Gonçalo Faria – Fabio Verona

Forecasting the equity risk premium with frequency-decomposed predictors

Bank of Finland Research
Discussion Paper
1 • 2017

Forecasting the equity risk premium with frequency-decomposed predictors*

Gonçalo Faria[†] Fabio Verona[‡]

November 14, 2016

Abstract

We show that the out-of-sample forecast of the equity risk premium can be significantly improved by taking into account the frequency-domain relationship between the equity risk premium and several potential predictors. We consider fifteen predictors from the existing literature, for the out-of-sample forecasting period from January 1990 to December 2014. The best result achieved for individual predictors is a monthly out-of-sample R^2 of 2.98 % and utility gains of 549 basis points per year for a mean-variance investor. This performance is improved even further when the individual forecasts from the frequency-decomposed predictors are combined. These results are robust for different subsamples, including the Great Moderation period, the Great Financial Crisis period and, more generically, periods of bad, normal and good economic growth. The strong and robust performance of this method comes from its ability to disentangle the information aggregated in the original time series of each variable, which allows to isolate the frequencies of the predictors with the highest predictive power from the noisy parts.

Keywords: predictability, equity risk premium, frequency domain, discrete wavelets

JEL classification: C58, G11, G12, G17

* The authors thank Amit Goyal and David Rapach for kindly providing codes and data on their webpages; Gene Ambrocio, Eleonora Granziera, Mikael Juselius, Pedro Duarte Silva and seminar participants at the Bank of Finland, at the Helsinki macro research away day and at Católica Porto Business School for useful comments; and Riccardo Verona for IT support. The views expressed are those of the authors and do not necessarily reflect those of the Bank of Finland. Faria gratefully acknowledges financial support from Fundação para a Ciência e Tecnologia (through project UID/GES/00731/2013).

† Universidade Católica Portuguesa, Católica Porto Business School and CEGE (gfaria@porto.ucp.pt)

‡ Bank of Finland – Monetary Policy and Research Department, and University of Porto – *CEF.UP* (fabio.verona@bof.fi)

1 Introduction

The equity risk premium (ERP) plays a crucial role in economics and finance. On one side, it is an important determinant of the cost of capital for corporations and of savings decisions of individuals. On the other side, from a macroeconomic perspective, the ERP reflects a broad outlook for the whole economy. Factors shaping investors' views on market risk, and hence the ERP, include the outlook for economic growth, consumer demand, inflation, interest rates and geopolitical risks. Furthermore, the ERP has recently returned to the forefront as a leading indicator of the business cycle, a potential explanation for jobless recoveries and a gauge of financial stability (Duarte and Rosa, 2015).

These reasons stimulated an extensive research on the forecastability of the ERP, as reviewed by Rapach and Zhou (2013) and Harvey et al. (2016). In this paper we empirically re-evaluate the forecasting performance of several ERP predictors – some of which have been extensively used (and rejected) in the literature – by explicitly considering the frequency relationship between the ERP and those predictors. Concretely, the time series of each predictor used in Rapach et al. (2016) is first decomposed in its different frequencies by using the discrete wavelet transform multiresolution analysis (see *e.g.* Crowley, 2007 and the references in section 1.1). This method consists in decomposing a time series into n orthogonal time series components, each of them capturing the oscillations of the original variable within a specific frequency interval. The lower frequencies represent the long-term dynamics of the original time series, while the higher frequencies capture the short-term dynamics. Those n frequency-decomposed components are orthogonal so that, by adding them, it is possible to recover the original time series. Then, the frequency-decomposed predictors are evaluated as ERP predictors. As Rua (2011) shows, the wavelet multiresolution analysis is a useful tool for forecasting, as the forecast accuracy can be improved by first forecasting each frequency band separately, and then aggregating the individual forecasts to produce the forecast for the original time series. This method thus allows to use wavelets tools and, at the same time, to have a foot on traditional time series analysis.

When compared with the traditional time series forecast analysis, we find that, by selecting the proper frequencies, the statistical out-of-sample (OOS) performance is improved for *all* predictors. Although for some of the predictors this is still not enough to outperform the historical mean (HM) as an ERP predictor, there are 6 remarkable exceptions: the earnings-price ratio, the dividend-payout ratio, the inflation rate, the long-term government bond return, the term spread and the short interest index.[1] These 6 variables deliver positive and statistically significant out-of-sample R-squares (R^2_{OS}), with the dividend-payout ratio

[1] It has been extensively documented in the literature that the HM of the ERP outperforms a long list of potential predictors in an OOS forecasting exercise (*e.g.* Goyal and Welch, 2008, Ferreira and Santa-Clara, 2011 and Rapach et al., 2016).

being the best ERP predictor for the period under analysis (monthly R^2_{OS} of 2.98%). We also find that there are further significant advantages in combining the best individual forecasts from the frequency-decomposed predictors, as suggested by Rapach et al. (2010).

To give a flavor of our results, figure 1 plots, for the full OOS period (January 1990 to December 2014), the realized ERP (black solid line) together with the excess return predictions based on the HM (black dashed line) and on our best model, which combines three frequency-decomposed predictors (blue line).[2] The correlations between the realized S&P500 index excess returns and the excess returns forecasts from our best model and the HM forecast are 0.20 and almost zero, respectively. It is evident from figure 1 that our predictor captures not only the low-frequency dynamics of the ERP, as the HM does, but also some of the higher-frequency movement of the equity risk premium, which is not captured by the HM.

We then evaluate the economic significance of the forecasting performance of the frequency-decomposed predictors through an asset allocation analysis. We find that, for a mean-variance investor who allocates her wealth between equities and risk-free bills, there are significant utility gains when making the forecasts using the proper set of frequencies of the predictors. From an economic point of view, the best individual performance is achieved when only the lowest frequency component of the term spread is used as ERP predictor. In this case, the annual rate of return that an investor would be willing to accept instead of holding the risky portfolio is 549 basis points. This is further increased to 674 basis points when combining the forecasts from the frequency-decomposed term spread, earnings-price ratio, dividend-payout ratio and short interest index.

As a robustness exercise, we analyse the statistical and economic forecasting performance of the predictors for different subsamples. First, as in Rapach et al. (2016), we split the sample in two time windows, 1990:01 - 2006:12 and 2007:01 - 2014:12, corresponding to (part of) the great moderation period and to the great financial crisis, respectively. We find that results are robust when evaluated in those two subperiods. Second, we split the sample in periods of bad, normal and good economic growth. This is relevant, as it has been documented in the literature that the excess return predictability is concentrated in recessions or bad times (see *e.g.* Henkel et al., 2011). In line with previous literature, we find that the forecasting performance is i) superior during bad and normal growth periods, and ii) significantly improved by using the frequency-decomposed predictors. However, and differently from previous literature, we document that forecasting during good growth periods using some of the frequency-decomposed predictors *also* outperforms the HM benchmark.

[2] Explained in detail in section 2.2.

In the remainder of this section, we briefly review related literature. Section 2 presents the data and the methodology. Section 3 presents the out-of-sample results and the analysis of the economic value of the proposed excess return predictive methodology. In section 4 are reported the results of the robustness exercises. Section 5 concludes.

1.1 Related literature

This paper primarily relates to an extensive literature on the OOS forecasting of the ERP. Several studies find evidence of in-sample predictability using different predictors.[3] However, as firstly pointed out by Goyal and Welch (2008), most of those predictors perform very poorly OOS. As any predictive model requires OOS validation, the subsequent literature has focused on improving the OOS forecastability of the ERP. Two directions have been explored. The first one develops and tests new predictors, including macro variables,[4] investor sentiment indexes (Huang et al., 2015) and financial market variables.[5] The second one focuses on improving the forecasting strategy by considering, for example, dynamic factor models for large date sets to summarize a large amount of information by few estimated factors (Ludvigson and Ng, 2007 and Kelly and Pruitt, 2013), the combination of individual forecasts from different predictors (Rapach et al., 2010), the sum-of-the-parts method consisting in forecasting separately the components of stock market returns (Ferreira and Santa-Clara, 2011), predictive regressions with time-varying coefficients (Dangl and Halling, 2012) or with economic constrains on forecasts of the ERP (Pettenuzzo et al., 2014). We place our contribution in both strands of research, as the frequency decomposition of the predictors is not only a methodological contribution *per se*, but also represents an enlargement of the set of possible predictors, as each frequency of each predictor can be understood and potentially used as a new predictor. In Faria and Verona (2016) a frequency decomposition of several stock returns predictors is implemented in the context of the Ferreira and Santa-Clara (2011) sum-of-the-part method. Differently, in this paper we evaluate the ERP forecasting performance of the frequency-decomposed predictors within a standard OOS forecasting regression set-up.

[3] As regards the US stock market, the most popular predictors are the dividend-price ratio (Fama and French, 1988, Campbell and Shiller, 1988a, Cochrane, 2008 and Pastor and Stambaugh, 2009), the dividend yield (Campbell, 1987), the earning-price ratio (Campbell and Shiller, 1988b), the dividend-payout ratio (Lamont, 1998), the book-to-market ratio (Kothari and Shanken, 1997 and Pontiff and Schall, 1998), the treasury bill rate (Fama and Schwert, 1977), the inflation rate (Ferson and Harvey, 1991 and Campbell and Vuolteenaho, 2004), interest rate spreads (Fama and French, 1989) and the consumption-wealth ratio (Lettau and Ludvigson, 2001), among others. As regards the stock return forecastability in international markets, see *e.g.* Cutler et al. (1991), Harvey (1991), Bekaert and Hodrick (1992), Ferson and Harvey (1993), Ang and Bekaert (2007), Cooper and Priestley (2009), Hjalmarsson (2010), Engsted and Pedersen (2010), Rapach et al. (2013) and Jordan et al. (2014).

[4] Cooper and Priestley (2009, 2013) use the output gap and the world business cycle, respectively, Li et al. (2013) study the aggregate implied cost of capital and Moller and Rangvid (2015) study different macroeconomic variables by focusing on their fourth-quarter growth rate.

[5] This includes the variance risk premium (Bollerslev et al., 2009), the lagged US market returns for the OOS predictability of stock returns of other industrialized countries (Rapach et al. (2013)) and technical indicators (Neely et al., 2014).

The second stream of literature to which this paper is related respects to the application of wavelet methods in the analysis of economic and finance topics. Crowley (2007) and Aguiar-Conraria and Soares (2014) provide excellent reviews of economic and finance applications of wavelets. Ramsey and Lampart (1998a,b) applied for the first time wavelets to study the relationship between macroeconomic variables (consumption versus income and money supply versus income, respectively). More recently, wavelets methods have been applied to test for the (in-sample) frequency dependence between two (or more) variables (Kim and In, 2005, Gencay et al., 2005, Gallegati et al., 2011 and Gallegati and Ramsey, 2013) and to study the comovements and lead-lag relationship between variables at different frequencies (Rua and Nunes, 2009, Rua, 2010, Aguiar-Conraria and Soares, 2011 and Aguiar-Conraria et al., 2012). However, very few research has been done on applying wavelet methods for forecasting purposes. Besides the above-mentioned paper by Faria and Verona (2016), a few exceptions are Rua (2011) and Kilponen and Verona (2016), who propose a wavelet approach for factor-augmented forecasting of GDP growth and to forecast aggregate investment using the Tobin's Q theory of investment, respectively. Our paper applies the wavelet decomposition to forecast the ERP, documenting relevant statistical and economic gains derived from the use of this methodology with respect to some of the traditional predictors used in the literature.[6]

2 Data and methodology

We focus on the OOS predictability of monthly excess returns, measured by the difference between the log return on the S&P500 index and the log return on a one-month Treasury bill. As it has been emphasized in the literature (*e.g.* Goyal and Welch, 2008 and Huang et al., 2015), the OOS exercise is more relevant to evaluate effective return predictability in real time while avoiding the in-sample over-fitting issue, eventual small-sample size distortions and the look-ahead bias concern. As regards our choice of the forecasting horizon, we only consider a one-month period for two main reasons. First, it has been documented that return predictability with a short horizon is usually magnified at longer horizons (Campbell and MacKinlay, 1997 and Cochrane, 2001). Second, as we perform a business (financial) cycle analysis, multiple-horizons regressions could contaminate the results as they would include random combinations of expansions and recessions (or good and bad periods).

We use monthly data from January 1973 to December 2014 for the set of predictors from Rapach et al. (2016).

[6] Chaudhuri and Lo (2016) apply Fourier transform techniques to quantify the stock-return dynamics across multiple time horizons. The authors highlight that wavelets technique can provide substantial implementation benefits versus the Fourier transform technique they are using.

Specifically, we use the log dividend-price ratio (DP), the log dividend yield (DY), the log earnings-price ratio (EP), the log dividend-payout ratio (DE), the excess stock return volatility (RVOL), the book-to-market ratio (BM), the net equity expansion (NTIS), the Treasury bill rate (TBL), the long-term bond yield (LTY), the long-term bond return (LTR), the term spread (TMS), the default yield spread (DFY), the default return spread (DFR), the lagged inflation rate (INFL) and the short interest index (SII). In appendix 1 these predictors are briefly explained.

Our methodology to forecast the ERP applies, within the standard setting of OOS predictive regressions, the discrete wavelet transform decomposition of the different predictors, as described in section 2.1. The OOS procedure is explained in section 2.2.

2.1 Maximal Overlap Discrete Wavelet Transform MultiResolution Analysis (MODWT MRA)

Spectral analysis and Fourier transforms have been, for a long time, the most common frequency domain methods used in different areas. Wavelets are signal processing techniques that were developed to overcome some of the limitations of those traditional frequency domain tools, as they provide a more complete decomposition of the original time series without suffering their weaknesses. For instance, and differently from the Fourier analysis, wavelets are defined over a finite window in the time domain, with the size of that window being resized automatically according to the frequency of interest. This means that only high frequency features of the time series can be captured when using a short window, whereas by looking at the same signal with a large window, the low frequency features are revealed. Hence, it is possible to extract both time-varying and frequency-varying features simultaneously just by changing the size of the window. Wavelets are thus better tools to handle non-stationary time series as well as time series with structural breaks or jumps.

The decomposition process of a given time series into different time series is known as multiresolution analysis (MRA). By applying a maximal overlap discrete wavelet transform multiresolution analysis (MODWT MRA), a time series y_t is decomposed as:

$$y_t = y\left(D_1\right)_t + ... + y\left(D_J\right)_t + y\left(S_J\right)_t \quad , \tag{1}$$

where the $y\left(D_j\right)_t$, $j = 1, 2, \ldots, J$ are the wavelet details and $y\left(S_J\right)_t$ is the wavelet smooth. The original time series is therefore decomposed into orthogonal components ($y\left(D_1\right)_t$ to $y\left(D_J\right)_t$ and $y\left(S_J\right)_t$), called crystals, each defined in the time domain and representing the fluctuation of the original time series in a specific frequency band. For small j, the J wavelet details represent the higher frequency characteristics of the time series (*i.e.* its short-term dynamics) and, as j increases, the j wavelet details represent lower frequencies movements of the series, whereas the wavelet smooth captures the lowest frequency dynamics (*i.e.* its long-term behavior or trend).[7]

Given the sufficiently long data series, we are able to apply a $J = 6$ level MRA so that the wavelet decomposition delivers seven orthogonal crystals: six wavelet details ($y(D_1)_t$ to $y(D_6)_t$) and the wavelet smooth ($y(S_6)_t$). As we use monthly data, the first detail level $y(D_1)_t$ captures oscillations between 2 and 4 months, while detail levels $y(D_2)_t$, $y(D_3)_t$, $y(D_4)_t$, $y(D_5)_t$ and $y(D_6)_t$ capture oscillations with a period of 4-8, 8-16, 16-32, 32-64 and 64-128 months, respectively. Finally, the smooth component $y(S_6)_t$ captures oscillations with a period longer than 128 months (10.6 years).[8]

To illustrate the rich set of different dynamics aggregated (and therefore hidden) in the original time series, figure 2 plots the time series of the (log) excess returns (top left panel) and of its seven crystals (remaining panels). As expected, the lower the frequency, the smoother the resulting filtered time series.

2.2 Out-of-sample forecasts

The 1-step ahead forecasts are generated using a sequence of expanding windows. We use an initial sample (1973:01 to 1989:12) to make the first 1-step ahead OOS forecast. The sample is then increased by one observation and a new 1-step ahead OOS forecast is produced. This is the procedure until the end of the sample. The full OOS period therefore spans from 1990:01 to 2014:12.

[7] A more detailed analysis of wavelet methods can be found in appendix 2 and in Percival and Walden (2000). Papers using a similar wavelet decomposition includes *e.g.* Galagedera and Maharaj (2008), Xue et al. (2013), Barunik and Vacha (2015), Caraiani (2015) and Kilponen and Verona (2016).

[8] All the simulations were run using the WMTSA Wavelet Toolkit for Matlab available at http://www.atmos.washington.edu/~wmtsa/. In this paper we perform the MODWT MRA using the Haar wavelet filter (as in *e.g.* Manchaldore et al., 2010, Malagon et al., 2015 and Faria and Verona, 2016) with reflecting boundary conditions. As the wavelet family used in the MODWT may influence the results, we also run the simulations using the Daubechies wavelet filter with the filter length L = 4 (as in Barunik and Vacha, 2015) and the Coiflet wavelet filter with the filter length L = 6 (as done by Galagedera and Maharaj, 2008). Our results are robust to changes in the wavelet family. As regards the choice of J, the number of observations dictates the maximum number of frequency bands that can be used. In particular, if N is the number of observations in the in-sample period, then J has to satisfy the constraint $N \geq 2^J$.

2.2.1 Predictive regression model specifications and forecast: time series

Let r be the ERP. For each original predictor x_i, the predictive regression is

$$r_{t+1} = \alpha + \beta x_{i,t} + \varepsilon_{t+1} \ , \tag{2}$$

and the 1-step ahead OOS forecast of the excess returns, $\hat{r}_{t+1}$, is simply given by:

$$\hat{r}_{t+1} = \hat{\alpha} + \hat{\beta} x_{i,t} \ , \tag{3}$$

where $\hat{\alpha}$ and $\hat{\beta}$ are the OLS estimates of α and β, respectively. We run the forecast for each of the original predictors, which corresponds to the replication of Rapach et al. (2016) OOS forecasting exercise. We refer to it as TS (time series) in the following analysis.

2.2.2 Forecasting with wavelets

To forecast with wavelets we closely follow the method proposed by Rua (2011), which consists of fitting a model like (2) to each timescale component of the wavelet MRA decomposition of r (equation (1)), instead of fitting a model to the original variable r. The forecast for r can then be obtained by summing the forecasts for the orthogonal components using the corresponding estimated models.

Let us explain in more detail the steps involved. Firstly, a MODWT MRA decomposition is performed to the variable to be forecasted, r, as well as for all the 15 predictors. Second, we estimate a model like (2) for each resolution level and use the results to produce the 1-step ahead forecast of the corresponding component of the excess return r (as in (3)). Finally, the 1-step ahead forecast for r is obtained by adding up those forecasts. Importantly, as the MODWT MRA at a given point in time uses information of neighboring data points (both past and future), we recompute the crystals at each iteration of the OOS forecasting process in order to make sure that we only use current and past information when making the forecasts. This ensures that our method does not suffer from any look-ahead bias.

As an example, the 1-step ahead forecast of the ERP using TMS as a predictor, $\hat{r}_{t+1}^{TMS}$, is given by:

$$\hat{r}_{t+1}^{TMS} = \left[\hat{\alpha_1} + \hat{\beta}_1 \left(D_1\right)_t^{TMS}\right] + \left[\hat{\alpha_2} + \hat{\beta}_2 \left(D_2\right)_t^{TMS}\right] + \ldots + \left[\hat{\alpha_7} + \hat{\beta}_7 \left(S_6\right)_t^{TMS}\right] \ , \tag{4}$$

where $(D_j)_t^{TMS}$ and $(S_6)_t^{TMS}$ are the TMS j^{th} wavelet detail and wavelet smooth, respectively. As we use all the crystals to make the forecast of the equity premium, we denominate this specification as WAV_ALL.

This forecast exercise leads to the conclusion that some of the frequencies of the predictors carry a lot of noise to the forecast exercise. Hence, in order to improve the forecast, we take advantage of the flexibility offered by the MODWT MRA and propose a new and intuitive way of improving the forecast using wavelets. Namely, we search, for each individual predictor, for the combination of crystals that maximizes its R^2_{OS}. Taking again the TMS as an example, the forecasting "regression" is given by:

$$\hat{r}_{t+1}^{TMS} = \delta_1 \left[\hat{\alpha_1} + \hat{\beta}_1 \left(D_1\right)_t^{TMS}\right] + \delta_2 \left[\hat{\alpha_2} + \hat{\beta}_2 \left(D_2\right)_t^{TMS}\right] + \ldots + \delta_7 \left[\hat{\alpha_7} + \hat{\beta}_7 \left(S_6\right)_t^{TMS}\right] \ , \tag{5}$$

where we consider that the weights δ_j, j=*1,...,7*, can take 5 possible values: 0, 0.25, 0.50, 0.75 and 1.[9] A weight of 0 would therefore exclude a particular frequency from the forecast, thus allowing to remove the eventual noise carried by that frequency to the forecast exercise.

This is the WAV_I_BEST specification for each individual predictor and should inform about the relevant frequencies of each predictor for the ERP forecasting purposes.

2.2.3 Individual forecast combination

Even though numerous economic variables, when considered individually, fail to deliver consistent OOS forecasting gains relative to the HM benchmark, Rapach et al. (2010) show that it is possible to obtain statistically and economically significant OOS gains by combining their individual forecasts. As the authors emphasize, although the advantages of combining individual forecasts has been early pointed out by Bates and Granger (1969), applications in finance have been relatively rare (with a few exceptions such as *e.g.* Mamaysky et al., 2007, 2008). Accordingly, in this paper we also run a similar exercise and combine the individual 1-step ahead forecasts of the ERP to check whether we can improve our forecast results even further.

Concretely, the combination forecasts of r_{t+1} at time t, denoted as $\hat{r}_{c,\,t+1}$, is computed as the weighted-average of M individual forecasts based on equations (3) or (5):

[9] We consider this grid of five values, instead of a more granular one, exclusively due to the required computational effort. Although results are expected to improve as a more detailed grid is considered, we believe that this five value grid already spans a reasonable set of scenarios while being computationally tractable.

$$\hat{r}_{c,\,t+1} = \sum_{i=1}^{M} \phi_i \hat{r}_{i,\,t+1}, \tag{6}$$

where ϕ_i denotes the weight to attribute to the individual forecast based on predictor i.

Rapach et al. (2010) use several weight combination methods, which include simple averaging schemes (mean, median and trimmed mean) and the discount mean square prediction error from Stock and Watson (2004). However, the authors conclude that, in line with previous forecasting literature (Timmerman, 2006), simple combining methods typically outperform more complicated ones. In particular, Rapach et al. (2010) show that the results using the mean combination forecast are usually the strongest ones. Accordingly, in (6) we consider an equal-weight combination of the excess returns forecasts based on M predictors, that is $\phi_i = 1/M$. We denominate this specification as WAV_BEST. When useful for presentation purposes, we aggregate all forecasting models based on the wavelet decomposition of predictors (WAV_ALL, WAV_I_BEST and WAV_BEST) under the generic designation of WAV models.

2.2.4 Forecast evaluation

The forecasting performances of the TS and WAV models are evaluated using the Campbell and Thompson (2008) R^2_{OS} statistic. As standard in the literature, the benchmark model is the prevailing mean forecast $\overline{r}_s$, which is the average excess return up to time s. The R^2_{OS} statistic measures the proportional reduction in the mean squared forecast error for the predictive model ($MSFE_{PRED}$) relative to the historical mean ($MSFE_{HM}$) and is given by

$$R^2_{OS} = 1 - \frac{MSFE_{PRED}}{MSFE_{HM}} = 1 - \frac{\sum_{s=s_0}^{T-1} \left(r_{s+1} - \hat{r}_{t+1}\right)^2}{\sum_{s=s_0}^{T-1} \left(r_{s+1} - \overline{r}_s\right)^2} \ ,$$

where $\hat{r}_{t+1}$ is the excess return forecast for *t+1* from the TS or specific WAV model considered and r_{s+1} is the realized stock market return in *s+1*. A positive (negative) value of R^2_{OS} indicates that the predictive model outperforms (underperforms) the HM in terms of MSFE.

As in Rapach et al. (2010), Dangl and Halling (2012), Neely et al. (2014) and Rapach et al. (2016), among many others, the statistical significance of the results is evaluated using the Clark and West (2007) statistic. This statistic tests the null hypothesis that the MSFE of the HM model is less than or equal to the MSFE of the TS or specific WAV model against the alternative hypothesis that the MSFE of the HM model is greater than the MSFE of the TS or specific WAV model ($H_0: R^2_{OS} \leq 0$ against $H_0: R^2_{OS} > 0$).

2.3 Asset allocation

Finally, we analyse the economic value of the different models (time series and wavelets) from an asset allocation perspective. Following Kandel and Stambaugh (1996), Campbell and Thompson (2008), Ferreira and Santa-Clara (2011) and Huang et al. (2015), among others, we consider a mean-variance investor who allocates her wealth between equities and risk-free bills. At the end of month t, the investor optimally allocates

$$w_t = \frac{1}{\gamma}\frac{\hat{R}_{t+1}}{\hat{\sigma}^2_{t+1}} \tag{7}$$

of the portfolio to equity for period $t+1$. In (7), γ is the investor's relative risk aversion coefficient, $\hat{R}_{t+1}$ is the (TS or WAV) model prediction of excess return at time t for the period $t+1$, and $\hat{\sigma}^2_{t+1}$ is the forecast of the variance of the excess return.[10] As in Rapach et al. (2016), we assume a relative risk aversion coefficient of three, use a ten-year moving window of past excess returns to estimate the variance forecast and constrain the weights w_t to lie between -0.5 and 1.5. These constraints introduce realistic limits to the possibilities of short selling and leveraging the portfolio.

The realized portfolio return at time $t+1$, RP_{t+1}, is given by $RP_{t+1} = w_t R_{t+1} + RF_{t+1}$, where RF_{t+1} denotes the risk-free return from time t to $t+1$ (*i.e.* the market rate, which is know at time t). The average utility (or certainty equivalent return, CER) of an investor that uses the portfolio rule (7) is given by $CER = \overline{RP} - 0.5\gamma\sigma^2_{RP}$, where $\overline{RP}$ and σ^2_{RP} are the sample mean and variance of the portfolio return, respectively. We report the annualized utility gain from using the TS, WAV_I_BEST and WAV_BEST models. The utility gain is computed as the difference between the CER for an investor that uses the TS or the specific WAV model to forecast excess returns and the CER for an investor who uses the HM benchmark for forecasting. The difference is multiplied by 12, which allows to interpret it as the annual portfolio management fee that an investor would accept to pay to have access to the alternative forecasting model versus the historical average forecast.

[10] As in Rapach et al. (2016), among many others, the asset allocation exercise is done with the excess return in level (R_t) and not in logs (r_t).

3 Out-of-sample forecasting results

3.1 Statistical analysis

3.1.1 Individual predictors

The statistical results of the forecasting performance of individual predictors using different model specifications versus the HM, for the entire OOS period (1990:01 - 2014:12), are reported in table 1.

The time series analysis (second column) confirms Goyal and Welch (2008) results (*i.e.* that traditional predictors perform badly OOS) and shows that the Rapach et al. (2016) SII is a good predictor of excess stock market returns. As regards the forecast made with wavelets, there is no value added by considering simultaneously all the frequencies (third column, WAV_ALL model, equation 4). Except for INFL and TMS, all the $R^2_{OS}s$ are lower than in the time series analysis. This suggests that there could be excessive noise when considering the information from all frequencies. However, when the frequencies are optimally chosen (fourth column, WAV_I_BEST, equation 5), for all predictors the $R^2_{OS}s$ are higher than the respective TS R^2_{OS}, that is, the OOS forecasting performance always increases. As regards the traditional predictors, for some of them the improved OOS performance is still not enough to outperform the HM model ($R^2_{OS} < 0$). However, there are 5 cases (EP, DE, LTR, TMS and INFL) where the $R^2_{OS}s$ become positive and statistically significant. This means that some of the ERP that have been rejected in the literature have nevertheless predictability power, as long as their proper frequencies are chosen. Consider, for example, the case of the term spread (TMS). The R^2_{OS} is -0.76 when considering its original time series. However, after decomposing it into its different frequencies and applying the WAV_I_BEST model, we find that its lowest frequency alone (the trend or long-run component) has a very strong OOS predictive power, yielding a positive R^2_{OS} of 1.95 (with significance at the 1% level). To put this result into perspective, this is similar to the R^2_{OS} of 1.94 of the SII variable, which is "*the* strongest predictor of the equity risk premium identified to date" (Rapach et al., 2016, pag. 46). Using our method, the best individual predictor is the dividend-payout ratio (DE), with a monthly R^2_{OS} of 2.98 (with a 5% significance level). This is achieved when considering a weight equal to 1 for its highest and lowest frequencies (D_1 and S_6, respectively) and weights lower than 1 for some of its intermediate frequencies. As regards the SII (the only predictor with positive and statistically significant R^2_{OS} in the time series analysis), there is also a substantial improvement in its OOS performance, with the R^2_{OS} increasing to 2.55% when using the optimally-chosen frequencies.

To complement this analysis, and following Goyal and Welch (2008), Rapach et al. (2010) and Huang et al.

(2015), among others, we analyze the dynamics of the difference between the cumulative square forecasting error for the HM forecasting model and the cumulative square forecasting error when the TS or the WAV_I_BEST model for each predictor is used. Results, plotted in figure 3, should be read as follows. When the line increases/decreases, the predictive regression WAV model (in blue) or TS model (in black) outperforms/underperforms the HM. A forecasting model that constantly outperforms the HM will thus always have a positively sloped curve. The graph therefore allows to evaluate the consistency over time of the OOS performance of the forecasting model, as the less changes in the slope of the curve the more consistent is the forecasting performance.

From the plots in figure 3, when predictors are considered in their original time series (black lines), it is clear that, with the exception of the SII, for all the other predictors the corresponding lines are almost always below zero, *i.e.* underperforming the HM. As regards the WAV_I_BEST models (blue lines), it is possible to broadly classify the individual frequency-decomposed predictors into four different groups as regards the consistency of their OOS performance. The first group, which includes the DFR, the DFY, the DP, the DY, the NTIS and the TBL, respects to those predictors with an OOS performance close to that of the HM during the sample period (*i.e.* the graphs are relatively stable around zero). There are then 2 predictors (RVOL and LTY) which have an erratic forecasting performance, as the slopes swing between positive and negative values. A third group includes those predictors that, suddenly during the last NBER-dated recession, post a strong OOS outperformance versus the HM. This group includes the BM, the EP, the DE, the INFL and the SII. Finally, two predictors (LTR and TMS) post a consistent positive outperformance during the entire period (except for the first 5 years), with their corresponding lines featuring smooth and positively sloped trends.

At last, we analyze the level of complementarity in each pair of ERP predictors regarding their OOS forecasting performance. The more complementary (or less redundant) two predictors are, the higher is the expected benefit from combining their individual forecasts. Concretely, we use the forecast encompassing tests regarding the one-month ahead market excess return forecasts from the frequency decomposed individual predictors (WAV_I_BEST). Those tests provide the econometric sources of the eventual benefits of forecast combination: if the null hypothesis of encompassing between predictors i and j is rejected, then it is useful to combine forecasts from predictors i and j. Table 2 provides p-values corresponding to the Harvey and Newbold (1998) forecast encompassing test statistic (MHLN). The p-values correspond to an upper tail test of the null hypothesis that the forecast from the column predictor encompasses the forecast from the row predictor against the alternate hypothesis that it does not. Results show that there are many cases where the null hypothesis of forecast encompassing cannot be rejected at 10%, 5% and 1% levels. This suggests the existence of benefits from combining the individual forecasts of different predictors. We therefore analyse the forecast combination in the next section.

3.1.2 Forecast combination

In table 3 we report the results when combining (equal-weight) the individual forecasts from different predictors, both when their original time-series (TS) and wavelet decomposition series (WAV_I_BEST) are taken into account. As regards the time series analysis, there are no gains in combining forecasts from different predictors, as the SII alone outperforms all the alternative combinations. Moreover, any combination with more than two predictors is no longer statistically significant. These results differ from the ones reported by Rapach et al. (2010), as we are using monthly (instead of quarterly) data over a more recent (and shorter) period.

Differently, with the optimally-chosen frequency-decomposed predictors, there are gains in combining their individual forecasts. The best OOS performance (R^2_{OS} of 3.75%) is achieved when combining the excess returns forecasts from the optimally-chosen frequency-decomposed EP, DE and SII. Moreover, although the statistical performance of the forecast combination decreases when additional predictors are added to the combination EP, DE and SII, it is noticeable that all the $R^2_{OS}s$ remain statistically significant. So, overall, by combining the forecasts from the frequency-decomposed predictors it is possible to further improve the forecasting exercise versus both the TS result and the best result with the individual frequency-decomposed predictor (DE with R^2_{OS} of 2.98%).

Additionally, from the blue line in the lower right graph in figure 3, it is immediate that the OOS outperformance from the best combination of the individual forecasts of the frequency-decomposed predictors (EP, DE and SII) emerges on early 2000s and has a strong boost since the last NBER-dated recession. This is consistent with the evolution of the individual OOS performances of EP, DE and SII. Until the early 2000s there is in fact no clear outperformance of this combination, versus both the HM benchmark (as the blue line is around zero) and the SII (black line in the same graph).

3.2 Economic analysis

So far we have shown that the frequency decomposition of the different predictors delivers statistically significant gains. We now quantify the economic value of our method for excess return forecasting from an asset allocation perspective.

Reported results in the third and sixth columns of table 1 show that there are CER gain improvements for all predictors when they are frequency decomposed (WAV_I_BEST). In fact, under the standard time series analysis (third column) only four predictors have positive CER gains, while under the WAV_I_BEST prediction regressions (sixth column), 10 out of 15 predictors have positive CER gains. The highest utility gains are obtained when using the TMS and the SII, with 549 and 535 basis points, respectively. These gains are higher than the largest CER gain under the time series analysis (417 basis points using the SII).

The fifth column in table 3 shows that, from an economic point of view, there are no utility gains in combining individual forecasts of excess returns from the TS predictors. However, the utility gains increase when combining the forecasts from the individual frequency-decomposed predictors (eighth column), with a maximum CER gain of 674 basis points being achieved when combining the frequency-decomposed EP, DE, SII and TMS (while the combination that maximizes the R^2_{OS} delivers a CER gain of 621 basis points). Note also that 10 combinations of frequency-decomposed predictors deliver CER gains higher than the largest gain in the time domain.

Figure 4 provides a dynamic perspective of the portfolio and cumulative wealth for an investor that uses the HM model, the original SII (the best predictor in the time domain), the optimally-chosen frequency-decomposed SII (to compare with the time series SII), the optimally-chosen frequency-decomposed TMS (which obtains approximately the same R^2_{OS} of the original SII) and the combination of the optimally-chosen frequency-decomposed predictors of the EP, the DE and the SII (which is the combination that yields the higher R^2_{OS}).

Panel A presents the dynamic equity weights, constrained to lie between -0.5 and 1.5, for those alternative portfolios. The first result that stands out is that the equity exposure of the HM portfolio (black dash line) is much smoother than any other of the alternatives under analysis. Second, the dynamics of the equity exposure of an investor using the original SII as a predictor of future excess returns (black line) follows very closely that of an investor using the WAV_I_BEST SII as a predictor (blue line). An explanation for this may be the fact that the WAV_I_BEST SII is obtained by considering a weight of one for four (out of seven) frequencies of the original SII variable. Interestingly, changes in the equity allocation in a portfolio constructed based on the WAV_I_BEST TMS (red line) are much smoother. This is explained by the fact that the WAV_I_BEST TMS only considers the wavelet smooth frequency (the long run) of the original TMS variable. At last, when using the combination of forecasts from the frequency-decomposed EP, DE and SII (yellow line), the portfolio's equity weight has huge swings. In this case, and differently from the alternative models, the constraints on the weight ($-0.5 \leq w \leq 1.5$) are strongly binding. It is also evident, considering the NBER-dated recessions, that a strategy based on the combination of forecasts has a strong market timing. In particular, with the exception of the recession in the early 90's, there is a strong reduction of the equity exposure (including shorting) before and in the very early stages of the recessions, and the opposite occurs in the late stage of the recessions and early stages of expansions.

Panel B in figure 4 shows the log cumulative wealth for an investor that begins with 1$ and reinvests all proceeds. Between the mid 90s until the early 2000's recession, the strategies based on the original SII and on the WAV models benefit from the higher exposure to the bull equity market. During this period, the strategy based on the combination of individual forecasts has the strongest cumulative performance as it is most of the time with maximum leverage on equity. During the recession on early 2000s, the HM, the original SII and the WAV_I_BEST SII based strategies are outperformed by the WAV_I_BEST TMS and by the combination of forecasts. This reflects the excellent market timing of a strategy based on the latter and, in the case of WAV_I_BEST TMS, the smooth and continuous reduction of exposure to equity markets before the beginning of the recession. Furthermore, until the recession in late 2000s, the much higher exposure to equity markets of those two strategies versus those based on the SII, WAV_I_BEST SII and HM, explains the divergence of the respective cumulative wealth. Interestingly, during the late 2000s recession, the strategies based on the original SII, on the WAV_I_BEST SII and on the combination of individual forecasts, clearly outperform the strategy based on the WAV_I_BEST TMS. The reason is that the strategy based on the latter reduces much less the exposure to equity during the recession. Immediately after the recession, the strategies based on the SII and on the combination of individual forecasts quickly increase exposure to equity, which then stays at maximum level until the end of the sample period. A strategy based on the WAV_I_BEST TMS also reaches this point, although following a slower path. The HM based portfolio keeps a much lower exposure to equity markets, justifying its strong underperformance.

Overall, the cumulative wealth of an investor adopting a trading rule based on the combination of individual forecasts of the frequency-decomposed predictors EP, DP and TMS is clearly above any of the alternatives strategies. Interestingly, when using the WAV_I_BEST TMS or WAV_I_BEST SII as the predictors driving the equity allocation, the cumulative wealth of the investor at the end is approximately the same. The evolution is however quite different: the cumulative wealth from WAV_I_BEST TMS is higher than that of the WAV_I_BEST SII almost since the beginning of the sample period until the end of the late 2000s recession. After that, there is a strong outperformance of the strategy based on the WAV_I_BEST

SII, leading to the convergence of cumulative wealth. This is consistent with the dynamics of the difference between the cumulative square forecasting error for the HM forecasting model and the cumulative square forecasting error for the WAV_I_BEST SII and TMS models reported in figure 3, as it is evident that the strong forecasting performance of the WAV_I_BEST SII is a recent phenomena (post late 2000s recession), whereas that of the WAV_I_BEST TMS is robust throughout the entire sample period.[11] The strategy based on the original SII is dominated, from a cumulative wealth perspective, by all WAV models analysed.

4 Robustness analysis

In this section we test the robustness of our proposed methodology by running two robustness checks. The first one respects to the analysis of the forecast performance in different sample periods. The second one changes the objective function, centering the attention in maximizing the economic performance (*i.e.* the CER gains) instead of maximizing the statistical performance (*i.e.* the R^2_{OS}). Results are reported in sections 4.1 and 4.2, respectively.

4.1 Different sample periods

4.1.1 Great Moderation and Great Financial Crisis

Following Rapach et al. (2016), we divide the OOS period into two subperiods: from 1990:01 to 2006:12, which is included in the so-called Great Moderation period, and from 2007:01 to 2014:12, which corresponds to the Great Financial Crisis and aftermath.

In table 4 are reported the R^2_{OS} and the CER gains for all individual predictors, based on their original time series (TS) and on the WAV_I_BEST frequency decomposition. The OOS predictability during the 1990-2006 period is usually weaker than in the period 2007-2014. It is noticeable that, for both sample periods, using the wavelet decomposition there are significant improvements of the OOS forecasting performance for almost all predictors. In the first period, four predictors (EP, RVOL, LTR, TMS) yield positive and

[11] When comparing the WAV_I_BEST TMS and the WAV_I_BEST SII, there are three other potential advantages in using the former. The first one respects to the construction of the predictor. From the detailed explanation in Rapach et al. (2016), it is clear that the construction of the original SII requires much more assumptions than the construction of the term spread. Second, the WAV_I_BEST TMS only contains one frequency, while the WAV_I_BEST SII considers four frequencies which raises the topic of the sensitivity to the weighting scheme of the different frequencies. Third, the much smoother dynamics of the equity allocation associated with the WAV_I_BEST TMS versus the WAV_I_BEST SII, suggests that a strategy based on the former may imply less transaction costs. If this is the case then, post transaction costs, the cumulative wealth associated with the WAV_I_BEST TMS for the period under analysis may be higher that of the WAV_I_BEST SII.

statistically significant R^2_{OS} when frequency decomposed, while in the TS analysis no predictor outperforms the HM benchmark. In the second period, the SII is the best predictor both in the time series analysis and with the wavelet decomposition. However, while in the time series analysis it is the only statistically significant predictor, with the wavelet decomposition other 4 predictors (DE, BM, INFL, TMS) also yield positive and statistically significant R^2_{OS}. Moreover, in both subsample periods, the CER gains also increase significantly for almost all predictors when the wavelet decomposition method is applied. The maximum utility gains obtained are 442 and 1375 basis points in the first and second subsample periods (TMS and SII with WAV_I_BEST model, respectively), which are significantly above the gains achievable with the standard time series analysis (115 and 1118 basis points for TBL and SII, respectively).[12]

4.1.2 Bad, normal and good growth periods

There is a debate in the literature about the OOS excess returns predictability during expansions and recessions. On the one hand, Henkel et al. (2011) find evidence for complete absence of return predictability during expansions. Similar empirical results are reported by Ferreira and Santa-Clara (2011) and Neely et al. (2014), while Cujean and Hasler (2016) explore, through a general equilibrium model, why the return predictability concentrates in bad times. Interestingly, it has also been reported in the literature that fund managers perform (statistically and economically) better during recessions than during expansions (*e.g.* Kacperczyk et al., 2016). On the other hand, Dangl and Halling (2012) find statistically significant levels of OOS predictability during expansions, but only for models including time-varying coefficients. Likewise, Huang et al. (2016) conclude that the OOS forecasting in good times is also possible as long as one uses state-dependent predictive regressions.

Accordingly, and following Rapach et al. (2010), we evaluate the individual forecasts during periods of bad, normal and good economic growth. Those regimes are defined as the bottom, middle, and top third of sorted growth rates of industrial production in the US, respectively.[13] This guarantees a sufficient number of observations in each regime (100 observations each), which is difficult to achieve using NBER-dated recessions (as only 34 observations are classified as recession period) during the OOS period under analysis.[14] We report the $R^2_{OS}s$ and the CER gains for each regime in table 5 .

[12] Regarding the WAV_I_BEST specification, for each predictor and for each subsample period the optimal weights for different frequencies are recomputed. Those weights are not reported but are available upon request from the authors.

[13] The data for the industrial production in the US was downloaded from Federal Reserve Economic Data at http://research.stlouisfed.org/fred2/.

[14] We also evaluate the forecast performance during (i) good and bad times using the good time indicator from Huang et al. (2016), which is based on the past six-month excess market return and (ii) NBER-dated recessions and expansions. The results are qualitatively similar to those obtained in this section.

Looking at the R^2_{OS} of each individual predictor during bad growth periods, the number of statistically significant predictors increases from one (SII) when using the original time series, to four (DE, LTR, TMS and SII) when using the WAV_I_BEST models, with two of them significant at the 1% level. The maximum R^2_{OS} is 8.46% (DE with WAV_I_BEST model) which compares with a 2.82% for the SII (TS model). From an utility gain perspective, the CER gains also increase significantly for the DE, the LTR, the TMS and the SII when the wavelet decomposition method is applied. The maximum utility gain is of 1318 basis points (DE with WAV_I_BEST model), which is significantly above the maximum CER gain attained using the original time series of predictors (494 basis points for SII).

The same qualitative conclusions can be extended to the normal growth period, even if for this regime the only statistically significant predictor using the original time series is INFL, while using the WAV_I_BEST models the set of statistically significant predictors now includes the DY, the RVOL and the TMS. Although the attainable R^2_{OS} and CER gains using the WAV_I_BEST model during normal periods are usually lower than during bad periods, the levels are still high: the maximum R^2_{OS} and CER gains are 5.92% and 572 basis points using the RVOL and the DY, respectively.

Looking at the good period regime, there is a remarkable improvement in forecasting when using the frequency-decomposed predictors. In fact, in good periods none of the predictors is statistically significant when the TS model is used. This is aligned with the findings of some of the above-mentioned literature. However, when the proper frequencies of the predictors are considered, three predictors (EP, TMS and SII) become statistically significant. Moreover, the OOS performance is rather good, as the $R^2_{OS}s$ using these three predictors are 3.66%, 1,25% and 3.24%, respectively. From an utility perspective, results are also very strong, as the annualized CER gains are 632, 465 and 770 basis points, respectively.

Overall, the frequency decomposition of the predictors improves significantly their OOS forecast performance also when considering subsample periods corresponding to bad, normal *and* good growth.

4.2 Maximizing CER gains

It is well known from the literature on forecasting excess returns OOS that maximizing the statistical performance (*i.e.* the R^2_{OS}) does not always imply maximizing the utility of the representative investor (*i.e.* the CER gain). Bearing this in mind, we run a robustness exercise that consists in changing the objective function when measuring the forecasting performance of different predictors. In particular, for the full OOS period, we look for the weights of the different frequencies of the individual predictors that maximize the

CER gains (instead of the R^2_{OS}) and document the increase in the achieved utility gains.

The results are reported in the fourth column of table 6. Two results stand out. First, the set of predictors obtaining the largest CER gains (EP, DE, LTR, TMS, INFL and SII) is the same as in the base case scenario of section 3.1. The differences respect to the optimal combination of frequencies of each predictor. Overall, when the objective is to maximize the CER gains, the optimal number of frequencies to be included in each of those six predictors is equal or higher than when the objective is to maximize the R^2_{OS}. In other words, more information is required when maximizing the CER gains. Second, the improvements in the CER gains are particularly impressive for the EP and the DE, while being substantially smaller for the LTR and the INFL, negligible for the TMS and null for the SII. The maximum annualized CER gain is 615 basis points (using DE), which is slightly better than the best result achieved when choosing the weights that maximize the R^2_{OS} (549 basis points using the TMS).

Overall, although there are differences regarding the utility gains associated with OOS excess returns forecasting, the main insights do not change when adopting this alternative objective function of maximizing the CER gain instead of the R^2_{OS}.

5 Concluding remarks

Goyal and Welch (2008) and posterior research have documented the poor OOS ERP forecasting performance of an extensive list of predictors. In this paper we propose a new method for forecasting excess returns in equity markets, which is based on the wavelet decomposition of several predictors considered so far in the literature. We forecast excess returns of the S&P500 index, for the OOS period from January 1990 to December 2014 and for different subsamples. Regardless of the sample period, the proposed method delivers statistically and economically significant gains for investors and clearly outperforms both the traditional HM benchmark and the individual predictors when considered in their original time series. For the full OOS period, the best result achieved for an individual frequency-decomposed predictor is a monthly OOS R^2 of 2.98% and utility gains of 549 basis points per year for a mean-variance investor. This good performance is further improved when the individual forecasts are combined.

The strong and robust performance of the proposed method is essentially attributable to the fact that, by first decomposing the original time series of each predictor in its different frequencies and then choosing the frequencies that are relevant for the forecasting exercise, the accuracy of the forecast remarkably improves.

In doing so we show that some of the variables considered to be bad predictors are indeed good predictors of the ERP. The key step to capture their effective forecasting power is to eliminate the noise aggregated and embedded in the time series so as to extract the relevant frequencies for forecasting purposes.

References

Aguiar-Conraria, L., M. M. Martins, and M. J. Soares: 2012, 'The yield curve and the macro-economy across time and frequencies'. *Journal of Economic Dynamics and Control* **36**(12), 1950–1970.

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

Aguiar-Conraria, L. and M. J. Soares: 2014, 'The Continuous Wavelet Transform: Moving Beyond Uni- And Bivariate Analysis'. *Journal of Economic Surveys* **28**(2), 344–375.

Ang, A. and G. Bekaert: 2007, 'Stock Return Predictability: Is it There?'. *Review of Financial Studies* **20**(3), 651–707.

Barunik, J. and L. Vacha: 2015, 'Realized wavelet-based estimation of integrated variance and jumps in the presence of noise'. *Quantitative Finance* **15**(8), 1347–1364.

Bates, J. M. and C. W. J. Granger: 1969, 'The Combination of Forecasts'. *Operational Research Quarterly* **20**(4), 451–468.

Bekaert, G. and R. J. Hodrick: 1992, 'Characterizing Predictable Components in Excess Returns on Equity and Foreign Exchange Markets'. *The Journal of Finance* **47**(2), 467–509.

Bollerslev, T., G. Tauchen, and H. Zhou: 2009, 'Expected Stock Returns and Variance Risk Premia'. *Review of Financial Studies* **22**(11), 4463–4492.

Campbell, J. Lo, A. and A. MacKinlay: 1997, *The Econometrics of Financial Markets*. Princeton University Press.

Campbell, J. Y.: 1987, 'Stock returns and the term structure'. *Journal of Financial Economics* **18**(2), 373–399.

Campbell, J. Y. and R. J. Shiller: 1988a, 'The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors'. *Review of Financial Studies* **1**(3), 195–228.

Campbell, J. Y. and R. J. Shiller: 1988b, 'Stock Prices, Earnings, and Expected Dividends'. *Journal of Finance* **43**(3), 661–676.

Campbell, J. Y. and S. B. Thompson: 2008, 'Predicting Excess Stock Returns Out of Sample: Can Anything Beat the Historical Average?'. *Review of Financial Studies* **21**(4), 1509–1531.

Campbell, J. Y. and T. Vuolteenaho: 2004, 'Inflation Illusion and Stock Prices'. *American Economic Review* **94**(2), 19–23.

Caraiani, P.: 2015, 'Estimating DSGE models across time and frequency'. *Journal of Macroeconomics* **44**(C), 33–49.

Chaudhuri, S. and A. Lo: 2016, 'Spectral Portfolio Theory'. mimeo.

Clark, T. and K. West: 2007, 'Approximately normal tests for equal predictive accuracy in nested models'. *Journal of Econometrics* **138**(1), 291 – 311.

Cochrane, J.: 2001, *Asset Pricing*. Princeton University Press.

Cochrane, J. H.: 2008, 'The Dog That Did Not Bark: A Defense of Return Predictability'. *Review of Financial Studies* **21**(4), 1533–1575.

Cooper, I. and R. Priestley: 2009, 'Time-Varying Risk Premiums and the Output Gap'. *Review of Financial Studies* **22**(7), 2601–2633.

Cooper, I. and R. Priestley: 2013, 'The World Business Cycle and Expected Returns'. *Review of Finance* **17**(3), 1029–1064.

Crowley, P. M.: 2007, 'A Guide To Wavelets For Economists'. *Journal of Economic Surveys* **21**(2), 207–267.

Cujean, J. and M. Hasler: 2016, 'Why Does Return Predictability Concentrate in Bad Times?'.

Cutler, D. M., J. M. Poterba, and L. H. Summers: 1991, 'Speculative Dynamics'. *Review of Economic Studies* **58**(3), 529–546.

Dangl, T. and M. Halling: 2012, 'Predictive regressions with time-varying coefficients'. *Journal of Financial Economics* **106**(1), 157–181.

Duarte, F. M. and C. Rosa: 2015, 'The equity risk premium: a review of models'. Staff Reports 714, Federal Reserve Bank of New York.

Engsted, T. and T. Q. Pedersen: 2010, 'The dividend-price ratio does predict dividend growth: International evidence'. *Journal of Empirical Finance* **17**(4), 585–605.

Fama, E. F. and K. R. French: 1988, 'Dividend yields and expected stock returns'. *Journal of Financial Economics* **22**(1), 3–25.

Fama, E. F. and K. R. French: 1989, 'Business conditions and expected returns on stocks and bonds'. *Journal of Financial Economics* **25**(1), 23–49.

Fama, E. F. and G. Schwert: 1977, 'Asset returns and inflation'. *Journal of Financial Economics* **5**(2), 115–146.

Faria, G. and F. Verona: 2016, 'Forecasting stock market returns by summing the frequency-decomposed parts'. mimeo.

Ferreira, M. A. and P. Santa-Clara: 2011, 'Forecasting stock market returns: the sum of the parts is more than the whole'. *Journal of Financial Economics* **100**(3), 514–537.

Ferson, W. and C. Harvey: 1991, 'The Variation of Economic Risk Premiums'. *Journal of Political Economy* **99**(2), 385–415.

Ferson, W. E. and C. R. Harvey: 1993, 'The Risk and Predictability of International Equity Returns'. *Review of Financial Studies* **6**(3), 527–566.

Galagedera, D. and E. Maharaj: 2008, 'Wavelet timescales and conditional relationship between higher-order systematic co-moments and portfolio returns'. *Quantitative Finance* **8**(2), 201–215.

Gallegati, M., M. Gallegati, J. B. Ramsey, and W. Semmler: 2011, 'The US Wage Phillips Curve across Frequencies and over Time'. *Oxford Bulletin of Economics and Statistics* **73**(4), 489–508.

Gallegati, M. and J. B. Ramsey: 2013, 'Bond vs stock market's Q: Testing for stability across frequencies and over time'. *Journal of Empirical Finance* **24**(C), 138–150.

Gencay, R., F. Selcuk, and B. Whitcher: 2005, 'Multiscale systematic risk'. *Journal of International Money and Finance* **24**(1), 55–70.

Goyal, A. and I. Welch: 2008, 'A Comprehensive Look at The Empirical Performance of Equity Premium Prediction'. *Review of Financial Studies* **21**(4), 1455–1508.

Harvey, C. R.: 1991, 'The World Price of Covariance Risk'. *The Journal of Finance* **46**(1), 111–157.

Harvey, C. R., Y. Liu, and H. Zhu: 2016, '... and the Cross-Section of Expected Returns'. *Review of Financial Studies* **29**(1), 5–68.

Harvey, David I., S. J. L. and P. Newbold: 1998, 'Tests for Forecast Encompassing'. *Journal of Business and Economic Statistics* **16**(2), 254–259.

Henkel, S. J., J. S. Martin, and F. Nardari: 2011, 'Time-varying short-horizon predictability'. *Journal of Financial Economics* **99**(3), 560–580.

Hjalmarsson, E.: 2010, 'Predicting Global Stock Returns'. *Journal of Financial and Quantitative Analysis* **45**, 49–80.

Huang, D., F. Jiang, J. Tu, and G. Zhou: 2015, 'Investor Sentiment Aligned: A Powerful Predictor of Stock Returns'. *Review of Financial Studies* **28**(3), 791–837.

Huang, D., F. Jiang, J. Tu, and G. Zhou: 2016, 'Forecasting Stock Returns in Good and Bad Times: The Role of Market States'.

Jordan, S. J., A. J. Vivian, and M. E. Wohar: 2014, 'Forecasting returns: New European evidence'. *Journal of Empirical Finance* **26**, 76–95.

Kacperczyk, M., S. Van Nieuwerburgh, and L. Veldkamp: 2016, 'A Rational Theory of Mutual Funds' Attention Allocation'. *Econometrica* **84**(2), 571–626.

Kandel, S. and R. F. Stambaugh: 1996, 'On the Predictability of Stock Returns: An Asset-Allocation Perspective'. *Journal of Finance* **51**(2), 385–424.

Kelly, B. and S. Pruitt: 2013, 'Market Expectations in the Cross-Section of Present Values'. *Journal of Finance* **68**(5), 1721–1756.

Kilponen, J. and F. Verona: 2016, 'Testing the Q theory of investment in the frequency domain'. mimeo.

Kim, S. and F. In: 2005, 'The relationship between stock returns and inflation: new evidence from wavelet analysis'. *Journal of Empirical Finance* **12**(3), 435–444.

Kothari, S. P. and J. Shanken: 1997, 'Book-to-market, dividend yield, and expected market returns: A time-series analysis'. *Journal of Financial Economics* **44**(2), 169–203.

Lamont, O.: 1998, 'Earnings and Expected Returns'. *Journal of Finance* **53**(5), 1563–1587.

Lettau, M. and S. Ludvigson: 2001, 'Consumption, aggregate wealth, and expected stock returns'. *Journal of Finance* **56**(3), 815–849.

Li, Y., D. T. Ng, and B. Swaminathan: 2013, 'Predicting market returns using aggregate implied cost of capital'. *Journal of Financial Economics* **110**(2), 419–436.

Ludvigson, S. C. and S. Ng: 2007, 'The empirical risk-return relation: A factor analysis approach'. *Journal of Financial Economics* **83**(1), 171 – 222.

Malagon, J., D. Moreno, and R. Rodriguez: 2015, 'Time horizon trading and the idiosyncratic risk puzzle'. *Quantitative Finance* **15**(2), 327–343.

Mamaysky, H., M. Spiegel, and H. Zhang: 2007, 'Improved Forecasting of Mutual Fund Alphas and Betas'. *Review of Finance* **11**(3), 359–400.

Mamaysky, H., M. Spiegel, and H. Zhang: 2008, 'Estimating the Dynamics of Mutual Fund Alphas and Betas'. *Review of Financial Studies* **21**(1), 233–264.

Manchaldore, J., I. Palit, and O. Soloviev: 2010, 'Wavelet decomposition for intra-day volume dynamics'. *Quantitative Finance* **10**(8), 917–930.

Mele, A.: 2007, 'Asymmetric stock market volatility and the cyclical behavior of expected returns'. *Journal of Financial Economics* **86**(2), 446–478.

Moller, S. V. and J. Rangvid: 2015, 'End-of-the-year economic growth and time-varying expected returns'. *Journal of Financial Economics* **115**(1), 136 – 154.

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

Pastor, L. and R. Stambaugh: 2009, 'Predictive Systems: Living with Imperfect Predictors'. *The Journal of Finance* **64**(4), 1583–1628.

Percival, D. and A. Walden: 2000, *Wavelet methods for time series analysis*. Cambridge University Press.

Pettenuzzo, D., A. Timmermann, and R. Valkanov: 2014, 'Forecasting stock returns under economic constraints'. *Journal of Financial Economics* **114**(3), 517–553.

Pontiff, J. and L. D. Schall: 1998, 'Book-to-market ratios as predictors of market returns'. *Journal of Financial Economics* **49**(2), 141–160.

Ramsey, J. B. and C. Lampart: 1998a, 'The Decomposition of Economic Relationships by Time Scale Using Wavelets: Expenditure and Income'. *Studies in Nonlinear Dynamics & Econometrics* **3**(1), 1–22.

Ramsey, J. B. and C. Lampart: 1998b, 'Decomposition Of Economic Relationships By Timescale Using Wavelets: Money and Income'. *Macroeconomic Dynamics* **2**(1), 49–71.

Rapach, D. and G. Zhou: 2013, *Forecasting Stock Returns*, Vol. 2 of *Handbook of Economic Forecasting*, pp. 328–383. Elsevier.

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

Rapach, D. E., J. K. Strauss, and G. Zhou: 2010, 'Out-of-Sample Equity Premium Prediction: Combination Forecasts and Links to the Real Economy'. *Review of Financial Studies* **23**(2), 821–862.

Rapach, D. E., J. K. Strauss, and G. Zhou: 2013, 'International Stock Return Predictability: What Is the Role of the United States?'. *Journal of Finance* **68**(4), 1633–1662.

Rua, A.: 2010, 'Measuring comovement in the time-frequency space'. *Journal of Macroeconomics* **32**(2), 685–691.

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

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

Stock, J. H. and M. W. Watson: 2004, 'Combination forecasts of output growth in a seven-country data set'. *Journal of Forecasting* **23**(6), 405–430.

Timmerman, A.: 2006, *Forecast Combinations*. Amsterdam, The Netherlands: Elsevier. in G. Elliott, C. W. J. Granger, and A. Timmermann (eds.), Handbook of Economic Forecasting.

Xue, Y., R. Gencay, and S. Fagan: 2013, 'Jump detection with wavelets for high-frequency financial time series'. *Quantitative Finance* **14**(8), 1427–1444.

	TS		WAV_ALL	WAV_I_BEST								
	R^2_{OS}	CER gains	R^2_{OS}	R^2_{OS}	CER gains	δ_1	δ_2	δ_3	δ_4	δ_5	δ_6	δ_7
DP	-2.06	-3.19	-41.24	-0.36	-0.68	0	0	0	0	0	0	0.5
DY	-2.20	-2.96	-25.82	-0.37	-0.69	0	0	0	0	0	0.25	0.5
EP	-1.14	-0.34	-51.04	2.64*	3.23	0	0.5	1	0.25	0	0	0.75
DE	-2.27	-1.13	-3.22	2.98**	3.87	1	0	0.25	0.5	0	0	1
RVOL	-0.56	-1.82	-3.95	-0.01	0.27	0	0	0.25	0	0.5	0	0.75
BM	-0.56	-0.78	-15.79	0.21	0.25	0	0	0	0	1	1	0.5
NTIS	-3.23	-2.57	-3.77	-0.03	0.10	1	1	0	0	0	0	0.5
TBL	-0.38	0.66	-3.50	-0.26	-0.38	0	0	0	0	0	1	0.75
LTY	-0.31	-0.05	-1.65	-0.19	-0.49	0	0	0	1	0.25	1	0.5
LTR	-0.51	-0.95	-2.31	1.00*	1.67	0.25	0.5	0	0.25	0	0	1
TMS	-0.76	0.25	-0.52	1.95***	5.49	0	0	0	0	0	0	1
DFY	-3.07	-4.90	-12.77	-0.64	-1.32	0	0	0	0	0	0	0.5
DFR	-1.75	1.08	-6.63	0.55	0.64	1	0	0	0	0.75	1	0.25
INFL	-0.64	-0.55	0.44	1.00*	2.45	0	1	0	0.25	1	1	0.75
SII	1.94***	4.17	1.16**	2.55**	5.35	1	0	0	0	1	1	1

Table 1: Out-of-sample R-squares $\left(R^2_{OS}\right)$ and annualized CER gains

This table reports the out-of-sample R-squares (in percentage) for the excess returns forecasts at monthly (nonoverlapping) frequencies from the model as given by equation (3) for each of the original predictors (TS, second column), from WAV_ALL model specification (equation (4), fourth column) and from the WAV_I_BEST model (equation (5), fifth column) for each predictor where the crystals used and corresponding weights (δ_j, $j = 1, 2, \ldots, 7$) are listed in the last seven columns. The out-of-sample R-squares $\left(R^2_{OS}\right)$ measures the proportional reduction in the mean squared forecast error for the predictive model relative to the forecast based on the historical mean (HM). The 1-month ahead out-of-sample forecast of excess return is generated using a sequence of expanding windows. In columns three and six are reported the annualized certainty equivalent return (CER) gain (in percent) for an investor who allocates her wealth between equities and risk free bills according to the rule (7), using stock return forecasts from models in equations (3) and (5) instead of the forecasts based on the HM. The sample period is from 1973:01 to 2014:12. The full out-of-sample forecasting period is from 1990:01 to 2014:12, monthly frequency. Asterisks denote significance of the out-of-sample *MSFE*-adjusted statistic of Clark and West (2007). ***, ** and * denotes significance at the 1%, 5% and 10% levels respectively.

	Column predictor														
Row predictor	DP	DY	EP	DE	RVOL	BM	NTIS	TBL	LTY	LTR	TMS	DFY	DFR	INFL	SII
DP		0.42	0.49	0.66	0.36	0.93	0.39	0.52	0.43	0.77	0.92	0.17	0.46	0.84	0.78
DY	0.53		0.49	0.67	0.37	0.94	0.41	0.56	0.45	0.78	0.93	0.16	0.47	0.83	0.78
EP	0.02	0.02		0.18	0.02	0.03	0.02	0.02	0.02	0.05	0.10	0.01	0.03	0.06	0.09
DE	0.00	0.00	0.09		0.01	0.01	0.00	0.00	0.00	0.02	0.05	0.00	0.01	0.02	0.04
RVOL	0.16	0.15	0.28	0.47		0.29	0.17	0.16	0.16	0.42	0.62	0.09	0.34	0.44	0.47
BM	0.03	0.02	0.41	0.61	0.18		0.21	0.06	0.12	0.62	0.92	0.02	0.33	0.69	0.65
NTIS	0.19	0.19	0.35	0.49	0.18	0.35		0.20	0.23	0.49	0.59	0.12	0.34	0.34	0.50
TBL	0.28	0.25	0.50	0.69	0.28	0.85	0.35		0.36	0.75	0.97	0.07	0.43	0.81	0.84
LTY	0.22	0.21	0.36	0.58	0.25	0.59	0.34	0.27		0.73	0.83	0.07	0.41	0.58	0.68
LTR	0.03	0.03	0.21	0.39	0.07	0.08	0.06	0.03	0.04		0.46	0.01	0.16	0.21	0.38
TMS	0.00	0.00	0.18	0.31	0.01	0.00	0.00	0.00	0.00	0.03		0.00	0.05	0.05	0.33
DFY	0.66	0.70	0.50	0.68	0.46	0.90	0.60	0.74	0.73	0.81	0.92		0.54	0.73	0.78
DFR	0.16	0.16	0.32	0.42	0.19	0.22	0.16	0.17	0.17	0.26	0.36	0.13		0.24	0.42
INFL	0.01	0.01	0.26	0.43	0.06	0.05	0.05	0.02	0.02	0.21	0.61	0.01	0.14		0.48
SII	0.00	0.00	0.09	0.09	0.01	0.00	0.00	0.00	0.00	0.01	0.09	0.00	0.01	0.02	

Table 2: Forecast encompassing test results, MHLN statistic p-values

This table reports p-values of the forecasting encompassing test statistic of Harvey and Newbold (1998) (MHLN statistic). The statistic corresponds to a one-sided (upper-tail) test of the null hypothesis that the forecast from the column predictor encompasses the forecast from the row predictor against the alternative hypothesis that the forecast from the column predictor does not encompass the forecast from the row predictor. The dependent variable in these regressions is the 1-month ahead market excess returns. Predictors are from the WAV_I_BEST (equation (5)) using the crystals listed in table 1. The sample period is from 1973:01 to 2014:12. The full out-of-sample forecasting period is from 1990:01 to 2014:12, monthly frequency.

Number of predictors	Number of combinations	TS R^2_{OS}	TS predictor	TS CER gains	WAV_I_BEST R^2_{OS}	WAV_I_BEST predictor	WAV_I_BEST CER gains
1	15	1.94***	SII	4.17	2.98**	DE	3.87
2	105	1.34**	SII, RVOL	2.39	3.52**	EP, SII	5.66
3	455	1.16	SII, LTR, DFR	2.48	3.75***	EP, SII, DE	6.21
4	1365	0.98	as in 3 + RVOL	1.92	3.49***	as in 3 + TMS	6.74
5	3003	0.86	as in 4 + TBL	1.89	3.21***	as in 4 + LTR	6.43
6	5005	0.73	as in 5 + TMS	1.79	2.98***	as in 5 + DFR	5.81
7	6435	0.62	as in 6 + LTY	1.48	2.80***	as in 6 + INFL	5.51
8	6435	0.50	as in 7 + BM	1.16	2.59***	as in 7 + RVOL	5.41
9	5005	0.40	as in 8 + INFL	1.05	2.40***	as in 8 + NTIS	5.12
10	3003	0.32	as in 9 + EP	0.84	2.21***	as in 9 + BM	4.71
11	1365	0.20	as in 10 + DP	0.55	2.03***	as in 10 + LTY	4.32
12	455	0.08	as in 11 + DE	0.37	1.85***	as in 11 + TBL	3.95
13	105	-0.02	as in 12 + DY	0.18	1.70***	as in 12 + DP	3.58
14	15	-0.16	as in 13 + NTIS	-0.13	1.56**	as in 13 + DY	3.26
15	1	-0.29	as in 14 + DFY	-0.50	1.43**	as in 14 + DFY	2.99

Table 3: Forecast combination: out-of-sample R-squares $\left(R^2_{OS}\right)$ and annualized CER gains

This table reports the out-of-sample R-squares $\left(R^2_{OS}\right)$ and annualized certainty equivalent return (CER) gain (in percent) obtained when combining the excess returns forecasts at monthly (nonoverlapping) frequencies from the model as given by equation (3) for each of the original predictors (TS, columns three to five) and from the WAV_I_BEST model (equation (5), columns six to eight). The out-of-sample R-squares $\left(R^2_{OS}\right)$ measure the proportional reduction in the mean squared forecast error for the predictive model relative to the forecast based on the historical mean (HM). The 1-month ahead out-of-sample forecast of excess return is generated using a sequence of expanding windows. The CER gain (in percent) is computed for an investor who allocates her wealth between equities and risk free bills according to the rule (7), using the stock return forecasts from models in equations (3) and (5) instead of the forecast based on the HM. The sample period is from 1973:01 to 2014:12. The full out-of-sample forecasting period is from 1990:01 to 2014:12, monthly frequency. Asterisks denote significance of the out-of-sample *MSFE*-adjusted statistic of Clark and West (2007). ***, ** and * denotes significance at the 1%, 5% and 10% levels respectively.

	1990:01 - 2006:12				2007:01 - 2014:12			
	TS		WAV_I_BEST		TS		WAV_I_BEST	
	R^2_{OS}	CER gains	R^2_{OS}	CER gains	R^2_{OS}	CER gains	R^2_{OS}	CER gains
DP	-3.27	-4.47	0.10	-0.64	-0.15	-0.48	-0.04	-0.57
DY	-3.59	-4.35	-0.06	-0.83	0.00	-0.02	-0.12	-0.05
EP	-1.05	-1.02	1.53*	2.28	-1.27	1.10	5.54	7.03
DE	-1.81	-1.58	1.74	2.65	-2.99	-0.16	6.41**	7.61
RVOL	-1.73	-3.03	2.28**	3.77	1.28	0.70	-0.46	-1.71
BM	-0.86	-1.14	0.06	-0.05	-0.09	-0.01	0.54 *	1.14
NTIS	-2.84	-1.58	0.67	1.39	-3.84	-4.67	0.04	-1.09
TBL	-0.49	1.15	-0.34	-0.79	-0.21	-0.40	0.04	0.86
LTY	-0.41	-0.33	-0.19	-0.41	-0.16	0.57	0.60	1.68
LTR	-0.83	-0.98	1.54**	1.74	-0.01	-0.92	2.69	2.56
TMS	-1.11	1.05	1.60***	4.42	-0.20	-1.48	2.63***	7.08
DFY	-3.37	-4.43	0.04	-0.09	-2.58	-5.93	-0.88	-2.24
DFR	-2.64	0.96	0.50	1.88	-0.35	1.34	2.95	2.46
INFL	-0.03	0.62	0.36	0.86	-1.61	-3.01	4.39 *	8.59
SII	-0.15	0.88	0.19	1.11	5.23***	11.18	6.81 **	13.75

Table 4: Out-of-sample R-squares $\left(R^2_{OS}\right)$ and annualized CER gains

This table reports between column two and five the out-of-sample R-squares (in percentage) for excess returns forecasts at monthly (nonoverlapping) frequencies from the model as given by equation (3) for each of the original predictors (TS) and from the WAV_I_BEST model in equation (5) for each predictor. The out-of-sample R-squares $\left(R^2_{OS}\right)$ measures the proportional reduction in the mean squared forecast error for the predictive model relative to the forecast based on the historical mean (HM). The 1-month ahead out-of-sample forecast of excess return is generated using a sequence of expanding windows. From columns six to nine is reported the annualized certainty equivalent return (CER) gain (in percent) for an investor who allocates her wealth between equities and risk free bills according to the rule (7), using stock return forecasts from above mention models in equations (3) and (5) instead of forecasts based on the HM. The sample period is from 1973:01 to 2014:12. Two out-of-sample forecasting periods are considered: from 1990:01 to 2006:12 and from 2007:01 to 2014:12, monthly frequency. Asterisks denote significance of the out-of-sample *MSFE*-adjusted statistic of Clark and West (2007). ***, ** and * denotes significance at the 1%, 5% and 10% levels respectively.

	Bad				Normal				Good			
	TS		WAV_I_BEST		TS		WAV_I_BEST		TS		WAV_I_BEST	
	R^2_{OS}	CER gains	R^2_{OS}	CER gains	R^2_{OS}	CER gains	R^2_{OS}	CER gains	R^2_{OS}	CER gains	R^2_{OS}	CER gains
DP	-1.08	-2.07	-0.11	-0.28	-2.85	-3.81	1.05	1.53	-2.60	-3.68	0.11	-1.88
DY	-1.14	-1.62	-0.14	-0.31	-3.06	-3.83	3.86*	5.72	-2.78	-3.41	0.03	-1.66
EP	-1.86	-0.64	7.16	8.61	0.21	0.47	1.82	0.59	-1.28	-0.82	3.66**	6.32
DE	-2.80	0.85	8.46***	13.18	-3.13	-3.65	-0.92	-1.13	-1.07	-0.60	2.59	3.97
RVOL	-0.15	-1.86	0.18	0.58	-1.08	-2.64	5.92**	5.53	-0.65	-1.00	0.11	1.22
BM	-0.77	-1.81	0.80	2.02	0.05	-0.22	0.39	0.34	-0.77	-0.30	0.16	-0.54
NTIS	-5.51	-6.04	1.67	1.17	-0.37	0.46	0.91	2.79	-2.70	-2.08	0.39	3.50
TBL	-0.24	-0.28	0.22	0.77	0.13	1.69	0.17	-0.02	-0.90	0.52	-0.32	-0.15
LTY	-0.43	-0.90	0.42	1.09	0.42	0.61	0.61	0.01	-0.69	0.16	-0.20	-0.42
LTR	-0.62	-1.36	2.56*	5.17	-0.59	-1.00	0.75	0.54	-0.34	-0.52	1.76	2.57
TMS	0.56	0.60	2.71***	7.17	-2.65	-0.32	2.08**	4.77	-0.90	0.43	1.25*	4.65
DFY	-2.89	-5.04	-0.31	-0.40	-4.48	-5.92	0.39	0.09	-2.26	-3.75	-0.67	-0.87
DFR	-10.58	-1.30	0.28	0.56	-1.05	-0.71	0.67	-0.60	7.55	5.26	7.94	3.83
INFL	-1.32	-1.46	1.98	2.80	1.34*	1.40	1.22	2.22	-1.28	-1.60	1.46	3.21
SII	2.82**	4.94	3.78 **	7.34	0.29	1.96	1.14	1.99	2.12	5.57	3.24*	7.70

Table 5: Out-of-sample R-squares $\left(R^2_{OS}\right)$ and annualized CER gains

This table reports between column two and seven the out-of-sample R-squares (in percentage) for excess returns forecasts at monthly (nonoverlapping) frequencies from the model as given by equation (3) for each of the original predictors (TS) and from the WAV_I_BEST model in equation (5) for each predictor. The out-of-sample R-squares $\left(R^2_{OS}\right)$ measures the proportional reduction in the mean squared forecast error for the predictive model relative to the forecast based on the historical mean (HM). The 1-month ahead out-of-sample forecast of excess return is generated using a sequence of expanding windows. From columns eight to thirteen is reported the annualized certainty equivalent return (CER) gain (in percent) for an investor who allocates her wealth between equities and risk free bills according to the rule (7), using stock return forecasts from above mention models in equations (3) and (5) instead of forecasts based on the HM. The sample period is from 1973:01 to 2014:12. Three out-of-sample forecasting periods are considered, each with 100 monthly observations: bad growth, normal growth and good growth. Those regimes are defined as the bottom, middle and top third of sorted growth rates of industrial production in the US, respectively. Asterisks denote significance of the out-of-sample *MSFE*-adjusted statistic of Clark and West (2007). ***, ** and * denotes significance at the 1%, 5% and 10% levels respectively.

	TS	Max R^2_{OS}	Max CER gains	δ_1	δ_2	δ_3	δ_4	δ_5	δ_6	δ_7
DP	-3.19	-0.68	0.21	0	0	0	0	0	0	1
DY	-2.96	-0.69	0.17	0.25	0	0	0	0	0	1
EP	-0.34	3.23	6.15	0	0.25	1	1	1	1	1
DE	-1.13	3.87	5.48	0.5	0	0.75	0.5	0	1	1
RVOL	-1.82	0.27	1.45	1	0	0	0	0.25	0	1
BM	-0.78	0.25	0.93	0	0	0	0	1	1	1
NTIS	-2.57	0.10	0.60	1	1	1	0	0	1	0.75
TBL	0.66	-0.38	-0.16	1	0	0	0	0	1	1
LTY	-0.05	-0.49	-0.11	0	0	0	1	0	1	1
LTR	-0.95	1.67	2.22	1	0.75	0	0	0	0	1
TMS	0.25	5.49	5.53	0	0	0	0	0	0.5	1
DFY	-4.90	-1.32	-1.20	0	0	0	0	0	0	1
DFR	1.08	0.64	1.65	1	0	0	0	0.75	1	0.5
INFL	-0.55	2.45	3.03	0	1	0.75	1	1	1	1
SII	4.17	5.35	5.35	1	0	0	0	1	1	1

Table 6: Annualized CER gains

This table reports the annualized certainty equivalent return (CER) gain (in percent) for an investor who allocates her wealth between equities and risk free bills according to the rule (7), using stock return forecasts from model (3) using the original time series (TS) of each predictor, model (5) using frequency decomposed predictors where the respective crystals are chosen to maximize the out-of-sample R-squares or to maximize the CER gains, instead of using forecasts based on the HM. The optimal combination of crystals for each predictor that maximizes the CER gains are reported in the last seven columns. The sample period is from 1973:01 to 2014:12. The full out-of-sample forecasting period is from 1990:01 to 2014:12, monthly frequency.

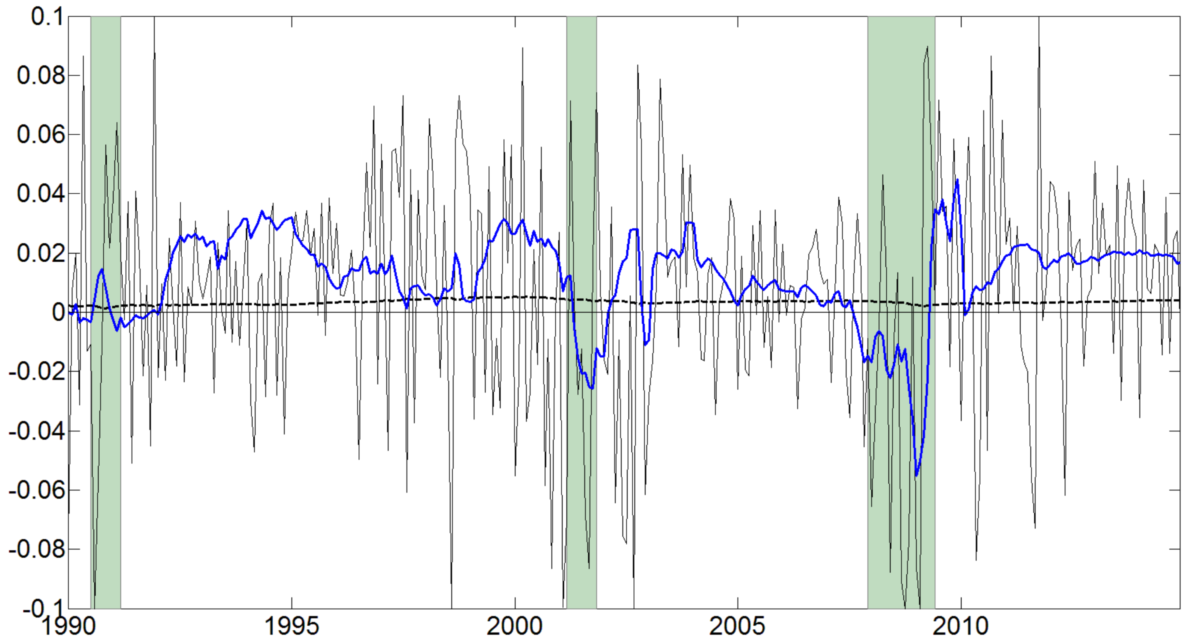


Figure 1: Realized and predicted excess returns

The black solid line corresponds to the log realized excess return as proxied by the log S&P 500 index return minus the log return on a one-month Treasury bill. The remaining lines represent the one-month ahead out-of-sample excess return forecast based on the historical mean (HM) of excess returns (dashed back line) and on the optimal combination of individual forecasts from frequency decomposed predictors (blue line). Monthly frequency, from 1990:01 to 2014:12.

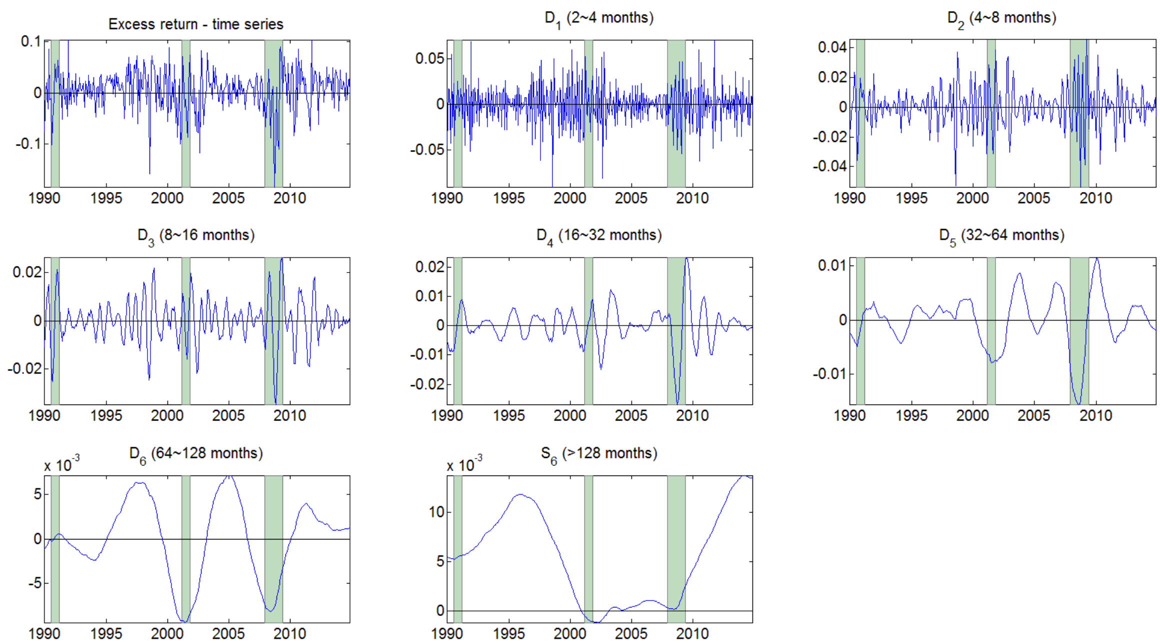


Figure 2: Excess return, time series and wavelet decomposition

The time series of the (log) excess return as proxied by the log S&P 500 index return minus the log return on a one-month Treasury bill is presented in the top left panel. From top to bottom and from left to right are displayed the seven orthogonal crystals into which the excess return time series is decomposed. It is applied a $J = 6$ level wavelet decomposition which leads to six wavelet details (D_1, $D_2, \ldots,$ D_6), representing the higher-frequency characteristics of the series, plus a wavelet smooth (S_6), that captures the low-frequency dynamics of the series. See section 2.1 and appendix 2 for full technical details on the wavelet decomposition. Sample period from 1973:01 to 2014:12, monthly frequency.

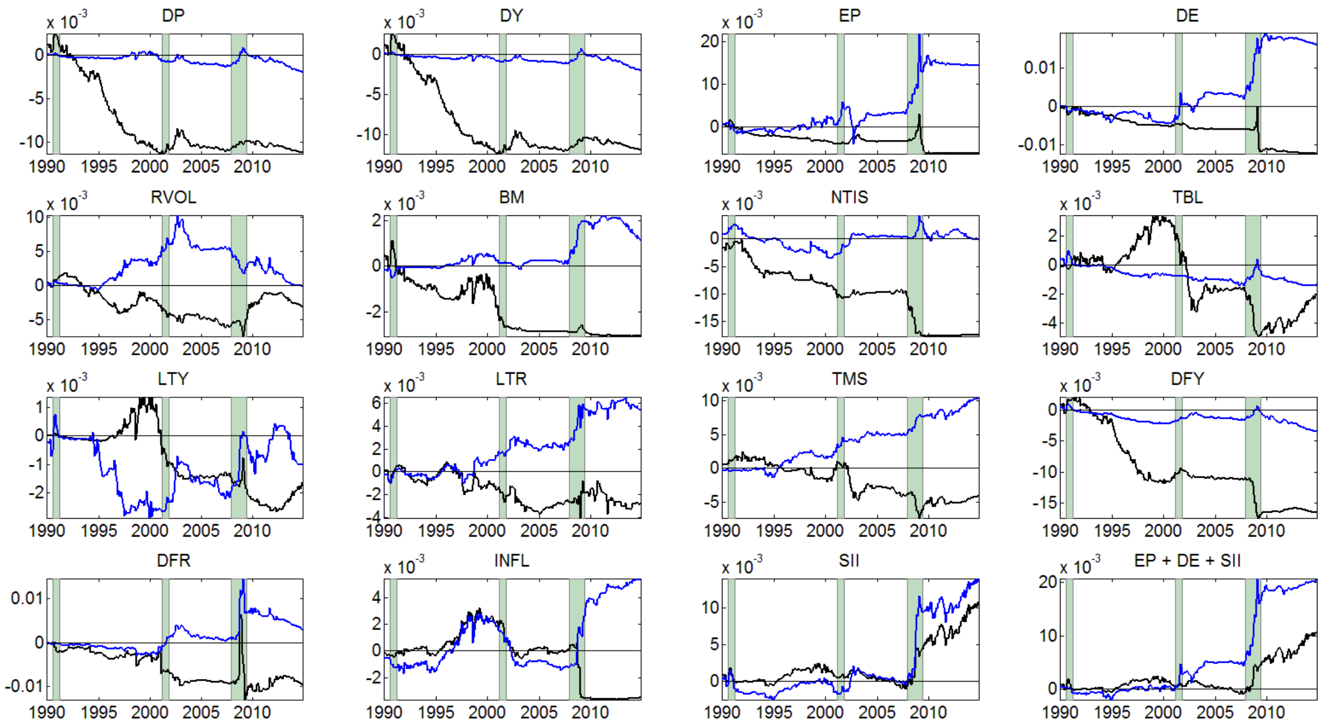
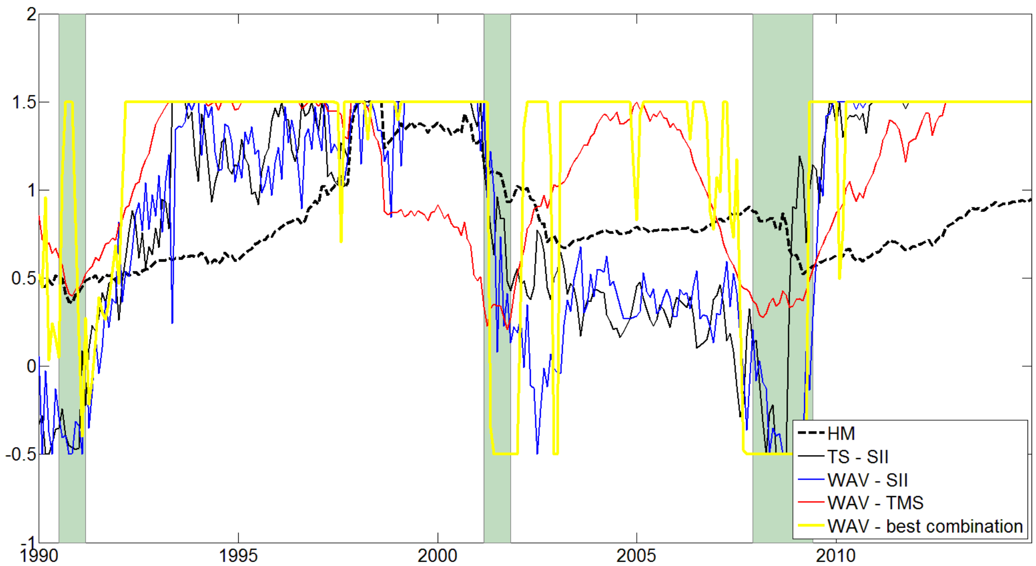


Figure 3: Difference between cumulative square forecasting error for the HM forecasting model and the cumulative square forecasting error for the individual predictive regression forecasting model

This figure reports, for each of the 15 equity premium predictors described in appendix 1, the difference between the cumulative square forecasting error for the HM forecasting model and the cumulative square forecasting error for the individual predictive regression forecasting based on the WAV_I_BEST model (5) with crystals reported in table 5 of appendix 3, line in blue, and when each individual predictor is considered in its original time series (TS), line in black.The sample period is from 1973:01 to 2014:12. The full out-of-sample forecasting period is from 1990:01 to 2014:12, monthly frequency.

A. Equity weights



B. Log cumulative wealth

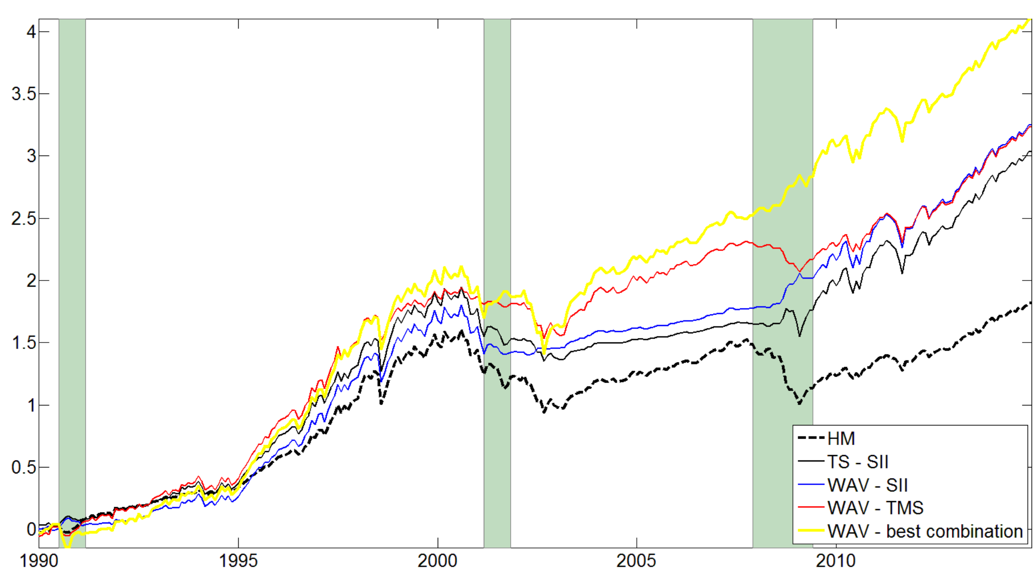


Figure 4: Equity weights and log cumulative wealth

Panel A plots the dynamics of the equity weight for a mean-variance investor who allocates monthly her wealth between equities and risk free bills according to the rule (7), using stock return forecasts based on the HM benchmark (dashed black line), on the SII as the predictor considering its original time series (TS - SII, solid black line) and its frequency decomposition (WAV - SII, blue line), on the frequency decomposed TMS (WAV - TMS, red line) and on the best combination of individual forecasts from frequency decomposed predictors (WAV - Best, yellow line). The frequency decomposition of individual predictors is made according to the WAV_I_BEST model (5) using crystals as reported in table 5 of appendix 3. The combination of forecasts is made according to the WAV_BEST model (6) and includes to EP, DE and SII. The equity weight is constrained to lie between -0.5 and 1.5. Panel B delineates the corresponding log cumulative wealth for the investor, assuming that she begins with 1$ and reinvests all proceeds. Green bars denote NBER-dated recessions. The investor is assumed to have a relative risk aversion coefficient of three. Sample period from 1990:01 to 2014:12, monthly frequency.

Appendix 1. Definition of predictors of excess returns

The excess returns predictors analysed are:

- Log dividend-price ratio (DP): difference between the log of dividends (12-month moving sums of dividends paid on S&P 500) and the log of prices (S&P 500 index).
- Log dividend yield (DY): difference between the log of dividends (12-month moving sums of dividends paid on S&P 500) and the log of lagged prices (S&P 500 index).
- Log earnings-price ratio (EP): difference between the log of earnings (12-month moving sums of earnings on S&P 500) and the log of prices (S&P 500 index price).
- Log dividend-payout ratio (DE): difference between the log of dividends (12-month moving sums of dividends paid on S&P 500) and the log of earnings (12-month moving sums of earnings on S&P 500).
- Excess stock return volatility (RVOL): calculated using a 12-month moving standard deviation estimator, as in Mele (2007).
- Book-to-market ratio (BM): ratio of book value to market value for the Dow Jones Industrial Average.
- Net equity expansion (NTIS): ratio of 12-month moving sums of net equity issues by NYSE-listed stocks to the total end-of-year NYSE market capitalization.
- Treasury bill rate (TBL): three-month Treasury bill rate.
- Long-term yield (LTY): long-term government bond yield.
- Long-term return (LTR): long-term government bond return.
- Term spread (TMS): difference between the long-term government bond yield and the T-bill.
- Default yield spread (DFY): difference between Moody's BAA- and AAA-rated corporate bond yields.
- Default return spread (DFR): difference between long-term corporate bond and long-term government bond returns.
- Inflation rate (INFL): calculated from the Consumer Price Index (CPI) for all urban consumers.
- Short interest index (SII): standardized detrended series measuring total short selling in the US exchanges, as constructed by Rapach et al. (2016).

The time series of predictors are obtained from David Rapach website.

Appendix 2. Discrete wavelet transform (DWT)

A wavelet is a function of finite length which oscillates around the time axis and looses power as it moves away from the center. The name wavelet originates from the admissibility condition, which requires the (mother) wavelet to be of finite support (small) and of oscillatory (wavy) behavior, hence wavelet (small wave).

The DWT allows to decompose a time series into its constituent multiresolution components. High-frequency components reflect the short-term behavior of the variable, whereas the low-frequency component captures its long-term dynamics. There are two distinct wavelets: the father wavelets ϕ, which captures the smooth and low-frequency part of the series, and mother wavelets ψ, that captures the detail and high-frequency components of the series:

$$\int \phi_t dt = 1 \quad and \quad \int \psi_t dt = 0 \ .$$

Given a time series y_t, with the number of observations equal to N, its orthogonal wavelet approximation is defined by

$$y_t = \sum_k s_{J,k}\phi_{J,k,t} + \sum_k d_{J,k}\psi_{J,k,t} + \sum_k d_{J-1,k}\psi_{J-1,k,t} + \cdots + \sum_k d_{1,k}\psi_{1,k,t} \ , \tag{8}$$

with J representing the number of multi-resolution levels (or frequencies) and k ranging between one and the number of coefficients in the corresponding component.[15] The maximum number of frequencies that can be considered in the analysis is driven by the number of observations as $N \geq 2^J$.

In equation (8) there are two families of inputs: the approximating wavelet functions $\phi_{J,k,t}$ and $\psi_{j,k,t}$, and the wavelet transform coefficients $s_{J,k}$, $d_{J,k}$, $d_{J-1,k}, \ldots,$ $d_{1,k}$. The approximating wavelet functions are generated from the father ϕ and mother ψ wavelets through scaling and translation in the following way:

$$\begin{aligned} \phi_{J,k,t} &= 2^{-J/2}\phi\left(\frac{t-2^J k}{2^J}\right), \\ \psi_{j,k,t} &= 2^{-j/2}\psi\left(\frac{t-2^j k}{2^j}\right), \quad j = 1, 2, \ldots J. \end{aligned}$$

The wavelet transform coefficients represent the contribution of the respective wavelet function to the signal

[15] When the number of observations is divisible by 2^J there are N wavelet coefficients. See Rua (2011) for further details.

and are given by

$$
\begin{aligned}
s_{J,k} &= \int y_t \phi_{J,k,t} dt, \\
d_{j,k} &= \int y_t \psi_{j,k,t} dt \quad j = 1, 2, \ldots, J.
\end{aligned}
$$

Those wavelet transform coefficients are obtained using the DWT method, which maps the vector $y = (y_1, y_2, \ldots, y_N)'$ to a vector of N wavelet coefficients that includes the smooth coefficients $s_{J,k}$ and the detail coefficients $d_{j,k}$. The DWT method therefore maps the original time series y_t in the time domain to a representation in the time-frequency domain $(y_{1,t}, y_{2,t}, \ldots, y_{N,t})'$. Equation (8) can therefore be rewritten as:

$$
y_t = S_{J,t} + D_{J,t} + D_{J-1,t} + \ldots + D_{1,t}, \tag{9}
$$

where $S_{J,t} = \sum_k s_{J,k} \phi_{J,k,t}$ is the single wavelet smooth and $D_{j,t} = \sum_k d_{j,k} \psi_{j,k,t}$ for $j = 1, 2, \ldots, J$ are the J wavelet details. A J level wavelet decomposition of the variable y_t therefore consists of J wavelet details, which represent the higher-frequency characteristics of y_t, and a single wavelet smooth that captures the low-frequency dynamics. Equation (9) represents the time-frequency decomposition of y_t and is the so-called wavelet multiresolution decomposition: the original series y_t – exclusively defined in the time domain – is decomposed in orthogonal components (or crystals), $S_{J,t}$, $D_{J,t}$, $D_{J-1,t}, \ldots, D_{1,t}$, each defined in the time domain and representing the fluctuation of the original time series in a specific frequency band.

Bank of Finland Research Discussion Papers 2017

ISSN 1456-6184, online

1/2017 Gonçalo Faria – Fabio Verona
Forecasting the equity risk premium with frequency-decomposed predictors
ISBN 978-952-323-144-3, online

http://www.suomenpankki.fi/en • email: Research@bof.fi
ISBN 978-952-323-144-3, ISSN 1456-6184, online