Erasmus University Rotterdam

Erasmus School of Economics

Bachelor Thesis Econometrie & Operationele Research
(Major Quantitative Finance)

Forecasting applications of the persistence-based Wold decomposition

Author:
Wouter de Voogd

Student number:
452840

Supervisor:
Dr. M. Grith

Second assessor:
Prof. dr. M. van der Wel

Date final version: July 2, 2020

Abstract

Stationary time series show dependence across different time scales, which can be modelled by the persistence-based Wold decomposition. This paper studies forecasting applications of this decomposition. We apply the methodology to forecast bond returns and realized volatility, where we provide an adjustment of the used autoregressive model to provide better estimation results for larger time horizons. We make a forecast combination with the HAR-RV model to improve realized volatility forecasting and extend the analysis with a high-frequency realized volatility data. The model does not show improvements for short-term forecasting in terms of the RMSE and Mincer-Zarnowitz R^2. For the long-term, the error variance decreases slightly, which also applies for high-frequency data. We construct portfolios with realized volatility forecasts and show that these results improve minimum variance portfolios based on moving windows. We finally apply the decomposition on a factor model. We show that for the chosen stock indices, the portfolio weights and spectral betas do not change significantly, which indicates equal systematic risk across components.

Contents

1 Introduction

A well-known characteristic of financial data is the complexity of its volatility, as this can be related to multiple periods. Therefore, it is informative for investors to know whether pricing depends on long-term relations, to avoid suffering from mispricings in the short term. Thus, time series analysis based on long-term information is an important field of analysis in finance. Ortu et al. [2020] propose a method to decompose a stationary time series into orthogonal components of different durations: the persistence-based Wold decomposition (PWD). This decomposition is an extension of the classical Wold decomposition, constructed by Wold [1938]. The PWD is based on white noise processes and decompose them such that persistence components over increasing time periods are created. This method can be used to investigate whether a certain time series depends on long-term innovations and also provides a way to decompose the variance of a time series into short-term and long-term scales. Ortu et al. [2020] use these scales to construct a forecasting model with persistence scales. They compare their forecasts to another volatility forecasting method, but do not investigate combinations of the methods. Also, this volatility forecasting method can contain useful information for investment issues.

In this paper, we investigate in which ways PWD can be used to construct volatility forecasts in different applications. We start by replicating some results of Ortu et al. [2020] and perform a simulation, for which we show the reconstruction of a time series. Next, we apply PWD to forecast bond returns and predict realized volatility. The forecasting method starts by estimating an autoregressive model for the data, from which the coefficients and error terms are used to construct the scales. We investigate if the forecasting performance is sensitive for the choice of the autoregressive level and if the forecasting results can be improved by using another criterion for the level choice.

Ortu et al. [2020] compare their realized volatility forecasts to the results of the Heterogeneous Autoregressive model of Realized Volatility (HAR-RV), proposed by Corsi [2009]. This model forecasts realized volatility with different time period components, but uses, in contrast to Ortu et al. [2020], aggregated realized volatilities of days, weeks, and months. This relatively simple model shows reasonably good performance in forecasting. Ortu et al. [2020] find that the models give different estimates and differ in positive and negative errors. Forecasts of models can be combined, which can help to reduce the adverse effects of misspecification. Timmermann [2006] remarks that especially models with different approaches can obtain diversification gains from forecast combinations. Therefore, we compare PWD and HAR-RV in forecasting performance and investigate if we can improve the performance of the single models by combining both models, using the method of Timmermann [2006]. To investigate the forecasting performance of the single models and forecast combination model for high-frequency RV, we apply the models on two different RV data sets: one of 2-hour RV and one of 10-minute RV.

Furthermore, the volatility forecasts based on the PWD could also contain useful information for the construction of portfolios. The estimation of realized volatility is useful in portfolio selection, which is a famous research field in finance, founded by Markowitz [1952]. The estimation of asset volatilities is an essential element in determining asset weights for a portfolio. Much literature has been written about this subject. Studies like De Pooter et al. [2008] use high-frequency data for the covariance matrix and a simple estimation for the expected return. Halbleib and Voev [2016] show that a mixed approach of high-frequency data for the realized volatility and lower-frequency data for the returns performs well compared to other high-frequency approaches, whereas a mixed approach has computational advantages. To investigate if the realized volatility forecasts improve the performance of portfolios, we construct a portfolio for stock indices where the volatility forecasts of the PWD are used to estimate the covariance matrix by the mixed approach of Halbleib and Voev [2016], using high-frequency data

and lower-frequency data. We also explore another way to estimate portfolio weights using scale components and use the spectral factor model, described by Bandi et al. [2019]. This model applies the multivariate form of the PWD to decompose a factor model. We use this model to investigate if we can observe differences in systematic risk across scale components. We summarize this all into the following research question: **What is the added value of the persistence-based Wold decomposition for forecasting?**

This paper makes several contributions to the literature. The first contribution is the replication of the research of Ortu et al. [2020] for the forecasting of bond returns and realized volatility. We find in line with Ortu et al. [2020] that the bond returns can be predicted by short-term level components and long-term slope components, and that the results for the realized volatility are generally accurate. Our second contribution is that we show that forecasts by PWD can be improved by using different criteria for different forecast horizons, to choose the level of the autoregressive model on which the scale decomposition is based. Specifically, the AIC turns out to be a good criterion for long-term forecasting. Our third contribution is that we find that forecast combinations of the PWD and HAR-RV model do not improve forecasts in the short term. However, forecast combinations for larger time horizons give indications they can reduce the error variance. Our fourth contribution is that we show that the results of the forecast combination for 10-minute RV data are generally similar to the 2-hour RV results. The fifth contribution of this paper is a method to estimate portfolio weights for stock indices using realized volatility forecasts by PWD. We show that we can improve minimum variance portfolios by making use of realized volatility forecasts in a mixed approach. Our sixth contribution is that we show in our application of the spectral factor model based on the PWD, that the systematic risk of the portfolio is approximately equal across components for our stock index data. The results show that the portfolio weights do not change significantly from a simple factor model and the spectral betas are relatively constant over scale components.

This paper proceeds as follows. In Section 2, we present the methodology. Here we describe the persistence-based Wold decomposition and reconstruction of a time series. Next, we explain how this model can be used to construct forecasts, forecast combinations and portfolios by realized volatility estimations and the spectral factor model. In Section 3, we describe the used data sets for the applications. Section 4 presents the results of our empirical applications. Section 5 concludes.

2 Methodology

2.1 The persistence-based Wold decomposition

Consider a weakly stationary stochastic process with zero-mean

$$x_t = \sum_{h=0}^{\infty} \alpha_h \varepsilon_{t-h}, \quad t \in \mathbb{Z}, \tag{1}$$

where the process ε_t is white noise and denotes the so-called fundamental innovations. Assume further that the coefficients $\alpha_h = \mathbb{E}[x_t \varepsilon_{t-h}]$ are square-summable and independent from t. Ortu et al. [2020] show that Equation 1 can be decomposed in the following way, that is called the **persistence-based Wold decomposition**:

$$x_t = \sum_{j=1}^{\infty} \sum_{k=0}^{\infty} \beta_k^{(j)} \varepsilon_{t-k2^j}^{(j)}, \quad \text{for } j \in \mathbb{N}. \tag{2}$$

This decomposition consists of two groups of terms. The first group is the **detail process at scale j**, denoted by the white noise process $\varepsilon_t^{(j)}$, which is a MA(2^j-1) process with respect to the fundamental innovations:

$$\varepsilon_t^{(j)} = \frac{1}{\sqrt{2^j}}\Big(\sum_{i=0}^{2^{j-1}-1} \varepsilon_{t-i} - \sum_{i=0}^{2^{j-1}-1} \varepsilon_{t-k2^{j-1}-i}\Big), \tag{3}$$

The second group of terms is formed by the **scale-specific moving average coefficients** $\beta_k^{(j)}$, that are also square-summable for all $j \in \mathbb{N}$ and independent of t, and are calculated by

$$\beta_k^{(j)} = \frac{1}{\sqrt{2^j}}\Big(\sum_{i=0}^{2^{j-1}-1} \alpha_{k2^j+i} - \sum_{i=0}^{2^{j-1}-1} \alpha_{k2^j+2^{j-1}+i}\Big). \tag{4}$$

The PWD is split into **persistence components at scale j**, whose summation results in the process x_t. They are defined as

$$g_t^{(j)} = \sum_{k=0}^{\infty} \beta_k^{(j)} \varepsilon_{t-k2^j}^{(j)}. \tag{5}$$

In practice, we also need a finite-scale version of the extended Wold decomposition, where the maximum scale J is fixed. In this case, x_t is decomposed as respectively the sum of the persistent components at the scales $j = 1, ..., J$ and a certain residual component. This results in the following decomposition:

$$x_t = \sum_{j=1}^{J} \sum_{k=0}^{\infty} \beta_k^{(j)} \varepsilon_{t-k2^j}^{(j)} + \pi^{(J)}. \tag{6}$$

2.2 Reconstruction of a time series from its components

Assume that for each scale j a detail process is driven by its underlying innovations as a MA(2^j-1) process, as follows:

$$\varepsilon_t^{(j)} = \sum_{i=0}^{2^j-1} \delta_i^{(j)} \varepsilon_{t-i}, \quad \text{where all } \delta_i^{(j)} \in \mathbb{R}. \tag{7}$$

Consider a process x_t, which is the sum of the stochastic processes $g_t^{(j)}$, that are built up from coefficients $\beta_k^{(j)}$. Under some assumptions, for which we refer to Ortu et al. [2020], the process x_t is weakly stationary and can be decomposed as

$$x_t = \sum_{h=0}^{\infty} \alpha_h \varepsilon_{t-h}, \quad \text{where} \quad \alpha_h = \sum_{j=1}^{\infty} \beta_{\lfloor \frac{h}{2^j} \rfloor}^{(j)} \delta_{h-2^j \lfloor \frac{h}{2^j} \rfloor}^{(j)}, \tag{8}$$

where $\lfloor \frac{h}{2^j} \rfloor$ returns the greatest integer less than or equal to $\frac{h}{2^j}$.

2.2.1 Simulation of a time series reconstruction

To illustrate the methodology of a multiscale construction, we perform in line with Ortu et al. [2020], a simulation. We start with detail processes $\varepsilon_t^{(j)}$, $j = 1, ..., J$, with coefficients $\delta_i^{(j)}$ as follows: $\frac{1}{\sqrt{2^j}}$ for $0 \le i \le 2^{j-1}-1$ and $-\frac{1}{\sqrt{2^j}}$ for $2^{j-1}-1 \le i \le 2^j-1$. We consider two weakly stationary stochastic processes x_t and y_t, that are each defined by a sequence of scale-specific

moving average coefficients. We take the following extended Wold coefficients for the AR(1) process x_t with parameter ρ_x, where $|\rho_x| < 1$:

$$\beta_{x,k}^{(j)} = \frac{\rho_x^{k2^j}}{\sqrt{2^j}} \frac{(1-\rho_x^{2j-1})^2}{1-\rho_x} \quad \text{where } j \in \mathbb{N},\ k \in \mathbb{N}_0. \tag{9}$$

The scale-specific coefficients of y_t, an AR(1) process with horizon 2^J and parameter ρ_y, where $|\rho_y| < 1$, are defined as

$$\beta_{y,k}^{(j)} = \begin{cases} 0 & \text{for } 1 \le j \le J \\ \dfrac{\rho_y^{k2^{j-J}}}{\sqrt{2^{j-J}}} \dfrac{(1-\rho_y^{2j-J-1})^2}{1-\rho_y} & \text{for } j \ge J+1 \end{cases} \quad \text{where } j \in \mathbb{N}, k \in \mathbb{N}_0. \tag{10}$$

We construct the coefficients of the process z_t as the sum of the coefficients of x_t and y_t, i.e. $\beta_{z,k}^{(j)} = \beta_{x,k}^{(j)} + \beta_{y,k}^{(j)}$. Therefore, z_t is expressed as follows:

$$z_t = \sum_{h=0}^{\infty} \alpha_h \varepsilon_{t-h}, \quad \text{where}$$

$$\alpha_h = \begin{cases} \displaystyle\sum_{j=1}^{\infty} \Big(\frac{\rho_x^{k2^j}}{\sqrt{2^j}} \frac{(1-\rho_x^{2j-1})^2}{1-\rho_x} \Big) \delta_{h-2^j\lfloor \frac{h}{2^j} \rfloor}^{(j)} & \text{for } 1 \le j \le J \\ \displaystyle\sum_{j=1}^{\infty} \Big(\frac{\rho_x^{k2^j}}{\sqrt{2^j}} \frac{(1-\rho_x^{2j-1})^2}{1-\rho_x} + \frac{\rho_y^{k2^{j-J}}}{\sqrt{2^{j-J}}} \frac{(1-\rho_y^{2j-J-1})^2}{1-\rho_y} \Big) \delta_{h-2^j\lfloor \frac{h}{2^j} \rfloor}^{(j)} & \text{for } j \ge J+1 \end{cases} \tag{11}$$

for $j \in \mathbb{N}$, $k \in \mathbb{N}_0$. By construction, z_t is also weakly stationary. In Section 4, we show simulation results for some parameter choices and compare them to an AR(1) process.

2.3 Forecasting with the persistence-based Wold decomposition

2.3.1 Choice of autoregressive model

To construct forecasts from the PWD, we first approximate our data x_t by an AR(M) model. Ortu et al. [2020] use the Bayesian Information Criterion (BIC) to choose the autoregressive level, assuming a Gaussian distribution for the errors to calculate the likelihood function. The BIC is generally defined as

$$\text{BIC}(p) = p \ln n - 2 \ln \hat{L}, \tag{12}$$

where $\hat{L}$ denotes the log-likelihood, p the number of parameters and n the sample size. We use two alternative criteria to estimate the autoregressive level. Ortu et al. [2020] estimate the BIC for a varying number of observations $n = T - p$, where T denotes the whole time series. As this method can overestimate p, we also use a BIC based on the same number of observations. For this purpose, we choose a maximum number of parameters P and calculate all BICs for the same $n = T - P$ observations. The second alternative criterion we use to determine the autoregressive level is the Akaike Information Criterion (AIC):

$$\text{AIC}(p) = \frac{2p}{n} - \frac{2}{n} \ln \hat{L}. \tag{13}$$

This criterion has the advantage that it aims to minimize prediction error and also tends to estimate a higher autoregressive level than the BIC. Especially for long-term forecasting, this

could give us better results. We evaluate the forecasting performance for all criteria in the results. After determining the autoregressive order, we estimate the following linear regression for a moving window:

$$x_t = \sum_{m=1}^{M} b_m x_{t-m} + \varepsilon_t. \tag{14}$$

2.3.2 Forecasting

The autoregressive model of Equation 14 allows us to construct coefficients and forecasts. In line with Ortu et al. [2020], we iteratively construct the impulse response functions as described by Di Virgilio et al. [2019]:

$$\alpha_0 = \mathbb{E}\big[\varepsilon_t^2\big]^{1/2}, \quad \alpha_m = \sum_{h=\max\{m-M,0\}}^{m-1} \alpha_h b_{m-h}, \quad \forall m \in \mathbb{N}. \tag{15}$$

From the estimated values of α_h from Equation 15, we construct the moving average coefficients $\beta_k^{(j)}$ as described in Equation 4. With the values of $\beta_k^{(j)}$ and the normalized ε_t obtained in Equation 14 until time t, we construct the persistence components $g_t^{(j)}$, $j = 1, ..., J$, as described in Equation 5.

Suppose we want to perform a p-step ahead forecast, where $p \in \mathbb{N}$. We assume that the relationship between the main process x_t and the scale components $g_t^{(j)}$ is constant over time in the moving window. To estimate the dependence of the main process on the scale components, we perform the following linear regression:

$$x_t = a^{(0)} + \sum_{j=1}^{J} a^{(j)} g_t^{(j)} + \nu_t. \tag{16}$$

From the estimated coefficients $\hat{a}^{(j)}$, we construct a forecast for x_t. First, we estimate the expected persistence components for time $t+p$. With the information at time t, the conditional expectation of the persistence components for time $t+p$ are given by

$$\mathbb{E}_t[g_{t+p}^{(j)}] = \sum_{k=0}^{K_j-1} \beta_{k,p}^{(j)} \varepsilon_{t-k2^j}^{(j)}, \tag{17}$$

where K_j denotes the number of moving average coefficients that are calculated for each scale j. These coefficients are calculated by

$$\beta_{k,p}^{(j)} = \frac{1}{\sqrt{2^j}} \Big(\sum_{i=0}^{2^{j-1}-1} \alpha_{k2^j+i+p} - \sum_{i=0}^{2^{j-1}-1} \alpha_{k2^j+2^{j-1}+i+p} \Big). \tag{18}$$

Using the coefficients $\hat{a}^{(j)}$ obtained in Equation 16, we obtain for x_{t+p} the following forecast:

$$\mathbb{E}_t[x_{t+p}] = \hat{a}^{(0)} + \sum_{j=1}^{J} \hat{a}^{(j)} \mathbb{E}_t[g_{t+p}^{(j)}]. \tag{19}$$

2.4 Forecast combination

We examine if forecasting by a combination of two models produces more accurate estimations. First, we estimate two models for a comparison sample, using a moving window. With the results of the errors for this comparison sample, we calculate the weights for which the variance is minimized. Based on these weights, we estimate a combined model for the estimation sample, again with a moving window. We combine the persistence-based Wold decomposition by Ortu et al. [2020], described in Subsection 2.3, and the HAR-RV model by Corsi [2009]. This method forecasts realized volatility (RV) in the following way:

$$x_{t+1}^{(d)} = c + \beta^{(d)} x_t^{(d)} + \beta^{(w)} x_t^{(w)} + \beta^{(m)} x_t^{(m)} + \eta_{t+1}, \tag{20}$$

where $x_t^{(d)}$ denotes the RV of the current day, $x_t^{(w)}$ denotes the average RV of the last 5 week days (the current day included) and $x_t^{(m)}$ denotes the average RV of the last 22 month days (the current day included).

For the forecast combination, we use the method described by Timmermann [2006] where the variance of the forecast errors is minimized. First we estimate x_t with two different models and obtain the forecasts $\hat{x}_{1,t+p|t}$ and $\hat{x}_{2,t+p|t}$. Assume that the forecasts are unbiased and the forecast errors have variance of σ_1^2 and σ_2^2, and a covariance of $\sigma_{12} = \rho_{12}\sigma_1\sigma_2$. Then we can combine forecasts in the form of

$$\hat{x}_{t+p|t}^{c} = \omega \hat{x}_{1,t+p|t} + (1-\omega)\hat{x}_{2,t+p|t}, \quad \text{where } \omega \in \mathbb{R}. \tag{21}$$

Provided the assumptions are true, this forecast also has zero mean and a variance of

$$\sigma_c^2(\omega) = \omega^2\sigma_1^2 + (1-\omega^2)\sigma_2^2 + 2\omega(1-\omega)\sigma_{12}. \tag{22}$$

Minimizing $\sigma_c^2(\omega)$ with respect to ω yields the following optimal weights:

$$\omega^* = \frac{\sigma_2^2 - \sigma_{12}}{\sigma_1^2 + \sigma_2^2 - 2\sigma_{12}}, \quad 1-\omega^* = \frac{\sigma_1^2 - \sigma_{12}}{\sigma_1^2 + \sigma_2^2 - 2\sigma_{12}}. \tag{23}$$

2.5 Portfolio selection

2.5.1 Realized volatility forecasting

We use the RV forecasts by PWD from Equation 19 for optimal portfolio construction. For this purpose, we use daily reported stock index data. Let $r_{i,t}$ denote the close-to-close returns and $x_{i,t}$ denote the 10-minute RV, for stock indices $i = 1, ..., N$ and t in days. We first estimate the daily return forecasts $\hat{r}_{i,t+1}$ by taking the mean of the assets using a short moving window. Next, we forecast the RV of each stock index i using the method of Ortu et al. [2020] described in Subsection 2.3, and obtain daily RV forecasts $\hat{x}_{i,t+1}$ as described in Equation 19. Here we use a large moving window, that is updated every day.

We use these forecasts to construct the matrices in Equation 24. First, we construct the diagonal matrix $\hat{D}_{t+1}$ for each predicted day, where the diagonal elements are the squared roots of the predicted daily high-frequency-based RVs. Next, we estimate for each day in the estimation sample the correlation matrix $\hat{R}_{t+1}$ of the stock index returns, where we use a shorter moving window on each stock that is also updated every day. This results in the following matrices, where $\rho_{X,Y} = \text{Corr}(r_{X,t}, r_{Y,t})$:

$$\hat{D}_{t+1} = \begin{pmatrix} \sqrt{\hat{x}_{1,t+1}} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sqrt{\hat{x}_{N,t+1}} \end{pmatrix}, \quad \hat{R}_{t+1} = \begin{pmatrix} 1 & \rho_{1,2} & \dots & \rho_{1,N} \\ \rho_{1,2} & 1 & \dots & \rho_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{1,N} & \rho_{2,N} & \dots & 1 \end{pmatrix}. \tag{24}$$

To estimate the covariance matrix, we use the Mixed Approach (MA), proposed by Halbleib and Voev [2016]. This method is based on a decomposition of the covariance matrix, which is used by Engle [2002]. In the MA, the covariance matrix is estimated as follows:

$$\hat{\Sigma}_{t+1}^{MA} = \hat{D}_{t+1}^{F^D} \hat{R}_{t+1}^{F^R} \hat{D}_{t+1}^{F^D}, \tag{25}$$

where F^D and F^R denote two different information sets on which the forecasts are based. For this purpose, we use the estimated matrices from Equation 24, that are indeed based on different information sets. An advantage of using the diagonal matrix $\hat{D}_{t+1}$ in the Mixed Approach is that we do not need covariance estimations from Ortu et al. [2020], such that we focus on the performance of the RV estimations.

From the obtained estimations, we construct optimal portfolio weights. We consider two methods: global minimum variance (GMV) and tangency. The portfolio weight vectors are constructed as follows, where we denote $\hat{r}_{t+1} = (\hat{r}_{1,t+1}, ..., \hat{r}_{N,t+1})'$ and ι as a vector of ones:

$$w_{\tan} = \frac{(\hat{\Sigma}_{t+1})^{-1}\hat{r}_{t+1}}{\iota'(\hat{\Sigma}_{t+1})^{-1}\hat{r}_{t+1}}, \quad w_{\text{GMV}} = \frac{(\hat{\Sigma}_{t+1})^{-1}\iota}{\iota'(\hat{\Sigma}_{t+1})^{-1}\iota}. \tag{26}$$

As an unrestricted tangency portfolio can return very extreme weights, we rescale the tangency portfolios on days where one or more weights w_i are calculated for which $|w_i| > 1$, for some $1 \leq i \leq N$, as follows:

$$w_i^{\text{new}} = \frac{w_i - 1/N}{\max\limits_{1 \leq i \leq N} |w_i|} + 1/N \quad \text{for } i = 1, ..., N, \tag{27}$$

such that the weights are bounded between $1 + 1/N$ and $-1 + 1/N$. To investigate the power of using the RV estimations of Ortu et al. [2020], we compare the results with the portfolio selection where $\hat{\Sigma}_{t+1}$ is calculated as the covariance matrix of the stocks based on a shorter moving window (which we denote as MW):

$$\hat{\Sigma}_{t+1}^{MW} = \begin{pmatrix} \text{Var}(r_1) & \text{Cov}(r_1, r_2) & \dots & \text{Cov}(r_1, r_N) \\ \text{Cov}(r_1, r_2) & \text{Var}(r_2) & \dots & \text{Cov}(r_2, r_N) \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}(r_1, r_N) & \text{Cov}(r_2, r_N) & \dots & \text{Var}(r_N) \end{pmatrix}. \tag{28}$$

Moreover, we use the $1/N$ portfolio as a benchmark. We compare the returns of the methods with some performance measures: the mean return, the Sharpe Ratio, the Sortino Ratio, the Value-at-Risk at 5% and the variance.

2.5.2 Spectral factor model

Next, we use the persistence-based Wold decomposition to construct portfolios by factor models. We use the spectral factor model of Bandi et al. [2019], that is based on the multivariate form of the PWD. We first construct a simple factor model. Assume that we have stock returns $r_{i,t}$

for stocks $i = 1, ..., N$ and days t. The vector $\beta_i = (\beta_{i,1}, ..., \beta_{i,M})$ denotes the sensitivities of asset i to the M factors $f_t = (f_{1,t}, ..., f_{M,t})$. The standard factor decomposition of the asset returns has the following form:

$$r_{i,t} = \alpha_i + \beta_i f_t' + u_{i,t}. \tag{29}$$

Here the covariance matrix of the returns is denoted as

$$\Sigma_r = BVB' + D, \tag{30}$$

where B is an $N \times M$ matrix of the factor betas that are obtained by a regression, V is the $M \times M$ covariance matrix of the factors f_t and D is assumed to be the diagonal matrix $D = \mathbb{E}[u_t' u_t]$.

Now we construct a spectral factor model. First, we split both the returns and the factors into J components by the multivariate form of the persistence-based Wold decomposition. We obtain the scales $r_{i,t}^{(j)}$, for $j = 1, ..., J+1$, $i = 1, ..., N$, and the scale vectors $f_t^{(j)}$, $j = 1, ..., J+1$. We perform the following $N(J+1)$ linear regressions:

$$r_{i,t}^{(j)} = \alpha_i + \beta_i^{(j)} (f_t^{(j)})' + u_{i,t}^{(j)}, \quad \text{for } j = 1, ..., J+1 \text{ and } i = 1, ..., N. \tag{31}$$

Here scale $J+1$ represents the residual component, such that $f_t^{(J+1)} = f_t - \sum_{j=1}^{J} f_t^{(j)}$. Now, the covariance matrices for scale j, denoted as $V^{(j)}$, are calculated as the $M \times M$ covariance matrices of $f_t^{(j)}$, the factor scale components at scale j. The matrices $B^{(j)}$, that are $N \times M$ matrices of the betas for scale j, and the $N \times N$ matrices $D^{(j)}$ are calculated as

$$B^{(j)} = (\beta_1^{(j)\prime}, ..., \beta_N^{(j)\prime})', \quad D^{(j)} = \mathbb{E}[u_t^{(j)\prime} u_t^{(j)}]. \tag{32}$$

The covariance matrix $\Sigma_{R,s}$ for the spectral factor model is calculated by constructing the covariance matrices for each scale and calculate the sum of all scale covariance matrices:

$$\Sigma_{r,s} = \sum_{j=1}^{J+1} \Sigma^{(j)}, \quad \text{where} \quad \Sigma^{(j)} = B^{(j)} V^{(j)} B^{(j)\prime} + D^{(j)}. \tag{33}$$

Bandi et al. [2019] note that $\Sigma_r = \Sigma_{r,s}$ if $\beta_i^{(j)} = \beta_i$ for all j. Therefore, we compare the spectral factor model with the simple factor model as described in Equation 29, to investigate if the systematic risk differs across the components and the decomposition better explains the returns and factors. Also, we compare the results with the Moving Window method. For both the MW and the simple factor model, we calculate the optimal portfolio weights by GMV and the rescaled tangency. Again, the expected returns are estimated as the mean of a moving window and the $1/N$ portfolio is used as a benchmark. We also use the mean return, Sharpe Ratio, Sortino Ratio, 5% Value-at-Risk and variance as performance measures to evaluate the results.

3 Data

The first data set we use concerns treasury yield curve estimates of the Federal Reserve Board, that is constructed by Gürkaynak et al. [2007]. This data set contains end-of-month data from January 1962 until December 2016, that we retrieved from the Supplementary Material of Ortu et al. [2020]. We use the period until September 1971 to initialize the components and obtain the first innovation. Therefore, we restrict the regression data sample to the period from September 1971 until December 2016. The yield-to-maturities in the data are calculated for 1 to 120 months in the future. In our application, we use the maturities for 12, 24, 36, 48, 60, 84 and 120 months.

In line with Ortu et al. [2020], we further make use of daily realized volatility weekday data from the USD/CHF exchange rate, constructed by Corsi [2009]. This data set is constructed from tick-by-tick data from December 1989 to December 2003 and has been made available by Ortu et al. [2020] in their Supplementary Material. The realized volatility is constructed as described in Andersen et al. [2003], based on time intervals of two hours. The realized volatility in the weekend is excluded to avoid seasonal behaviour, which results in 3599 observations.

To investigate the performance of the forecast combination model for high-frequency RV measures. For this purpose, we use daily weekday data of stock indices, containing returns and 10-minute RV from a few markets, that we retrieved from Heber et al. [2009]. Again, the realized volatility is calculated as in Andersen et al. [2003], with a higher amount of time intervals. To test the forecast combination on high-frequency realized volatility data, we use daily 10-minute RV of the AEX from January 2000 until May 2020, and annualize this data by $\sqrt{\text{RV} \cdot 252}$. We compare the RVs of the USD/CHF exchange rate and the AEX index in Table 1. The RVs of USD/CHF and AEX have a similar mean and median, but the standard deviation and the maximum of AEX are more than twice as high as the values of USD/CHF. Therefore, we conclude that the RV of the AEX index, based on a higher frequency, is more volatile. We investigate if this influences the forecasting performance.

Type of RV	Obs.	Mean	Median	St. Dev.	Max.	Skewness	Kurtosis
USD/CHF	3599	12.39	11.59	3.90	41.95	1.59	8.07
AEX	5186	14.47	11.80	9.53	95.28	2.71	14.44

Table 1: Descriptive statistics of the RV of the USD/CHF exchange rate and the AEX index

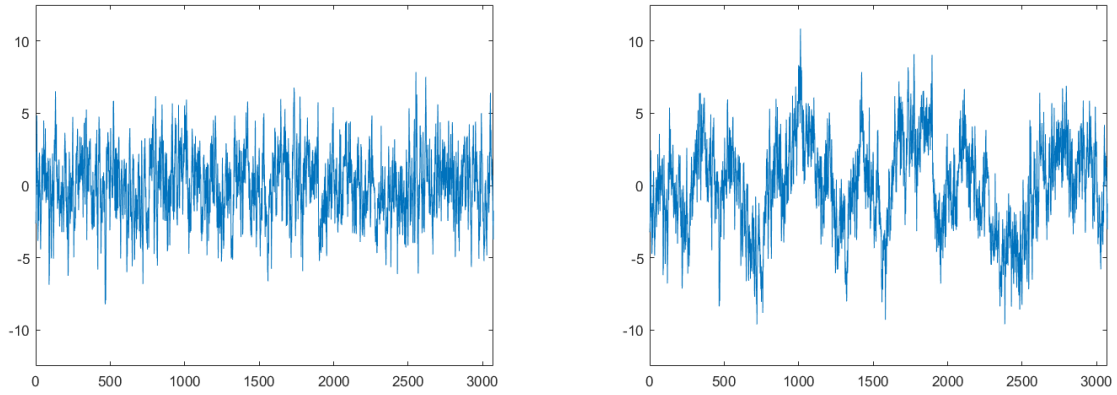
We also use the data set made by Heber et al. [2009] to construct a portfolio. We consider the close-to-close daily return and the 10-minute RV of a few stock indices, in the period from January 2000 until February 2020. We use the following stock indices: All Ordinaries, CAC 40, DAX, Dow Jones and Nasdaq. As risk-free rate for performance measures, we use in line with Fama and French [2015] the 1-month treasury bill rate, converted to daily rates, retrieved from the website of Kenneth French at Dartmouth. For the factor model, we use the close-to-close returns of Heber et al. [2009]. We model these indices by the five-factor model of Fama and French [2015], containing the market portfolio (Mkt-RF), Small-minus-Big (SMB), High-minus-Low (HML), Robust-minus-Weak (RMW) and Conservative-minus-Aggressive (CMA). We use daily data for the appropriate period.

4 Results

4.1 Simulation

We start by presenting the results for the simulation described in Subsubsection 2.2.1. In line with Ortu et al. [2020], we choose the parameter values $\rho_x = 0.7$, $\rho_y = 0.9$ and $J = 3$. We simulate the processes x_t and z_t and plot them in Figure 1. The process x_t is a simple autoregressive model, which is shown in panel (a) of Figure 1. However, we see in panel (b) for the process z_t, where a persistent component has been added, that the process looks like a mean reverting process.

Figure A.1 in the Appendix shows the autocorrelation function of process z_t, as described in Equation 11. This function is for some parts approximated by AR(1) processes for values of the parameter $\phi = 0.9, 0.97, 0.98$.



(a) (b)

Figure 1: Comparison of the simulated paths for the basic autoregressive process x_t in panel (a) and the process z_t in panel (b).

4.2 Yield-to-maturity data

In this application, we use the extended Wold decomposition to predict bond returns. We follow Ortu et al. [2020] and fit a VAR(p) model on the yield-to-maturity data, where we consider the yields $y_t^{(n)}$ for the maturities of $n = 1, 2, 3, 4, 5, 7, 10$ years. Using the methodology described in Subsection 2.1 for multivariate processes, we obtain the scale-specific innovations, Wold coefficients and Wold yield components. We choose $p = 24$ and take a maximum scale of $J = 6$. We calculate the level of the yield curve as $L = \sum_{n>1} y_t^{(n)}/6$ and we compute the slope as $S = L - y_t^{(1)}$.

We follow Ortu et al. [2020] and regress the average excess bond returns on the level and its components, the slope and its components and some other predictors. We report the parameter estimations, Newey-West corrected t-statistics and R^2 values of the regressions. We remark that a constant always is included in the regression. However, as these coefficients are insignificant for all regressions (as expected), we decide to exclude them from the tables.

Table 2 shows the results for the regression on the level L, the level components $L^{(i)}, i = 1, ..., 6$ and the residual component $\pi^{(6)}$. Model a) shows that the level is an insignificant predictor for bond returns. However, model b) shows that the PWD detects multiple cycles in the level that have significant power to predict bond returns. The results for model c) show that ignoring the residual component does not affect the results significantly. Model d) is constructed by dropping insignificant components. In model e), we regress the bond returns on one single level cycle, that is calculated as $L^{(1)} + L^{(2)} + L^{(3)} + L^{(5)}$, which turns out to produce almost the same result.

The results for the regression on the slope S, the slope components $S^{(i)}, i = 1, ..., 6$ and the corresponding residual component $\pi^{(6)}$ are shown in Table 3. Model a) is the regression of the average bond returns on the slope in model a). According to the t-statistic, the slope has significant predictive power for the bond returns. Model b) shows that the slope contains cycles, which results in an increase of the R^2 in comparison to model a). However, only the t-statistics of scale 5 and 6 are significant, so the scale components contain information on the long term, which is remarkably different from the level. Using only scales 5 and 6 in a regression, we obtain the surprising result that the R^2 even rises by 0.01. Again, taking scales 5 and 6 in a single slope cycle does not lead to a large accuracy decrease.

Table 4 compares the level and slope factors for the PWD with other factors that

	Predictors								Wold Level	R^2
	L	$L^{(1)}$	$L^{(2)}$	$L^{(3)}$	$L^{(4)}$	$L^{(5)}$	$L^{(6)}$	$\pi^{(6)}$		
a)	0.40									0.06
	(1.46)									
b)		3.40	3.57	4.21	1.66	5.91	-1.49	0.29		0.29
		(5.53)	(4.58)	(4.50)	(1.07)	(5.55)	(-1.03)	(1.10)		
c)		3.38	3.61	4.27	1.64	5.92	-1.15			0.27
		(5.45)	(4.39)	(4.20)	(1.04)	(5.28)	(-0.75)			
d)		3.37	3.54	4.18		5.86				0.26
		(5.33)	(4.13)	(4.06)		(5.36)				
e)									5.16	0.25
									(6.49)	

Table 2: Regression of average excess bond returns to the Level factor and its component. The Newey-West t-statistics are reported between brackets.

	Predictors								Wold Slope	R^2
	S	$S^{(1)}$	$S^{(2)}$	$S^{(3)}$	$S^{(4)}$	$S^{(5)}$	$S^{(6)}$	$\pi^{(6)}$		
a)	2.09									0.10
	(2.67)									
b)		-0.99	-1.12	-0.87	-0.53	5.04	2.49	0.10		0.15
		(-0.71)	(-0.69)	(-0.42)	(0.33)	(3.00)	(2.65)	(0.05)		
c)		-0.99	-1.12	-0.87	-0.53	5.04	2.49			0.15
		(-0.71)	(-0.69)	(-0.42)	(-0.33)	(3.00)	(2.63)			
d)						5.01	2.48			0.16
						(2.96)	(2.64)			
e)									3.02	0.14
									(3.11)	

Table 3: Regression of average excess bond returns to the Slope factor and its component. The Newey-West t-statistics are reported between brackets.

have been proposed in earlier literature to estimate bond returns. Model a) shows the model where the Wold level from Table 2, model e) and the Wold slope from Table 3, model e) are used. We observe that the coefficients in the multiple regression just decline a little. The R^2 for the multiple regression is approximately equal to the sum of the R^2s from the single models. Therefore, the level and slope seem to be almost orthogonal. Model b) shows the performance of the model proposed by Cochrane and Piazzesi [2005], where we regress the bond returns on a linear combination of the forward rates. The R^2 of their model is clearly lower, so the PWD fits the data better. Model c) is related to the level factor of Cieslak and Povala [2015], where we remove the inflation trend from the level curve. However, model d) shows that after adding the Wold level factor, the level factor of Cieslak and Povala [2015] becomes insignificant. Model e) uses the level factor and slope factor from Cieslak and Povala [2015]. Despite a different built-up of the models a) and e), as the PWD does not use long term inflation, the models have equal fit according to the R^2.

	Wold factors		Ci&Po (2015)		Co&Pi (2005)					R^2
	Level	Slope	Level	Slope	f_t^1	f_t^2	f_t^3	f_t^4	f_t^5	
a)	5.04	2.89								0.38
	(6.25)	(3.35)								
b)					-3.35	6.69	-6.15	-1.27	4.88	0.26
					(-3.44)	(0.90)	(-0.27)	(-0.04)	(0.37)	
c)			1.41							0.10
			(3.45)							
d)	4.61		0.56							0.27
	(5.56)		(1.56)							
e)			2.72	4.06						0.38
			(6.50)	(6.45)						

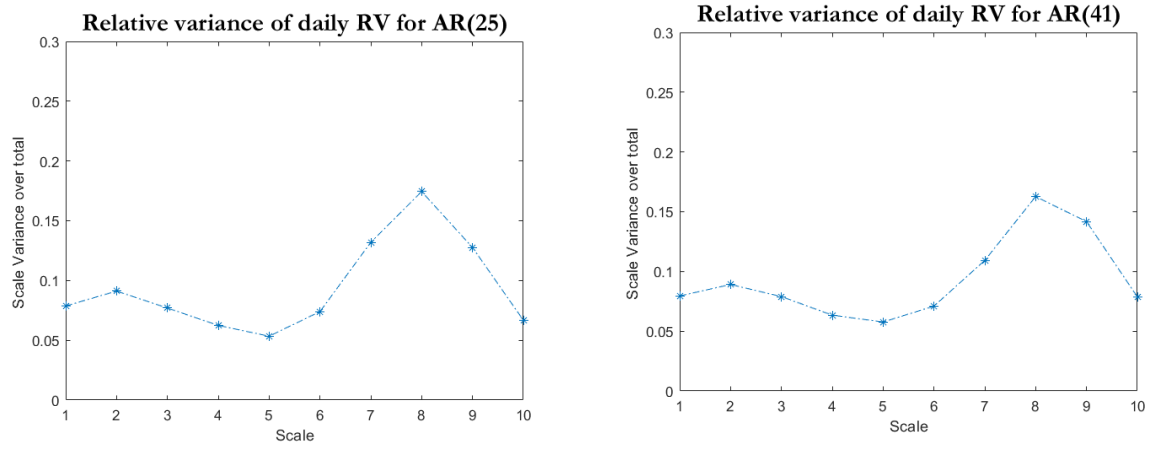
Table 4: Comparison of the PWD with other bond return estimation methods

4.3 Realized volatility of USD/CHF exchange rate

4.3.1 Decomposition and forecasting

In this subsection, we analyze the realized volatility of the daily USD/CHF exchange rate and start by estimating an autoregressive model. The BIC that is used by Ortu et al. [2020] suggests an AR(25) model. Using the corrected BIC for the number of parameters, we obtain an AR(10) model. The AIC suggests an AR(41) model.

We compute the innovations, scale-specific coefficients and components as described in Section 2.1. For the maximum scale, we choose $J = 10$. We compute the explained variance per scale j, given by $\sum_k (\beta_k^{(j)})^2$, and calculate the fraction in terms of the total sample variance. Figure 2 shows the results for AR(25) and AR(41). We observe that for AR(25) the variance is mainly explained by scales $j = 7, 8, 9$, that correspond to shocks that involve up to 512 workdays. The relative variance per scale for AR(41) gives the same conclusion. The results for AR(10) are shown in Figure A.2 in the Appendix. In this model, the variance is mainly explained by the scales $j = 6, 7, 8$.



(a) (b)

Figure 2: Variance ratios of the total variance that is explained by each scale for the PWD, based on an AR(25) model for panel (a) and on an AR(41) model for panel (b).

We forecast the realized volatility as described in Subsection 2.3, deviating a little of this methodology. We start by regressing the RV on the component that explains the largest part of the variance, and forecast the RVs with one component. Next, we add one component at a time, again based on the explained variance. We end up with results from models with 1 to 10 components and calculate some performance measures; the Root Mean Squared Error (RMSE), Mean Absolute Error (MAE) and the R^2 of a Mincer-Zarnowitz regression. The results for AR(25) and AR(41) are shown in Figure 3. For the AR(25) model, we conclude from all three performance measures that the prediction does not improve significantly after including the first three scales, $j = 7, 8, 9$. For the AR(41) model, we observe that especially the scales $j = 8, 9$ improve the predictions and the contribution of scale $j = 7$ is relatively small. The forecast error measures for the AR(10) are shown in Figure A.2 in the Appendix. For this model we observe for the first, second and sixth added scale a big increase in the performance of the models, that correspond to the scales $j = 7, 8, 9$. This indicates that the decompositions based on the AR(25) and AR(41) models make more sense in practice. Therefore, we use the model with scales $j = 7, 8, 9$ to construct forecasts.

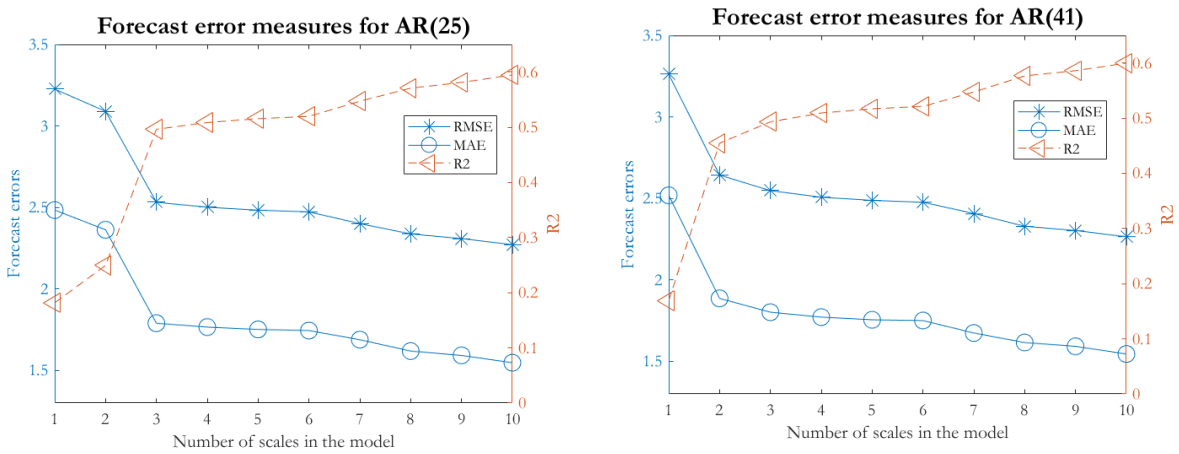


Figure 3: Panel (a) shows the performance measures for PWD on an AR(25) model, panel (b) for PWD on an AR(41) model. RMSE denotes the Root Mean Squared Error, MAE the Mean Absolute Error and R^2 is obtained from a Mincer-Zarnowitz regression.

Table 5 evaluates the performance of the PWD forecasts for 3 scales by an AR(25) model and compares the results with the HAR-RV model by Corsi [2009]. The forecasts are daily re-estimated on a moving window of 2600 observations and are compared over two different time horizons: 1 day and 3 months. The performance of the PWD forecasts with 9 scales is evaluated. Scale 10 is excluded, because this scale requires a very large estimation window. With 2600 observations, we can construct 4 coefficients for scale $j = 9$ and we can build 528 realizations of $g_t^{(j)}$. We see that the PWD with 3 scale components in the long term performs approximately equal to HAR-RV for the performance measures and the PWD with 9 scale components shows a clearly higher R^2.

To investigate the influence of the choice of autoregressive lags, we forecast the exchange rate using $m = 1, ..., 50$ autoregressive lags and calculate the performance measures. Again, we use the time horizons of $p = 1, 66$ days and scales $j = 7, 8, 9$. Figure 4 shows the results. The performance measures indicate that estimations for a larger time horizon are improved by a higher autoregressive level. For $p = 1$, the forecasts are approximately constant and high for the levels 15 up to 24. Therefore, the BIC used by Ortu et al. [2020] seems to be

Panel A: 1-day-ahead

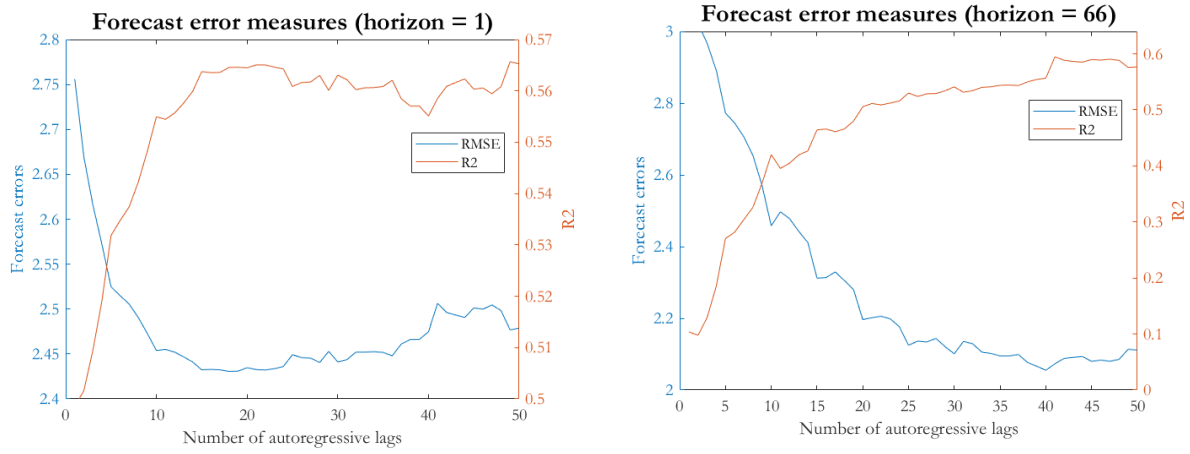
	RMSE	MAE	R^2
HAR-RV	**2.144**	**1.548**	**0.658**
PWD (3 components)	2.449	1.873	0.561
PWD (9 components)	2.300	1.705	0.627

Panel B: 66-day-ahead

	RMSE	MAE	R^2
HAR-RV	2.110	1.693	0.523
PWD (3 components)	2.125	1.646	0.530
PWD (9 components)	**2.106**	**1.620**	**0.602**

Table 5: Forecasting performance of the HAR-RV and PWD with 3 and 9 components, based on an AR(25) model. RMSE denotes the Root Mean Squared Error, MAE denotes the Mean Absolute Error and the R^2 is obtained from a Mincer-Zarnowitz regression.

a suitable criterion for 1-day-ahead prediction. The forecasts for $p = 66$ are optimal around the autoregressive level of 41. This level was suggested by the AIC, which indicates that the AIC gives better estimation results for long-term prediction. These results support the earlier statement that the AR(25) and AR(41) model are more appropriate than the AR(10) model. In our next applications, we use therefore the BIC for $p = 1$ and the AIC for $p = 66$ to construct the coefficients for the decomposition.



(a) (b)

Figure 4: Panel (a) shows the performance measures for 1-day-ahead forecasts for the number of autoregressive lags. Panel (b) shows the performance measures for 66-day-ahead forecasts. For clarity, the MAE is excluded from this graph, as it follows the same pattern as the RMSE.

We conclude that our results validate the findings of Ortu et al. [2020] for the AR(25) model. In Figure 4, we observe that the results of the PWD for 3 components are also accurate for 1-day-ahead forecasting. However, we observe for 66-day-ahead forecasts that we can improve the performance measures. Choosing an AR(41) model by AIC, the PWD with 3 components returns better results than the HAR-RV model, especially for the R^2.

4.3.2 Forecast combination

We start the forecast combination by constructing p-ahead forecasts by PWD and HAR-RV for the RV of the USD/CHF exchange rate. For the comparison step, we use a moving window of 2500 observations. We choose maximum scale $J = 9$ to construct forecasts, such that we can estimate 4 coefficients $\beta_k^{j)}$ for $j = 9$ for the PWD. We estimate $529 - p$ daily realized volatilities for the moving window with HAR-RV and PWD. According to those estimations, we construct weights to estimate the second estimation sample. This sample includes $528 - p$ estimations, again obtained by a moving window of 2500 observations. We use for PWD an AR(25) model for the 1-day-ahead estimations and an AR(41) model for the 66-day-ahead estimations.

The results are shown in Table 6. To compare the models for the 66-day-ahead forecasts, we add two other measures: the Mean Error (ME) and Error Variance (EV). We observe that for the 1-day-ahead forecasts, the PWD is not able to improve the results of the HAR-RV model. The forecasts of PWD have higher bias and are more volatile. For the 66-day-ahead forecasts, only the R^2 is slightly increasing, which indicates a little improvement. According to the ME and EV, this comes from a decreasing error variance. However, the improvements are not clear, as the RMSE and MAE do not show improvements.

Panel A: 1-day-ahead ($\omega_{PWD} = 0.919$, $\omega_{HAR} = 0.081$)

	RMSE	MAE	R^2
HAR-RV	**1.816**	**1.376**	**0.763**
PWD (3 components)	2.342	1.853	0.669
Combined model	2.279	1.796	0.686

Panel B: 66-day-ahead ($\omega_{PWD} = 0.465$, $\omega_{HAR} = 0.535$)

	RMSE	MAE	R^2	ME	EV
HAR-RV	**2.655**	**2.280**	0.666	**-1.983**	3.123
PWD (3 components)	2.771	2.350	0.665	-2.140	3.107
Combined model	2.700	2.309	**0.679**	-2.056	**3.070**

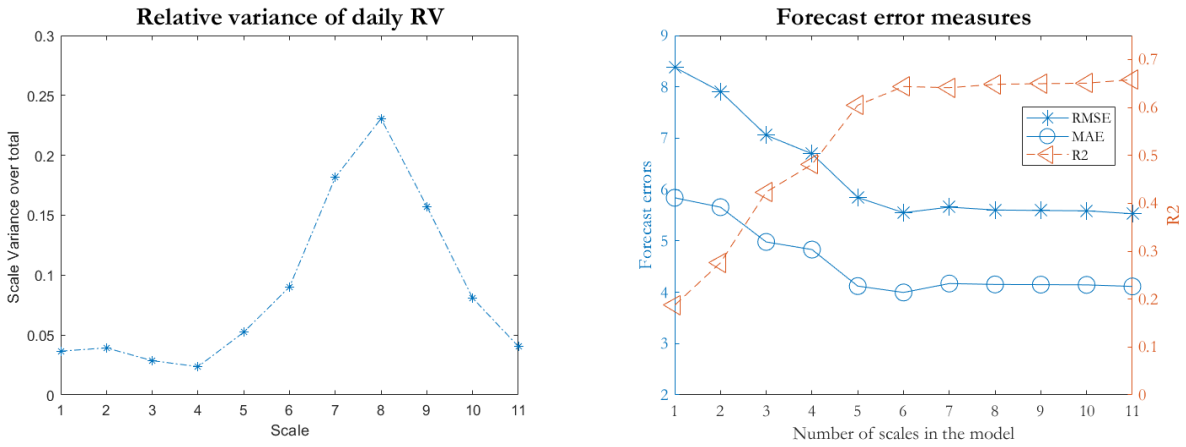
Table 6: Performance for the USD/CHF exchange rate RV of the forecast combination of PWD and HAR-RV, compared with the forecasts made by the single models.

4.4 Realized volatility of AEX index

4.4.1 Decomposition and forecasting

We apply the PWD for the realized volatility of the AEX index, for which the 10-minute RV is reported. The BIC used by Ortu et al. [2020] suggests an AR(20) model, while the AIC suggests an AR(36) model. We compute the innovations and the scale-specific coefficients and components. As our data set is larger, we can choose $J = 11$ as maximum scale. The explained variance per scale j is shown in Figure 5. We observe that the variance is especially explained in the scales $j = 6, ..., 10$. Again, we add the scales one at a time and calculate each time performance measures. We observe that the performance measures do not improve very much after having added 5 scales. Therefore, we prefer a model including the scales $j = 6, ..., 10$. However, choosing maximum scale $J = 10$ for forecasting, provided that we can estimate 4 coefficients $\beta_k^{(j)}$, requires at least 4096 observations and, in addition, also about 500 realizations. As this window would be large in comparison to the estimation sample and we need even more

observations for forecast combination, we choose to drop scale $j = 10$ for forecasting.



(a) (b)

Figure 5: Panel (a) shows the percentage of the variance explained by each scale for the daily 10-minute RV of the AEX index. Panel (b) shows the performance measures for all numbers of scales in the model. These figures are based on an AR(20) model.

Table 7 compares the performance of the HAR-RV model to the forecasts of the PWD with 4 and 9 scales included. The forecasts are estimated on a moving window of 4100 observations. The full data set contains 5186 observations, where the observations from March up to May 2020 concern data of the starting coronavirus recession. We observe that the 1-day-ahead forecasts are not influenced, but the 66-day-ahead forecasts show a clear decrease in performance. We conclude that the HAR-RV and PWD at least for long-term estimations are sensitive for large changes in realized volatility. The forecasts for 5130 observations show results that are, especially for 66-day-ahead estimations, easier to compare. Therefore, we decide to use 5130 observations for the forecast combination and leave the performance of RV prediction in a recent crisis for further research. For the data set of 5130 observations, we again observe

Panel A: 1-day-ahead

	$N = 5130$			$N = 5186$		
	RMSE	MAE	R^2	RMSE	MAE	R^2
HAR-RV	**3.280**	**2.280**	**0.361**	**4.489**	**2.681**	**0.705**
PWD (4 components)	4.092	3.148	0.138	5.719	3.720	0.533
PWD (9 components)	4.022	3.102	0.176	5.332	3.503	0.596

Panel B: 66-day-ahead

	$N = 5130$			$N = 5186$		
	RMSE	MAE	R^2	RMSE	MAE	R^2
HAR-RV	2.764	2.325	**0.266**	**4.139**	2.816	**0.022**
PWD (4 components)	2.763	2.308	0.112	4.261	2.831	0.0001
PWD (9 components)	**2.717**	**2.277**	0.140	4.241	**2.803**	0.0004

Table 7: Forecasting performance of the HAR-RV and PWD with 4 and 9 components for the AEX index realized volatility. Both the results for the full sample and the restricted sample of 5130 observations are reported.

that the RMSE and MAE differ for 1-day-ahead forecasts and the models converge to each other in the long term in performance. The R^2 for the 66-day-ahead forecasts is for this data set still low, which differs from the analysis of the USD/CHF exchange rate. This shows that the high-frequency RV is more difficult to forecast. This can be explained by Table 1 in Section 3, where was shown that the RV of the AEX index is more volatile.

4.4.2 Forecast combination

We investigate if forecast combinations of HAR-RV and PWD are useful for high-frequency RV. We use a moving window of 3400 observations for the comparison step and choose maximum scale $J = 9$ to construct forecasts. In the first step, we estimate $829 - p$ daily realized volatilities with HAR-RV and PWD. The second estimation sample includes $859 - p$ estimations, where again a moving window of 3400 observations is used. We use an AR(20) model for the 1-day-ahead estimations and an AR(36) model for the 66-day-ahead estimations.

Table 8 shows the results. For 1-day-ahead estimations, we again observe no improvements from the HAR-RV model, as PWD performs clearly worse than HAR-RV. However, for 66-day-ahead estimations, we observe again that the MAE and R^2 are slightly improved. The ME and EV show again that these improvements can be explained by the decreased error variance. We conclude that forecast combination does not solve bias, but is able to decrease the prediction error variance. This is in line with the analysis of the 2-hours RV from the USD/CHF exchange rate. In that sense, the frequency of the realized volatility does not affect the conclusions.

Panel A: 1-day-ahead ($\omega_{PWD} = 0.864$, $\omega_{HAR} = 0.136$)

	RMSE	MAE	R^2
HAR-RV	**2.984**	**2.160**	**0.314**
PWD (4 components)	3.926	3.172	0.141
Combined model	3.690	2.949	0.173

Panel B: 66-day-ahead ($\omega_{PWD} = 0.830$, $\omega_{HAR} = 0.167$)

	RMSE	MAE	R^2	ME	EV
HAR-RV	**2.592**	2.235	0.121	**1.911**	3.070
PWD (4 components)	2.794	2.225	0.209	2.065	3.548
Combined model	2.672	**2.171**	**0.224**	2.038	**2.985**

Table 8: Performance for the AEX index RV of forecast combination of PWD and HAR-RV, compared with the forecasts made by the single models. ME denotes the mean error, EV denotes the error variance.

4.5 Portfolio selection

4.5.1 Realized volatility forecasting

For the portfolio selection, we use the close-to-close returns and 10-minute RVs of five stock market indices. We apply PWD for the stock indices mentioned in Section 3. We choose a maximum scale of $J = 9$ and use a moving window of 3400 observations. For the MW estimation of the mean returns and covariance matrix of the returns, we use a moving window of 500 days. The RVs of the stock indices are estimated over a horizon of 1 day. We construct portfolios for 1500 days.

The results are presented in Table 9. We observe that the RV forecasts by the PWD improve the results of global minimum variance portfolios, such that it outperforms the other portfolios for the mean return, the Sharpe Ratio and the Sortino Ratio. The VaR 5% for the PWD-GMV portfolio is lower than the values of the MW-GMV and $1/N$ portfolios, which indicates that the risk also increases slightly. The mean daily return of 0.046% is rather good, as this corresponds to an annual return of 12.29%. The tangency portfolio is slightly improved by the PWD, as the mean return, Sharpe Ratio and Sortino Ratio increase. However, the risk also increases, according to the Value-at-Risk and variance. Also, the PWD-Tangency portfolio still does not outperform the $1/N$ portfolio. We conclude that the use of realized volatility forecasts mainly improves the GMV portfolio returns, while keeping the risk on a reasonably good level. In addition, it is also a consistent method, as it beats the $1/N$ portfolio on different subsamples (excluded for the Value-at-Risk and variance). For the tangency portfolio, we observe some improvements, but they are less clear.

Portfolio	Mean return (%)	Sharpe Ratio	Sortino Ratio	Value-at-Risk	Variance
PWD-Tangency	0.030	0.474	0.607	-1.744	1.012
PWD-GMV	**0.046**	**0.863**	**1.159**	-1.444	0.704
MW-Tangency	0.024	0.441	0.552	-1.488	0.754
MW-GMV	0.018	0.491	0.655	**-0.945**	**0.345**
$1/N$	0.030	0.648	0.829	-1.245	0.534

Table 9: Performance measures for portfolios of PWD, Moving Window and $1/N$, based on RV forecasts. For the mean return, Value-at-Risk and the variance, the daily returns in percents are used. The Sharpe Ratio and Sortino Ratio are annualized by multiplication with $\sqrt{252}$.

4.5.2 Spectral factor model

For the prediction of the spectral factor model, we first determine the scales of both the stock index returns and the five Fama-French factors, using in line with Bandi et al. [2019] as maximum scale $J = 6$. To determine the scale components of the stock indices and factors, we use a VAR(p) model with level $p = 9$, where we note that the results are robust for choices of higher p. We use a moving window of 700 observations to estimate the index returns and the covariance matrix. From these estimations, we determine the portfolio weights and re-estimate them for computational reasons every year (i.e. every 252 days). In addition, we use the first 100 observation days to initialize the scale components. We thus construct portfolios for 4075 observations. Again, we estimate the $1/N$ portfolio as a benchmark. Also, we compare the performance of the model with moving window portfolios and investigate the systematic risk across portfolios by estimating the simple factor model, using the same moving window. We use as well for these models a moving window of 700 observations, use the first 100 observations for initialization and re-estimate the weights every year.

The results are shown in Table 10. We observe that the GMV portfolios have approximately equal performance. For the comparison between the spectral factor model and the simple factor model, this means that the systematic risk does not differ significantly across components. Therefore, we conclude that the risk and dependence between stock indices and that the variance is not explained better by larger time spans of the returns. In comparison with the $1/N$ portfolio, we see that the minimum variance portfolios succeed in keeping the risk low, as the GMV portfolios have a higher Value-at-Risk than the $1/N$ portfolio. However, the GMV portfolios suffer from a low expected return. Therefore, the Sharpe Ratio and Sortino Ratio are lower for the global minimum variance portfolios. For the tangency portfolios, we also

observe that the difference between the spectral factor model and the simple factor model is relatively small. Interestingly, the moving window method for the tangency scores even better and outperforms all other methods for the mean return, the Sharpe Ratio and the Sortino Ratio. The mean return of 0.042% corresponds to an annualized return of 11,16%.

Portfolio	Mean return (%)	Sharpe Ratio	Sortino Ratio	Value-at-Risk	Variance
SFM-Tangency	0.032	0.474	0.628	-1.761	1.168
SFM-GMV	0.018	0.384	0.485	-1.184	**0.582**
MW-Tangency	**0.042**	**0.600**	**0.803**	-1.770	1.260
MW-GMV	0.018	0.370	0.471	**-1.120**	0.598
Fac-Tangency	0.033	0.467	0.609	-1.811	1.230
Fac-GMV	0.017	0.340	0.430	-1.208	0.625
$1/N$	0.027	0.462	0.593	-1.463	0.849

Table 10: Performance measures for the spectral factor model (SFM), compared with the model where the returns and variance are based on a moving window (MW), the simple factor model (Fac) and the $1/N$ portfolio. For the mean return, Value-at-Risk and the variance, the daily returns in percents are used. The Sharpe Ratio and Sortino Ratio are annualized by multiplication with $\sqrt{252}$.

The performance measures for our data show that the portfolio returns do not improve by allowing different levels of systematic risk over scales. A possible reason is that stock indices are already diversified, such that the systematic risk is already more stable. To get more insight in this result, we regress for the full data set the scales $j = 1, ..., 6$ of the stock index returns on the scales of the factors. These results show indeed that the spectral betas do not show large differences across scales, which supports the conclusion that the risk is equal across the scale components. In general, we observe that the stock indices have a high dependency on the Mkt-RF, except for the first scale components. All other scale components show very similar results. The details of the spectral betas estimations for the scale components are shown in the Appendix in Figure A.3. From both the performance measures and the spectral betas, we conclude that the performance of the portfolios for the used stock index data is not clearly improved by the spectral factor model.

5 Conclusion

In this paper, we investigated the forecasting power of the persistence-based Wold decomposition and explored several forecasting applications. We started by applying the PWD to bond return forecasting and found, in line with the findings of Ortu et al. [2020], that short-term level scales and long-term slope scales have good predicting power for bond returns. Next, we applied the PWD on the realized volatility of the USD/CHF exchange rate and compared the results with the HAR-RV model of Corsi [2009]. Here we observed convergence between the two models for forecasting on a long time horizon. We investigated the robustness of the PWD for the autoregressive model on which the coefficients are based. An important finding is that we can improve predictions for a large time horizon by choosing a high autoregressive level, for which the AIC is a suitable criterion. We also applied the PWD on 10-minute RV data of the AEX index. A remarkable result is that both the HAR-RV and PWD models show a clear decrease in forecasting performance for the recent coronavirus recession. For further research, it is useful to investigate methods to mitigate this problem.

Next, we made forecast combinations between the PWD and HAR-RV model. To investigate if the performance of RV forecasting depends on frequency, we applied the PWD

on on 2-hour realized RV data and 10-minute RV data. We concluded for both data sets that forecast combinations of the PWD and HAR-RV for short-term forecasting do not improve the results, mainly because the PWD performs poor in comparison to the HAR-RV model, having a larger bias and more error variance. For further research of short-term forecasting, it is useful to combine two forecasting methods with approximately equal predicting power. For the 66-day-ahead estimations, we observe slight improvements for the error variance, but those improvements are currently not clearly better. Again, combining forecast methods that have similar performance would shed some light on the value of forecast combinations. In addition, a larger data set of realized volatility would also allow to choose a larger scale, which could capture more information in the model.

Considering the outcomes for the data of 2-hour RV and 10-minute RV, we observe that the used 10-minute RV data is more difficult to predict. However, we found that the performances for the PWD forecasts and the forecast combination still show a similar pattern. Both data sets contain daily reported RVs. For future research, one could investigate how the PWD and forecast combinations perform for predicting realized volatilities that are calculated for a lower frequency, for example, weekly or monthly.

We applied estimations of realized volatility by the PWD to portfolio selection and show that a simple moving window portfolio construction is improved for minimum variance portfolios, so forecasting by the PWD is certainly useful here. For this application, we used stock indices, which are in general relatively stable. An interesting extension is to apply the PWD to the RV of single stocks, to investigate the performance on these assets. We adjusted the tangency portfolios by a simple rescaling method to avoid extreme weights. An alternative way is to solve a quadratic programming problem with restrictions for the weights. Finally, we estimated a factor model with scale components. Here we concluded from the performance measures and spectral betas that the systematic risk is approximately equal across the scales. Also here it would be interesting to investigate the performance of this model for single stocks, for which the risk often more fluctuates over time, such that the systematic risk is more likely to vary across scales.

References

T. G. Andersen, T. Bollerslev, F. X. Diebold, and P. Labys. Modeling and forecasting realized volatility. *Econometrica*, 71(2):579–625, 2003.

F. M. Bandi, S. Chaudhuri, A. W. Lo, and A. Tamoni. Spectral factor models. *Research Paper 18-17*, John Hopkins Carey Business School, 2019.

A. Cieslak and P. Povala. Expected returns in treasury bonds. *The Review of Financial Studies*, 28(10):2859–2901, 2015.

J. H. Cochrane and M. Piazzesi. Bond risk premia. *American Economic Review*, 95(1):138–160, 2005.

F. Corsi. A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, 7(2):174–196, 2009.

M. De Pooter, M. Martens, and D. Van Dijk. Predicting the daily covariance matrix for s&p 100 stocks using intraday data — but which frequency to use? *Econometric Reviews*, 27(1-3): 199–229, 2008.

D. Di Virgilio, F. Ortu, F. Severino, and C. Tebaldi. Optimal asset allocation with heterogeneous persistent shocks and myopic and intertemporal hedging demand. In I. Venezia, editor, *Behavioural Finance: the Coming of Age*, chapter 4, pages 57–108. World Scientific, 2019.

R. Engle. Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroscedasticity models. *Journal of Business and Economic Statistics*, 20(3):339–350, 2002.

E. F. Fama and K. R. French. A five-factor asset pricing model. *Journal of Financial Economics*, 116(1):1–22, 2015.

R. S. Gürkaynak, B. Sack, and J. H. Wright. The us treasury yield curve: 1961 to the present. *Journal of Monetary Economics*, 54(8):2291–2304, 2007.

R. Halbleib and V. Voev. Forecasting covariance matrices: A mixed approach. *Journal of Financial Econometrics*, 14(2):383–417, 2016.

G. Heber, A. Lunde, N. Shephard, and K. K. Sheppard. *Oxford-Man Institute's realized library (0.3)*. Oxford-Man Institute, University of Oxford, 2009.

H. Markowitz. Portfolio selection. *The Journal of Finance*, 7(1):77–91, 1952.

F. Ortu, F. Severino, A. Tamoni, and C. Tebaldi. A persistence-based wold-type decomposition for stationary time series. *Quantitative Economics*, 11(1):203–230, 2020.

A. Timmermann. Forecast combinations. In G. Elliott, C. W. Granger, and A. Timmermann, editors, *Handbook of Economic Forecasting*, chapter 4, pages 135–196. Elsevier B.V., 2006.

H. Wold. *A Study in the Analysis of Stationary Time Series*. Almqvist & Wiksells Boktryckeri, 1938.

A Additional figures

A.1 Autocorrelation function of simulation

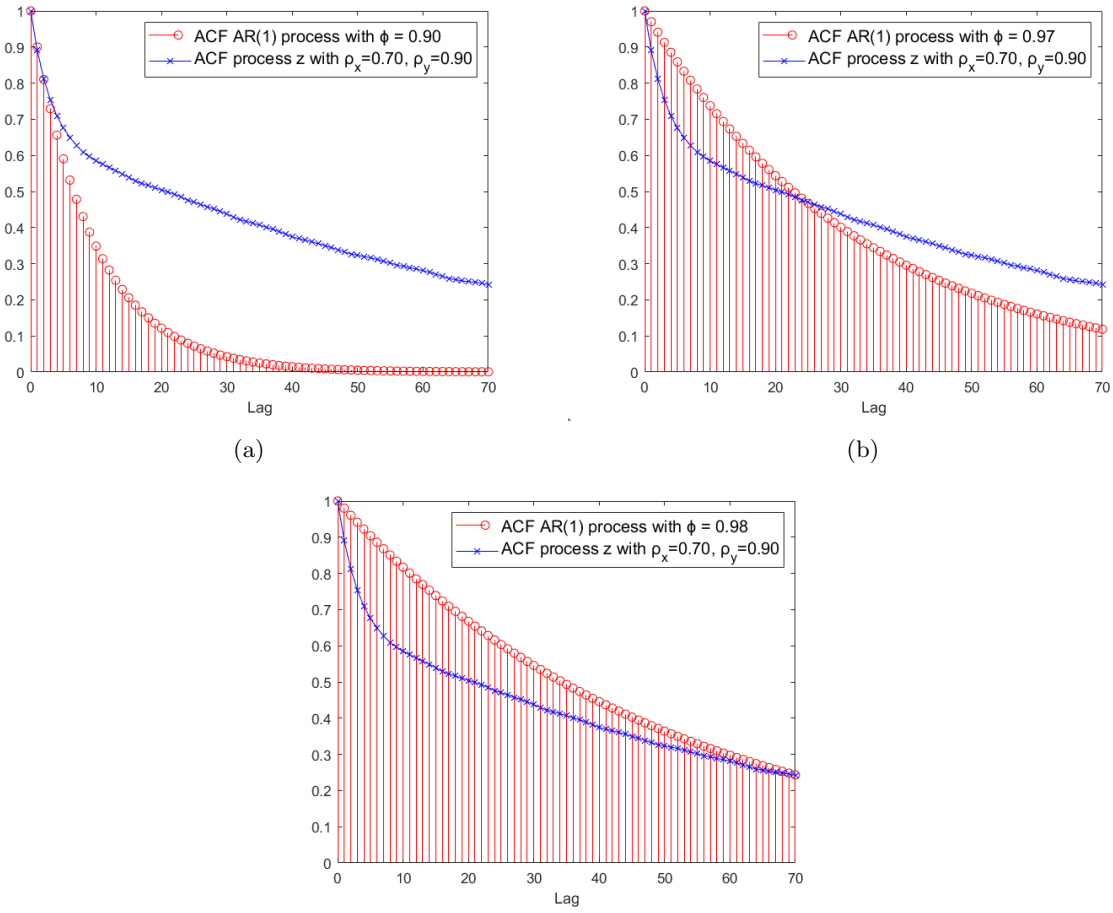
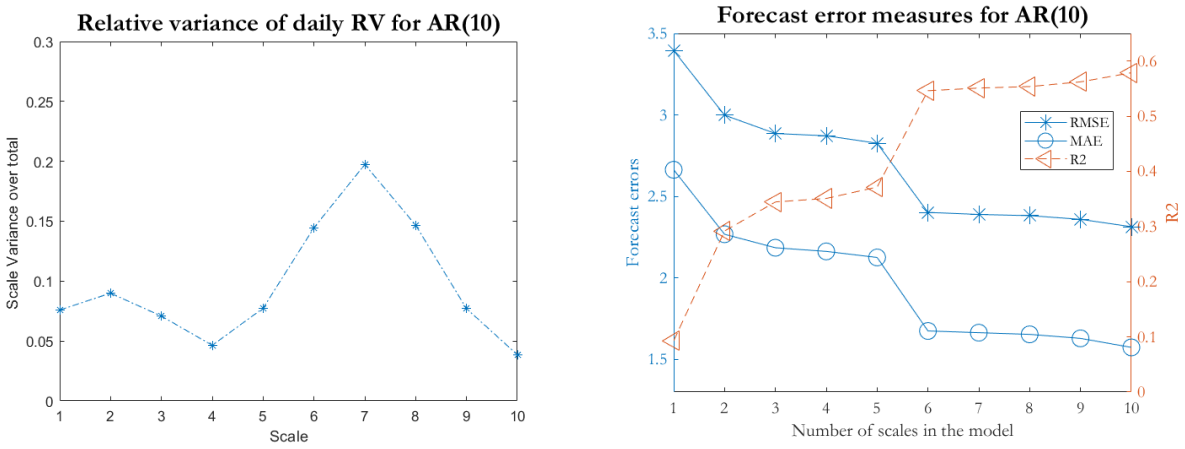


Figure A.1: Comparisons between the autocorrelation function of z_t with $\rho_x = 0.7$, $\rho_y = 0.9$, $J = 3$ and the autocorrelation function of an AR(1) process with parameter $\phi = 0.9$ in panel (a), 0.97 in panel (b) and 0.98 in panel (c).

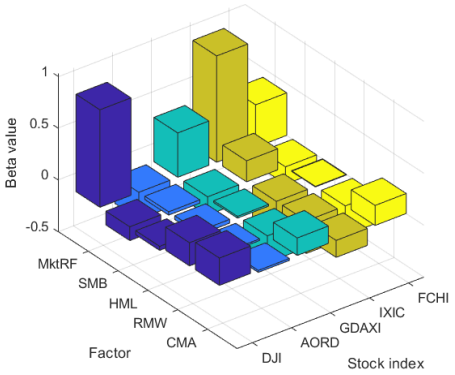
A.2 Results for the USD/CHF realized volatility with AR(10)



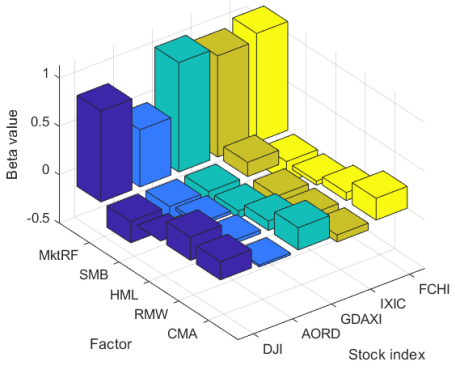
(a) (b)

Figure A.2: Panel (a) shows the relative variance of the daily RV, based on the AR(10) model. Panel (b) shows the performance measures for the PWD on an AR(10) model. RMSE denotes the root mean squared error, MAE the mean absolute error and R^2 is obtained from a Mincer-Zarnowitz regression.

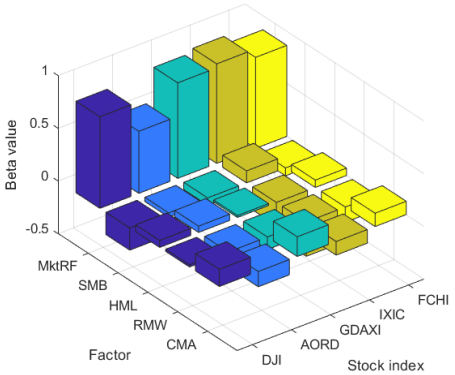
A.3 Spectral betas per scale in the spectral factor model



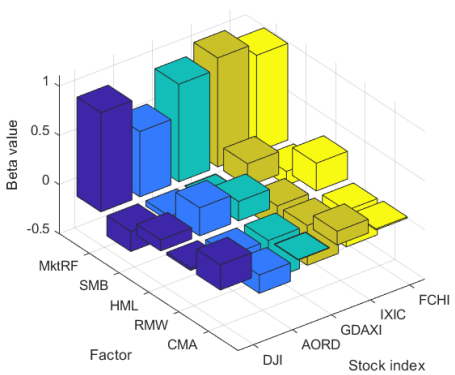
(a) Betas at scale $j = 1$, $\beta_i^{(1)}$.



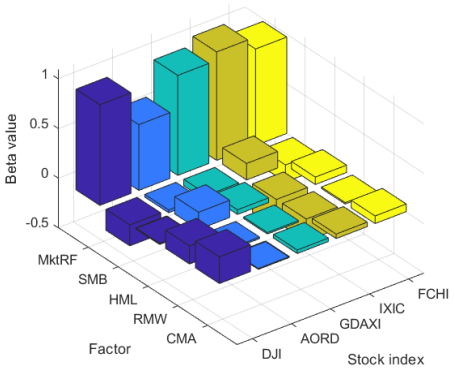
(b) Betas at scale $j = 2$, $\beta_i^{(2)}$.



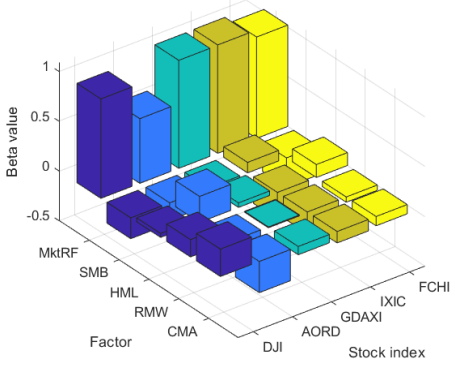
(c) Betas at scale $j = 3$, $\beta_i^{(3)}$.



(d) Betas at scale $j = 4$, $\beta_i^{(4)}$.



(e) Betas at scale $j = 5$, $\beta_i^{(5)}$.



(f) Betas at scale $j = 6$, $\beta_i^{(6)}$.

Figure A.3: Spectral betas for each scale. Each figure shows the betas for the regression where scale j of a certain stock index is regressed on all scales j of the factors.

B Description of used MATLAB code

The following MATLAB files are used for this thesis, that are included in the zip-file.

The folder **1 Simulation** contains the code used for Subsection 4.1, where the simulation of Ortu et al. [2020] is replicated. These files are made by Ortu et al. [2020].

- The file **MainCodeFigure1andA1** performs the simulation and returns Figure 1b and Figure A.1. By adjusting a parameter, Figure 1a is obtained.
- The folder **LibOSTT** contains the used functions in the code of Figure 1 and A.1.

The folder **2 Bond returns** contains the code that produces the results for Subsection 4.2, where the results for the bond returns are replicated. The codes are made by Ortu et al. [2020].

- The file **ReplicationBondReturnsTable2_3_4** returns the output for Table 2, 3 and 4. This code is written by Ortu et al. [2020].
- The main replication code uses the functions of the folder **LibraryWoldComponents**. This folder contains the functions to perform the multivariate form of the PWD.
- The data file **YieldGSWDataSet** contains the yield-to-maturity data that is used.
- The data file **CPoTrendforRebonato** contains the CPI data that is used to construct the model of Cieslak and Povala [2015], where the bond returns are regressed on the inflation.

The folder **3_4 RV Forecasting** contains the code to produce the figures and tables for Subsection 4.3 and 4.4, for the USD/CHF exchange rate RV and AEX index RV.

- The file **CodeTable1** contains the code for Table 1.
- The file **USDCHFCodeFig2_3_A2** produces Figure 2, 3 and A.2. The code is made by Ortu et al. [2020]. We added two alternative criteria to determine the autoregressive level.
- The file **USDCHFCodeTable5** produces Table 5, that replicates the results of the RV forecasting for the USD/CHF exchange rate. This file is made by Ortu et al. [2020].
- The file **USDCHFCodeTable6** contains the code to produce Table 6, where the forecast combination for the USD/CHF exchange rate RV is displayed.
- The file **USDCHFCodeFigure4** produces Figure 4. The first part of the code is from the file that produces Table 5 and is made by Ortu et al. [2020], which is adjusted for the loop over the autoregressive lags. The second part makes the figures to visualize the performance measures.
- The file **AEXCodeFigure5** produces Figure 5, where the relative variance and performance measures per scale for the AEX index RV is displayed. The code is based on the code of **USDCHFCodeFig2_3_A2**, that is made by Ortu et al. [2020] and adjusted for the new data set.
- The file **AEXCodeTable7** produces the estimations of the HAR-RV model and PWD (4 and 9 components). It is based on the code of Table 5.

- The file **AEXCodeTable8** contains the code to produce Table 8, where the performance measures of the forecast combination for the AEX index RV are shown.
- The data files **USD_CHF_RV_89_93** and **AEXindices** contain the realized volatilities of respectively the USD/CHF exchange rate and the AEX index.
- The functions **IRFforecast**, **IRFforecast_horizon** and **IRFscale** are made by Ortu et al. [2020] and are used in the files to construct scales and forecasts for the PWD.
- The following functions are used in the file **USDCHFCodeTable6** and **AEXCodeTable8**:
  - The function **OrtuCorsiWeights** estimates the weights for the HAR-RV and PWD based on the first estimation round.
  - The function **OrtuCorsiWeightsRound2** makes the forecast combination for the second subsample.
  - The function **out_of_sample_forecasting_horizon_rounds_USDCHF** estimates HAR-RV and PWD, for the subsample that is chosen. This function is based on the file made by Ortu et al. [2020] that produces Table 5.
  - The function **out_of_sample_forecasting_horizon_rounds_AEX** is the same function, adjusted for the application on the AEX RV.

The folder **5 Portfolio selection** contains the codes used for the portfolio selections in Subsection 4.5.

- The file **MainPortfolioSelectionTable9** is used to produce Table 9 and selects portfolios based on realized volatility estimates.
- The file **MainBandiTable10_FigA3** produces Table 10 and estimates portfolio weights based on the spectral factor model. It also produces Figure A.3, where the spectral betas for the whole data set are reported.
- The data files **AORDrate**, **DJIrate**, **FCHIrate**, **GDAXIrate** and **IXICrate** are the close-to-close returns of five stock indices, respectively All Ordinaries, Dow Jones, CAC 40, DAX and Nasdaq.
- The data files **FF5factors** and **rf_rate** are the five Fama-French factors and the riskfree rate, retrieved from the website of Kenneth French at Dartmouth.
- The functions **IRFforecast_horizon** and **IRFscale** are made by Ortu et al. [2020] and are used in the files to make scales and forecasts.
- The function **performanceMeasures** calculates some performance measures for the portfolio excess returns and is used in the files that produce Table 9 and 10.
- The function **out_of_sample_forecasting_horizon** is based on the code of Table 5, that is made by Ortu et al. [2020] and is used to predict the RVs to construct portfolios for Table 9.
- The folder **LibraryWoldComponents**, that contains the functions to perform the multivariate form of the PWD, is used to construct the spectral factor model for Table 9.