Quantitative Finance

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/rquf20

Short-term volatility forecasting with kernel support vector regression and Markov switching multifractal model

Khaldoun Khashanah & Chenjie Shao

To cite this article: Khaldoun Khashanah & Chenjie Shao (2022) Short-term volatility forecasting with kernel support vector regression and Markov switching multifractal model, Quantitative Finance, 22:2, 241-253, DOI: 10.1080/14697688.2021.1939116

To link to this article: https://doi.org/10.1080/14697688.2021.1939116

Published online: 12 Jul 2021.

Quantitative Finance, 2022
Vol. 22, No. 2, 241–253, https://doi.org/10.1080/14697688.2021.1939116

Short-term volatility forecasting with kernel support vector regression and Markov switching multifractal model

KHALDOUN KHASHANAH* and CHENJIE SHAO

Stevens Institute of Technology, Hoboken, NJ, USA

(*Received 2 June 2019; accepted 28 May 2021; published online 12 July 2021*)

In volatility forecasting literature, Markov switching multifractal (MSM) models are well known for capturing many important stylized facts such as long memory and fat tails. MSM delivers stronger performance both in- and out-of-sample than GARCH-type models in long-term forecasts. However, the literature shows that MSM forecasts only slightly improve on GARCH(1,1) at short-term intervals. This indicates that there may exist certain patterns to be discovered in the innovation part ε_t. To enhance MSM's prediction accuracy at the short-term level with higher frequency data, a hybrid model of the MSM model and support vector regression (SVR) is proposed, in which a particle swarm optimization (PSO) algorithm is applied to optimize hyperparameters of the support vector regression in the scope of constraint permission. The method is referred to as MSM-PSO-SVR. Further, we introduce the Fourier kernel MSM-PSO-SVR and evaluate the performance of various MSM-PSO-SVR models in terms of mean absolute error (MAE) and the mean squared error (MSE) with one-minute data of the exchange traded fund (ETF) SPDR S&P 500 Trust ETF (ticker symbol: SPY). The experimental results show that the proposed approach outperforms the other competing peer models and in particular, the selection of SVR kernel might yield significant boosts in forecasting ability. Results of Hansen's Superior Predictive Ability test further validate the conclusion.

Keywords: Short-term volatility forecasting; MSM-PSO-SVR; Fourier kernel; Kernel selection

JEL Classification: C53, C14, C45, C61

1. Introduction

Financial volatility forecasting is one of the core research questions in finance. Accurate modeling of volatility dynamics plays a major role in portfolio allocation, derivative pricing and risk management (Granger and Poon 2003, Andersen *et al.* 2005, 2006). Volatility forecasting at higher-than-daily frequency informs on intraday volatility fluctuations with an impact on trading and pricing. As explained subsequently, the contribution of this paper is to construct a hybrid† model of Markov switching multifractal (MSM) and support vector regression (SVR) optimized by particle swarm optimization (PSO) to better-forecast higher frequency volatility time series.

Volatility forecasting is challenging due to various stylized facts including the long-memory property, thick tails, volatility persistence, and outliers. In the huge literature on volatility forecasting, the vast majority of research uses variants of the GARCH family models (Granger and Poon 2003). The ARFIMA model is able to generate autocorrelation, which tends to decay rapidly (Granger and Joyeux 2008, Brailsford and Faff 1996). However, these authors do not capture the long memory existing in volatility (Lux and Segon 2018). Fractionally Integrated GARCH (FIGARCH) was proposed by Baillie *et al.* (1996) to reproduce the long memory property of financial return volatility. Although some results speak in favor of FIGARCH models in volatility forecasting (Kang *et al.* 2009), studies have shown that the MSM models can forecast future volatility more accurately than FIGARCH (Calvet and Fisher 2003, Lux and Kaizoji 2006, Lux 2006). Also with Mandelbrot's famous work on the fluctuations of cotton prices in the early sixties (Mandelbrot 1963), researchers had learned that the standard Geometric Brownian motion proposed by Bachelier (1900) is unable to reproduce the stylized facts of financial time series (Lux and Segon 2018). In particular, the fat tails and the strong correlations observed in volatility are in sharp contrast to the 'mild',

*Corresponding author. Email: kkhashan@stevens.edu
† A hybrid model here also means a *model of models*.

uncorrelated fluctuations implied by models with Brownian random terms (Lux and Segon 2018, Calvet and Fisher 2013).

Multifractal processes have been proposed as a new method for modeling financial volatility because both volatility time series and multifractal processes share certain properties such as long memory and fat tails (i.e., Mandelbrot 2001a, 2001b, 2001c, 2001d, Di Matteo 2007, Lux and Morales-Arias 2013, Calvet and Fisher 2002. In contrast to GARCH family models and stochastic volatility models, the source of heterogeneity of volatility in multifractal models stems from the time-variation of local regularity in the price path (Fisher *et al.* 1997)).

Multifractal models are based on power-law scaling thereby generating the power-law decay in autocorrelation and capturing the long memory property. The theory of multifractal measures was originally developed by Mandelbrot (1974) for the distribution of energy in turbulent dissipation and thereafter initiating a literature on multifractal processes in statistical physics such as Kahane and Peyriére (1976), Molchan (1996), Arbeiter and Patzschke (1996) and Barral (1999). A number of early contributions have indeed pointed out certain similarities between volatility and fluid turbulence (Galluccio *et al.* 1997, Schmitt *et al.* 1999), while theoretical modeling in finance using the concept of multifractality started with its adaptation to an asset-pricing framework in Mandelbrot *et al.* (1997) and Lux and Segon (2018). Calvet and Fisher (2002) provided one of the first thorough investigations of multifractal moment scaling in financial returns by confirming that the moments of empirical currency and equity returns vary as a power law of the time horizon. Earlier, Vassilicos and Hunt (1993) presented empirical evidence of multifractality in finance data, but (Mandelbrot *et al.* 1997) were the first to develop a multifractal model of asset returns (MMAR). MMAR assumes that returns follow a compound process in which an incremental Brownian motion is subordinate to the cumulative distribution function of a multifractal measure (Wang *et al.* 2013).

To overcome MMAR's limitation of being nonstationary, Calvet and Fisher (2001, 2004) introduced the Markov switching multifractal (MSM) models; see also Calvet and Fisher (2008). In MSM, total volatility is the multiplicative product of a large but finite number of random components that have identical marginal distributions and differ only in their switching probabilities (Calvet and Fisher 2004). By taking multifractality into consideration, one advantage of MSM models lies in their ability to capture the stylized facts of volatility such as outliers, volatility clustering and asymptotic power-law behavior of the autocovariance function (long-term memory). Another advantage of MSM is that the parsimonious parameterization of the transition matrix allows MSM to be estimated with reasonable precision even with a very large state space (Calvet and Fisher 2004). Many empirical studies suggest that the MSM model provides considerable gains in long-term forecasting (10–15 days) while it only slightly improves on GARCH(1,1) at daily and weekly intervals (Engle 1982, Calvet and Fisher 2004, Lux and Kaizoji 2007, Idier 2008). In terms of modeling, it has become common practice to propose a hybrid model or a model of models with the hypothesis that the combination of model strengths can enhance forecast accuracy (Makridakis *et al.* 1982, Palm and Zellner 1992, Wedding II and Cios 1996, Bildirici and Ersin 2009, Barucci *et al.* 2012, Wang *et al.* 2013, Ortega and Khashanah 2014, Reher and Wilfling 2016, Yang *et al.* 2019). In this paper, to enhance MSM's short-term prediction accuracy, we propose a volatility forecasting framework based on the MSM model with support vector regression (MSM-SVR). Since the performance of SVR is highly sensitive with respect to its hyperparameters, the hyperparameters of the SVR are optimized using the particle swarm optimization (PSO) algorithm. We evaluate the model performance with the exchange traded fund (ETF) SPDR S&P 500 Trust ETF with the ticker symbol SPY tick data and with one-minute data. Instead of stochastic volatility models, we use GARCH(1,1) as the benchmark model because studies have shown that stochastic volatility models perform worse than GARCH family models (Gospodinov *et al.* 2006). However, since the GARCH-SVR model is available in Peng *et al.* (2018), it makes more sense to benchmark MSM-SVR to GARCH-SVR as a hybrid model should be benchmarked to a more basic hybrid model.

Support vector regression (SVR) is another popular method for financial time series forecasting (Vapnik 2000, Tay and Cao 2002, Kim 2003, Cao and Tay 2003, Siao *et al.* 2016, Law and Shawe-Taylor 2017). SVR's primary advantage over traditional statistical models lies in its capability to model nonlinear processes without assuming any prior knowledge about the data. Compared with other common neural network models, SVR obtains better performance since it is based on the structural risk minimization principle and linearly constrained quadratic programming theory (Tang *et al.* 2009). Also, the solution of SVR is globally optimal. The nonlinear capability of SVR is achieved through kernel mapping. The kernel function must satisfy the condition in Mercer's theorem. Kernel functions are used to map the data in the input space to a high-dimensional feature space in which the problem becomes linearly separable (Schölkopf *et al.* 1998).

The prediction performance of SVR is greatly dependent upon the selection of kernel functions. There exist various support vector kernels such as linear, polynomial and Gaussian kernels. As any continuously differentiable periodic function may be expanded into a Fourier series, the Fourier kernel makes the performance of SVR optimal for cyclical components. It is valuable for us to consider whether desirable features can be achieved if we integrate Fourier analysis into MSM-PSO-SVR to forecast volatility. We also evaluate the performance of wavelet kernel MSM-PSO-SVR since wavelet functions not only describe time series at various locations but also at varying time granularities. Multiresolution and wavelet analysis are discussed in Mallat (1989), Daubechies (1990), Daubechies (1992), Mallat (1999); and for wavelet SVR, see Zhang *et al.* (2004). Considering that most papers in the literature of MSM test their model performance with daily data rather than higher frequency data (i.e., Calvet and Fisher 2004, Wang *et al.* 2013), we forecast one-minute realized volatility of SPY. Furthermore, we compare the forecasting performance of MSM-PSO-SVR with different kernels including the Fourier kernel and a wavelet kernel.

The selection of parameters has a strong effect on SVR's general performance. The particle swarm optimization algorithm (PSO) provides an excellent solution (Wu 2009). Hence, we propose a framework integrating PSO-optimized SVR into the MSM model (MSM-PSO-SVR) and analyze the proposed model performance under different kernels. It is shown that the Fourier kernel MSM-PSO-SVR has the best performance in terms of mean squared error (MSE) when forecasting one-minute realized volatility. It may be due to the strong sinusoidal pattern shown in the input data of SVR. The parameters of the MSM model are estimated by maximum likelihood and we use Bayesian updating to forecast volatility. In terms of mean absolute error (MAE), the performance of Fourier kernel MSM-PSO-SVR still outperforms other kernels (including linear kernel and polynomial kernel) and is very close to that of wavelet kernel MSM-PSO-SVR. The results further validate that the selection of the kernel function is important for the performance of MSM-PSO-SVR. We also compare the performance of MSM-PSO-SVR under different kernels with GARCH (Bollerslev 1986) and the GARCH-SVR model proposed by Peng *et al.* (2018) using the data of SPY and AAPL. The results show that MSM-PSO-SVR achieves a clear improvement in short-term volatility forecasting.

As a hybrid model, the MSM-SVR model proposed by Wang *et al.* (2013) is similar to ours with some key differences distinguishing our proposed model from their model. First, Wang *et al.* (2013) propose a hybrid model of MSM and SVR for daily volatility prediction while we propose another hybrid model of MSM and SVR with particle swarm optimization (PSO). Wang *et al.* (2013) mainly focus on demonstrating that MSM-SVR provides a promising alternative model for volatility forecasting and compare the model performance with GARCH using daily closing price data for two indexes, SSEC and SZSC in the Chinese A-share stock markets. There is no discussion of MSM-SVRs with different kernel functions. However, we mainly focus on showing that the proposed model MSM-PSO-SVR outperforms the benchmark models including GARCH with different kernel functions such as the Fourier kernel, wavelet kernel, Gaussian kernel and polynomial kernel. We also rank the performance of MSM-PSO-SVR with different kernels using classical statistical techniques and evaluate their forecasting ability based on statistical tests.

Second, Our experiments aim at higher frequency time series at one-minute frequency volatility forecasting while existing studies have not addressed higher-then-daily volatility fluctuations. Furthermore, we have used statistical techniques to check the existence of multifractality in the dataset while this point is not mentioned in Wang *et al.* (2013).

Third, for future research, Wang *et al.* (2013) conclude that their proposed MSM-SVR provides a promising alternative for the volatility forecasting problem without further evidence of outperformance. We further conclude that our proposed MSM-PSO-SVR outperforms benchmark models in volatility forecasting under different kernels and point to future research on kernel performance ranking. Kernel selection and its impact on accommodating market microstructure effects requires further analysis of the performance of the Fourier and wavelet kernels. Finally, in this paper, we point out and offer initial assessment of model robustness with respect to out-of-sample proportion, datasets, and error criteria.

The key contribution of this paper is to combine different models that can forecast volatility in general. The MSM-PSO-SVR is a model of models or hybrid model that combines the MSM model volatility decomposition with nonlinearities preserved by the kernel function mapping, achieved via the PSO-optimized SVR. Becker and Clements (2008) argued that forecasts by model integration has the potential to generate forecasts of superior predictive ability since different models capture different dynamics in volatility.

Another contribution of this paper is to offer a hybrid model for volatility forecasting with higher-than-daily frequency data. The literature shows that MSM forecasts only slightly improve on GARCH(1,1) at short-term intervals, which indicates that there may exist certain patterns to be discovered in the innovation part ε_t. One-minute data is used in this paper to test the performance of MSM-PSO-SVR. The need to study higher frequency volatility is motivated by higher frequency trading and the need for short-term risk assessment as evidenced, for example, by the Flash Crash of May 6, 2010. Here the financial return is specified as the product of an MSM volatility term and a stochastic innovation ε_t. Since the absolute residuals ε_t have an unspecified and potentially rich dependence structure, SVR is used to produce forecasts of $\mid \varepsilon_t \mid$. The performance of the proposed MSM-PSO-SVR with different kernels is evaluated in terms of MSE and MAE using empirical data. Subject to the data set and the MSM-PSO-SVR method in this paper, there is no claim of a clear kernel winner, perhaps suggesting that a different selection of kernel functions is predicated on the specific financial application and further performance ranking.

The rest of this paper is organized as follows. Section 2 discusses the necessary theoretical background of MSM, SVR and kernels used (Fourier and wavelet), and PSO, benchmark models of GARCH and GARCH-SVR. Section 2 also offers our algorithm design. Section 3 presents an application of the proposed methodology with experimental results on the SPY index including the multifractality detection test as well as various performance criteria and robustness checks. Section 4 summarizes the conclusions.

2. Theory

2.1. *Markov switching multifractal*

As in Calvet and Fisher (2004), consider the log-prices p_t of a financial asset defined at $t = 0, 1, 2, \ldots, \infty$. Define the returns $r_t \equiv p_t - p_{t-1}$, the MSM models r_t by an MSM volatility term times an innovation term ϵ_t so that

$$r_t \equiv \sigma_t \cdot \varepsilon_t \tag{1}$$

$$\sigma_t^2 = \sigma^2 \prod_{k=1}^{\bar{k}} M_{k,t} \tag{2}$$

where σ is a positive constant, $M_{1,t}, M_{2,t}, \ldots, M_{\bar{k},t}$ are $\bar{k}$ volatility components decaying at heterogeneous frequencies

$\gamma_1, \ldots, \gamma_{\bar{k}}$ and ε_t is a sequence of standard normal random variables. The volatility components $M_{k,t}$ satisfy $E(M_{k,t}) = 1$. Also they are persistent and non-negative. The MSM assumes that the multipliers $M_{1,t}, M_{2,t}, \ldots, M_{\bar{k},t}$ at a given time t are statistically independent. The return process has volatility $\sigma_t = \sigma\sqrt{\prod_{k=1}^{\bar{k}} M_{k,t}}$. Define a first-order Markov vector process of the volatility state in terms of components at a given time t

$$M_t = (M_{1,t}, M_{2,t}, \ldots, M_{\bar{k},t}) \tag{3}$$

The returns r_t are observed directly while the vector M_t is latent. Thus MSM is a hidden Markov chain model of volatility and M_t must be inferred recursively by Bayesian updating. The next period multiplier $M_{k,t}$ is either drawn from a fixed distribution of M with probability γ_k, or kept at its previous value $M_{k,t} = M_{k,t-1}$ with probability $1 - \gamma_k$ with $M > 0$ and $E(M) = 1$, identical for all $k = 1, \ldots, \bar{k}$. The distribution of M is a Bernoulli with $p = \frac{1}{2}$ and values of m_0 or $2 - m_0$, $m_0 \in (0, 2]$ and $E(M) = 1$. The switching events and updates from M are assumed to be independent across k and t. Thus the volatility components $M_{k,t}$ are different in their transition probabilities $\gamma_k = 1 - (1 - \gamma_1)^{(b^{(k-1)})}$ though not in their marginal distribution M. It is assumed that $\gamma_1 \in (0, 1)$ and $b \in (1, \infty)$.

In summary, the full parameter vector is $\psi \equiv (m_0, \sigma, b, \gamma_1) \in \mathbf{R}_+^4$, where m_0 characterizes the distribution of the volatility component multipliers, σ is the unconditional standard deviation of returns, and b and γ_1 determine the Markov switching probabilities.

The MSM parameters vector can be estimated by the maximum-likelihood estimation (MLE) given by

$$\ln L(r_1, \ldots, r_T; \psi) = \sum_{t=1}^{T} \ln [\omega(r_t) \cdot (\Pi_{t-1} A)] \tag{4}$$

where $\omega(r_t)$ is the conditional density function of ε_t, $\Pi_t = (\Pi_t^1, \ldots, \Pi_t^d)$ is the conditional probability vector with $\Pi_t^k \equiv \mathbf{P}(M_t = m_k | r_1, \ldots, r_t)$, $A = (a_{i,j})_{1 \le i,j \le d}$ and $d = 2^{\bar{k}}$ is the transition matrix with components $a_{i,j} = \mathbf{P}(M_{t+1} = m^j | M_t = m^i) = \prod_{k=1}^{\bar{k}} [(1 - \gamma_k) 1_{\{m_k^i = m_k^j\}} + \gamma_k \mathbf{P}(M = m_k^j)]$ in which m_k^i denotes the kth component of vector m^i, and $1_{\{m_k^i = m_k^j\}}$ is the dummy variable equal to 1 if $m_k^i = m_k^j$, and 0 otherwise.

Using the Bayesian method, the future volatility $\hat{\sigma}_t$ is predicted with the estimated parameters. Given the conditional probability vector Π_t, the probability distribution of future volatility state is $\mathbf{P}(M_{t+h} = m^i | r_1, \ldots, r_t) = (\Pi_t A^h)_i, 1 \le i \le d$ where h is the forecasting horizon. Thus the future volatility is calculated according to

$$\mathbf{E}(\sigma_{t+h}^2) = \sum_{i=1}^{d} \sigma_t^2(m^i) \cdot (\Pi_t A^h)_i \tag{5}$$

where $\sigma_t^2(m^i) = \sigma^2 \prod_{k=1}^{\bar{k}} M_{k,t}^i$ with $M_{k,t}^i = m_0$ or $2 - m_0$.

2.2. *Kernels and support vector regression*

A support vector kernel function can be described either as a function of the inner product $K(x, x') = K(\langle x, x' \rangle)$ or as the translation invariant function $K(x, x') = K(x - x')$. Mercer's Theorem† generalizes the conditions of Gram matrix (symmetric semi-positive definite matrix) singular value decomposition to infinite-dimensional spaces of functions and spectral decomposition of compact integral operators. Mercer's theorem characterizes those kernel functions that can be expressed in terms of an inner product over a Hilbert space. The kernel decomposition relative to the eigen-functions of the operator enables the *kernel trick* via expressing the kernel function as an inner product of a feature function. Mercer's Theorem in functional analysis enables the construction of *reproducing kernel Hilbert spaces* (RKHS). Some references on Mercer's theorem can be found in Courant and Hilbert (1953), Ash (1965) and Boser *et al.* (1992). With $L^2(\Omega)$ denoting the Hilbert space of square integrable functions on Ω, a version of Mercer's Theorem in data science is as follows

THEOREM 2.1 Mercer's Theorem *Suppose* $K(x, x') \in L^2(R^N \times R^N)$ *(R^N denotes the input space), for all* $g(x) \in L^2(R^N)$, *if*

$$\int \int_{L^2 \otimes L^2} K(x, x') g(x) g(x') \, \mathrm{d}x \, \mathrm{d}x' \ge 0$$

holds, we can write $K(x, x')$ *as a dot product* $K(x, x') = \langle \phi(x), \phi(x') \rangle$ *in some feature space.*

The translation invariant (stationary) type of kernels appeared in the work of Bochner on harmonic analysis with a characterization in terms Fourier transform of a candidate kernel function (Bochner 1933). For kernel classification and translation invariant support vector kernels in machine learning, see Smola *et al.* (1998), Genton (2001) and Schölkopf and Smola (2002).

2.2.1. Support vector regression (SVR). Similar to other regressors, SVR aims to fit a line/curve to data by minimizing a cost function. However, what differentiates SVR is that a nonlinear mapping can be applied when fitting SVR to find a best-fit hypersurface (manifold). This process is based on the so-called kernel trick and the representation of the solution. In the meanwhile, the basic algorithm for best-fit regression essentially remains similar except for enabling a nonlinear mapping via the kernel function. The choice of the kernel function varies from a simple inner product to a complicated nonlinear function. Therefore, SVR minimizes the l^2 norm of the *coefficients* vector subject to the predictor-target error being *uniformly* bounded by a maximum threshold of $\epsilon > 0$. The output of SVR is a best-fit hypersurface along with boundaries that are $\epsilon > 0$ away from the hypersurface and, as a result, its algorithm maximizes the number of data points in the ϵ-tube around the support vector. The reader is referred to Smola and Schölkopf (2004) for a complete formulation. The SVR function is formulated as

$$y_t = f(X_t) = \omega \phi(X_t) + b \tag{6}$$

† James Mercer's paper titled 'Functions of positive and negative type, and their connection with the theory of integral equations' in Phil. Trans. R. S. (A), 1909.https://doi.org/10.1098/rspa.1909.0075

where ϕ is a nonlinear mapping, which maps the original input space to a high-dimensional feature space. The coefficients ω and b are estimated by solving the convex optimization problem:

$$\min \frac{1}{2}||\omega||^2$$

subject to

$$|y_t - \omega\phi(X_t) - b| \leq \epsilon \tag{7}$$

in which $\frac{1}{2}||\omega||^2$ is the regularization term which measures the flatness of the function. ϵ is a user-prescribed parameter measuring the maximum deviation from the actual values y_t for all estimated values $\hat{y}_t$.

To obtain the estimations of ω and b, the constraint in (7) is transformed to the primal constraints given in (8) by introducing the positive slack variables ξ and $\xi^\star$. The slack variables represent the distance from the actual values to the corresponding boundary values of ϵ-tube and the problem becomes

$$\min \frac{1}{2}||\omega||^2 + C\sum_{t=1}^{l}(\xi_t + \xi_t^\star)$$

subject to

$$\begin{cases} y_t - \omega\phi(X_t) - b & \leq \epsilon + \xi_t \\ \omega\phi(X_t) + b - y_t & \leq \epsilon + \xi_t^\star \\ \xi_t, \xi_t^\star \geq & 0 \end{cases} \tag{8}$$

where C is a positive regularization parameter. Upon a Lagrangian transformation, the optimization problem takes the form

$$\max \left\{ -\frac{1}{2}\sum_{t,t'=1}^{l}(\alpha_t - \alpha_t^\star)(\alpha_{t'} - \alpha_{t'}^\star)K(X_t, X_{t'}) - \epsilon\sum_{t=1}^{l}(\alpha_t + \alpha_t^\star) + \sum_{t=1}^{l} y_t(\alpha_t - \alpha_t^\star) \right\} \tag{9}$$

subject to $\sum_{t=1}^{l}(\alpha_t - \alpha_t^\star) = 0$ and $\alpha_t, \alpha_t^\star \in [0, C]$ where α_t and $\alpha_t^\star$ are Lagrangian multipliers and $K(X_t, X_{t'})$ is called the kernel function and $K(X, X') = \langle\phi(X), \phi(X')\rangle = \phi(X) \cdot \phi(X')$. Therefore, the decision function given in (6) has the following explicit form

$$f(X_t) = \sum_{t'=1}^{l}(\alpha_{t'} - \alpha_{t'}^\star)K(X_t, X_{t'}) + b \tag{10}$$

Comparing this solution to the decision function in (6), ω must have the form

$$\omega = \sum_{t'=1}^{l}(\alpha_{t'} - \alpha_{t'}^\star)\phi(X_{t'})$$

The choice of the kernel K is user-prescribed. Any function satisfying Mercer's theorem can be used as the kernel function. Equation (10) is a support vector expansion of the function f with a support vector kernel K. As in equation (11) in Smola and Schölkopf (2004), only in the linear case can ω be determined by the training patterns completely.

For this paper, in order to use SVR in the MSM model to forecast volatility $\hat{\sigma}_t$, we first see that, from (1), $\varepsilon_t = r_t/\hat{\sigma}_t$. Let $y_t = \sqrt{\varepsilon_t^2}$, $t = 0, 1, \ldots, l$, which is then modeled with SVR built upon a vector of historical absolute residuals $X_t = (\sqrt{\varepsilon_{t-1}^2}, \ldots, \sqrt{\varepsilon_{t-\tau}^2})$, where τ indicates the number of look-back time intervals. In the experiment of Section 3; we set $\tau = 15$. The hypothesis here is that the previous absolute innovation values X_t contain some information about the future value or that the process has memory.

2.2.2. Fourier kernel. Consider mapping the input variable x into the $(2N + 1)$-dimensional vector $u = (\frac{1}{\sqrt{2}}, \sin x, \ldots, \sin Nx, \cos x, \ldots, \cos Nx)$. Then for any fixed x, the Fourier expansion can be considered as the inner product in this $(2N + 1)$-dimensional feature space:

$$f(x) = (a \cdot u) = \frac{a_0}{\sqrt{2}} + \sum_{k=1}^{N}(a_k \sin kx + b_k \cos kx) \tag{11}$$

Therefore the inner product of two vectors in this space has the form

$$\begin{aligned} K_N(x, x') = (u \cdot u) &= \frac{1}{2} + \sum_{k=1}^{N}(\sin kx \sin kx' + \cos kx \cos kx') \\ &= \frac{1}{2} + \sum_{k=1}^{N}\cos k(x - x') \end{aligned} \tag{12}$$

Vapnik (1998). The Dirichlet kernel is obtained with some algebraic manipulations $K_N(x, x') = [\sin\frac{2N+1}{2}(x - x')]/[2\sin\frac{(x-x')}{2}]$. However the Dirichlet kernel $K_N(x, x')$ does not have good approximation properties. Therefore, we consider the following regularized Fourier expansion:

$$f(x) = \frac{a_0}{\sqrt{2}} + \sum_{k=1}^{\infty} q^k(a_k \sin kx + b_k \cos kx), \quad 0 < q < 1 \tag{13}$$

where a_k and b_k are coefficients of the Fourier expansion. This expansion differs from the expansion in (11) by multipliers q^k that provide a mode of regularization. As in Vapnik (1998), it follows that for the d-dimensional vector space $x = (x^1, \ldots, x^d)$, the support vector kernel becomes

$$K(x, x') = \prod_{k=1}^{d}\frac{1 - q^2}{2(1 - 2q\cos(x^k - x'^k) + q^2)} \tag{14}$$

Setting the Fourier kernel function as the support vector's kernel function, the estimation function of FSVR is defined as

$$f(x) = \sum_{t=1}^{l}(\alpha_t - \alpha_t^\star)\prod_{k=1}^{d}\frac{1 - q^2}{2(1 - 2q\cos(x^k - x_t^k) + q^2)} + b \tag{15}$$

2.2.3. Wavelet kernel. Wavelet analysis approximates a signal or function by a family of functions generated by dilations

and translations of a mother wavelet function. For this paper, we note that not every wavelet function produces an allowable support vector kernel function. The Morlet wavelet function satisfies Lemma 2.2, as in Wu (2011), and produces an allowable support vector kernel function.

The Morlet wavelet kernel function is defined as

$$K(x,x') = \prod_{k=1}^{d} \cos\left(1.75 \times \frac{x^k - x'^k}{a}\right) \exp\left(-\frac{||x^k - x'^k||^2}{2a^2}\right),$$
$$x \in R^{l\times d},\ x^k \in R^d \tag{16}$$

which is an admissible support vector kernel function. For further details, see Zhang and Benveniste (1992).

2.2.4. The proposed optimization algorithm. SVR's performance is sensitive to the choice of its hyperparameters. Particle Swarm Optimization (PSO) algorithm is considered to be an excellent optimization technique (Krusienski and Jenkins 2006, Lin *et al.* 2008, Yamaguchi and Yasuda 2006, Wu 2009). Introduced by Kennedy and Eberhart (1995), PSO uses a set of particles to indicate potential solutions.

The swarm consists of m particles, in which each particle has a position $X_i = \{x_{i1}, x_{i2}, \ldots, x_{id}, \ldots, x_{im}\}$ and a velocity $V_i = \{v_{i1}, v_{i2}, \ldots, v_{id}, \ldots, v_{im}\}$, $i = 1, 2, \ldots, n$ and $d = 1, 2, \ldots, m$. All particles move through an n-dimensional search space. Each particle moves from its best previous position and towards the particle g with the best position in the swarm at that iteration. In our context, positions refer to coordinates in the hyperparameters space of SVR while the velocity vector measures the rate of change of positions with respect to a time step.

Denote the best previously visited position of the ith particle with the best fitness value as $P_i = \{p_{i1}, p_{i2}, \ldots, p_{id}, \ldots, p_{im}\}$, and the best previously visited position of the swarm that gives best fitness as $P_g = \{P_{g_1}, P_{g_2}, \ldots, P_{g_d}, \ldots, P_{g_m}\}$. The change of each particle's position from one iteration to another is computed based on both the distance between the current position and its previous best position and the distance between the current position and the best position of swarm:

$$v_{id}^{k+1} = \omega v_{id}^k + c_1 r_1 (p_{id} - x_{id}^k) + c_2 r_2 (P_{g_d} - x_{id}^k)$$
$$x_{id}^{k+1} = x_{id}^k + v_{id}^{k+1} \tag{17}$$

where ω controls the trade-off between the global and local exploration abilities of the swarm. $k = 1, 2, \ldots, k_{max}$ denotes the iteration number. c_1 and c_2 are learning factors. r_1 and r_2 are random numbers uniformly distributed in $[0, 1]$. Thus, the particle advances through potential solutions towards P_{ik} and P_{gk} in a navigated way and at the same time exploring new areas by the stochastic mechanism to escape from local optima. In the PSO algorithm, the position and velocity are usually initialized randomly. The algorithm needs to input the objective function such as the mean squared error MSE or the mean absolute error MAE. The objective function is evaluated and the position and velocity is updated repeatedly until a pre-specified stopping criterion is satisfied. Here in this paper, the objective function is either MSE or MAE and the length of iterations is around 500. Table 1 summarizes the hyperparmaters for different SVR kernels: Fourier-SVR (FSVR), wavelet-SVR (WSVR), Gaussian-SVR (GSVR), and polynomial-SVR (PSVR).

Table 1. Hyperparameters for different kernel.

	Kernel parameters
Fourier Kernel SVR (FSVR)	q
Wavelet Kernel SVR (WSVR)	a
Gaussian Kernel SVR (GSVR)	γ
Polynomial Kernel SVR (PSVR)	d

2.3. Benchmark models: GARCH and GARCH-SVR

2.3.1. GARCH. The common approach to modeling volatility is the generalized autoregressive conditional heteroskedasticity (GARCH) model proposed by Bollerslev (1986). Given P_t the observed price at time $t = 1, \ldots, T$, the linear standard GARCH(1,1) can be summarized as

$$r_t = \mu_t + \varepsilon_t \tag{18}$$

where $r_t = \log(\frac{P_t}{P_{t-1}})$ is the log return and ε_t is a stochastic term with zero mean. The mean equation of the return in equation (18) is defined by an AR(1) model

$$\mu_t = \gamma_0 + \gamma_1 r_{t-1} \tag{19}$$

and the volatility equation is given by

$$\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2 \tag{20}$$

in which σ_t denotes conditional volatility, $\omega > 0, \alpha, \beta \geq 0$ and $\alpha + \beta < 1$.

The GARCH(1,1) model can be estimated by the maximum-likelihood estimation (MLE) in which the log-likelihood function is

$$L(\theta_{GARCH}; \{r_t\}_{t=1,\ldots,T}) = -\frac{T}{2}\ln(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\left(\ln\sigma_t^2 - \frac{\varepsilon_t^2}{\sigma_t^2}\right) \tag{21}$$

where T is the sample size, σ_t^2 follow equation (20), and θ_{GARCH} refers to the parameters ω, α and β in GARCH.

2.3.2. GARCH-SVR. Peng *et al.* (2018) proposed to combine the GARCH model with Support Vector Regression (SVR) for volatility forecasting using high-frequency data. By using SVR to the estimation of mean and volatility equations in GARCH (equations (19) and (20)), GARCH-SVR provides a good fit to the data and thus here it is adopted as a benchmark model. The specification of GARCH-SVR(1,1) is the same of the standard GARCH(1,1), but the mean and volatility equations are estimated via SVR using the functions

$$r_t = f_m(r_{t-1}) + \varepsilon_t$$
$$h_t = f_v(h_{t-1}, \varepsilon_{t-1}^2) \tag{22}$$

where $f_m(.)$ is the SVR decision function for the mean equation (19), and $f_v(.)$ is the SVR decision function for the volatility equation (20). Following Peng *et al.* (2018), the Gaussian kernel SVR is used.

2.4. Algorithm design

Our algorithm is designed as described in the following steps

(i) Conduct multifractal detrending moving average algorithm to detect the existence of multifractality in the 1-min log returns of the exchange traded fund (ETF) SPDR S&P 500 Trust ETF with the ticker symbol SPY.
(ii) Perform MLE on the SPY 1-min log returns r_t and estimate the parameters of the MSM model.
(iii) Estimate the volatility σ_t according to the MLE result in the training set and forecast $\hat{\sigma}_t$ in the testing set.
(iv) Calculate $y_t = \sqrt{\varepsilon_t^2} = \sqrt{r_t^2/\sigma_t^2}$.
(v) The variable $\{\hat{y}_t\}$ is predicted by an SVR model built upon historical residuals vector $\{y_t\}$ and the hyperparameters of SVR optimized by PSO.
(vi) Predict one-minute volatility $\hat{rv}_t = \hat{\sigma}_t \times \hat{y}_t$, in which $\hat{\sigma}_t$ is obtained from Step (iii) and $\hat{y}_t$ is derived in Step (v).

A flow chart depicting our algorithm is shown in figure 1.

3. Experimental results

This section reports the experimental results of the proposed algorithm applied to the SPY data.

3.1. Data

The experiment uses one-minute historical trading price of the SPY ETF starting from August 1, 2016 and ending on February 28, 2017. The choice of the SPY data is due to its high liquidity, wide equity representation in its components, and for future research initiatives. We use the first 90% data as training set and leave the rest 10% for testing purposes. This gives about 50 895 observations in training set and 5655 entries in the test set.

3.2. Parameter estimation

Table 2 reports on the maximum-likelihood estimation for the multifractal parameters $m_0, \sigma, b, \gamma_{\bar{k}}$ for $2 \leq \bar{k} \leq 8$ in the training set. The columns of the table correspond to the number of frequencies $\bar{k}$ varying from 2 to 8. The data generate parameter estimates sharing some properties with those in Wang *et al.* (2013) and Calvet and Fisher (2004). When $\bar{k}$ increases, the multiplier parameter $\hat{m}_0$ tends to decline and the switching probability of the highest-frequency multiplier ($\hat{\gamma}$) increases. At the same time, the estimate $\hat{b}$ decreases steadily with $\bar{k}$. Table 2 suggests that additional high-level multipliers make a smaller and smaller contribution to the moments. The increasingly small influence of additional elements in the hierarchy of multipliers also suggests that it should be hard or impossible to distinguish between multifractal models with different numbers of components once $\bar{k}$ increases beyond a certain threshold (Wang *et al.* 2013, Calvet and Fisher 2004)). Tables 2 and 3 show that the likelihoods of MSM model are higher than the likelihood of GARCH for any value of $\bar{k}$ within the range of $2 \leq \bar{k} \leq 8$. To reduce the computation complexity, we produce forecasting

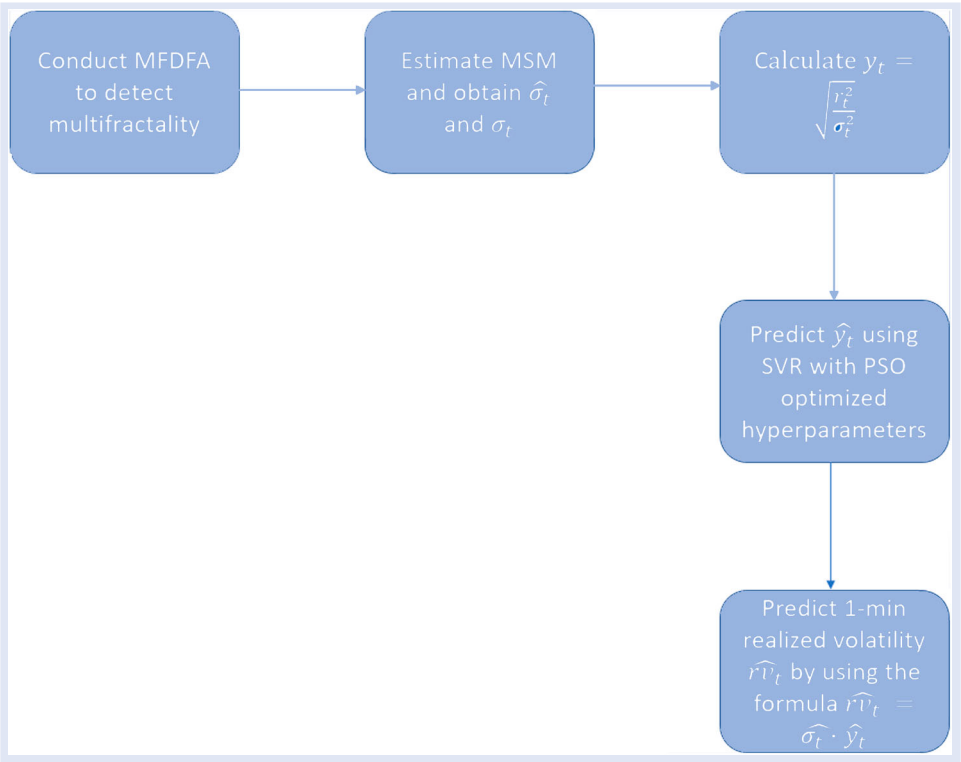

Figure 1. The flowchart of MSM-PSO-SVR.

Table 2. Maximum-likelihood estimation results of the MSM model.

$\bar{k}$	b	m_0	$\gamma_{\bar{k}}$	σ	$\ln L$
2	11.97	1.59	0.08	0.08	360 994.07
3	13.60	1.52	0.20	0.08	361 474.33
4	9.03	1.47	0.38	0.08	361 612.10
5	7.39	1.43	0.45	0.08	361 709.01
6	7.12	1.41	0.55	0.08	361 761.75
7	5.07	1.37	0.72	0.07	361 789.05
8	4.87	1.36	0.74	0.09	361 805.27

Table 3. Maximum-likelihood estimation results of GARCH.

	ω	α	β	d	$\ln L$
GARCH	0.00	0.05	0.9	–	354 430.5

Table 4. Future one-step ahead one-minute volatility forecasting results.

	Kernel parameters	MSE (%)	MAE (%)
MSM-PSO-FSVR	q = 0.8	11.18	2.46
MSM-PSO-WSVR	a = 0.001	11.19	2.46
MSM-PSO-GSVR	γ = 10	11.19	2.46
MSM-PSO-PSVR	d = 3	11.53	2.42
MSM-PSO-LSVR	–	11.79	2.44
GARCH-FSVR	q = 0.6	12.03	2.46
GARCH-WSVR	a = 0.01	12.06	2.47
GARCH-GSVR	γ = 10	12.06	2.47
GARCH-PSVR	d = 3	13.32	2.96
GARCH-LSVR	–	13.52	3.03
GARCH	–	12.29	2.58
MSM	–	13.31	2.95

results with components $\bar{k} = 5$ in the experimental section. To facilitate the comparison, we fix the regularization parameter $C = 1$ and the parameter $\epsilon = 1$. Parameters of the SVR kernel function are obtained by the particle swarm optimization (PSO).

3.3. *Performance criteria*

We aim to predict the one-minute volatility with prediction performance evaluated according to two standard measures of goodness-of-fit: mean squared error (MSE) and mean absolute error (MAE).

Here r_t^2 is taken as the proxy of volatility and the performance criteria definitions are

$$MSE = \frac{1}{N}\sum_{t=1}^{N}(rv_t - \hat{rv}_t)^2$$
$$MAE = \frac{1}{N}\sum_{t=1}^{N} \mid rv_t - \hat{rv}_t \mid \qquad (23)$$

where rv_t is the actual one-minute volatility at time t, $\hat{rv}_t$ is the forecasting value of the one-minute volatility, and N is the number of time intervals for forecasting. Statistically, the MSE is a measure of variance of the deviations between values of two time series of (actual and predicted) while the MAE is a measure of the mean distance between the same time series. The two measures contain different information of location (mean for MAE) and dispersion (variance for MSE). Location of mean deviations that is closer to zero values indicates higher performance and low dispersion indicates low number of outliers and more number of accurate predictions, i.e. higher performance. If two methods produce the same MAE, the one with minimum MSE has higher performance while if the two methods have the same MSE, the one with minimum MAE has higher performance. If both methods produce equal MAE and MSE, the two methods are said to have statistically equivalent performance relative to the data sample.

The prediction results of the SPY volatility based on different kernel SVRs are shown in table 4. Compared with GARCH-SVR models, MSM-PSO-SVR models with the same kernel has better performance in terms of lower MSE and MAE. Figure 2 shows plots of the differences between actual volatility and the forecasts of various MSM-PSO-SVR models and benchmark models using part of the testing dataset. Qualitatively, all methods seem to capture the same essential error process with some methods like MSM in this sample producing apparent overestimates. Quantitatively, we follow the ranking implied by the MSE and MAE summarized in table 4. For the purpose of conveying the idea neatly and considering that the performance GARCH-GSVR is good compared with other GARCH-SVR models under different kernels, only GARCH-GSVR is plotted. The Fourier kernel MSM-PSO-SVR has the best performance in terms of MSE slightly better than the performance of wavelet kernel and Gaussian kernel MSM-PSO-SVR's as well as the benchmark models of GARCH and GARCH-SVR. Although the forecasts of Fourier kernel MSM-PSO-SVR have very similar dynamics as the forecasts of wavelet kernel or Gaussian kernel MSM-PSO-SVR, some of the predictions of Fourier kernel MSM-PSO-SVR are closer to actual values, resulting in lower MSE. All MSM-PSO-SVR models outperform the two benchmark models and MSM. Generally speaking, hybrid MSM-PSO-SVR models perform better than the other peer competing models. There is no clear winner among the various MSM-PSO-SVR models; however, suggesting that the selection of kernel functions is based on the specific financial application to be used.

In table 4, the MAE of MSM-PSO-FSVR is equal to those of MSM-PSO-WSVR or MSM-PSO-GSVR; however, MSM-PSO-FSVR has the lowest MSE. It implies that MSM-PSO-FSVR, for this dataset, has higher forecasting performance for the SPY index.

Therefore, the following results stand out: our results essentially confirm the findings that MSM-PSO-SVR outperforms GARCH, GARCH-SVR and MSM in both indicators for the SPY index. Furthermore, the proposed PSO-SVR based MSM models have achieved considerable accuracy improvement in forecasting short-term volatility and show better performance to capture the short-term volatility movement.

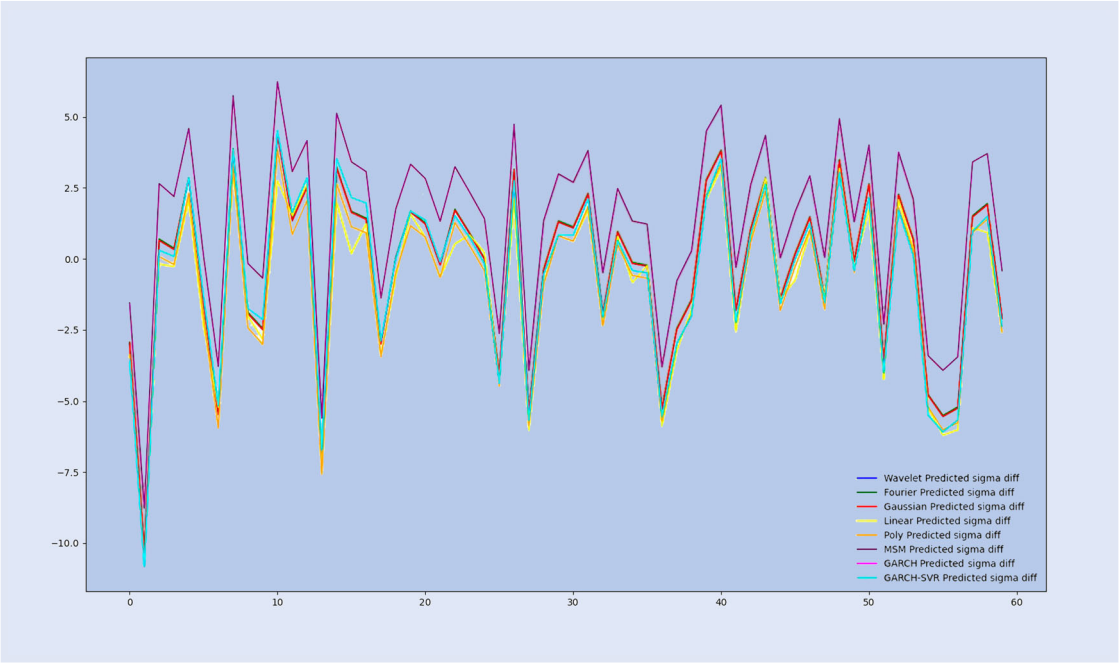

Figure 2. The difference between predicted volatility and actual volatility from the 1900th to the 1960th in the testing data set.

3.3.1. Tests for forecasting performance. The loss functions described above allow forecasts to be ranked according to their out-of-sample forecasting performance. However, they give no indication whether the forecasting losses across various models are significantly different. Therefore, we employ the superior predictive ability (SPA) test of Hansen (2005), which allows for comparing the forecasting performance of two or more models at a time. Consider N different models with forecasts being compared against a benchmark model M_0 based on a predefined loss function (i.e., MSE or MAE) as follows:

$$f_{i,t} = L_{0,t} - L_{i,t}, i = 1, \ldots, N \tag{24}$$

where $L_{i,t}$ refers to the predefined loss function computed for the i^{th} model.

The null hypothesis is that the performance of any of the other N competitive models is worse than that of the benchmark model M_0 and can be formulated as

$$H_0 : \max_{i=1,\ldots,N} \mu_i \leq 0 \tag{25}$$

where $\mu_i = E[f_{i,t}]$ for $i = 1, \ldots, N$ and μ_i denotes the expected performance of model i. The P-value of the SPA test is calculated by using the stationary bootstrap procedure in Politis and Romano (1994).

In table 5, forecasting results of the method MSM-PSO-SVR with five kernels: Fourier kernel MSM-PSO-FSVR, wavelet kernel MSM-PSO-WSVR, Gaussian kernel MSM-PSO-GSVR, and linear kernel MSM-PSO-LSVR, are compared to each of the benchmark models (GARCH, MSM, GARCH-FSVR, GARCH-WSVR and GARCH-GSVR), respectively. The SPA test is employed on the loss functions of MSE and MAE defined in the performance criteria section. The null hypothesis is that none of the models is better than the benchmark. Rejection of the null hypothesis implies that the best one from the models can outperform the benchmark. We can see that at the significance level of 0.1, the tests can be rejected, indicating that the proposed MSM-PSO-SVR models have better predictive ability than the benchmark models.

Table 5. Future one-step ahead one-minute volatility forecasting results: SPA Test.

	MSE	MAE
GARCH	0.02	0.03
MSM	0.03	0.02
GARCH-FSVR	0.02	0.01
GARCH-WSVR	0.03	0.01
GARCH-GSVR	0.03	0.01

3.3.2. Robustness checks. We also conduct a number of robustness checks to ensure the conclusion is consistent across the following five different cases: (1) applying the t-test of 20 non-overlapping data groups to rank the MSM-PSO-SVR models with different kernels based on MSE and MAE; (2) testing a longer out-of-sample period using SPY data; (3) Comparing the performance of the proposed model with other financial data samples; (4) testing results with lower frequency to include 5-minute and 15-minute SPY data; and (5) using deseasonalized one-minute SPY data.

- *t-test of 20 non-overlapping samples* . In order to further understand the performance differences between various MSM-PSO-SVR with different kernels, we partition the test data set into 20 (non-overlapping) sub-samples and calculate the MSE for each sub-sample using MSM-PSO-SVR under different kernel functions. Then for each model, we

Table 6. *P*-values from *t*-test of 20 sub-samples MSE with one-minute SPY data.

	MSM-PSO-PSVR	MSM-PSO-LSVR	MSM-PSO-WSVR	MSM-PSO-GSVR	MSM-PSO-FSVR
MSM-PSO-PSVR	–	0.00	0.00	0.00	0.00
MSM-PSO-LSVR	–	–	0.00	0.00	0.00
MSM-PSO-WSVR	–	–	–	Insignificant	0.03
MSM-PSO-GSVR	–	–	–	–	0.03

Table 7. *P*-values from *t*-test of 20 sub-samples MAE with one-minute SPY data.

	MSM-PSO-PSVR	MSM-PSO-LSVR	MSM-PSO-WSVR	MSM-PSO-GSVR	MSM-PSO-FSVR
MSM-PSO-PSVR	–	0.0	0.0	0.0	0.0
MSM-PSO-LSVR	–	–	0.99	0.99	0.99
MSM-PSO-WSVR	–	–	–	Insignificant	1.00
MSM-PSO-GSVR	–	–	-	–	1.00

have 20 MSE values. In the *t*-test, the null hypothesis (H_0) assumes that the difference between the mean of the model in the corresponding row and the mean of the Model in the corresponding column in terms of MSE/MAE equal to zero. The alternative hypothesis (H_1) assumes that the difference is greater than zero. If the *P*-value is less than 0.05, then the alternative hypothesis is true. After running the *t*-test between every two models MSE's, the results are summarized in table 6, which imply that Fourier $\gg$\{Wavelet, Gaussian\} $\gg$ \{Linear, Polynomial\}. Although MAE is not as sensitive to outliers as MSE, we still calculate MAE for each sub-sample and summarize the *t*-test results in table 7. In terms of MAE, we can see that it is hard to rank the performance of Fourier kernel MSM-PSO-FSVR, MSM-PSO-WSVR, MSM-PSO-GSVR and MSM-PSO-LSVR, while all of them outperform polynomial kernel MSM-PSO-SVR (MSM-PSO-PSVR). The results validate that kernel selection is important.

- *Out-of-sample robustness*. For the same dataset, does out-of-sample proportion change the ranking results? For example, initially we used the first 90% of the SPY dataset as the training sample and the rest 10% as the testing sample. Here in our first robustness checks, we shorten the training sample size to the first 80% data and the rest 20% for testing purpose. As is shown in table 8, the SPA test is employed on the loss functions of MSE and MAE with texts in boldface denote the data frequency and the values denote that the corresponding *P*-values of Hansen's SPA test between the MSM-PSO-SVR models and each of the 3 benchmark models (GARCH-SVR, GARCH or MSM) using the corresponding data under the specific loss function (MSE or MAE). The null hypothesis is that none of the MSM-PSO-SVR models with different kernels is better than the benchmark model. Rejection of the null hypothesis implies that the best one from the models can outperform the benchmark. The results in table 8 under 'longer-out-sample' show the *P*-values of Hansen's SPA test. We can see that at the significance level of 0.1, the MSM-PSO-SVR models increase the predictive ability of MSM and outperform other models.
- *Dataset Robustness*. Fixing other control parameters, we use different datasets with the same frequency and compare the results to those of the historical one-minute SPY dataset. As an example for this paper, in our dataset robustness checks, we use the historical one-minute AAPL dataset (1/4/2016 to 12/29/2017) to see whether the conclusion still holds. As before, the first 90% AAPL data is the training sample and the rest 10% is for testing purpose. Similar to the prior analysis, we then run Hansen's SPA test and the results are in table 8, which again validates the conclusion that the MSM-PSO-SVR models increase the predictive ability of MSM and outperform the peer models.
- *Data frequency robustness*. Fixing other control parameters, we use the historical 5-minute and

Table 8. Future one-step-ahead forecasting performance evaluation using various datasets: SPA Test

	1-min SPY									
	deseasonalized		longer out-of-sample		5-min SPY		15-min SPY		1-min AAPL	
Benchmark Model	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
GARCH	0.01	0.03	0.02	0.05	0.00	0.02	0.03	0.08	0.02	0.03
MSM	0.05	0.03	0.04	0.02	0.04	0.03	0.02	0.01	0.03	0.01
GARCH-FSVR	0.04	0.03	0.03	0.01	0.04	0.03	0.02	0.01	0.03	0.01
GARCH-WSVR	0.05	0.03	0.04	0.01	0.03	0.02	0.03	0.01	0.02	0.01
GARCH-GSVR	0.05	0.03	0.04	0.01	0.03	0.02	0.03	0.01	0.02	0.01

15-minute SPY dataset to see whether the conclusion still holds. The test results are summarized in table 8. The null hypothesis is that none of the MSM-PSO-SVR models with different kernels is better than the benchmark model. Rejection of the null hypothesis implies that the best one from the models can outperform the benchmark. We can see that at the significance level of 0.1, the tests can be rejected for all benchmark models, indicating that the proposed MSM-PSO-SVR models have better predictive ability than the benchmark models.

- *Error criterion robustness*. Although MAE is not as sensitive to outliers as MSE, we still calculate MAE for each group and summarize the *t*-test results in tables 6 and 7. In this case, we can easily find that Fourier kernel MSM-PSO-SVR does not outperform wavelet kernel MSM-PSO-SVR, which leads to the interesting question: under what scenarios does the Fourier kernel perform better than the wavelet kernel? Since this paper focuses more on proposing a promising new methodology for volatility forecasting, it will be another paper to answer this question. As a starting point, based on the work of Nielsen and Frederiksen (2008), although both the Fourier method and the wavelet methods are robust towards irregularities such as long memory, non-stationarity, and jumps, the Fourier method is superior compared to the wavelet method in the presence of market microstructure effects such as bid-ask bounce effects. This superiority is attributed to the decomposition of the integrated variance into components of varying frequencies by the Fourier transform. In order to ensure that the proposed MSM-PSO-SVR models also perform well for deseasonalized data, the robustness check is presented in the following 'Deseasonalization' part.
- *Deseasonalization*. The intraday seasonal patterns in the volatility of stock markets have important implications for modeling the volatility of high-frequency returns. Andersen and Bollerslev (1997) suggest seasonality filtering procedures that can eliminate the seasonal component from high-frequency financial data after comparing certain intraday ARCH model estimates with theoretical results using the same data. Taylor and Xu (1997) proposed to model the intraday seasonality by averaging sums of squared returns across similar time periods and a deseasonalized return series is calculated by dividing each return by its seasonal multiplier. We adopt a similar method to adjust for intraday volatility seasonality. Let $R_{d,n}$ be the n^{th} intraday return on day d, and suppose we have D days and N intraday periods. Then the seasonal variance estimate is given by

$$s_n^2 = \frac{1}{D} * \sum_{d=1}^{D} R_{d,n}^2$$

$$r_{d,n} = \frac{R_{d,n}}{s_n} \tag{26}$$

where $n = 1, 2, \ldots, N$ and $r_{d,n}$ is the deseasonalized return on day d of the intraday period n.

To take account of day-of-the-week effects, we estimate the seasonality component by week and then remove the seasonal component from the one-minute SPY data. Similar to the earlier analysis, we conduct Hansen's SPA test and the results are presented in table 8. At the significance level of 0.1, the tests can be rejected for all three benchmark models, implying that the proposed MSM-PSO-SVR models have better forecasting ability than the benchmark models.

Although the choice of method benchmark is important, it is not considered a robustness check. In general, the initial performance of a new volatility forecasting model can be compared to the performance of the GARCH model as the benchmark. Since the model is hybrid, it makes sense to compare to another hybrid model of GARCH. To this end, we conduct a comparison analysis between the hybrid model of GARCH and SVR (GARCH-SVR) and our proposed MSM-PSO-SVR models. To make its structure consistent with the proposed MSM-SVR model, the distribution of the return series r_t in GARCH-SVR is also assumed to be $r_t \equiv \sigma_t \cdot \varepsilon_t$ where σ_t now follows a GARCH process and the absolute value of ε_t will be predicted with SVR.

Similar to the prior analysis, we run Hansen's SPA test on the out-of-sample data and the test results are in table 8. We can see that at the significance level of 0.1, the tests can be rejected for the benchmark models of Fourier kernel GARCH-SVR (GARCH-FSVR), wavelet kernel GARCH-SVR (GARCH-WSVR) and Gaussian kernel GARCH-SVR (GARCH-GSVR). It implies that the proposed MSM-PSO-SVR models have better predictive ability than the GARCH-SVR models.

4. Conclusions

Volatility forecasts based on advanced combination techniques, which can be seen as a model of models, are usually superior to individual forecasts and simple combination techniques. In order to complement each other and make the fullest use of each model's unique characteristics, we propose a Markov Switching Multifractal (MSM) model integrated with particle swarm optimization (PSO) with support vector regression (SVR). Thus, obtaining the MSM-PSO-SVR to predict realized volatility, which exploits the MSM model to forecast σ_t and PSO-optimized SVR to model the absolute value of innovation. Multifractal models introduce a new way to volatility forecasting. Compared with standard Markov switching models, the MSM model captures the multifractality of asset returns with a very limited number of parameters. Support vector regression (SVR) is another promising way to predict financial volatility. The primary advantage of SVR is that SVR learns a high-dimensional function via a sequence

of nonlinear transformations and is able to capture the nonlinearities without assuming any prior knowledge about the underlying data-generating process. Since any continuously differentiable periodic function may expand into a Fourier series, the Fourier kernel can make the performance of SVR optimal for cyclical components. Hence we introduce Fourier kernel SVR whose hyperparameters are optimized by particle swarm optimization (PSO). Wavelet kernel MSM-PSO-SVR is also presented here for their multiresolution property.

The proposed model MSM-PSO-SVR has been evaluated with empirical financial data and the results demonstrate that it is effective in improving predictive ability. Generally speaking, MSM-PSO-SVR models outperform other peer models while the kernel selection is important. Compared to MSM-PSO-GSVR and MSM-PSO-WSVR, the MSM-PSO-FSVR has the same forecasting performance in terms of MAE while it dominates them in terms of MSE. The performance of MSM-PSO-FSVR, MSM-PSO-WSVR and MSM-PSO-GSVR outperform the benchmark models and they are close to each other in terms of MAE. Examples for robustness checks with respect to out-of-sample proportion, different financial dataset, deseasonalized dataset and error criterion were conducted to further validate the conclusion. MSM-PSO-FSVR not only takes advantages of the Fourier kernel's capability to approximate cyclical components but also has some other attractive properties, such as generalization performance and the automatic optimal parameters. Hence, in the process of establishing the forecasting models, much uncertain information of noise data is not neglected but considered wholly into the Fourier kernel function. The forecasting accuracy is improved by means of adopting the Fourier technique and the PSO algorithm.

Disclosure statement

No potential conflict of interest was reported by the author(s).

References

Andersen, T. and Bollerslev, T., Intraday periodicity and volatility persistence in financial markets. *J. Empir. Financ.*, 1997, **4**, 115–158.

Andersen, T.G., Bollerslev, T., Christoffersen, P.F. and Diebold, F.X., Volatility forecasting. Working Paper 11188, National Bureau of Economic Research, 2005.

Andersen, T.G., Bollerslev, T., Christoffersen, P.F. and Diebold, F.X., Volatility and correlation forecasting. In *Handbook of Economic Forecasting*, Vol. 1, edited by G. Elliott, C.W.J. Granger and A. Timmermann, pp. 777–878, 2006 (Elsevier: Amsterdam).

Arbeiter, M. and Patzschke, N., Random self-similar multifractals. *Math. Nachr.*, 1996, **181**, 5–42.

Ash, R.B., *Information Theory*, 1965 (Dover).

Bachelier, L., Theory of speculation. The random character of stock market prices, 1900.

Baillie, R., Bollerslev, T. and Mikkelsen, H.O., Fractionally integrated generalized autoregressive conditional heteroskedasticity. *J. Econom.*, 1996, **74**, 3–30.

Barral, J., Moments, continuité, et analyse multifractale des martingales de Mandelbrot. *Probab. Theory. Relat. Fields*, 1999, **113**, 535–569.

Barucci, E., Magno, D. and Mancino, M.E., Fourier volatility forecasting with high-frequency data and microstructure noise. *Quant. Finance*, 2012, **12**, 281–293.

Becker, R. and Clements, A., Are combination forecasts of S&P 500 volatility statistically superior?. *Int. J. Forecast.*, 2008, **24**, 122–133.

Bildirici, M. and Ersin, O., Improving forecasts of GARCH family models with the artificial neural networks: An application to the daily returns in Istanbul Stock Exchange. *Expert. Syst. Appl.*, 2009, **36**, 7355–7362.

Bochner, S., Monotone Functions, Stieltjessche Integrale und Harmonische Analyse. *Math. Ann.*, 1933, **108**, 378–410.

Bollerslev, T., Generalized autoregressive conditional heteroskedasticity. *J. Econom.*, 1986, **31**, 307–327.

Boser, B.E., Guyon, I.M. and Vapnik, V.N., A training algorithm for optimal margin classifiers. In *Proceedings of the fifth annual workshop on Computational learning theory*, New York, pp. 144–152, 1992.

Brailsford, T.J. and Faff, R.W., An evaluation of volatility forecasting techniques. *J. Bank. Financ.*, 1996, **20**, 419–438.

Calvet, L.E. and Fisher, A.J., Forecasting multifractal volatility. *J. Econom.*, 2001, **105**, 27–58.

Calvet, L.E. and Fisher, A.J., Multifractality in asset returns: Theory and evidence. *Rev. Econ. Stat.*, 2002, **84**, 381–406.

Calvet, L.E. and Fisher, A.J., Regime-switching and the estimation of multifractal processes. *National Bureau of Economic Research, Inc, NBER Working Papers*, Vo. 2, 2003 (Working Paper 9839: Cambridge, MA). http://www.nber.org/papers/w9839

Calvet, L.E. and Fisher, A.J., How to forecast long-run volatility: Regime switching and the estimation of multifractal processes. *J. Financ. Econom.*, **2**, 200449–83.

Calvet, L.E. and Fisher, A.J., *Multifractal Volatility*, 2008 (Academic Press).

Calvet, L.E. and Fisher, A.J., Extreme risk and fractal regularity in finance. *Contemp. Math.*, 2013, **601**.

Cao, L.J. and Tay, F.E.H., Support vector machine with adaptive parameters in financial time series forecasting. *IEEE Trans. Neural Netw.*, 2003, **14**, 1506–1518.

Courant, R. and Hilbert, D., *Methods of Mathematical Physics*, Vol. 1, p. 112, 1953 (Interscience: New York).

Daubechies, I., The wavelet transform, time-frequency localization and signal analysis. *IEEE Trans. Inform. Theory*, 1990, **36**, 961–1005.

Daubechies, I., *Ten Lectures on Wavelets*, Vol. 61, 1992 (Siam).

Di Matteo, T., Multi-scaling in finance. *Quant. Finance*, 2007, **7**, 21–36.

Engle, R.F., Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econ.: J. Econ. Soc.*, **51**, 1982987–1007.

Fisher, A.J., Calvet, L.E. and Mandelbrot, B., Multifractality of Deutschemark/US Dollar Exchange Rates. Cowles Foundation Discussion Papers 1166, Cowles Foundation for Research in Economics, Yale University, 1997.

Galluccio, S., Caldarelli, G., Marsili, M. and Zhang, Y.C., Scaling in currency exchange. *Phys. A: Stat. Mech. Appl.*, 1997, **245**, 423–436.

Genton, M.G., Classes of kernels for machine learning. *J. Mach. Learn. Res.*, 2001, **2**, 299–312.

Gospodinov, N., Gavala, A. and Jiang, D., Forecasting volatility. *J. Forecast.*, 2006, **25**, 381–400.

Granger, C.W.J. and Joyeux, R., An introduction to long memory time series models and fractional differencing. *J. Time Seri. Anal.*, 2008, **1**, 15–29.

Granger, C.W.J. and Poon, S.H., Forecasting volatility in financial markets: A review. *J. Econ. Lit.*, 2003, **41**, 478–539.

Hansen, P.R., A test for superior predictive ability. *J. Bus. Econ. Stat.*, 2005, **23**, 365–380.

Idier, J., Long term vs. short term comovements in stock markets: The use of Markov-switching multifractal models. Working papers 218, Banque de France, 2008.

Kahane, J.P. and Peyriére, J., Sur certaines martingales de Benoit Mandelbrot. *Adv. Math. (N. Y)*, 1976, **22**, 131–145.

Kang, S.H., Kang, S.M. and Yoon, S.M., Forecasting volatility of crude oil markets. *Energy Econ.*, 2009, **31**, 119–125.

Kennedy, J. and Eberhart, R., Particle swarm optimization. In *Proceedings of ICNN'95 – International Conference on Neural Networks*, Vol. 4, pp. 1942–1948, Nov. 1995 (IEEE conference publication: Perth, WA).

Kim, K., Financial time series forecasting using support vector machines. *Neurocomputing*, 2003, **55**, 307–319.

Krusienski, D. and Jenkins, K., A modified particle swarm optimization algorithm for adaptive filtering. *ISCAS 2006: 2006 IEEE International Symposium on Circuits and Systems*, Kos, Greece, IEEE, 2006.

Law, T. and Shawe-Taylor, J., Practical Bayesian support vector regression for financial time series prediction and market condition change detection. *Quant. Finance*, 2017, **17**, 1403–1416.

Lin, S.W., Ying, K.C., Chen, S.C. and Lee, Z.J., Particle swarm optimization for parameter determination and feature selection of support vector machines. *Expert Syst. Appl.*, 2008, **35**, 1817–1824.

Lux, T., The Markov-Switching multifractal model of asset returns: GMM estimation and linear forecasting of volatility. Economics Working Papers 2006-17, Christian-Albrechts-University of Kiel, Department of Economics, 2006.

Lux, T. and Kaizoji, T., Forecasting volatility and volume in the Tokyo stock market: Long memory, fractality and regime switching. Economics Working Papers 2006-13, Christian-Albrechts-University of Kiel, Department of Economics, 2006.

Lux, T. and Kaizoji, T., Forecasting volatility and volume in the Tokyo Stock Market: Long memory, fractality and regime switching. *J. Econ. Dyn. Control*, 2007, **31**, 1808–1843.

Lux, T. and Morales-Arias, L., Relative forecasting performance of volatility models: Monte Carlo evidence. *Quant. Finance*, 2013, **13**, 1375–1394.

Lux, T. and Segon, M., Multifractal models in finance: Their origin, properties, and applications. In *The Oxford Handbook of Computational Economics and Finance*, edited by S.-H. Chen, M. Kaboudan, and Y.-R. Du, pp. 204–248, 2018 (Oxford University Press: Oxford). DOI:10.1093/oxfordhb/9780199844371.013.8

Makridakis, S., Andersen, A., Carbone, R., Fildes, R., Hibon, M., Lewandowski, R., Newton, H., Parzen, E. and Winkler, R., The accuracy of extrapolation (time series) methods: Results of a forecasting competition. *J. Forecast.*, 1982, **1**, 111–153.

Mallat, S., A Theory of Multiresolution Signal Decomposition: The Wavelet Representation. *IEEE Trans. Pattern Anal. Mach. Intel.-PAMI*, **11**, 1989, 674–693.

Mallat, S., *A Wavelet Tour of Signal Processing*, 1999 (Elsevier).

Mandelbrot, B., Intermittent turbulence in self-similar cascades: Divergence of high moments and dimension of the carrier. *J. Fluid Mech.*, 1974, **62**, 331–358.

Mandelbrot, B., Scaling in financial prices: I. Tails and dependence. *Quant. Finance*, 2001a, **1**, 113–123.

Mandelbrot, B., Scaling in financial prices: II. Multifractals and the star equation. *Quant. Finance*, 2001b, **1**, 124–130.

Mandelbrot, B., Scaling in financial prices: III. Cartoon Brownian motions in multifractal time. *Quant. Finance*, 2001c, **1**, 427–440.

Mandelbrot, B., Scaling in financial prices: IV. Multifractal concentration. *Quant. Finance*, 2001d, **1**, 641–649.

Mandelbrot, B., Fisher, A.J. and Calvet, L.E., A multifractal model of asset returns. Cowles Foundation Discussion Papers 1164, Cowles Foundation for Research in Economics, Yale University, 1997.

Mandelbrot, B., The variation of certain speculative prices. *J. Business*, 1963, **36**, 394–394.

Molchan, G.M., Scaling exponents and multifractal dimensions for independent random cascades. *Comm. Math. Phys.*, 1996, **179**, 681–702.

Nielsen, M.r. and Frederiksen, P., Finite sample accuracy and choice of sampling frequency in integrated volatility estimation. *J. Empir. Financ.*, 2008, **15**, 265–286.

Ortega, L. and Khashanah, K., A neuro-wavelet model for the short-term forecasting of high-frequency time series of stock returns. *J. Forecast.*, 2014, **33**, 134–146.

Palm, F. and Zellner, A., To combine or not combine? Issues of combining forecasts. *J. Forecast.*, 1992, **11**, 687–701.

Peng, Y., Albuquerque, P.H.M., de Sá, J.M.C., Padula, A.J.A. and Montenegro, M.R., The best of two worlds: Forecasting high frequency volatility for cryptocurrencies and traditional currencies with Support Vector Regression. *Expert. Syst. Appl.*, 2018, **97**, 177–192.

Politis, D.N. and Romano, J.P., The stationary bootstrap. *J. Am. Stat. Assoc.*, 1994, **89**, 1303–1313.

Reher, G. and Wilfling, B., A nesting framework for Markov-switching GARCH modelling with an application to the German stock market. *Quant. Finance*, 2016, **16**, 411–426.

Schmitt, F., Schertzer, D. and Lovejoy, S., Multifractal analysis of foreign exchange data. *Appl. Stoch. Models Data Anal.*, 1999, **15**, 29–53.

Schölkopf, B., Burges, C. and Smola, A.J., *Advances in Kernel Methods: Support Vector Machines*, 1998 (MIT Press: Cambridge, MA).

Schölkopf, B. and Smola, A.J., Support vector machines and kernel algorithms. In *The Handbook of Brain Theory and Neural Networks*, edited by M. Arbib, 2002 (MIT Press: Cambridge, MA).

Siao, J.S., Hwang, R.C. and Chu, C.K., Predicting recovery rates using logistic quantile regression with bounded outcomes. *Quant. Finance*, 2016, **16**, 777–792.

Smola, A.J. and Schölkopf, B., A tutorial on support vector regression. *Stat. Comput.*, 2004, **14**, 199–222.

Smola, A.J., Schölkopf, B. and Müller, K.R., The connection between regularization operators and support vector kernels. *Neural Netw.*, 1998, **11**, 637–649.

Tang, L.B., Tang, L.X. and Sheng, H.Y., Forecasting volatility based on wavelet support vector machine. *Expert. Syst. Appl.*, 2009, **36**, 2901–2909.

Tay, F. and Cao, L., Modified support vector machines in financial time series forecasting. *Neurocomputing*, 2002, **29**, 847–861.

Taylor, S.J. and Xu, X., The incremental volatility information in one million foreign exchange quotations. *J. Empir. Financ.*, 1997, **4**, 317–340.

Vapnik, V., *Statistical Learning Theory*, 1998 (Wiley: New York).

Vapnik, V., *The Nature Of Statistical Learning Theory*, 2000 (Springer).

Vassilicos, J. and Hunt, J., Turbulent flamelet propagation. *Combustion Sci. Tech.*, 1993, **87**, 291–327.

Wang, B., Huang, H. and Wang, X., A support vector machine based MSM model for financial short-term volatility forecasting. *Neural Comput. Appl.*, 2013, **22**, 21–28.

Wedding II, D.K. and Cios, K., Time series forecasting by combining RBF networks, certainty factors, and the Box-Jenkins model. *Neurocomputing*, 1996, **10**, 149–168.

Wu, Q., The forecasting model based on wavelet ν-support vector machine. *Expert. Syst. Appl.*, 2009, **36**, 7604–7610.

Wu, Q., Hybrid model based on wavelet support vector machine and modified genetic algorithm penalizing Gaussian noises for power load forecasts. *Expert. Syst. Appl.*, 2011, **38**, 379–385.

Yamaguchi, T. and Yasuda, K., Adaptive particle swarm optimization; self-coordinating mechanism with updating information. In *Proceedings of the Systems, Man and Cybernetics, 2006. SMC'06*, Vol. 3, pp. 2303–2308, 2006 (IEEE: Taipei, Taiwan).

Yang, H., Shao, C. and Khashanah, K., Multi-scale economic dynamics: The micro–macro wealth dynamics and the two-level imbalances of the Euro crisis. *Comput. Econ.*, 2019, **53**, 587–616.

Zhang, L., Zhou, W. and Jiao, L., Wavelet support vector machine. *IEEE Trans. Syst., Man, Cybern Part B, Cybern.: A Publ. IEEE Syst., Man, Cybern. Soc.*, 2004, **34**, 34–39.

Zhang, Q. and Benveniste, A., Wavelet networks. *IEEE Trans. Neural Netw.*, 1992, **3**, 889–898.