International Journal of Forecasting 35 (2019) 601–615

Contents lists available at ScienceDirect

International Journal of Forecasting

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

Combining wavelet decomposition with machine learning to forecast gold returns

Marian Risse[1]

Department of Economics, Helmut Schmidt University, Holstenhofweg 85, P.O.B. 700822, 22008 Hamburg, Germany

ARTICLE INFO

Keywords:
Real-time forecasting
Discrete wavelet transform
Support vector regression
Trading rule

ABSTRACT

This paper combines the discrete wavelet transform with support vector regression for forecasting gold-price dynamics. The advantages of this approach are investigated using a relatively small set of economic and financial predictors. I measure model performance by differentiating between a statistically-motivated out-of-sample forecasting exercise and an economically-motivated trading strategy. Disentangling the predictors with respect to their time and frequency domains leads to improved forecasting performance. The results are robust compared to alternative forecasting approaches. My findings on the relative importances of such wavelet decompositions suggest that the influences of short-term and long-term trends are not stable over the full evaluation period.

1. Introduction

Once the target series is expected to depend on a broad information set, forecasting in economics and finance requires a researcher to use a large set of predictors. In order to cover important predictors, researchers then use various techniques to compress or combine the predictor space. Examples include principal component analysis (PCA, see Forni, Hallin, Lippi, & Reichlin, 2005; H.S.tock & W.W.atson, 2002; McCracken & Ng, 2016; Stock & Watson, 2002), (Bayesian) model averaging and selection (Avramov, 2002; Koop & Korobilis, 2012; Prüser, 2019; Raftery, Kárnỳ, & Ettler, 2010; Risse & Kern, 2016), or shrinking the predictor space using penalty terms, as it is done by shrinkage regression or least angle regression (Efron, Hastie, Johnstone, & Tibshirani, 2004; Tibshirani, 1996). Recent applications, such as those of Medeiros and Vasconcelos (2016), Garcia, Medeiros, and Vasconcelos (2017), and Koop, Korobilis, and Pettenuzzo (in press), for example, open the broad field of machine learning to decrease the model dimension (for applications of machine learning algorithms in the field of economic and financial forecasting, see Döpke, Fritsche, & Pierdzioch, 2017; Mittnik, Robinzonov, & Spindler, 2015; Pierdzioch, Risse, & Rohloff, 2016b, among others).[2]

One drawback of such approaches is that researchers have to build a large database to control for model uncertainty and to cover the maximum movement of the target series. Building such a database can be challenging, as data often are not available over the full evaluation period.[3] As

E-mail address: marian.risse@hsu-hh.de.

1 *Acknowledgments:* I thank Joscha Beckmann, Christian Pierdzioch, and participants of the Annual Conference of the European Economics and Finance Society (Ljubljana, 2017). I am grateful to participants of the 12th ifo Workshop on Macroeconomics and Business Cycles (Dresden, 2018) and the 20th Workshop on International Economics (Göttingen, 2018). The usual disclaimer applies.

2 Reducing the predictor space when applying PCA can also be done using hard thresholding (Bai & Ng, 2008). However, model selection generally increases the probability that the *true* forecasting model is not included, and thus, may worsen the forecasting performance. One solution for attenuating this problem is a combination of selection techniques using *bridge regressions*, as has been proposed by Bulligan, Marcellino, and Venditti (2015) and Konzen and Ziegelmann (2016), for example. However, Elliott, Gargano, and Timmermann (2013, 2015) contradict this by claiming the superiority of randomly-combined forecasting models via complete subset regressions that explicitly neglect the true model and outperform static factor models.

3 It is worth noting that compression techniques such as PCA also work with periods of zero observations and without interpolation (see for example Koop & Korobilis, 2014).

https://doi.org/10.1016/j.ijforecast.2018.11.008
0169-2070/

an example, the Federal Reserve has recently made available to researchers a real-time database that features more than one hundred time series (see McCracken & Ng, 2016). Principal component analysis is expected to work well in this environment, as it benefits from a dense correlation structure of related time series. However, model-averaging techniques are applicable only if the predictor space is reduced by, say, PCA (Koop & Potter, 2004) or supervised learning techniques (Onorante & Raftery, 2016; Risse & Ohl, 2017).

The approach in this paper differs from those in the literature above because I assume a relatively compact dataset, leading to a good forecasting performance, but with a specific focus on the properties of the individual predictor series.[4] I extend the predictor space by extracting the time-based *and* frequency-based properties of the original time series using the discrete wavelet transform (DWT). The resulting multiresolution decomposition (MRD) of the original time series creates an expanded database that consists of the MRD-related subseries. Creating such a database allows a clear separation between the short- and long-run trends of the original time series without making use of additional data. In this regard, one advantage over the classical Fourier transform, for example, is the extraction of a coherent set of trend components that summarize back to the original time series. Having identified those components, a forecaster is able to focus, on the one hand, on rather heavily fluctuating market signals, as could be the case in an emerging crisis period, or, on the other hand, on gradually varying fundamental values, as was observed in the time of the Great Moderation.

The use of wavelet functions in the economics and finance literature is not new (for an introduction, see Ramsey, 2002, and Crowley, 2007; for applications, see, among others, Aguiar-Conraria & Soares, 2011a,b; Asgharian, Christiansen, & Hou, 2016; Berger, 2016; Gallegati, 2012; Rua & Nunes, 2009; and Beckmann, Berger, & Czudaj, 2017). However, analyses of the usefulness of the wavelet transform in terms of economic forecasting are very scarce, and are typically constrained to the decomposition of the target variable only (Conejo, Contreras, Espinola, & Plazas, 2005; Wong, Ip, Xie, & Lui, 2003). One possible reason for this could be that decomposing even a small dataset leads to dozens of new predictors needing to be analyzed. Thus, standard approaches such as model averaging over the full predictor space are not applicable. Two notable exceptions are the studies by Rua (2011, 2017), who uses a large economic database and reduces the dimension *within* the decompositions via PCA. The resulting wavelet-based factor models work well relative to standard, factor-augmented forecasting models. However, using such a large database complicates the evaluation of individual predictors, which is the goal of this paper.

Recently, Faria and Verona (2017, 2018) used the DWT for decomposing predictors of the Welch and Goyal (2008) dataset in order to forecast the S&P 500 index. Thus, the authors focus on the decompositions of the individual series without using a combined evaluation strategy. I abstract from this single-predictor evaluation by using support vector regression (SVR) to extract the most important components from the decomposed predictor space.[5] I also compute results for other machine-learning approaches in order to highlight the advantages of SVR.[6]

I study the advantages of combining the DWT with SVR (in what follows, DWT-SVR) by using gold excess returns as the target variable to be predicted out of sample. Gold is particularly useful in this context because there does not exist either a generally accepted theoretical pricing model or an empirically justified and generally accepted database. As a result, the price of gold has been the subject of extensive research. Researchers expect it to be influenced by a broad range of predictors, covering the fields of commodities, financial variables, macroeconomic data, and interest rates (for a survey, see O'Connor, Lucey, Batten, & Baur, 2015). The advantage of applying the DWT-SVR approach is that a researcher can consider a broad range of influences without explicitly including a vast number of additional time series in the original dataset. As an example, to help in understanding the economic intuition behind the DWT, one would expect that a fast-moving component of the consumer price index (CPI), such as commodity prices, might have a stronger impact in terms of predictability in turbulent economic times than a slower-moving component such as highly industrialized goods (for example, cars), or vice versa in times of a moderate economic development. Likewise, the influence of a commodity price index could also depend on faster-moving commodities such as crude oil, or on commodities that depend instead on the cyclical behavior of the economy. In this regard, the advantage of the DWT is that these dynamics can be analyzed by adding only two original time series to the forecasting model.[7]

I evaluate the performance of the forecasting environment by differentiating between a statistical and an economic evaluation, as is standard for finance-related time series (see Campbell & Thompson, 2008; Rapach, Ringgenberg, & Zhou, 2016; Rapach, Strauss, & Zhou, 2010; Welch & Goyal, 2008, among others). Unfortunately, the literature on the evaluation of the out-of-sample performance of gold is scarce. For example, Aye, Gupta, Hammoudeh, and Kim (2015), Baur, Beckmann, and Czudaj (2016), Pierdzioch, Risse, and Rohloff (2014), and Pierdzioch et al. (2016b) use out-of-sample forecast evaluation techniques to identify the most important drivers of gold returns. The first two publications only evaluate the statistical accuracy compared to a random walk model, but Pierdzioch et al. (2014,

[4] The use of a relatively small predictor space is also supported by De Mol, Giannone, and Reichlin (2008) and Bulligan et al. (2015), who report evidence of the superiority of a small but well-behaved dataset over the whole predictor space after dimension reduction.

[5] A similar approach was also proposed by Sujjaviriyasup (2017) and Seo, Choi, and Choi (2017), but not in the field of economics and finance.

[6] The choice of the SVR machine-learning algorithm is empirically-motivated. In principle, researchers can use any dimension-reducing algorithm they prefer, depending on the specific research question.

[7] In fact, one disadvantage of this approach is that short-term and long-term decompositions are not related directly to observable time series. This is also a problem for PCA, where the economic meaning of the extracted factors cannot be addressed directly.

2016b) focus on the economic value added by those forecasts.[8] Risse and Ohl (2017) provide an extensive evaluation of static and dynamic forecasting models. However, the authors report good performances only in terms of a trading simulation, not in terms of statistical accuracy. This is not surprising, since improving on naive benchmarks, such as the historical mean, for predicting financial time series is still a challenging task (Campbell & Thompson, 2008; Rapach et al., 2016; Welch & Goyal, 2008).

The outline of this paper is as follows. Section 2 explains the shortcuts of the underlying econometric approaches. Section 3 describes the empirical framework and Section 4 lays out the results. Finally, Section 5 concludes.

2. Econometric framework

This section briefly presents the DWT and SVR approaches. Both approaches on their own are well-established in the literature, and applications in this context rely on standard specifications. Therefore, I only provide an intuitive introduction.[9] In general, estimation follows a two-step procedure where the DWT is first used to create an extended database that includes all decompositions of the original predictor series; then, in a second step, this database is used as the input for the SVR in order to identify the most important trend components for forecasting gold returns. Since the evaluation includes a true out-of-sample forecasting exercise, the procedure is replicated in every period of time in which new data become available.

2.1. Discrete wavelet transform

The principle of wavelet decomposition allows a researcher to zoom in on a time series by scaling and stretching a predefined wavelet function over all observations so as to extract abrupt short-term and long-term changes given a predefined frequency band. Wavelets are defined as either low-pass or high-pass filters, where the low-pass filter is also known as the father wavelet and the high-pass filter is known as the mother wavelet. For the discrete wavelet filter, three fundamental conditions must hold. Let h denote the mother wavelet, L the (even) width of the filter, and $j = 1, \ldots, J$ the level of decomposition; then, it follows that

$$\sum_{h=0}^{L-1} h_{j,l} = 0, \qquad \sum_{h=0}^{L-1} h_{j,l}^2 = 1, \qquad \sum_{h=0}^{L-1} h_{j,l}h_{j,l+2n} = 0, \quad \text{with } n \neq 0. \tag{1}$$

The meanings of these three conditions are: (i) all wavelet coefficients sum to zero, (ii) all wavelet coefficients have unit energy, and (iii) the filter is orthogonal to its even shifts. The second condition leads wavelets to be finite in time, which is one of the most important differences from the classical Fourier transform. The father wavelet can be constructed from the mother wavelet via the quadrature mirror filter relationship

$$g_{j,l} = (-1)^{j,l+1} h_{J,L-1-l}. \tag{2}$$

The filter has to be specified and depends on the properties of the time series. In what follows, I use the symmetric filter as proposed by Haar (1910), where $g_0 = g_1 = 1/\sqrt{2}$ and $h_0 = 1/\sqrt{2}$, $h_1 = -1/\sqrt{2}$ (for $L = 2$). In this case, the low-pass filter is a pairwisely weighted average, whereas the high-pass filter is the weighted difference between the corresponding observations.[10]

Wavelet coefficients for the next level can be computed using the pyramid algorithm (Mallat, 1989); i.e., the smooth component of the current level is the time series that is to be filtered at the next level. As a consequence, each level halves the number of wavelet coefficients. Hence, a drawback of the standard DWT is that the number of observations must be divisible by 2^L. I overcome this drawback by using the maximum overlap discrete wavelet transform (MODWT), where the number of coefficients is equal to the sample size across all levels. The MODWT filters $(\tilde{h}, \tilde{g})$ are obtained from the DWT directly by shifting the input vector circularly[11] and adjusting the filter with

$$\tilde{h}_{j,l} = h_{j,l}/2^{\frac{j}{2}}, \tag{3}$$

$$\tilde{g}_{J,l} = g_{J,l}/2^{\frac{J}{2}}. \tag{4}$$

Accordingly, the wavelet coefficients $\tilde{w}$, $\tilde{v}$ are computed as

$$\tilde{w}_{j,t} = \sum_{l=0}^{L-1} \tilde{h}_{j,l} x_{t-l \bmod N}, \tag{5}$$

$$\tilde{v}_{J,t} = \sum_{l=0}^{L-1} \tilde{g}_{J,l} x_{t-l \bmod N}, \tag{6}$$

where $L_j = (2^j - 1)(L - 1) + 1$ and the modulus (mod) operator controls for the boundary of the finite length, N, of the observation's vector. At the first level, the pyramid algorithm starts with decomposing the original series, x, with the MODWT filters, such that

$$\tilde{w}_{1,t} = \sum_{l=0}^{L-1} \tilde{h}_l x_{t-l \bmod N}, \tag{7}$$

8 A brief comparison of statistical accuracies in terms of component-wise boosting is given by Pierdzioch, Risse, and Rohloff (2016a).

9 For a technical introduction, the interested reader is referred to the work of Percival and Walden (2000, DWT) and Vapnik (1995, SVR). Non-technical introductions are given by Crowley (2007, DWT) and Smola and Schölkopf (2004, SVR).

10 The Haar wavelet is used frequently in the economics and finance literature (see for example Beckmann et al., 2017; Berger, 2016), and also leads to a sufficient performance in this forecasting exercise. An alternative filter is that proposed by Daubechies (1992), which did not lead to a superior performance, and therefore is not presented in the empirical analysis.

11 See Percival and Walden (2000, Chapter 5.3) for a detailed description. It is worth noting that circular shifting influences the original order of the time series (at its boundaries) for computing the wavelet coefficient, which complicates a structural analysis as well as an application of a full-fledged time-varying parameter model. For an example of a forecasting exercise on financial markets using dynamic recurrent neural networks, see Murtagh, Starck, and Renaud (2004).

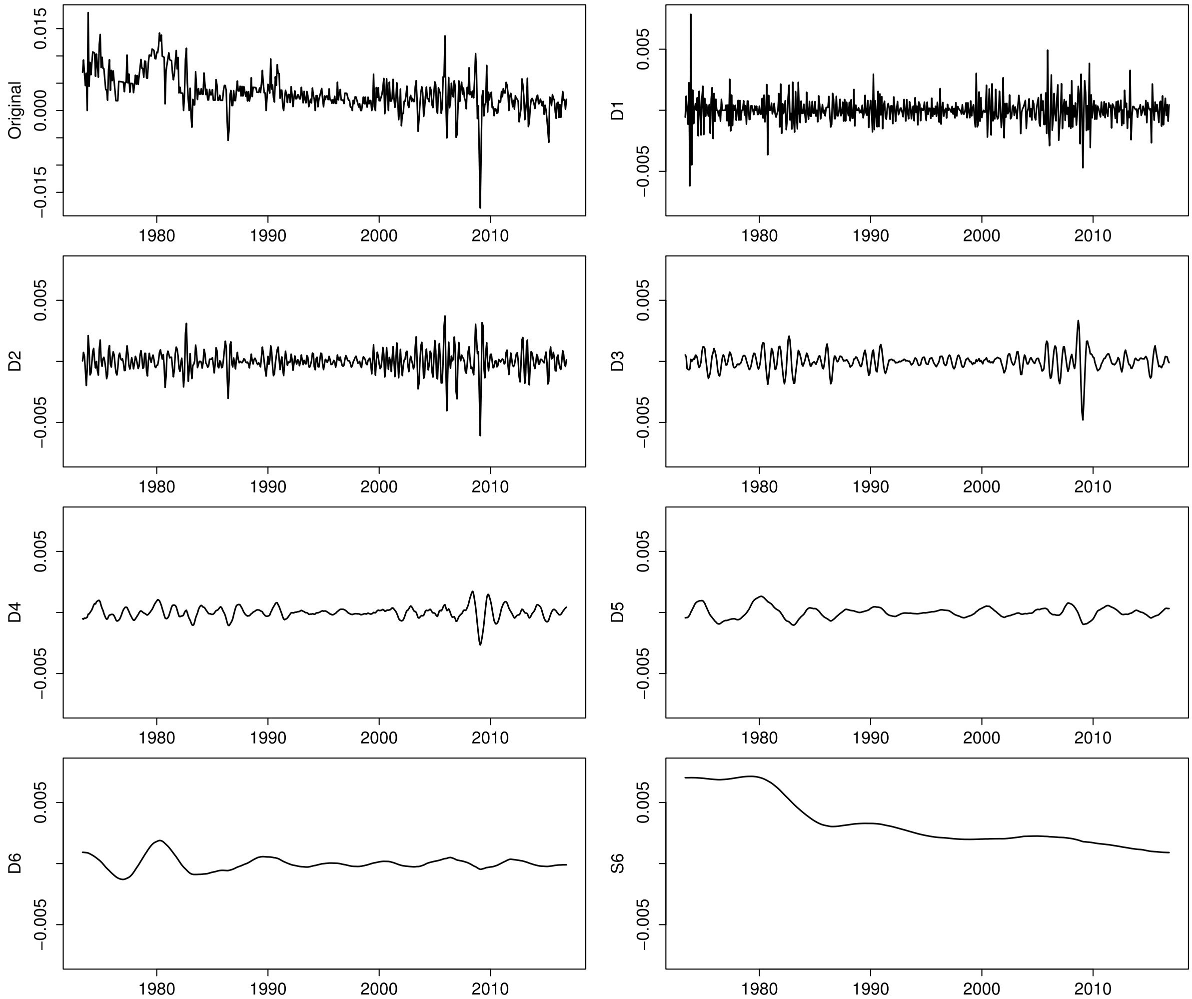

Fig. 1. Decomposition of CPI. Notes: The figure shows the decomposition of month-on-month changes in the log CPI index using the MODWT.

$$\tilde{v}_{1,t} = \sum_{l=0}^{L-1} \tilde{g}_l x_{t-l \bmod N}, \tag{8}$$

and it then continues with the wavelet smooth at the next level

$$\tilde{w}_{2,t} = \sum_{l=0}^{L-1} \tilde{h}_l \tilde{v}_{1,t-l \bmod N}, \tag{9}$$

$$\tilde{v}_{2,t} = \sum_{l=0}^{L-1} \tilde{g}_l \tilde{v}_{1,t-l \bmod N}, \tag{10}$$

and so on. Finally, a multiresolution analysis captures the wavelet transform at each level such that

$$x_t = x(\tilde{D}_1) + \cdots + x(\tilde{D}_J)_t + x(\tilde{S}_J)_t, \tag{11}$$

where $\tilde{D}_j$ denotes the wavelet detail that covers variations in x for the j^{th} level and $\tilde{S}_J$ represents the final wavelet smooth. As a consequence, the lower-level decompositions filter out the higher-frequency fluctuations, whereas filtering becomes smoother as j increases.

As an example, I consider the consumer price series that is depicted in Fig. 1. The top left graph shows the series in its log-differenced original transformation, while the following graphs plot the decomposition of six wavelet details (D1–D6) and the wavelet smooth component (S6).

While the first wavelet detail uses a bandwidth from two to four months, the sixth wavelet detail uses a bandwidth from 64 to 128 months. Thus, a seventh detail would require a training period of at least 256 months. In the following empirical analysis, I use a training period of approximately 20 years (236 months), so no more than six wavelet details are considered.

2.2. Support vector regression

SVR belongs to the class of supervised machine-learning algorithms where a non-linear relationship in the data is modeled by mapping a linear function into a high dimensional, kernel induced feature space.

I formalize SVR by starting with linear functions of the form

$$f(x) = \langle w, x_t \rangle + b, \tag{12}$$

where $\langle \cdot, \cdot \rangle$ is the dot product, w denotes some weight vector, $t = 1, \ldots, T$ is a time index, and b is a bias term.[12] The goal of SVR is to find a function $f(x)$ that has a maximum deviation of ϵ from the true values y_t, making w as flat as possible.[13] Analytically, this can be achieved by minimizing the norm $\|w\|^2 = \langle w, w \rangle$. The objective function, expressed in terms of a convex optimization problem, is then

$$\min \frac{1}{2}\|w\|^2 + c\sum_{t=1}^{T}(\xi_t + \xi_t^*), \tag{13}$$

subject to

$$y_t - f(x) \le \epsilon + \xi_t,$$
$$f(x) - y_t \le \epsilon + \xi_t^*,$$
$$\xi_t, \xi_t^* \ge 0.$$

The second term in Eq. (13) relaxes the model framework to a soft margin loss setting using slack variables ξ_t, ξ_t^*, which express the difference between an observation and ϵ. Values above ϵ are tolerated, but are penalized with the hyperparameter c.[14] Observations below ϵ are not penalized, resulting in a so-called ϵ-insensitive loss function. Thus, adding more training data within the threshold does not change the SVR outcome. Changing the hyperparameter c results in a trade-off between the flatness of $f(x)$ and the penalized loss.

Solving the Lagrange dual formula of the objective function leads to

$$f(x) = \sum_{t=1}^{T}(\alpha_t - \alpha_t^*)\langle x_t, x \rangle, \tag{14}$$

where α_t and α_t^* denote non-negative multipliers. To make the SVR nonlinear, a kernel function, $K(x_t, x')$, is introduced to replace the second term in Eq. (14). Popular kernel functions include the radial basis kernel (RBF), the polynomial kernel, and the sigmoid kernel. The RBF kernel, $K(x_t, x') = exp\left(-\delta\|x_t - x'\|^2\right)$, is used widely for the SVR, and also led to the best performances in this research. By choosing the RBF kernel, an additional hyperparameter, δ, is used to control for the bias-variance trade-off of the individual support vectors. Finally, Eq. (14) expands to

$$f(x) = \sum_{t=1}^{T}(\alpha_t - \alpha_t^*)exp\left(-\delta\|x_t - x\|^2\right). \tag{15}$$

In summary, the formal description of SVR requires three hyperparameters (ϵ, c, δ) to be specified by the researcher. In a static framework, a natural choice is to use cross-validation and some grid recommendations. However, since the model is estimated recursively, this procedure has to be repeated in every period of time in which the outcome can be influenced substantially by the variation of hyperparameters. For example, without prior knowledge, widely accepted grids for c and δ are $\{2^{-5}, 2^{15}, \ldots, 2^3\}$ and $\{2^{-15}, 2^{-13}, \ldots, 2^3\}$, respectively.[15] Unfortunately, training the model on these grids did not help to improve the performance. For this reason, I follow standard recommendations as they are proposed in the popular scikit-learn environment (Pedregosa, Varoquaux, Gramfort, Michel, Thirion, Grisel, et al., 2011) and set $\epsilon = 0.1$ and $\delta = 1/K$, where K equals the number of predictors. With respect to c, I follow Cherkassky and Ma (2004), who report a good approximation of $c = 3 * \sigma_y$, where σ_y is the standard deviation of gold returns (for training data only). I also used the initial training data to verify these recommendations by applying ten-fold cross validation, which led to similar results. Hence, I use these rules of thumb for the whole estimation process. As a consequence, c varies over time and δ is model-specific, depending on the number of predictors.

3. Empirical framework

3.1. The data

The sorts of predictors that are often studied as determinants of gold returns include the property of the precious metal acting as a hedging/safe haven instrument in terms of inflation (Beckmann & Czudaj, 2013; Blose, 2010; Mahdavi & Zhou, 1997), currencies (Capie, Mills, & Wood, 2005; Pierdzioch et al., 2016b; Pukthuanthong & Roll, 2011; Reboredo, 2013), and the stock market (Baur & Lucey, 2010; Baur & McDermott, 2010; Beckmann, Berger, & Czudaj, 2015). A theoretical justification for the link between gold and interest rates is given by Fortune (1987) and tested empirically by Batten, Ciner, and Lucey (2014). Since gold is also a commodity that is used for the production of industrial goods, comovements with other commodity prices can be expected.

I cover these influences by using a compact dataset that consists of nine time series, namely the consumer price index (**cpi**), the trade-weighted exchange rate (**twex**), the long-term yield on government bonds (**tbond**), the term spread (**tms**), the S&P500 composite index (**sp500**), the GSCI commodity index (**gsci**), the CRB commodity index (**crb**), global foreign exchange reserves (**reserves**), and the adjusted realized volatility (**volatility**), as the preferred choice of predictors. The realized volatility acts as a proxy for uncertainty and is computed from daily data, adjusted for autocorrelation (see Marquering & Verbeek, 2004). The series, along with their sources and transformations, are presented in Table 1.

12 The data are typically standardized to a zero mean, making b equal to zero.

13 A flat w implies that the predictions are less sensitive to random shocks (Alexandridis, Kampouridis, & Cramer, 2017).

14 Without slack variables, one achieves a hard margin loss setting that does not tolerate any values above ϵ. As a result, it may not be possible to optimize the objective function in some cases.

15 See Shynkevich, McGinnity, Coleman, Belatreche, and Li (2017) for a financially-driven application.

Table 1
The data.
Source: FRED, https://fred.stlouisfed.org/.

#	Series	Abbreviation	LAG	Source
0	Gold excess returns, see Eq. (16)	gold	–	Datastream
1	Inflation (CPI in logs, month-on-month change)	cpi	1	FRED
2	Trade-weighted U.S. dollar index: Broad (in logs, month-on-month change)	twex	0	FRED
3	Ten-year treasury bond	tbond	0	FRED
4	Term spread (ten-year treasury bond minus three-month treasury bill)	tms	0	FRED
5	S&P 500 composite index (in logs, month-on-month change)	sp500	0	Datastream
6	GSCI commodity index (in logs, month-on-month change)	gsci	0	Datastream
7	CRB commodity index (in logs, month-on-month change)	crb	0	Datastream
8	IMF-IFS foreign exchange reserves (in logs, month-on-month change)	reserves	3	Datastream
9	Adjusted realized volatility of gold (see Marquering & Verbeek, 2004)	volatility	0	–

Notes: The table reports sources and transformations of the time series used in this research. The LAG column determines the numbers of publication lags.

One-month-ahead[16] gold excess returns, r_{t+1}, are computed by subtracting the risk-free rate i_t from the log-difference of the price series P_t, such that

$$r_{t+1} = [log(P_{t+1}) - log(P_t)] - i_t. \tag{16}$$

For the daily available predictor data, I use end-of-month values. Monthly-based time series consider publication lags in order to create a realistic out-of-sample forecasting exercise where a forecaster uses only data that would have been available at time t to make predictions for $t + 1$.[17] Following the efficient market hypothesis, all available information in t is assimilated by the data series, and thus, any additional lag structure should be superfluous.[18] Due to availability, and considering the breakdown of the Bretton Woods System, the first observation of the transformed data starts in May 1973.

I use an environment that is comparable to that of Risse and Ohl (2017) and differentiate between two evaluation periods. First, I consider an out-of-sample evaluation period (the *full sample*) that ranges from January 1992 to December 2016. Hence, the training period comprises roughly 20 years of observations. As a second evaluation period, I also consider the impact of the recent financial crisis on the forecasting performance (the *crisis period*). This can also be treated as a worst case scenario, where an investor starts trading immediately after the collapse of Lehman Brothers in September 2007. The evaluation stops when the S&P 500 exceeds its pre-crisis level.[19] Thus, the crises period is from October 2007 to March 2013.

[16] The forecast horizon is limited to one month, allowing for a flexible application of the trading rule.

[17] It is worth noting that time series such as the consumer price index experience data revisions, making this application a *pseudo* real-time application.

[18] Note that, from a technical perspective, every additional lag increases the predictor space by 63, making it much harder for the SVR to isolate the most important influences. Even one additional lag leads to an enormous decrease in performance. Note also that, from a theoretical perspective, the efficient market hypothesis does not necessarily require the predictor space to be limited to the current time period.

[19] In this scenario, the training period is extended to 413 months, making it theoretically possible to add a seventh decomposition to the database. However, the results are not presented, to allow a direct comparison to the full period and because the performance did not improve.

Fig. 2 visualizes the price series in levels and excess returns. With respect to the crisis period, following the relatively small decline at the beginning of 2008, the gold-price dynamics are characterized by a sharp increase over the subsequent years, leading to peaks in mid-2011 and at the end of 2012. From an investor's perspective, beating a simple buy-and-hold strategy during that period should be very challenging for all types of forecasting models. The return series shows usual clustering.

3.2. Forecast evaluation

I asses the performance of the DWT-SVR approach relative to a simple forecasting scheme by following Welch and Goyal (2008) and Rapach et al. (2016) and, in a first step, running univariate regressions on each predictor. In a second step, I decompose the predictor series and use the SVR for estimation. Finally, as a last application, I estimate a forecasting model that includes all decomposed time series from all predictors and let the algorithm choose the best decompositions. I present various alternative forecasting approaches, including a naive historical mean forecast, to highlight the advantage of using SVR.[20]

The statistical accuracy is measured by computing the Campbell and Thompson (2008) out-of-sample R^2 (R^2_{oos}) that compares the forecasts of the underlying forecasting model, $\hat{r}_{t+1}$, with those of the historical mean benchmark, $\bar{r}_{t+1}$. The R^2_{oos} is computed as

$$R^2_{I,oos} = 100 * \left(1 - \frac{\sum_{t=0}^{T-1}(r_{t+1} - \hat{r}^I_{t+1})^2}{\sum_{t=0}^{T-1}(r_{t+1} - \bar{r}_{t+1})^2}\right), \tag{17}$$

where I is a model index and r_{t+1} denotes the actual values of gold excess returns. The test statistic describes the gain (in percentage) of the forecasting model over the historical mean benchmark. I test for significance of the improvement in performance by also applying the Clark and West (2007, CW) test for the forecast error comparison of nested forecasting models.

In addition to the statistical forecasting accuracy, researchers such as Leitch and Tanner (1991) have found that

[20] The historical mean forecast for period $t + 1$ is computed as the arithmetic mean of the excess return series using observations up to and including period t.

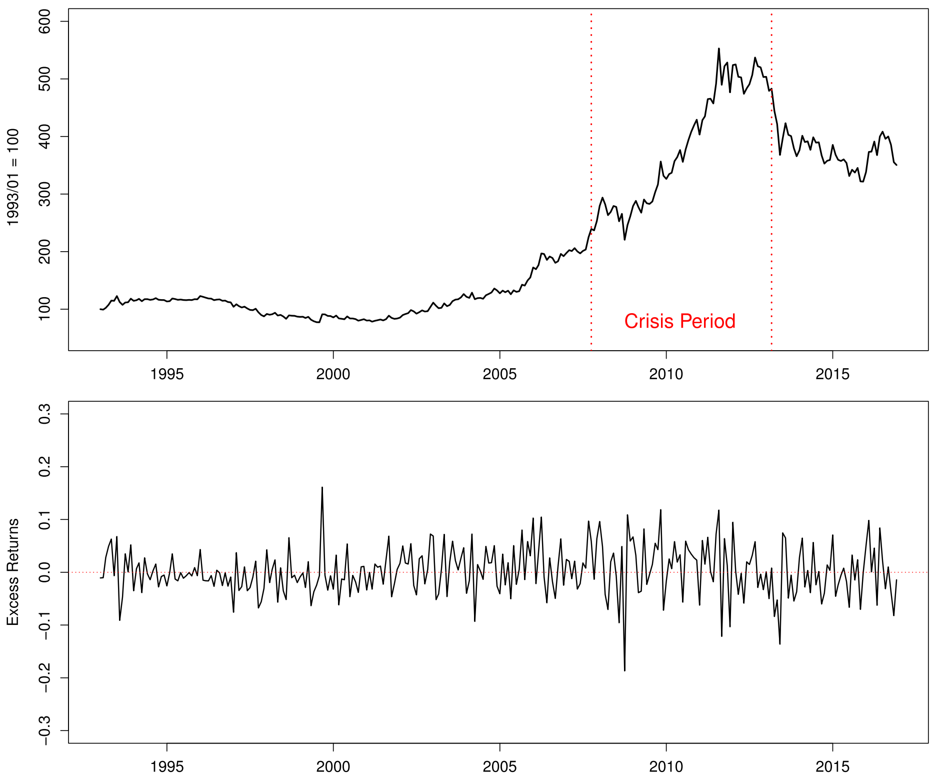

Fig. 2. Gold levels and returns. Notes: The figure shows the price of gold in levels (scaled to 1992/01 = 100) and the excess return series.

the statistical value added by a forecast does not necessarily imply a comparable increase in economic value. Since gold can be treated easily as a financially traded asset, I compute the additional economic value of these forecasts by means of a simple portfolio evaluation.

I apply a trading strategy where a mean–variance investor reallocates his investment in every period of time by computing portfolio weights, ϖ, between gold and a risk-free asset such that

$$\varpi_t = \frac{1}{\gamma}\frac{\hat{r}_{t+1}}{\hat{\sigma}^2_{t+1}}, \tag{18}$$

leading to a portfolio return of

$$R_{p,t+1} = \varpi_t r_{t+1} + (1-\varpi_t)(rf_{t+1}). \tag{19}$$

The risk-free return is denoted by rf_{t+1}, which is assumed to be the three-month treasury bill rate, and $\hat{r}_{t+1}$ denotes the excess return forecast for gold. The $t+1$ variance of gold returns, $\hat{\sigma}^2$, is defined as a ten-year rolling window forecast (see also Neely, Rapach, Tu, & Zhou, 2014, and Rapach et al., 2016, for an application to the S&P 500 return series) and weight-scaling is based upon a coefficient of relative risk-aversion $\gamma = 3$.[21]

Because the portfolio outcome depends strongly on the length of the trading period, I also compute the certainty equivalent return of the investment, CER_p. The CER_p depends on the average portfolio return μ_p and the portfolio variance σ_p^2, and is computed as

$$CER_p = \mu_p - \frac{1}{2}\gamma\sigma_p^2. \tag{20}$$

The CER reflects the risk-free rate that an investor is willing to accept in lieu of investing in the portfolio, and thus, is comparable across subsamples. In line with the statistical evaluation, I compute the portfolio performance in excess of a benchmark portfolio performance derived from the historical mean. The simulation of the portfolio strategy is initialized with one wealth unit. The terminal level of wealth (TW) and its Sharpe ratio (SR), measured as the quotient of the annualized mean return and standard deviation, are computed in every period.

In analogy to the CW test, the significance of the economic value added by forecasts is controlled by applying the GRS test for portfolio efficiency (Gibbons, Ross, & Shanken, 1989). The F-distributed test statistic is defined as

$$\frac{T(T-N-1)}{N(T-2)}W \sim F_{N,(T-N-1)},$$
$$\text{with } W = \left(\frac{\sqrt{1+SR_I^2}}{\sqrt{1+SR_{MEAN}^2}}\right)^2 - 1. \tag{21}$$

SR_I denotes the Sharpe ratio of the Ith forecasting model and SR_{MEAN} denotes the Sharpe ratio of the historical mean benchmark. T denotes the number of observations and N

21 Reducing the size of the window to, say, five years, as applied in Campbell and Thompson (2008), does not affect the results qualitatively. The same holds for a variation of γ.

the number of assets. Since the Sharpe ratio can also be negative, I adjust the test statistic accordingly. A rejection of the null hypothesis of equal performances implies a significant higher financial performance of the underlying forecasting model than of the historical mean.

3.3. Relative importance of predictors

The previous section limited model evaluation to the response variable (the target series). However, it is interesting to figure out whether the influence of the individual predictors changes after the original time series were decomposed. Since kernel-induced SVR is nonlinear, there does not exist a direct relationship between the right- and left-hand sides of the equation. Cortez and Embrechts (2013) propose evaluating such a black-box model by fixing all predictors that are not of interest to their sample means and varying the interesting predictor over the entire sample. The so-called sensitivity analysis (see also Kewley, Embrechts, & Breneman, 2000) then measures the impact of a certain predictor according to the response to the target series.

In what follows, I use a variance-based evaluation that measures the importance of the variable based on the amount of fluctuation in the response.[22] For any relevant predictor k, fluctuations are measured as

$$V_k = \frac{1}{N-1}\sum_{j=1}^{N}(\hat{r}_{k_j} - \tilde{r}_k)^2, \tag{22}$$

where $\tilde{r}_k$ is the mean of the response and $\hat{r}_k$ is the predicted value for the relevant predictor, k. N is the number of observations that are used for computing V_k. I use ten observations that are distributed equally across the training set.[23] The relative importance compared to the other M predictors, IMP_k, is then computed as

$$IMP_k = 100 \times \frac{V_k}{\sum_{m=1}^{M} V_m}. \tag{23}$$

4. Empirical results

4.1. Univariate evaluation

I start by evaluating the individual predictor series both on their own and by using the DWT-decomposed time series.[24] Table 2 presents the results, with the upper panel presenting the outcome for the full evaluation period and the lower panel presenting that for the crisis period. For both evaluation periods, I consider both statistical and economic forecast performances. When considering economic performances, I also present results of a simple buy-and-hold strategy. The first four columns of the left-hand side panels summarize the results derived from univariate regressions without decomposition. The right-hand side columns display the results of the wavelet decomposition, where the SVR is used for the selection of the best wavelet detail or the wavelet smooth. All results are expressed in terms of their outperformance of the naive historical mean benchmark. Trading results in terms of the latter are presented in the table notes.

Beating such a naive benchmark is not an easy task in terms of statistical forecast evaluation. When considering the full evaluation period, a significant improvement in terms of the univariate regression is observed for gsci and reserves (5% and 10% levels). For the decomposed time series, the R^2_{oos} is negative for all predictors, indicating quite a poor forecast accuracy relative to the historical mean. In fact, these results are not surprising, because similarly poor performances are reported for other finance-related time series (Campbell & Thompson, 2008).[25] However, in two cases (cpi and tbond), the R^2_{oos} increases after decomposition.

When we turn to the added economic value, the outcome changes drastically. While the results for the univariate regressions are not homogeneous, both the excess terminal wealth and the excess Sharpe ratio become positive for all time series after decomposition and become higher relative to the univariate regression for almost all series.[26] The highest excess level of wealth is observed for gsci (7.82); i.e., an investor more than quintuples his outcome relative to the historical mean. Similar results are observed for the CER, which exceeds that of the historical mean by 6.52 percentage points. Good performances in terms of wavelet decomposition are also reported for cpi, twex, and crb. Moreover, none of the decomposed outcomes are exceeded by the outcome of the original series. The GRS test highlights the significance of these findings for two-thirds of the predictors, thus doubling the number of significant predictors relative to the original time series. In addition, all significant trading results of the wavelet decomposition can exceed the outcomes of the buy-and-hold strategy. With respect to the univariate regressions, this is only true for gsci.

Next, I analyze whether these findings are robust to the worst case scenario: the recent financial crisis. In general, one would expect a relatively high economic outcome for a forecasting model that works well, as the gold series in levels experienced a sharp increase over the full evaluation period. The buy-and-hold strategy underlines this hypothesis because an investor approximately doubles his investment over the entire sample (TW: 2.14). The results for the decomposed time series also confirm these findings.

22 Cortez and Embrechts (2013) present three additional evaluation criteria that are based upon the range of the response, its median, and a gradient descent approach. All of these approaches led to similar results, and thus the results are not presented here. It is also worth noting that interaction effects can be visualized using this approach.

23 Increasing or decreasing the number of observations did not affect the results.

24 In terms of the practical implementation, I use the R packages *kernlab* (Karatzoglou, Smola, Hornik, & Zeileis, 2004), *rminer* (Cortez, 2016), and *waveslim* (Whitcher, 2015).

25 Statistical out-of-sample evaluations of excess gold return forecasts are relatively scarce, and typically rely only on a naive random walk (see Aye et al., 2015; Baur et al., 2016, among others). I do not present results for a random walk because the forecasting performance is less precise than that of the historical mean.

26 An exception is tbond, where the terminal wealth is lower but at a comparable level. However, the superiority of the Sharpe ratio and CER underline the certainty of this result relative to the original time series.

Table 2
Evaluation of the single time series in excess of the historical mean.

		R^2_{oos}	Δ_{CER}	Δ_{TW}	Δ_{SR}	R^2_{oos}	Δ_{CER}	Δ_{TW}	Δ_{SR}
		Univariate regression				**DWT-SVR**			
Full sample	**cpi**	−3.22	−0.64	−0.15	−0.01	−0.08	5.21	5.59	0.46***
	twex	−0.19	−0.43	−0.16	−0.09	−4.20	4.55	4.28	0.43***
	tbond	−2.64	1.42	2.28	0.21***	−0.75	2.68	2.11	0.29***
	tms	−2.00	−2.49	−0.75	−0.35	−2.52	−0.34	0.10	0.06
	sp500	−0.27	−1.06	−0.38	−0.20	−4.00	−0.14	0.23	0.08
	gsci	1.89**	3.10	2.50	0.33***	−0.58	6.52	7.82	0.59***
	crb	−0.05	−0.84	−0.22	−0.03	−2.44	3.67	3.38	0.35***
	reserves	0.33*	0.61	0.27	0.14***	−2.13	3.10	2.97	0.30***
	volatility	−0.27	−0.27	−0.11	−0.07	−3.73	−0.05	0.20	0.08
	buy/hold	–	2.02	1.82	0.24***				
Crisis period	**cpi**	−5.05	−7.93	−0.36	−0.85	2.71*	18.20	2.20	0.58***
	twex	−0.24	−0.30	−0.01	−0.20	−0.31	9.48	0.93	0.20***
	tbond	1.54	12.72	1.64	0.18***	0.21	4.23	0.43	−0.06
	tms	−1.61	−4.40	−0.23	−0.89	−1.65	−1.19	0.02	−0.32
	sp500	−0.35	−1.21	−0.06	−0.37	−4.70	−7.52	−0.28	−0.60
	gsci	2.72*	4.53	0.44	−0.04	5.18**	14.62	1.61	0.42***
	crb	1.09	0.44	0.09	−0.24	2.19	14.23	1.60	0.36***
	reserves	0.40	1.37	0.09	0.18***	−1.19	5.91	0.71	−0.03
	volatility	−1.42	−2.29	−0.12	−0.56	−4.07	1.66	0.19	−0.17
	buy/hold	–	10.36	1.03	0.25***				

Notes: The table reports statistical and economic forecasting performances using all predictors separately. R^2_{oos}: (Campbell & Thompson, 2008) out-of-sample R^2; CER: certainly equivalent return; TW: terminal wealth; SR: Sharpe ratio. *, ** and *** indicate significance at the 10%, 5% and 1% levels respectively. Asterisks are presented for the R^2_{oos} (CW test) and the Sharpe ratio (GRS test). All results are expressed in excess of a naive historical mean benchmark. The trading results for the historical mean are *CER*: 2.16, *TW*: 1.71, *SR*: −0.06 (full sample) and *CER*: 1.84, *TW*: 1.11, *SR*: 0.41 (crisis sample).

Regarding cpi, a maximum excess value of more than 18 percentage points is achieved for the CER. Considering the terminal wealth of the historical mean (1.11), an investor more than triples his initial investment (with a CER of approximately 20%). After decomposition, four predictor series lead to positive R^2_{oos} values, with two results (cpi and gsci) being significant at the 10% and 5% levels. For cpi, a sharp increase in accuracy of close to eight percentage points highlights the advantages of the DWT-SVR approach. In summary, even the four predictors with the best economic performances after decomposition do not lead to positive outcomes in the univariate regression framework (with respect to the Sharpe ratios).

Having highlighted the superiority of the wavelet approach, a natural question is, which fast- or slow-moving component dominates the forecasting performance? Thus, Table 3 shows the average relative importance of the decompositions for the entire evaluation periods.

For all predictors and both periods, the outcomes of the wavelet details and the smooth component are not distributed equally. An interesting finding is reported for both commodity indices: while gold excess returns tend to depend on the short-term component of the GSCI index, the opposite result is observed for the CRB index. This outcome can be explained by the determinants of the underlying commodity-price series. The CRB index includes a well-differentiated set of equally weighted commodity prices, reflecting the general level of commodity prices in the economy. In contrast, the weights for the GSCI index change over time and, over the later years, the index has preferred the (fast-moving) energy sector.

Bearing in mind the good performance of the GSCI index, one can argue that the price of gold responds strongly to fluctuations in this sector, and in particular to crude oil derivatives. These findings are also supported by the decomposition of the CPI index, where short-term components are more important than long-term components.

4.2. Multivariate evaluation

One drawback of the univariate evaluation is that, in real-time, a forecaster does not know which predictor will lead to the best forecasting performance in the next period. For this reason, I combine first all untransformed predictors, and second all wavelet decompositions, to obtain an overall forecasting model that includes 63 predictors. I assess the improvement of the DWT-SVR approach directly by also applying the SVR to the untransformed series. Moreover, since the outcome can be influenced by the machine learning algorithm, I also consider alternative classes of supervised learning. In addition to a linear regression (OLS), I estimate the least absolute shrinkage and selection operator (LASSO), component-wise boosting (BOOST), and a random forest (FOREST).[27] I also use equally weighted univariate regressions (OLS-EW) as a simple forecast combination scheme. While LASSO and BOOST are linear models that penalize the predictor space directly, FOREST is

27 Values for hyperparameters are given in the table notes. I use various alternative specifications that do not affect the results qualitatively. Combining the alternative machine learning approaches with DWT did not help to improve either the statistical or economic forecasting performances. Therefore, these results are not presented here, but are available upon request.

Table 3
Average relative importance of wavelet details and wavelet smooths.

	D_1	D_2	D_3	D_4	D_5	D_6	S_6	D_1	D_2	D_3	D_4	D_5	D_6	S_6
	Full sample							**Crisis period**						
cpi	0.23	0.22	0.14	0.11	0.24	0.03	0.04	0.19	0.21	0.22	0.17	0.13	0.03	0.04
twex	0.23	0.19	0.15	0.03	0.02	0.35	0.04	0.19	0.25	0.15	0.04	0.01	0.30	0.06
tbond	0.05	0.06	0.12	0.06	0.26	0.30	0.15	0.05	0.06	0.07	0.05	0.19	0.30	0.28
tms	0.10	0.27	0.07	0.02	0.04	0.44	0.06	0.11	0.23	0.06	0.02	0.03	0.50	0.05
sp500	0.05	0.15	0.06	0.10	0.15	0.06	0.42	0.05	0.09	0.05	0.14	0.16	0.03	0.49
gsci	0.11	0.25	0.17	0.06	0.24	0.09	0.07	0.21	0.27	0.19	0.08	0.06	0.14	0.06
crb	0.05	0.03	0.13	0.11	0.09	0.19	0.39	0.04	0.02	0.08	0.09	0.04	0.22	0.51
reserves	0.12	0.05	0.23	0.17	0.03	0.04	0.35	0.06	0.01	0.23	0.11	0.04	0.02	0.54
volatility	0.09	0.25	0.12	0.15	0.19	0.16	0.03	0.09	0.30	0.09	0.08	0.19	0.21	0.04

Notes: The table reports the average relative importance for all wavelet details (D1–D6) and the wavelet smooth (S6) for each entire evaluation period.

Table 4
Joint evaluation in excess of the historical mean.

	R^2_{oos}	Δ_{CER}	Δ_{TW}	Δ_{SR}	R^2_{oos}	Δ_{CER}	Δ_{TW}	Δ_{SR}
	Full sample				**Crisis period**			
OLS	−4.91	1.33	1.66	0.20***	−4.60	3.3	0.46	−0.13
OLS EW	0.36	1.23	0.62	0.23***	0.61	2.3	0.17	0.04
LASSO	−0.68	2.02	1.96	0.23***	−2.99	−1.2	0.08	−0.31
BOOST	−1.20	−0.51	0.27	0.08	−1.63	−4.7	−0.13	−0.45
FOREST	−7.33	0.46	0.80	0.14**	−2.99	−6.2	−0.21	−0.52
SVR	0.67**	2.30	2.29	0.25***	3.25*	7.8	0.89	0.04
DWT-SVR	0.93**	5.28	5.77	0.45***	2.64*	9.8	1.03	0.16***
Buy/hold	–	2.02	1.82	0.24***	–	10.4	1.03	0.25***

Notes: The table reports statistical and economic forecasting performances using all predictors. R^2_{oos}: (Campbell & Thompson, 2008) out-of-sample R^2; CER: certainly equivalent return; TW: terminal wealth; SR: Sharpe ratio. *, ** and *** indicate significance at the 10%, 5% and 1% levels, respectively. Asterisks are presented for the R^2_{oos} (CW test) and the Sharpe ratio (GRS test). All results are expressed in excess of a naive historical mean benchmark. Optimal panelization of the lasso is achieved by 10-fold cross validation. The stopping criterion for BOOST is the corrected AIC with a learning rate of 0.1 and a maximum number of iterations of 200. The maximum number of trees in terms of FOREST is set to 200. The trading results for the historical mean are *CER*: 2.16, *TW*: 1.71, *SR*: −0.06 (full sample) and *CER*: 1.84, *TW*: 1.11, *SR*: 0.41 (crisis sample).

a non-parametric decision tree-based approach that also considers interaction effects. Table 4 presents the results. Again, I differentiate between the full sample and the crisis period.

With respect to the alternative forecasting models, OLS-EW is the only approach that receives a positive R^2_{oos}, though it is not significant. The forecasting accuracy of the standard SVR is higher and is significant at the 5% level. The DWT-SVR approach leads to the highest accuracy (0.93). Turning to the recent financial crisis, the forecast accuracy increases for both SVR-based methods. In fact, the performance of the DWT-SVR cannot exceed that of the untransformed SVR model. The alternative machine learning approaches lead to relatively poor performances.

With respect to the economic value added by the forecasts over the full evaluation period, almost all models lead to a positive improvement over the historical mean, while the DWT-SVR approach more than doubles the excess outcome of the standard SVR model. However, the buy-and-hold strategy becomes difficult to beat when one turns to the crisis period. None of the alternative forecasting models achieve comparable outcomes. BOOST and FOREST even lead to negative outcomes in terms of all evaluation statistics. The results for the standard SVR are positive, and approximately double the TW excess outcome of the simple OLS model. The DWT-SVR approach is the only forecasting model that reaches the same level as the buy-and-hold strategy in terms of terminal wealth, and the only model that leads to a significant financial improvement at the 1% level as compared to the historical mean benchmark.

4.3. Evolution of wealth

I highlight the time-varying outcome of the investment decisions by plotting the evolution of terminal wealth over the full evaluation period. Fig. 3 presents the results for selected forecasting models. All series are transformed using the natural logarithm. I also display the evolution of the buy-and-hold strategy, which is equal to the original gold price series in log-levels.

The evolution of wealth then highlights the superior performance of the wavelet approach. After a short period at the beginning where there is very little variation in the price series, the DWT-SVR approach begins to dominate all forecasting models over almost the entire evaluation period. There is only a short period in which LASSO, as the only alternative forecasting approach, reaches a comparable level of wealth (around the beginning of the recent financial crisis). It is also worth noting that the wavelet approach detects the downturn around 2014, resulting in a very low proportion of investments in gold, and thus a stagnating but not decreasing level of wealth.

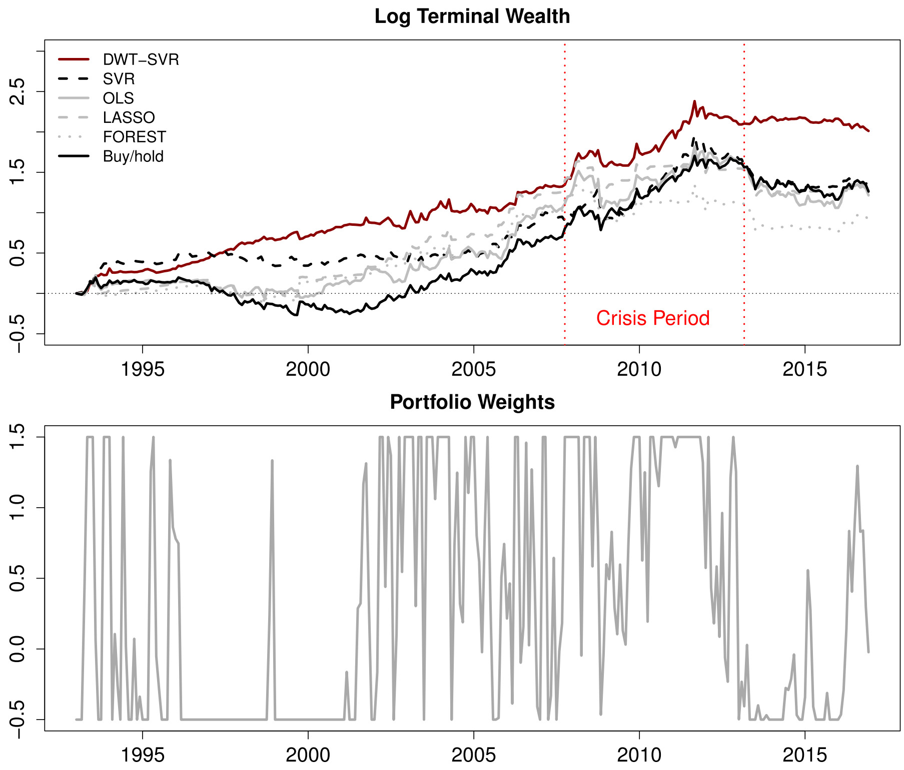

Fig. 3. Evolution of terminal wealth and portfolio weights. Notes: The figure presents the evolution of terminal wealth (top panel) for selected forecasting models and the corresponding portfolio weights for the wavelet decomposition (bottom panel).

The portfolio allocation scheme is depicted in the lower panel of Fig. 3. The investment in gold largely coincides with the upswing of the series. Portfolio weights tend to fluctuate in times of higher volatility. The gradual decline of the series, starting around 1996, leads to a long period of almost zero investment in gold.

4.4. Time-varying importance of predictors

Two last questions remain with respect to the relative importance of predictors. First, it is interesting to analyze whether the relative importances of the original time series differ from those derived from the decomposed time series. A large difference in relative importance would support the hypothesis that individual detail and smooth influences could lead to a different selection of predictors from the evaluation of the original untransformed series. For this reason, I compute the relative importance in real-time and display the proportions for each predictor in Fig. 4. The importance for the decomposed predictors is computed simply as the sum of all wavelet details and the wavelet smooth.

The general picture that emerges when we compare the relative importances over time and across predictors is not homogeneous.[28] Wavelet-based decompositions are more important for twex, sp500, crb, and reserves. Without decomposition, tbond is the predictor that gains by far the most importance over the entire sample. Similarly high influences are observed for cpi, but only in the first half of the evaluation period. With tms and volatility, there also exist two time series in which the relative importance remains at a relatively low level. However, the most interesting findings are those presented for cpi and gsci. In terms of the former, the relative importance of the wavelet decompositions exceeds the relative importance of the untransformed series at the end of the recent financial crisis. The opposite is observed for gsci, where the untransformed series gains momentum at the end of the sample. Hence, choosing the wavelet transform does not necessarily imply that the importance of the predictors remains stable over time.

The second question that must be addressed is whether the influences of wavelet details and wavelet smooths change over time, but independently from the predictors themselves. In other words, I visualize whether the influences of wavelet details and wavelet smooths are constant over time, or whether short- or long-term trends dominate in some periods. To do so, I summarize the relative importances within wavelet details and wavelet smooths across all predictors, in every period of time. The results are displayed in Fig. 5. The red dotted lines separate the faster details (D1–D3) from the slower-moving components, as well as the wavelet details from the wavelet smooth in general. It turns out that the influence of the smooth component is almost negligible at the beginning of the evaluation period, but the influences of the short-term components

28 An important remark at this point is that a high relative importance does not necessarily imply that the predictor on its own leads to an improved accuracy.

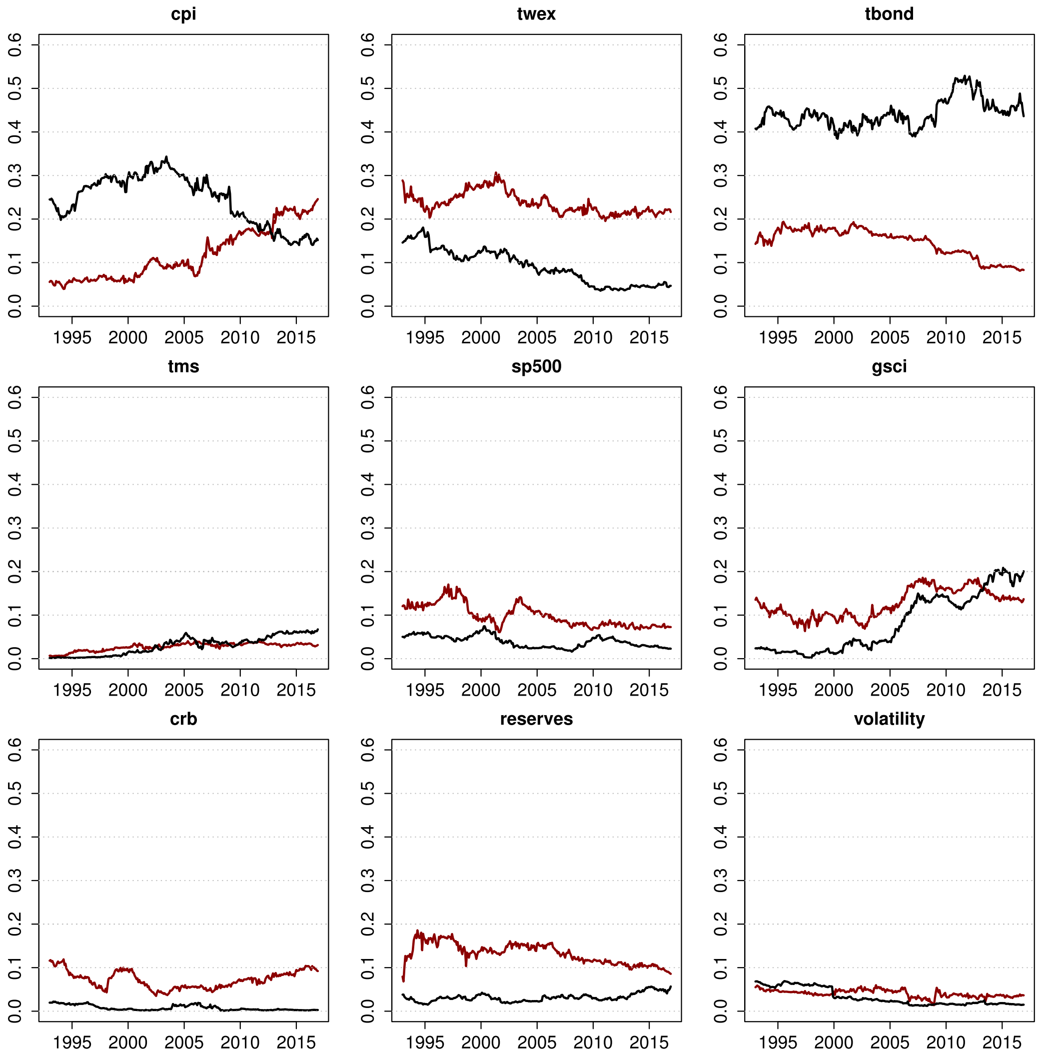

Fig. 4. Relative importance of different predictors over time. Notes: The figure presents the relative importance of each predictor over time. The black lines reflect the relative importance of the SVR model without decomposition. The red lines reflect the relative importance of the decomposed time series, where the value is computed, in every period of time, as the sum of all components.

and the smoothing component increase as time passes. At the end of the sample, more than two-thirds of the relative importance is addressed to the short-term components of the wavelet details. Bearing in mind the relatively good performance of the DWT-SVR approach at the end of the evaluation period, one could argue that higher-frequency decompositions helped an investor to retain his wealth. In contrast, the equally-distributed relative importance of wavelet details at the beginning of the evaluation period is in line with the good performance of the standard SVR during the period of the Great Moderation in the first half of the 1990s. Thus, decomposing is not always advantageous, particularly in times of calm fluctuations of the target series to be forecasted.

5. Conclusion

I have examined the importance of the DWT-SVR approach in an economic forecasting environment using the excess-return series of the price of gold as the underlying target series. With respect to the full evaluation period, results derived from statistical accuracy underline the difficulty of beating simple historical mean forecasts when considering only a single predictor. With respect to the economic value added by forecasts, DWT-SVR increases an investor's wealth significantly for two-thirds of the predictors being studied. Determinants that should be of particular interest in this context are inflation, the trade-weighted exchange rate, and both commodity indices.

Combining all predictors in a large forecasting model in order to reduce the predictor uncertainty leads to a decrease in statistical and economic performance. DWT-SVR has obtained the best economic outcome of all alternative forecasting models, as well as the highest statistical accuracy over the full sample period. Comparable findings are reported for the period of the recent financial crisis.

The analysis of individual trend components with respect to their relative importances has highlighted the

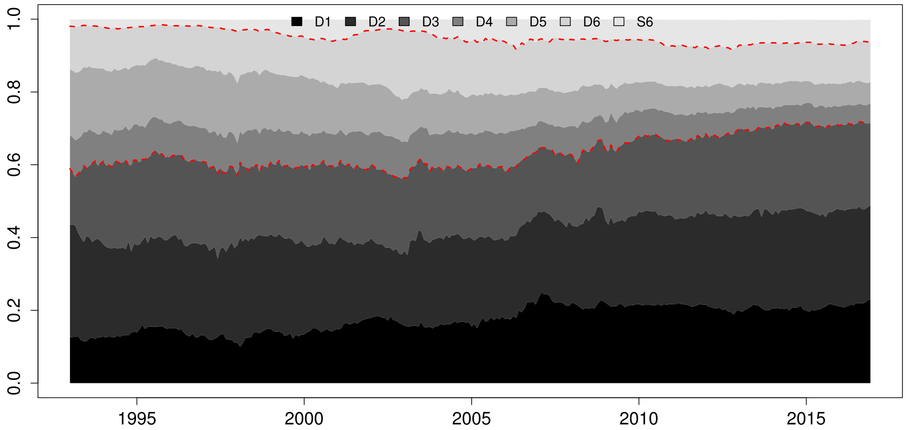

Fig. 5. Combined relative importance over time. Notes: The figure presents the relative importances of wavelet details and wavelet smooths summarized across all predictors. The lower red line represents the border between the faster-moving (D1–D3) and slower-moving (D4–D6) wavelet details. The upper red line separates the impact of the smoothing component (S6) from the wavelet details (D1–D6).

fact that there does not exist a single dominating faster or slower-moving wavelet detail. However, short-term wavelet coefficients have gained momentum over the full evaluation period, leading to a large proportion of high relative importances, particularly at the end of the sample. Therefore, the good performance of the DWT-SVR approach that has been realized in that time period leads to a preference for (short-term) time series decompositions instead of using the untransformed series. Moreover, as has been shown for every predictor individually, the relative importance differs between the original and transformed series.

As one possible avenue for future research, it might be interesting to try optimizing the machine learning framework so as to improve the selection of wavelet details and smooth components. In addition, the discussion of the influences of the wavelet function has been only rudimentary, though a more complex wavelet function would also be expected to improve the model performance.

References

Aguiar-Conraria, L., & Soares, M. J. (2011a). Business cycle synchronization and the Euro: A wavelet analysis. *Journal of Macroeconomics*, *33*(3), 477–489.

Aguiar-Conraria, L., & Soares, M. J. (2011b). Oil and the macroeconomy: Using wavelets to analyze old issues. *Empirical Economics*, *40*(3), 645–655.

Alexandridis, A. K., Kampouridis, M., & Cramer, S. (2017). A comparison of wavelet networks and genetic programming in the context of temperature derivatives. *International Journal of Forecasting*, *33*(1), 21–47.

Asgharian, H., Christiansen, C., & Hou, A. J. (2016). Macro-finance determinants of the long-run stock–bond correlation: The DCC-MIDAS specification. *Journal of Financial Economic*, *14*(3), 617–642.

Avramov, D. (2002). Stock return predictability and model uncertainty. *Journal of Financial Economic*, *64*(3), 423–458.

Aye, G., Gupta, R., Hammoudeh, S., & Kim, W. J. (2015). Forecasting the price of gold using dynamic model averaging. *International Review of Financial Analysis*, *41*, 257–266.

Bai, J., & Ng, S. (2008). Forecasting economic time series using targeted predictors. *Journal of Econometrics*, *146*(2), 304–317.

Batten, J. A., Ciner, C., & Lucey, B. M. (2014). On the economic determinants of the gold–inflation relation. *Resources Policy*, *41*, 101–108.

Baur, D. G., Beckmann, J., & Czudaj, R. (2016). A melting pot — Gold price forecasts under model and parameter uncertainty. *International Review of Financial Analysis*, *48*, 282–291.

Baur, D. G., & Lucey, B. M. (2010). Is gold a hedge or a safe haven? An analysis of stocks, bonds and gold. *Financial Review*, *45*, 217–229.

Baur, D. G., & McDermott, T. K. (2010). Is gold a safe haven? International evidence. *Journal of Banking and Finance*, *34*, 1886–1898.

Beckmann, J., Berger, T., & Czudaj, R. (2015). Does gold act as a hedge or a safe haven for stocks? A smooth transition approach. *Economic Modelling*, *48*, 16–24.

Beckmann, J., Berger, T., & Czudaj, R. (2017). Gold price dynamics and the role of uncertainty. *Tech. rep.*, Department of Economics, Chemnitz University of Technology.

Beckmann, J., & Czudaj, R. (2013). Gold as an inflation hedge in a time-varying coefficient framework. *North American Journal of Economics and Finance*, *24*, 208–222.

Berger, T. (2016). Forecasting based on decomposed financial return series: A wavelet analysis. *Journal of Forecasting*, *35*(5), 419–433.

Blose, L. E. (2010). Gold prices, cost of carry, and expected inflation. *Journal of Economics and Business*, *62*(1), 35–47.

Bulligan, G., Marcellino, M., & Venditti, F. (2015). Forecasting economic activity with targeted predictors. *International Journal of Forecasting*, *31*(1), 188–206.

Campbell, J. Y., & Thompson, S. B. (2008). Predicting excess stock returns out of sample: Can anything beat the historical average? *Review of Financial Studies*, *21*, 1509–1531.

Capie, F., Mills, T. C., & Wood, G. (2005). Gold as a hedge against the dollar. *Journal of International Financial Markets, Institutions and Money*, *15*, 343–352.

Cherkassky, V., & Ma, Y. (2004). Practical selection of SVM parameters and noise estimation for SVM regression. *Neural Networks*, *17*(1), 113–126.

Clark, T. E., & West, K. D. (2007). Approximately normal tests for equal predictive accuracy in nested models. *Journal of Econometrics*, *138*(1), 291–311.

Conejo, A. J., Contreras, J., Espinola, R., & Plazas, M. A. (2005). Forecasting electricity prices for a day-ahead pool-based electric energy market. *International Journal of Forecasting*, *21*(3), 435–462.

Cortez, P. (2016). rminer: Data mining classification and regression methods. R package version 1.4.2.

Cortez, P., & Embrechts, M. J. (2013). Using sensitivity analysis and visualization techniques to open black box data mining models. *Information Sciences*, *225*, 1–17.

Crowley, P. M. (2007). A guide to wavelets for economists. *Journal of Economic Surveys*, *21*(2), 207–267.

Daubechies, I. (1992). *CBMS-NSF regional conference series in applied mathematics*: *Vol. 61*, *Ten lectures on wavelets*. SIAM, Philadelphia, Pa, USA.

De Mol, C., Giannone, D., & Reichlin, L. (2008). Forecasting using a large number of predictors: Is Bayesian shrinkage a valid alternative to principal components? *Journal of Econometrics*, *146*(2), 318–328.

Döpke, J., Fritsche, U., & Pierdzioch, C. (2017). Predicting recessions with boosted regression trees. *International Journal of Forecasting*, *33*(4), 745–759.

Efron, B., Hastie, T., Johnstone, I., & Tibshirani, R. (2004). Least angle regression. *The Annals of Statistics*, *32*(2), 407–499.

Elliott, G., Gargano, A., & Timmermann, A. (2013). Complete subset regressions. *Journal of Econometrics*, *177*(2), 357–373.

Elliott, G., Gargano, A., & Timmermann, A. (2015). Complete subset regressions with large-dimensional sets of predictors. *Journal of Economic Dynamics & Control*, *54*, 86–110.

Faria, G., & Verona, F. (2017). Forecasting the equity risk premium with frequency-decomposed predictors. *Research Discussion Papers 01/2017*, Bank of Finland.

Faria, G., & Verona, F. (2018). Forecasting stock market returns by summing the frequency-decomposed parts. *Journal of Empirical Finance*, *45*, 228–242.

Forni, M., Hallin, M., Lippi, M., & Reichlin, L. (2005). The generalized dynamic factor model: One-sided estimation and forecasting. *Journal of the American Statistical Association*, *100*(471), 830–840.

Fortune, J. N. (1987). The inflation rate of the price of gold, expected prices and interest rates. *Journal of Macroeconomics*, *9*, 71–82.

Gallegati, M. (2012). A wavelet-based approach to test for financial market contagion. *Computational Statistics & Data Analysis*, *56*(11), 3491–3497.

Garcia, M. G., Medeiros, M. C., & Vasconcelos, G. F. (2017). Real-time inflation forecasting with high-dimensional models: The case of Brazil. *International Journal of Forecasting*, *33*(3), 679–693.

Gibbons, M. R., Ross, S. A., & Shanken, J. (1989). A test of the efficiency of a given portfolio. *Econometrica*, *57*, 1121–1152.

Haar, A. (1910). Zur Theorie der orthogonalen Funktionensysteme. *Mathematische Annalen*, *69*(3), 331–371.

H.S.tock, J., & W.W.atson, M. (2002). Macroeconomic Forecasting Using Diffusion Indexes. *Journal of Business & Economic Statistics*, *20*(2), 147–162.

Karatzoglou, A., Smola, A., Hornik, K., & Zeileis, A. (2004). Kernlab — An S4 Package for Kernel Methods in R. *Journal of Statistical Software*, *11*(9), 1–20.

Kewley, R. H., Embrechts, M. J., & Breneman, C. (2000). Data strip mining for the virtual design of pharmaceuticals with neural networks. *IEEE Transactions on Neural Networks*, *11*(3), 668–679.

Konzen, E., & Ziegelmann, F. A. (2016). Lasso-type penalties for covariate selection and forecasting in time series. *Journal of Forecasting*, *35*(7), 592–612.

Koop, G., & Korobilis, D. (2012). Forecasting inflation using dynamic model averaging. *International Economic Review*, *53*(3), 867–886.

Koop, G., & Korobilis, D. (2014). A new index of financial conditions. *European Economic Review*, *71*, 101–116.

Koop, G., Korobilis, D., & Pettenuzzo, D. Bayesian compressed vector autoregressions. *Journal of Econometrics* (in press).

Koop, G., & Potter, S. (2004). Forecasting in dynamic factor models using Bayesian model averaging. *The Econometrics Journal*, *7*(2), 550–565.

Leitch, G., & Tanner, J. (1991). Economic forecast evaluation: Profits versus the conventional error measures. *American Economic Review*, *81*(3), 580–590.

Mahdavi, S., & Zhou, S. (1997). Gold and commodity prices as leading indicators of inflation: Tests of long-run relationship and predictive performance. *Journal of Economics and Business*, *49*, 475–489.

Mallat, S. G. (1989). A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *11*(7), 674–693.

Marquering, W., & Verbeek, M. (2004). The economic value of predicting stock index returns and volatility. *Journal of Financial and Quantitative Analysis*, *39*(2), 407–429.

McCracken, M. W., & Ng, S. (2016). FRED-MD: A monthly database for macroeconomic research. *Journal of Business & Economic Statistics*, *34*(4), 574–589.

Medeiros, M. C., & Vasconcelos, G. F. (2016). Forecasting macroeconomic variables in data-rich environments. *Economics Letters*, *138*, 50–52.

Mittnik, S., Robinzonov, N., & Spindler, M. (2015). Stock market volatility: Identifying major drivers and the nature of their impact. *Journal of Banking and Finance*, *58*, 1–14.

Murtagh, F., Starck, J. L., & Renaud, O. (2004). On neuro-wavelet modeling. *Decision Support Systems*, *37*(4), 475–484.

Neely, C. J., Rapach, D. E., Tu, J., & Zhou, G. (2014). Forecasting the Equity Risk Premium: The Role of Technical Indicators. *Management Science*, *60*(7), 1772–1791.

O'Connor, F. A., Lucey, B. M., Batten, J. A., & Baur, D. G. (2015). The financial economics of gold – A survey. *International Review of Financial Analysis*, *41*, 186–205.

Onorante, L., & Raftery, A. E. (2016). Dynamic model averaging in large model spaces using dynamic Occam's window. *European Economic Review*, *81*, 2–14.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research (JMLR)*, *12*, 2825–2830.

Percival, D. B., & Walden, A. (2000). *Wavelet methods for time series analysis*. Cambridge, UK: Cambridge University Press.

Pierdzioch, C., Risse, M., & Rohloff, S. (2014). On the efficiency of the gold market: Results of a real-time forecasting approach. *International Review of Financial Analysis*, *32*, 95–108.

Pierdzioch, C., Risse, M., & Rohloff, S. (2016a). A boosting approach to forecasting gold and silver returns: economic and statistical forecast evaluation. *Applied Economics Letters*, *23*(5), 347–352.

Pierdzioch, C., Risse, M., & Rohloff, S. (2016b). A quantile-boosting approach to forecasting gold returns. *The North American Journal of Economics and Finance*, *35*, 38–55.

Prüser, J. (2019). Adaptive learning from model space. *Journal of Forecasting*, *38*, 29–38.

Pukthuanthong, K., & Roll, R. (2011). Gold and the Dollar (and the Euro, Pound, and Yen). *Journal of Banking and Finance*, *35*, 2070–2083.

Raftery, A. E., Kárnỳ, M., & Ettler, P. (2010). Online Prediction Under Model Uncertainty via Dynamic Model Averaging: Application to a Cold Rolling Mill. *Technometrics*, *52*(1), 52–66.

Ramsey, J. B. (2002). Wavelets in economics and finance: Past and future. *Studies in Nonlinear Dynamics & Econometrics*, *6*(3), 1–27.

Rapach, D. E., Ringgenberg, M. C., & Zhou, G. (2016). Short interest and aggregate stock returns. *Journal of Financial Economic*, *121*(1), 46–65.

Rapach, D. E., Strauss, J. K., & Zhou, G. (2010). Out-of-sample equity premium prediction: Combination forecasts and links to the real economy. *Review of Financial Studies*, *23*, 821–862.

Reboredo, J. C. (2013). Is gold a safe haven or a hedge for the U.S. dollar? Implications for risk management. *Journal of Banking and Finance*, *37*, 2665–2676.

Risse, M., & Kern, M. (2016). Forecasting house-price growth in the Euro area with dynamic model averaging. *The North American Journal of Economics and Finance*, *38*, 70–85.

Risse, M., & Ohl, L. (2017). Using Dynamic Model Averaging in State Space Representation with Dynamic Occam's Window and Applications to the Stock and Gold Market. *Journal of Empirical Finance*, *44*, 158–176.

Rua, A. (2011). A wavelet approach for factor-augmented forecasting. *Journal of Forecasting*, *30*(7), 666–678.

Rua, A. (2017). A wavelet-based multivariate multiscale approach for forecasting. *International Journal of Forecasting*, *33*(3), 581–590.

Rua, A., & Nunes, L. C. (2009). International comovement of stock market returns: A wavelet analysis. *Journal of Empirical Finance*, *16*(4), 632–639.

Seo, Y., Choi, Y., & Choi, J. (2017). River Stage Modeling by Combining Maximal Overlap Discrete Wavelet Transform, Support Vector Machines and Genetic Algorithm. *Water*, *9*(7), 525.

Shynkevich, Y., McGinnity, T., Coleman, S. A., Belatreche, A., & Li, Y. (2017). Forecasting price movements using technical indicators: Investigating the impact of varying input window length. *Neurocomputing*, *264*, 71–88.

Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and Computing*, *14*(3), 199–222.

Stock, J. H., & Watson, M. W. (2002). Forecasting Using Principal Components From a Large Number of Predictors. *Journal of the American Statistical Association*, *97*(460), 1167–1179.

Sujjaviriyasup, T. (2017). A new class of MODWT-SVM-DE hybrid model emphasizing on simplification structure in data pre-processing: A case study of annual electricity consumptions. *Applied Soft Computing*, *54*, 150–163.

Tibshirani, R. (1996). Regression shrinkage and selection via the LASSO. *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, *58*, 267–288.

Vapnik, V. (1995). *The nature of statistical learning theory*. Springer Science & Business Media.

Welch, I., & Goyal, A. (2008). A Comprehensive Look at The Empirical Performance of Equity Premium Prediction. *Review of Financial Studies*, *21*, 1455–1508.

Whitcher, B. (2015). Waveslim: Basic wavelet routines for one-, two- and three-dimensional signal processing. R package version 1.7.5.

Wong, H., Ip, W. C., Xie, Z., & Lui, X. (2003). Modelling and forecasting by wavelets, and the application to exchange rates. *Journal of Applied Statistics*, *30*(5), 537–553.

Marian Risse is a Postdoctoral Researcher, Department of Economics, Helmut Schmidt University, Hamburg, Germany. Research interests: forecasting, financial econometrics, macroeconomics, commodity markets, machine learning.