5 and 22). The forecasting models with EPA at 90 % confidence level are highlighted in table. The value 1 in the table means that the optimal model is chosen, the value 0 means the model is eliminated.

The expected value of Hit_t is zero and the sequence is uncorrelated with past information. In this case, there will be no autocorrelation in the hit sequence and the fraction of exceptions will be correct. The DQ test statistic is calculated as:

$$DQ = \frac{\widehat{\beta}' X' X \widehat{\beta}}{a(1-a)} \sim \chi^2(k) \tag{26}$$

where X is the explanatory variables and $\widehat{\beta}$ is the OLS estimates. The DQ test follows a χ^2 distribution with degree of freedom equal to the number of parameters.

Where VaR backtesting selects more than one model, the Quantile Loss (QL, Koenker & Bassett, 1978) is uded to rank the models and identify the best VaR performance. The QL for predicted VaR of confidence level α at time t given as:

$$QL_t(\alpha) = (\alpha - d_t)(r_t - VaR_t(\alpha)) \tag{27}$$

QL is an asymmetric loss function. As the weight $(1-\alpha)$ increases, the penalties will be heavier for returns that exceed the VaR. Models with lower average QL are preferred.

As opposed to VaR, there is no specific loss function for ES (see, for example, Bellini & Bignozzi, 2015; Ziegel, 2016). However, the loss function introduced by Fissler and Ziegel (2016, FZL) shows that VaR and ES can be assessed jointly. For predicted VaR and ES at risk level α for time t, the joint VaR and ES loss function of FZL given as:

$$FZL_t(\alpha) = \frac{1}{\alpha ES_t(\alpha)} d_t(r_t - VaR_t(\alpha)) + \frac{VaR_t(\alpha)}{ES_t(\alpha)} + log(-ES_t(\alpha)) - 1 \tag{28}$$

Table 10
Summary results of VaR backtesting at 1, 2.5 and 5 % levels.

	α = 0.01			α = 0.025			α = 0.05		
	LRuc	LRcc	DQ	LRuc	LRcc	DQ	LRuc	LRcc	DQ
				H = 1					
AR	0	0	0	0	0	0	0	0	0
HAR	0	0	0	0	0	0	0	0	0
ESTAR	0	0	0	0	0	0	0	0	0
LSTAR	0	0	0	0	0	0	0	0	0
EST-HAR	0	0	0	0	0	0	0	0	0
LST-HAR	0	0	0	0	0	0	0	0	0
MS-AR	0	0	0	0	0	0	0	0	0
MS-TVTP-AR	0	0	0	0	0	0	0	0	0
MSH-AR	0	0	0	0	0	0	0	0	0
MS-HAR	0	0	0	0	0	0	0	0	0
MS-TVTP-HAR	0	0	0	0	0	0	0	0	0
MSH-HAR	0	0	0	0	0	0	0	0	0
				H = 5					
AR	0	0	0	0	0	0	0	0	0
HAR	0	0	0	0	0	0	0	0	0
ESTAR	0	0	0	0	0	0	0	1	1
LSTAR	0	0	0	0	0	0	0	1	1
EST-HAR	0	0	0	0	0	0	0	0	0
LST-HAR	0	0	0	0	0	0	0	0	0
MS-AR	0	0	0	0	0	0	0	0	1
MS-TVTP-AR	0	0	0	0	0	0	0	0	0
MSH-AR	0	0	0	0	0	0	1	2	1
MS-HAR	0	0	0	0	0	0	0	0	0
MS-TVTP-HAR	0	0	0	0	0	0	1	1	1
MSH-HAR	0	0	0	0	0	0	0	0	0
				H = 22					
AR	1	1	1	1	2	2	6	7	7
HAR	0	0	0	0	0	0	0	1	3
ESTAR	1	1	1	1	1	3	6	7	6
LSTAR	1	1	1	1	1	3	6	7	6
EST-HAR	0	0	1	0	0	0	4	3	3
LST-HAR	0	0	1	0	0	0	1	1	2
MS-AR	0	0	0	0	1	2	4	4	5
MS-TVTP-AR	0	0	0	0	0	0	4	4	3
MSH-AR	1	1	1	0	1	2	6	6	7
MS-HAR	0	0	0	0	1	1	1	3	3
MS-TVTP-HAR	1	0	0	0	1	1	3	5	3
MSH-HAR	0	0	1	0	1	1	1	2	3

Notes: This table provides the number of indices that each model can pass the VaR backtesting, namely the Kupiec test (LRuc), Christoffersen test (LRcc) and Dynamic Quantile test (DQ), at the 1, 2.5, and 5 % levels for daily, weekly and monthly forecasting horizon.

where $ES_t(\alpha) \ll VaR_t(\alpha) < 0$. The losses of FZL are averaged over the forecasting period and the model with the lowest average value is preferred.

We use the same in-sample and out-of-sample periods to produce the daily, weekly, and monthly VaR and ES forecasts for three different risk levels: 1 %, 2.5 % and 5 %. Table 10 reports the summarized results of VaR backtesting applied to 1-day, 1-week and 1-month-ahead forecasts. For each risk level, this table presents the number of times each index passes the VaR backtesting, the unconditional and conditional coverage tests, as well as the DQ test. For daily VaR forecasting results, all the tests are statistically significant, and so none of the models pass the Kupiec, Christoffersen, and DQ backtesting tests across all risk levels. This means that over this horizon all models have excessive VaR violations and those VaR violations occur dependently and are correlated. For weekly VaR forecasts, equally, none of the models pass the backtesting at 1 and 2.5 % risk levels. But for a number of indices, backtesting is passed for the MS-HAR and MS-TVTP-HAR models. For monthly VaR forecasts, multiple models pass the backtesting at all risk levels, although this occurs with greater frequency at longer horizons.

Table 11 presents the average value of violation rate, QL and FZL across all stock indices to consider accurate joint predictions in terms of VaR and ES. In this table, we highlight the model with the closest violation rate to the nominal risk levels and the lowest QL and FSL value. The LSTAR model achieves the lowest VR, QL and FZL value at 1 and 2.5 % risk level at the day horizon. For the weekly results, in regard of the lowest average VR value at all risk levels, the MS-AR model is preferred while the MS-TVTP-HAR model offers the best performance for averaged QL and FZL values. Examining the monthly results, the MS-AR model generally performs well across the risk levels, although the difference in values between models is small.

To sum, all models do reject the null hypothesis of the Kupiec, Christoffersen, and DQ tests at the daily forecast horizon, which means all forecasting models exhibit excessive VaR violations that are dependent and correlated. Thus, while there is no clear evidence that the nonlinear regime-switching models improve forecasting ability at the daily horizon, although there is some support for the

Table 11
Model comparison summary of VaR and ES.

	α = 0.01			α = 0.025			α = 0.05		
	ave. VR	ave. QLoss $*10^{-4}$	ave. FSLoss	ave. VR	ave. QLoss $*10^{-4}$	ave. FSLoss	ave. VR	ave. QLoss $*10^{-3}$	ave. FSLoss
				H = 1					
AR	2.9307 %	3.5022	−3.2604	4.8277 %	6.5927	−3.5780	7.3128 %	1.0650	−3.7647
HAR	3.3082 %	3.5394	−3.2463	5.0229 %	6.5927	−3.5694	7.6357 %	1.0608	−3.7636
ESTAR	2.8600 %	3.4932	−3.2834	4.8178 %	6.5873	−3.5891	7.3173 %	1.0654	−3.7715
LSTAR	**2.8552 %**	**3.4891**	**−3.2857**	**4.7875 %**	**6.5834**	**−3.5907**	7.3074 %	1.0649	**−3.7725**
EST-HAR	3.2728 %	3.5459	−3.2527	4.9836 %	6.5975	−3.5708	7.4947 %	1.0613	−3.7641
LST-HAR	3.2423 %	3.5401	−3.2558	5.0783 %	6.5934	−3.5740	7.6050 %	1.0611	−3.7667
MS-AR	2.9955 %	3.4949	−3.2573	4.8065 %	6.5764	−3.5790	**7.2916 %**	1.0621	−3.7663
MS-TVTP-AR	2.9452 %	3.4980	−3.2681	4.7961 %	6.5840	−3.5835	7.3066 %	1.0639	−3.7689
MSH-AR	2.9755 %	3.5132	−3.2642	4.9826 %	6.6153	−3.5805	7.3723 %	1.0674	−3.7668
MS-HAR	3.2675 %	3.5386	−3.2497	5.0180 %	6.5922	−3.5714	7.6306 %	**1.0606**	−3.7651
MS-TVTP-HAR	3.2675 %	3.5480	−3.2483	5.0580 %	6.6037	−3.5700	7.6758 %	1.0620	−3.7643
MSH-HAR	3.2828 %	3.5358	−3.2527	5.0480 %	6.5951	−3.5716	7.6052 %	1.0608	−3.7644
				H = 5					
AR	2.5283 %	3.3661	−3.3325	4.2629 %	6.4538	−3.6102	6.5837 %	1.0547	−3.7813
HAR	2.8145 %	3.3812	−3.3254	4.4246 %	6.4456	−3.6085	6.9206 %	1.0481	−3.7821
ESTAR	2.4627 %	3.3546	−3.3584	4.2326 %	6.4405	−3.6220	6.5981 %	1.0543	−3.7882
LSTAR	2.4724 %	3.3548	−3.3586	4.2326 %	6.4432	−3.6224	6.5623 %	1.0544	−3.7883
EST-HAR	2.7185 %	3.3689	−3.3333	4.3141 %	6.4283	−3.6094	6.7187 %	1.0466	−3.7820
LST-HAR	2.7744 %	3.3796	−3.3382	4.3946 %	6.4418	−3.6139	6.8552 %	1.0481	−3.7858
MS-AR	**2.2892 %**	3.1052	−3.5053	**4.0761 %**	6.1295	−3.7083	**6.5372 %**	1.0210	−3.8503
MS-TVTP-AR	2.4163 %	3.1513	−3.4719	4.2748 %	6.1697	−3.6963	6.8218 %	1.0241	−3.8454
MSH-AR	2.3301 %	3.1323	−3.4861	4.1104 %	6.1533	−3.7024	6.5528 %	1.0234	−3.8482
MS-HAR	2.5337 %	3.0760	−3.5028	4.2888 %	6.0928	−3.7142	6.7254 %	1.0134	−3.8565
MS-TVTP-HAR	2.4984 %	**3.0521**	**−3.5173**	4.2139 %	**6.0551**	**−3.7226**	6.7104 %	**1.0081**	**−3.8635**
MSH-HAR	2.6839 %	3.2770	−3.3894	4.3382 %	6.3278	−3.6471	6.7991 %	1.0376	−3.8088
				H = 22					
AR	2.0437 %	3.2318	−3.4191	**3.5558 %**	6.3332	−3.6469	**5.6300 %**	1.0458	−3.7958
HAR	2.4373 %	3.3020	−3.3825	4.0821 %	6.3861	−3.6317	6.2498 %	1.0470	−3.7914
ESTAR	2.0138 %	3.2207	−3.4318	3.5564 %	6.3273	−3.6518	5.6957 %	1.0469	−3.7979
LSTAR	2.0188 %	3.2199	−3.4320	3.5362 %	6.3261	−3.6524	5.6356 %	1.0470	−3.7983
EST-HAR	2.3574 %	3.2751	−3.3916	3.9716 %	6.3660	−3.6362	6.1053 %	1.0449	−3.7922
LST-HAR	2.4024 %	3.2927	−3.3890	4.0716 %	6.3819	−3.6344	6.2914 %	1.0468	−3.7927
MS-AR	**1.9905 %**	**3.0318**	**−3.5368**	3.6401 %	**6.0743**	**−3.7187**	5.8333 %	1.0176	−3.8497
MS-TVTP-AR	2.1311 %	3.0635	−3.5272	3.7918 %	6.1048	−3.7155	6.0598 %	1.0192	**−3.8498**
MSH-AR	2.0360 %	3.0627	−3.5237	3.6151 %	6.1209	−3.7108	5.7540 %	1.0221	−3.8427
MS-HAR	2.2814 %	3.0423	−3.5223	3.9149 %	6.0956	−3.7124	6.1999 %	1.0169	−3.8473
MS-TVTP-HAR	2.2463 %	3.0469	−3.5192	3.9111 %	6.0926	−3.7114	6.1316 %	**1.0154**	−3.8464
MSH-HAR	2.3212 %	3.1482	−3.4646	3.9446 %	6.2132	−3.6822	6.2249 %	1.0288	−3.8251

Notes: this table provides the average value of VaR and ES comparison results at the 1, 2.5 and 5 % level for daily, weekly and monthly forecasting horizon. The average violation rate, the average asymmetric Quantile Loss function (QL) and the Average Fissler and Ziegel (2016) Loss function (FZL) for each model over each index. Bold fonts highlights the forecasting model with best performance.

smooth transition AR models, for weekly and monthly horizons, the MS-HAR and the MS-TVTP-HAR models do exhibit better performance.

6. Summary and conclusion

Previous work has shown that nonlinear RV dynamics can improve predictive performance. To systematically assess the predictive power of nonlinear regime-switching models in combination with AR and HAR models, we comprehensively integrate and compare forecasting performance across RV for eight stocks indices. Using the linear AR and HAR models as benchmarks, this paper considers both smooth transition and Markov-switching methods for different prediction time horizons. In addition, the standard Markov-switching model is extended to include time-varying transition probability and heteroscedasticity in the switching dynamics. Following convention, we employ two regimes representing low and high volatility respectively. In addition to standard statistical forecast evaluations, risk management applications are used through calculating Value-at-Risk (VaR) and Expected Shortfall (ES).

The in-sample results support the view that nonlinear regime-switching models exhibit a better goodness-of-fit than linear models. Moreover, the Markov-switching model that incorporates variance regime switching (i.e., the MSH-HAR model) is preferred. For the out-of-sample results, there is no consistent evidence that nonlinear regime-switching models outperform the HAR model at the daily forecasting horizon. However, for both the weekly and monthly horizons, the MS-TVTP-HAR model dominates forecasting performance. For the risk management applications, at the daily horizon, arguably the smooth transition AR model performs best, but all models fail tests associated with the violation rate. The MS-TVTP-HAR and MS-HAR models are preferred at the weekly and monthly horizons, respectively. Of note, there are no consistent differences in forecast performance between the developed and emerging country results.

In consideration of previous work, Cheikh et al. (2020) and Liu et al. (2020) examine two regimes linked to smooth transition variables and find that the high volatility regime tends to be transient. Other work considers the Markov-switching approach. For example, Zhang et al. (2020) utilize this model for international stock markets, while Luo et al. (2022) report enhanced portfolio performance using various Markov regime-switching structures. Studies by Lu et al. (2021) and Li et al. (2022) confirm that Markov-switching models outperform others in volatility forecasting accuracy for the US stock and crude oil markets, respectively. Wang et al. (2022) highlight dynamic transition functions in the Markov-switching process for the Chinese stock market. However, across all of these papers there is no comprehensive comparison of the smooth transition and Markov-switching methods, nor is there work that indicates which transition function provides a better understanding of RV dynamics. Therefore, the contribution presented here is that we provided a comprehensive assessment of nonlinear regime-switching approaches for a range of international stock markets.

It might not be surprising that our findings are more favourable to regime-switching frameworks than linear models, which has typically been found in previous papers (see, for example, Raggi & Bordignon, 2012, among others noted in the Introduction). While we present similar findings in respect of nonlinear models having a stronger forecasting performance, our results go further in reporting which regime-switching approach can generate more accurate predictions. The two nonlinear regime-switching models are quite different in nature; the smooth transition model allows for a gradual transition between two regimes, whereas the transition in the Markov switching model is abrupt and unobservable. Empirically, the abrupt transition technique is better suited to the markets considered here.

Author statement

We declare that this manuscript is original, has not been published before and is not currently being considered for publication elsewhere.

Declarations of interest

None.

References

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

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*, 701–720.

Andersen, T. G., Bollerslev, T., Diebold, F. X., & Labys, P. (2003). Modeling and forecasting realized volatility. *Econometrica, 71*, 579–625.

Andersen, T. G., Bollerslev, T., & Huang, X. (2011). A reduced form framework for modeling volatility of speculative prices based on realized variation measures. *Journal of Econometrics, 160*, 176–189.

Barndorff-Nielsen, O. E., Kinnebrock, S., & Shephard, N. (2008). Measuring downside risk - realised semivariance. *SSRN Electronic Journal*.

Bazzi, M., Blasques, F., Koopman, S. J., & Lucas, A. (2017). Time-varying transition probabilities for Markov regime switching models. *Journal of Time Series Analysis, 38*, 458–478.

Bellini, F., & Bignozzi, V. (2015). On elicitable risk measures. *Quantitative Finance, 15*, 725–733.

Bohl, M. T., Salm, C. A., & Wilfling, B. (2011). Do individual index futures investors destabilize the underlying spot market? *Journal of Futures Markets, 31*, 81–101.

Caporale, G. M., & Zekokh, T. (2019). Modelling volatility of cryptocurrencies using Markov-Switching GARCH models. *Research in International Business and Finance, 48*, 143–155.

Cheikh, N. B., Zaied, Y. B., & Chevallier, J. (2020). Asymmetric volatility in cryptocurrency markets: New evidence from smooth transition GARCH models. *Finance Research Letters, 35*, Article 101293.

Christoffersen, P. F. (1998). Evaluating interval forecasts. *International Economic Review*, 841–862.

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

Diebold, F. X., Lee, J.-., & Weinbach, G. C. (1994). Regime switching with time-varying transition probabilities. *Nonstationary Time Series Analysis and Cointegration*, 144–165.

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

Filardo, A. J. (1994). Business-cycie phases and their transitional dynamics. *Journal of Business & Economic Statistics, 12*(3), 299–308.

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

Gallo, G. M., & Otranto, E. (2015). Forecasting realized volatility with changing average levels. *International Journal of Forecasting, 31*(3), 620–634.

González-Rivera, G. (1998). Smooth-transition GARCH models. *Studies in Nonlinear Dynamics and Econometrics, 3*.

Guidolin, M., Hyde, S., McMillan, D. G., & Ono, S. (2009). Non-linear predictability in stock and bond returns: When and where is it exploitable? *International Journal of Forecasting, 25*, 373–399.

Guidolin, M., Hyde, S., McMillan, D. G., & Ono, S. (2014). Does the macroeconomy predict UK asset returns in a nonlinear fashion? Comprehensive out-of-sample evidence. *Oxford Bulletin of Economics & Statistics, 76*, 510–535.

Guidolin, M., & Timmermann, A. (2006). An econometric model of nonlinear dynamics in the joint distribution of stock and bond returns. *Journal of Applied Econometrics, 21*, 1–22.

Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica: Journal of the Econometric Society*, 357–384.

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

Izzeldin, M., Muradoğlu, Y. G., Pappas, V., & Sivaprasad, S. (2021). The impact of Covid-19 on G7 stock markets volatility: Evidence from a ST-HAR model. *International Review of Financial Analysis, 74*, Article 101671.

Khemiri, R. (2011). The smooth transition GARCH model: Application to international stock indexes. *Applied Financial Economics, 21*, 555–562.

Kim, C.-J., & Nelson, C. R. (1999). Has the US economy become more stable? A bayesian approach based on a markov-switching model of the business cycle. *The Review of Economics and Statistics, 81*, 608–616.
Koenker, R., & Bassett, G. (1978). Regression quantiles. *Econometrica: Journal of the Econometric Society*, 33–50.
Kupiec, P. H. (1995). Techniques for verifying the accuracy of risk measurement models. *Division of Research and Statistics, Division of Monetary Affairs, Federal*.
Li, X., Liang, C., Chen, Z., & Umar, M. (2022). Forecasting crude oil volatility with uncertainty indicators: New evidence. *Energy Economics, 108*, Article 105936.
Lieberman, O., & Phillips, P. C. (2008). Refined inference on long memory in realized volatility. *Econometric Reviews, 27*, 254–267.
Liu, L. Y., Patton, A. J., & Sheppard, K. (2015). Does anything beat 5-minute RV? A comparison of realized measures across multiple asset classes. *Journal of Econometrics, 187*, 293–311.
Liu, M., Taylor, J. W., & Choo, W.-C. (2020). Further empirical evidence on the forecasting of volatility with smooth transition exponential smoothing. *Economic Modelling, 93*, 651–659.
Lu, B., Ma, F., Wang, J., Ding, H., & Wahab, M. (2021). Harnessing the decomposed realized measures for volatility forecasting: Evidence from the US stock market. *International Review of Economics & Finance, 72*, 672–689.
Luo, J., Klein, T., Ji, Q., & Hou, C. (2022). Forecasting realized volatility of agricultural commodity futures with infinite Hidden Markov HAR models. *International Journal of Forecasting, 38*, 51–73.
Ma, F., Lu, X., Yang, K., & Zhang, Y. (2019). Volatility forecasting: Long memory, regime switching and heteroscedasticity. *Applied Economics, 51*, 4151–4163.
Ma, F., Wahab, M., Huang, D., & Xu, W. (2017). Forecasting the realized volatility of the oil futures market: A regime switching approach. *Energy Economics, 67*, 136–145.
McAleer, M., & Medeiros, M. C. (2008a). Realized volatility: A review. *Econometric Reviews, 27*, 10–45.
McAleer, M., & Medeiros, M. C. (2008b). A multiple regime smooth transition heterogeneous autoregressive model for long memory and asymmetries. *Journal of Econometrics, 147*(1), 104–119.
Mcmillan, D. G. (2007). Bubbles in the dividend–price ratio? Evidence from an asymmetric exponential smooth-transition model. *Journal of Banking & Finance, 31*, 787–804.
Patton, A. J., & Sheppard, K. (2015). Good volatility, bad volatility: Signed jumps and the persistence of volatility. *The Review of Economics and Statistics, 97*, 683–697.
Qu, H., Chen, W., Niu, M., & Li, X. (2016). Forecasting realized volatility in electricity markets using logistic smooth transition heterogeneous autoregressive models. *Energy Economics, 54*, 68–76.
Raggi, D., & Bordignon, S. (2012). Long memory and nonlinearities in realized volatility: A Markov switching approach. *Computational Statistics & Data Analysis, 56*, 3730–3742.
Taylor, J. W. (2004). Volatility forecasting with smooth transition exponential smoothing. *International Journal of Forecasting, 20*, 273–286.
Wang, J., Ma, F., Liang, C., & Chen, Z. (2022). Volatility forecasting revisited using Markov-switching with time-varying probability transition. *International Journal of Finance & Economics, 27*, 1387–1400.
Wang, X., Shrestha, K., & Sun, Q. (2019). Forecasting realised volatility: A Markov switching approach with time-varying transition probabilities. *Accounting and Finance, 59*, 1947–1975.
Zhang, Y., Lei, L., & Wei, Y. (2020). Forecasting the Chinese stock market volatility with international market volatilities: The role of regime switching. *The North American Journal of Economics and Finance, 52*, Article 101145.
Ziegel, J. F. (2016). Coherence and elicitability. *Mathematical Finance, 26*, 901–918.