

Forecasting volatility with macroeconomic and financial variables using Kernel Ridge Regressions

Felix C.A. Mourer
360518

Supervisor: Prof. dr. D.J. van Dijk

Bachelor thesis Econometrics and Operational Research

July 27th 2014

Abstract

This paper uses the nonlinear method kernel ridge regression to forecast volatilities with 38 macroeconomic and financial variables of four different asset classes, i.e. stocks, bonds, commodities and foreign exchanges. Kernels which are used in this paper are the linear, quadratic and the Gaussian kernel. Tuning parameters of this method are estimated in two ways, i.e. with principal components and on the full dataset of variables. Furthermore, the Least Angle Regression method is used to preselect variables, which are then used by the kernel ridge regression. Next to examining this nonlinear method, 3-month, 6-month and 12-month ahead forecasts are made for kernel ridge regressions and numerous of other linear models. Main findings are that kernel ridge regression performs better than simple linear models, and multi-step ahead forecasts with macroeconomic and financial variables are not better than a simple autoregressive model.

Keywords: Volatility, kernel ridge regression, multi-step ahead forecasting, nonlinear forecasting, high dimensionality

Table of Contents

1 Introduction.....	1
2 Data	3
3 Methodology	4
3.1 Kernel ridge regression.....	4
3.1.1 Polynomial kernel.....	6
3.1.2 Gaussian kernel	7
3.1.3 Tuning parameter selection	7
3.2 Longer step ahead forecasts	8
3.3 Statistical evaluation	9
3.4 Economical evaluation	10
4 Results	12
4.1 Kernel ridge regressions	12
4.1.1 Linear kernel.....	12
4.1.2 Quadratic kernel.....	13
4.1.3 Gaussian kernel	14
4.2 Multi-step ahead forecasts.....	16
4.2.1 Three months ahead forecasts.....	16
4.2.2 Six months ahead forecasts.....	17
4.2.3 Twelve months ahead forecasts.....	18
4.3 Economical evaluation	19
5 Conclusion	20
References.....	22
Appendices	24
Appendix A Statistical evaluation KRR.....	24
Appendix B Statistical evaluation multi-step ahead forecasts	28
Appendix C Diebold-Mariano statistics for multi-step ahead forecasts	33
Appendix D Economic evaluation.....	35

1 Introduction

Volatility has been one of the most successful areas of research in time series econometrics and forecasting in recent decades (Andersen, Bollerslev, Christoffersen & Diebold, 2006). Volatility is a statistical measure of the dispersion of returns for a given asset. It refers to the amount of uncertainty or risk about the size of changes in an underlying asset. Volatility has become an indispensable topic in financial markets for risk managers, portfolio managers, investors, academicians and all that have something to do with financial markets (Minkah, 2007). It is an essential input for risk management, asset pricing and portfolio management (Christiansen et al. 2012). Due to the increase in stock market uncertainty and the recent financial crises, there has been a growing interest in volatility as an input in asset allocation to determine an optimal portfolio. Additionally, financial risk management has taken a central role since a specific amendment of the Basel Accord was made in 1996 to allow banks to use their model together with the Value-at-Risk approach for calculating market risk related risk capital (McNeil & Frey, 2000). This amendment effectively makes volatility forecasting a compulsory risk-management exercise for many financial institutions around the world (Granger & Poon, 2003).

Next to the financial applications of volatility, it has wide repercussion on the economy as a whole. According to Granger and Poon (2003), there is clear evidence of an important link between financial market uncertainty and public confidence. For this reason, policy makers often rely on market estimates of volatility as an indicator for the vulnerability of financial markets and the economy.

During the last two decades, nonlinear relations in macroeconomic and financial time series have been getting more attention. However, methods such as regime-switching models and neural networks are only appropriate for a small number of predictors and the improvement over linear forecasting techniques is limited (Stock and Watson, 1999; Medeiros et al., 2006; Teräsvirta et al., 2005). Andersen et al. (2006) give an overview of several time series models for volatility, including GARCH models, stochastic volatility models and realized volatility models. Although these models are discussed in detail, they merely relate the current level of volatility to its past without including any other variables.

Christiansen, Schmeling and Schrimpf (2012) examined 38 macroeconomic variables with the use of forecast combinations and Bayesian Model Averaging. However, Bayesian Model Averaging does not take nonlinearity into account.

Earlier this year, the macroeconomic and financial variables which mainly affect the volatility of four different assets were investigated by Holtrop, Kers, Mourer and Verkuijlen (2014). In their research, the Least Angle Regression (LARS) method was used to preselect the variables and then, with the

most important variables, forecasts were constructed with forecast combinations and factor-based models. However, these forecasting methods are also of a linear type.

In the research of Exterkate, Groenen, Heij and van Dijk (2013), a technique is found that provides better forecasts than the traditional linear and nonlinear method. This technique is called kernel ridge regression (KRR) and can deal with the nonlinear relations between the (realized) volatility and a large number of macroeconomic and financial variables. The technique has been applied to forecast four key measures of real economic activity by Exterkate et al. (2013). Nevertheless, this method has not yet been applied to forecast realized volatility and not on (by LARS) preselected variables.

KRR is partly based on the standard linear ridge regression, which is commonly used in economic forecasting (Kim & Swanson, 2013), since it can be a better method than the standard ordinary least squares procedure, when the number of predictors is relatively larger than the number of observations. However kernel ridge regression uses a trick so that it has less computational drawbacks. The central idea is to use a set of nonlinear prediction functions and prevent overfitting by penalization. The set of predictors are transformed (or mapped) in to a high-dimensional space of nonlinear functions of the predictors. These transformations (or mappings) of the predictors are chosen in an efficient way leading to so-called kernel functions. Different kernel functions are known and the ones which are used in this research are the polynomial and the Gaussian kernel functions.

The goal of this research is to examine if this technique significantly produces more accurate forecasts than the linear type of predictions made by the linear methods of Holtrop, Kers, Mourer and Verkuijlen (2014). For easy comparison, the same dataset from Christiansen et al. (2012) is used which contains 38 macroeconomic and financial variables covering the period from January 1983 to December 2010.

The KRR forecast models are constructed with a moving window of 5 and 10 years, starting with forecasting in January 1993. KRR are performed on the full set of macroeconomic and financial variables and on 18 variables which are preselected by LARS first. In addition, the shrinkage parameters are estimated in two ways; first by regressing the volatility on the first four principal components of the variables and second by regressing it on the complete dataset of variables.

Furthermore, in this research the volatilities will be forecasted not only 1-month ahead, but also on a horizon of 3, 6 and 12 months. Investigations (Chen and Hong (2010)) have shown that on longer horizons returns can be better than on short horizons. If this also holds for volatilities will be checked in this paper. Forecasting over extended horizons will be done for both the models constructed by Holtrop et al. (2014) and the kernel ridge regressions. To avoid having to forecast all macroeconomic

and financial variables, so-called 'direct' forecasts are constructed. With the iterating forecast method we would have to presume that the variables follow a random walk.

The KRR forecasts will be evaluated statistically and economically. For the multi-step ahead forecasts will only a statistical analysis will be given.

In statistical terms can be seen that the kernel ridge regressions beat the linear benchmarks, although the ARX-model with 18 by LARS selected variables is still better. The type of kernel or the method of estimating the tuning parameters doesn't show significant differences, but a 10 year moving window generally provides better forecasts. This could be, because then more information can be taken in to account when estimating the models.

From the economic interpretation can be concluded that the Gaussian kernel using the principal component estimation method for the tuning parameters gives the investor the highest utility, based on her risk-averse behavior.

The statistical evaluation of the multi-step ahead forecasts shows that the macroeconomic and financial variables the forecasts over longer horizons. Better predictions can be made if all these variables are forecasted properly by other models

The research is organized as follows. Chapter 2 describes the data, which are used in this research. In Chapter 3, the methodology will be discussed followed by the results in Chapter 4. Finally in Chapter 5, a conclusion will be given.

2 Data

The data, which will be used in this research are retrieved from the 'short' dataset of Christiansen et al (2012) and consists of 336 monthly observations from January 1983 to December 2010. The dataset contains the volatility of four asset classes; stocks, commodities, bonds and exchange rates. The volatility is defined as the natural log of the square root of the squared daily returns. For the volatility of the asset class of stocks are based on the daily returns of the S&P 500. The 10-year Treasury note futures contract traded on the Chicago Board of Trade (CBOT) and the Standard & Poor's GSCI commodity index are used for the volatilities of the bonds and the commodities respectively. Finally, to determine the volatility of the exchange rates, an equally weighted basket of currencies from 49 countries against the US dollar is formed. With the daily spot rate changes of this aggregate foreign exchange portfolio, the realized volatility is constructed.

Next to these dependent variables, it also contains 38 macroeconomic and financial variables, which are classified in five economic categories. The first category contains equity market variables and risk factors, like the earnings price ratio and the dividend price ratio. In the second category, interest rates, spreads and bond market factors can be found. Examples of variables in this category are the T-bill rate and the Term Spread. The Dollar Risk Factor and the Carry Trade Factor are examples of variables which are included in the third category, foreign exchange variables and risk factors. The fourth category contains liquidity and credit risk variables. In this category, the Default Spread, TED spread and the Foreign Exchange Bid-Ask Spread can be found. The last category includes a large number of macroeconomic variables such as employment growth, industrial production growth and interest rates.

The stationarity of these variables was already investigated by Holtrop et al (2014) and appropriate transformations were performed. Christiansen et al. (2012) already adjusted the relevant variables for seasonality.

3 Methodology

As already stated in the introduction, this research consists of two parts. First, forecasting the volatility with use of the kernel ridge regression and second, performing longer step ahead forecasts for the models constructed by Holtrop et al. (2014) and also for the kernel ridge regressions.

3.1 Kernel ridge regression

This section describes the method of kernel ridge regression, which was proposed by Exterkate et al. (2013) as an approach for forecasting with many predictors that are related nonlinearly to the target variable.

Ordinary least squares (OLS) regression and ridge regression form the basis for the method of kernel ridge regression (KRR). Moreover, KRR is just an ordinary ridge regression on transformed regressors with a kernel trick, which improve computational efficiency. First, an introduction to ridge regression is given. Afterwards, the ridge regression will be extended to kernel ridge regression.

When performing an OLS regression, it is necessary that the number of predictors N is smaller or equal to the number of observations T . In general, N is much smaller than T to avoid overfitting. The in-sample-fit of overfitted models might be good, but forecasting out-of-sample predictions commonly leads to poor results.

Ridge regression tries to find a balance between the goodness-of-fit and the magnitude of the vector with parameters β .

The main difference between the OLS regression and the ridge regression is the penalization of the regression coefficients. The ridge regression estimate β is defined as the value of β that minimizes the ridge criterion

$$\sum_i (y_i - x_i' \beta)^2 + \lambda \sum_{j=1}^p \beta_j^2$$

The penalty term or the shrinkage parameter λ is chosen beforehand.

The solution of the minimization problem is given by:

$$\hat{\beta}_{linear\ ridge} = (X'X + \lambda I)^{-1} X'y$$

As can be seen, if the shrinkage parameter approaches zero, the ridge parameter estimate equals the OLS parameter estimate.

$$\lim_{\lambda \rightarrow 0} \hat{\beta}_{linear\ ridge} = \lim_{\lambda \rightarrow 0} (X'X + \lambda I)^{-1} X'y = (X'X)^{-1} X'y = \hat{\beta}_{OLS}$$

The forecast of the dependent variable is easily computed by

$$\hat{y}_{t+1} = \hat{x}'_{t+1} \hat{\beta}_{linear\ ridge} = \hat{x}'_{t+1} (X'X + \lambda I)^{-1} X'y \quad (1)$$

A great advantage of this forecast is that it can be computed when the number of observations T is smaller than the number of predictors N . However, when N becomes very large, the inversion of the $N \times N$ matrix $X'X + \lambda I$ can lead to computational problems.

To overcome this problem, kernel ridge regression is introduced, which also allows for nonlinear prediction functions. These nonlinear functions are made possible by so-called mappings of the N original predictor variables x in M transformed variables z ($z = \varphi(x)$). All transformed variables are collected in a matrix Z with rows $z_t^T = \varphi(x_t)^T$. In the ridge parameter estimates the X 's are simply replaced by Z 's, leading to the following kernel ridge regression estimate \hat{y}_{ridge} and forecast:

$$\hat{y}_{ridge} = \hat{z}'_{t+1} (Z'Z + \lambda I)^{-1} Z'y$$

$$\hat{y}_{t+1} = \hat{z}'_{t+1} \hat{y}_{ridge} = \hat{z}'_{t+1} (Z'Z + \lambda I)^{-1} Z'y$$

To allow for flexible forms of nonlinearity, the number of transformed predictor variables needs to be larger than the number of original predictor variables. However, calculating the matrix $Z'Z$ can lead to computational difficulties, since this matrix has dimensions $M \times M$ and $M \gg N$.

A solution for this problem is the so-called kernel trick. The basic idea is that, since the number of observations T is smaller than the number of transformed predictor variables M , working with T -

dimensional objects reduces the computational difficulties. This reduction of the dimensions can be shown by algebraic manipulations (Exterkate et al. (2013))

First, the ridge regression estimator is rewritten $\hat{y}_{ridge} = (Z'Z + \lambda I)^{-1}Z'y$ as

$$Z'Z \hat{y} + \lambda \hat{y} = Z'y \quad (2)$$

$$\text{or} \quad \hat{y} = \frac{1}{\lambda} Z'(y - Z\hat{y}).$$

Pre-multiplying (2) by Z gives $ZZ'Z\hat{y} + \lambda Z\hat{y} = ZZ'y$, or

$$Z\hat{y} = (ZZ' + \lambda I)^{-1}ZZ'y.$$

The forecast of the dependent variable can now be written as

$$\begin{aligned} \hat{y}_{t+1} &= \hat{z}'_{t+1} \hat{y}_{ridge} = \frac{1}{\lambda} \hat{z}'_{t+1} Z'(y - Z\hat{y}) = \frac{1}{\lambda} \hat{z}'_{t+1} Z'(y - (ZZ' + \lambda I)^{-1}ZZ'y) \\ &= \frac{1}{\lambda} \hat{z}'_{t+1} Z'(ZZ' + \lambda I)^{-1}(ZZ' + \lambda I - ZZ')y = \hat{z}'_{t+1} Z'(ZZ' + \lambda I)^{-1}y \\ &= \hat{k}'_{t+1}(K + \lambda I)^{-1}y \end{aligned} \quad (3)$$

where $\hat{k}_{t+1} = Z\hat{z}_{t+1}$ is a T x 1 vector and $K = ZZ'$ is called a T x T kernel matrix.

The (p,q) -th element of K equals $z'_p z_q = \varphi(x_p)' \varphi(x_q)$ and the q -th element of $\hat{k}_{t+1} = \varphi(x_t)' \varphi(x_{t+1})$.

It is important that a mapping ϕ is chosen, so $\kappa(a, b) = \varphi(a)' \varphi(b)$ can be computed without computing $\varphi(a)$ and $\varphi(b)$ separately. κ is the so-called kernel function and various types of these functions are known, such as the polynomial and the Gaussian kernel functions. Before computing the kernel matrix K, each observation x is divided by a positive scaling factor σ to control for the relative importance of the terms in $\varphi(x)$. For example, the weights of the different polynomial degrees are for the linear terms divided by σ and for the second-order terms by σ^2 . The way of selecting the scaling factor will be discussed in Section 3.1.3. Kernel functions which will be used in this research are the polynomial and the Gaussian kernel functions and these will be explained in the next two subsections.

3.1.1 Polynomial kernel

A polynomial function always has a degree, which is the highest degree of its terms. A polynomial kernel also has degrees. The first degree is a linear kernel function, hence $\varphi(a) = a$ and $\kappa(a, b) = \varphi(a)' \varphi(b) = a'b$. Then it holds that the transformed matrix Z is the same as the original matrix with predictors ($Z=X$). Likewise, $\hat{z}'_{t+1} = \hat{x}'_{t+1}$, which leads to the kernel matrix $K = ZZ' = XX'$ and

$\hat{k}_{t+1} = Z\hat{z}_{t+1} = X\hat{x}_{t+1}$. Inserting this in the forecast (3) gives $\hat{y}_{t+1} = \hat{x}'_{t+1}X'(XX' + \lambda I)^{-1}y = \hat{x}'_{t+1}(X'X + \lambda I)^{-1}X'y$, which is the same as the linear ridge regression forecast in (1).

A quadratic kernel function can be retrieved from a mapping which contains a constant term, variables, and their squares and cross products. This leads to the quadratic kernel function

$$\kappa(a, b) = \varphi(a)' \varphi(b) = (1 + a'b)^2$$

The derivation of this result is shown in Exterkate et al. (2013).

A generalization of the polynomial kernel functions is given by

$$\kappa(a, b) = \varphi(a)' \varphi(b) = (1 + a'b)^d,$$

where d is the maximum degree for which the mapping of a consists all polynomials in the elements of a .

3.1.2 Gaussian kernel

The Gaussian kernel function is given by

$$\kappa(a, b) = \exp\left(-\frac{1}{2}\|a - b\|^2\right)$$

An advantage is that this kernel can be used, even if the number of transformed variables M approaches infinity $M \rightarrow \infty$.

The elements of the corresponding mapping are the “dampened” polynomials

$$e^{-\frac{a'a}{2}} \prod_{n=1}^N \frac{a_n^{d_n}}{\sqrt{d_n!}}$$

A difficulty of performing KRR is how to select the parameters λ and σ , called respectively the shrinkage parameter and the scaling parameter. This selection will be examined in the next subsection.

3.1.3 Tuning parameter selection

For the implementation of KRR, two parameters need to be estimated. These are the shrinkage parameter λ and the scaling parameter σ .

This is done, as stated in Exterkate et al. (2013). They give estimates based on the signal-to-noise ratio (for λ) and the smoothness assumption (for σ). These estimates differ between the used kernel functions and are stated below

$$\sigma_0 = \sqrt{\frac{N}{2}} \quad \text{and} \quad \lambda_0 = \frac{(1 + N\sigma^{-2})(1 - \hat{R}^2)}{\hat{R}^2} \quad \text{for the Poly(1) kernel,}$$

$$\sigma_0 = \sqrt{\frac{N+2}{2}} \text{ and } \lambda_0 = \frac{(1 + 2N\sigma^{-2} + N(N+2)\sigma^{-4})(1 - \hat{R}^2)}{\hat{R}^2} \text{ for the Poly(2) kernel,}$$

$$\sigma_0 = \frac{\sqrt{c_N}}{\pi} \text{ and } \lambda_0 = \frac{(1 - \hat{R}^2)}{\hat{R}^2} \text{ for the Gaussian kernel,}$$

where c_N is the 95th percentile of the χ^2 distribution with N degrees of freedom, \hat{R}^2 can be computed in two ways. Exterkate et al. (2013) calculate the R-squared from the OLS regression of y on the first four principal components of X , but earlier Exterkate (2013) also stated that \hat{R}^2 can be obtained from linear OLS regression of y on a constant and X . In both cases the tuning parameters are re-estimated for each window.

The KRR will be applied on the full set of 38 predictors and on the best 18 macroeconomic and financial variables preselected by LARS, as was done by Holtrop et al (2014). In addition, the KRR models will be estimated on a moving window of 5 and 10 years.

3.2 Longer step ahead forecasts

As already stated in the introduction section, longer step ahead forecasts will be made for all models. According to Taieb et al. (2012) , there are a number of strategies to forecast h -step ahead. Two of these strategies will be discussed below. The recursive strategy is also the oldest and most intuitive forecasting strategy. With this strategy, the parameters of the model are estimated with the data until time period T . Then with this model and the parameters a one-step ahead prediction is made. Thereafter this prediction is used as input in the forecasting periods ahead keeping the parameters of the model the same. This strategy is sensitive to accumulation of errors. Errors in the intermediate forecasts are spread forward in the forecast horizon. Furthermore, for an ARX-model the exogenous variables are then presumed to follow a random walk, which is generally not the case.

Another strategy is the direct strategy, which forecasts each horizon independently from the others. An advantage is that does not use any approximated values to compute the forecasts. So there is no accumulation of errors. A drawback is that dependencies between the variables \hat{y}_{T+H} are not taken in to account and computational time increases since the model has to be re-estimated for each forecast horizon.

An example of such a 'direct' forecast is given for the ARX(1)-model, which is then specified as follows

$$y_t = \alpha + \beta y_{t-h} + \sum_{i=1}^p c_i x_{i,t-h} + \varepsilon_t$$

Now a forecast for y_{T+h} can be calculated without having to forecast the x -variables first. However, the coefficients α , β and c_i have to be estimated for each forecast horizon.

Multi-step ahead forecasts will be made on a quarterly (3 months, $H=3$), half-yearly (6 months, $H=6$) and yearly (12 months, $H=12$) horizon. Furthermore, more interesting is to predict the volatility over this period, and not specifically the volatility in month $T+h$. This is easily done by taking the sum of all the forecasted volatilities. ($\sum_{h=1}^H y_{t+h}$).

3.3 Statistical evaluation

The statistical evaluation of the kernel ridge regressions and the multi-step ahead forecasting will be done by analyzing the three factors accuracy, efficiency and unbiasedness of the forecasts of the constructed models. The statistical evaluation of the models of Holtrop et al. (2014) was also based on these three factors.

To evaluate the accuracy, the Mean Squared Prediction Error (MSPE) is used.

$$MSPE = \frac{1}{n} \sum_{t=1}^n \varepsilon_t^2$$

To compare the MSPEs of the forecast of the models and benchmarks, the out-of-sample R^2 is used.

The out-of-sample R^2 is defined by

$$R^2_{model\ x} = 1 - \frac{MSPE_{model\ x}}{MSPE_{benchmark}}$$

One can see that a negative out-of-sample R^2 corresponds with a lower performance in accuracy than the benchmark model. So positive values of this R^2 correspond with a more accurate forecast than the benchmark.

The MSPE's aren't directly comparable, since it is not necessary that a lower MSPE implies a better forecast. However, the Diebold Mariano test can be used to determine if the differences in the prediction errors are statistically significant. The null hypothesis of this test is that the sample mean of the difference in squared prediction errors of two different models i, j is not significantly larger than zero.

$$DM = \frac{\bar{d}}{\sqrt{\frac{V(d_{t+1})}{n}}} \approx N(0,1)$$

$$with\ d_{t+1} = e^2_{i,t+1|t} - e^2_{j,t+1|t} \quad and \quad V(d_{t+1}) = \frac{1}{n-1} \sum_{t=1}^n (d_{t+1} - \bar{d})^2$$

Where \bar{d} is the sample mean of d_{t+1} and n is the number of forecasts which is equal to 216. With a significance level of 5 percent, the critical value of this one-sided test is equal to 1.645. If the DM

statistic is positive, this means that \bar{d} is positive, since the (square root of the) variance in the DM-statistic is always positive. Then can be concluded that the sample mean of the squared prediction errors of model i $e_{i,t+1|t}^2$ are larger than the prediction errors of model j and so model i performs worse than model j in terms of accuracy.

To analyze the efficiency of the forecasts, a Mincer-Zarnowitz regression is used

$$y_{t+1} = \beta_0 + \beta_1 \hat{y}_{t+1|t} + \eta_{t+1}$$

The null hypothesis is that it should not be possible to forecast the forecast errors, based on the info at the time the forecast is made. In the regression above this null hypothesis is tested by a Wald test on the joint restrictions, $\beta_0 = 0$ and $\beta_1 = 1$.

Finally, the unbiasedness of the forecasts is checked. This unbiasedness means that the expected value of the forecast errors should be equal to zero. To evaluate if the sample mean of the forecast errors differ significantly from zero, a standard Z-test is performed.

3.4 Economical evaluation

Besides the statistical evaluation, a practical evaluation will be performed, known as the economical evaluation for the forecasts of the kernel ridge regression. This will be done by creating a fictive investor who selects his portfolio based on an investment function. Cakmakli and van Dijk (2013) also made use of this method and applied the following investment function of returns and volatility forecasts.

$$\max_{w_{t+1}} E(r_{p,t+1}) - \frac{1}{2} \gamma \text{Var}(r_{p,t+1}) \quad (?)$$

In this equation, γ is the rate of risk aversion. The portfolio return $r_{p,t+1}$ consists of the return in a risk-free investment $r_{f,t+1}$ and investing in the assets, which are considered in this paper $r_{s,t+1}$. w_{t+1} denotes the part of the portfolio which is invested in these assets.

$$r_{p,t+1} = (1 - w_{t+1})r_{f,t+1} + w_{t+1}r_{s,t+1}$$

The returns of the risk-free investment are retrieved from the 3-months T-bill rate, since no data is available for 1-month T-bill rates before 2001. The returns of the stocks consist of the monthly stock return on the S&P 500. For the other two assets, commodities and bonds, the data are derived respectively from the S&P GSCI Commodity Index and the monthly return on the 10-year T-bond.

The return forecast will be the average return of the past 5 and 10 years, depending on the window of the forecast. The volatility forecast will be the forecasts of the KRR models which are constructed in this paper. To obtain the optimal weights in the portfolio the variance of the portfolio is needed

$$\begin{aligned}
Var(r_{p,t+1}) &= Var\left((1 - w_{t+1})r_{f,t+1} + w_{t+1}r_{s,t+1}\right) \\
&= (1 - w_{t+1})^2 Var(r_{f,t+1}) + w_{t+1}^2 Var(r_{s,t+1}) + 2Cov(r_{f,t+1}, r_{s,t+1}) \\
&= w_{t+1}^2 Var(r_{s,t+1})
\end{aligned}$$

since the variance of the risk free return is equal to zero because it is assumed that $r_{f,t+1}$ is fixed at the end of month t . The covariance between the return of the risk-free investment and the investment in the relevant asset is equal to zero.

With the variance of the portfolio, the investment function can be maximized to obtain the optimal weights in the portfolio.

$$\begin{aligned}
\max_{w_{t+1}} E(r_{p,t+1}) - \frac{1}{2}\gamma Var(r_{p,t+1}) &= \max_{w_{t+1}} E\left((1 - w_{t+1})r_{f,t+1} + w_{t+1}r_{s,t+1}\right) - \frac{1}{2}\gamma w_{t+1}^2 Var(r_{s,t+1}) \\
&= \max_{w_{t+1}} (1 - w_{t+1})r_{f,t+1} + w_{t+1}E(r_{s,t+1}) - \frac{1}{2}\gamma w_{t+1}^2 Var(r_{s,t+1}) \\
&\Rightarrow -r_{f,t+1} + E(r_{s,t+1}) - \gamma Var(r_{s,t+1}) = 0
\end{aligned}$$

This leads to the optimal weights w_{t+1}^*

$$w_{t+1}^* = \frac{E(r_{s,t+1}) - r_{f,t+1}}{\gamma Var(r_{s,t+1})}$$

Two cases are considered; first, the weights are bounded between zero and one ($w_{t+1}^* \in [0,1]$). These weights imply that short selling and lending are not allowed. In the second case, short selling and lending are permitted ($w_{t+1}^* \in [-1,2]$). Transaction costs are neglected. To evaluate what an investor is willing to pay for using the volatility forecasts of this paper, the maximum performance fee is calculated. To be able to do this, a quadratic utility function is assumed (West, Edison & Cho, 1993). The average utility is given by

$$\bar{U} = \frac{W}{n} \sum_{t=0}^{n-1} \left(R_{p,t+1} - \frac{1}{2} \frac{\gamma}{(1 + \gamma)} R_{p,t+1}^2 \right) \quad \text{with } R_{p,t+1} = 1 + r_{p,t+1}$$

Here W is defined as the wealth to be invested and n is the number of time periods where the investing is analyzed. In order to calculate the maximum performance fee, the utility of a strategy arising from the forecast of the constructed models (*strategy a*) needs to be compared with an unsophisticated buy-and-hold strategy (*strategy b*). The buy-and-hold strategy consists of either only investing in the risk-free t-bonds, only investing in the market, or in an equally weighted combination of the two.

$$\sum_{t=0}^{n-1} \left((R_{p,t+1}^a - \Delta) - \frac{1}{2} \frac{\gamma}{(1 + \gamma)} (R_{p,t+1}^a - \Delta)^2 \right) = \sum_{t=0}^{n-1} \left(R_{p,t+1}^b - \frac{1}{2} \frac{\gamma}{(1 + \gamma)} (R_{p,t+1}^b)^2 \right)$$

From this equation the delta can be calculated, which is a fraction of the wealth that the investor is maximally willing to pay for this information. Hence, the performance of the model gets larger, if the delta of the relevant model rises.

Finally, after the forecasts have been evaluated statistically and economically, it can be seen if the non-linear forecasts, constructed by KRR, are better than the linear forecasts. Besides this, it can be concluded how the longer-step ahead forecasts perform.

4 Results

First, the results of the kernel ridge regressions will be evaluated. Thereafter, the results of the multi-step ahead forecasts will be examined. The final section gives the economical evaluation of the forecasts of KRR.

The tables that will be shown in the sections are all about the asset class stocks. The results of the other classes are shown in the appendices and are briefly summarized in the text.

4.1 Kernel ridge regressions

In this section, the results of the kernel ridge regression are discussed. As already explained in Section 3.1, different kernel functions are used. In Section 4.1.1, the linear kernel will be analyzed, followed by the quadratic kernel in Section 4.1.2 and finally the Gaussian kernel in Section 4.1.3. At the end of this section, the different kernels are compared with each other. The statistical results of the other asset classes can be found in Appendix A.

4.1.1 Linear kernel

First the linear kernel is statistically evaluated. The models are tested as stated in Section 3.3 for unbiasedness, accuracy and efficiency.

Model	Unbiasedness	MSPE	Mincer-Zarnowitz	Out-of-sample R ²		
				AR(1)	ARX(1)	RW
			p-value			
10y PC	-1.595	0.059	0.105	0.382*	0.372*	0.421*
10y OLS	-2.196*	0.061	0.090	0.369*	0.359*	0.409*
10y PC 18var	-1.843	0.061	0.036*	0.360*	-0.090	0.401*
10y OLS 18var	-2.010*	0.060	0.117	0.374*	-0.066	0.413*
5y PC	-1.595	0.059	0.292	0.354*	0.842*	0.344*
5y OLS	-2.196*	0.061	0.402	0.349*	0.841*	0.339*
5y PC 18var	-1.843	0.061	0.243	0.345*	0.048	0.334*

5y OLS 18var	-2.010*	0.060	0.380	0.365*	0.077	0.354*
---------------------	---------	-------	-------	--------	-------	--------

Table 4.1 Evaluation of linear kernel for the asset class stocks where the shrinkage parameters are estimated with an OLS regression of y on 4 principal components (PC) and a regression of y on the complete set of macroeconomic and financial variables (OLS). Furthermore this is done for a moving window of 5 and 10 years.

Note: In the column of unbiasedness the z-statistics are given, where the null hypothesis is that the forecast is unbiased. The column of Mincer-Zarnowitz shows the p-values of the Wald test using the F-test statistic. The null hypothesis is that the forecast is efficient.

The significance of the out-of-sample R^2 is determined by the Diebold-Mariano statistics. For each model the corresponding moving window length and the number of variables is chosen in the benchmark for the out-of-sample R^2 (for example the 10y PC 18var model is compared with the ARX(1) 10y 18 var). If there is an *, it means that there is a significant difference in accuracy

* denotes the rejection of the null hypothesis for a significance level of 5 percent

The table above shows that when the shrinkage parameters are estimated with a regression of the volatility on the complete set of macroeconomic and financial variables, biased forecasts are produced. In general the forecasts overestimate the true value in a negative way, since values below zero are found when analyzing the unbiasedness.

The mean squared prediction errors (MSPE) of all the forecasts are roughly equal, but not all benchmarks are beaten. For the linear kernel where 18 macroeconomic and financial variables are preselected by LARS, the ARX-model with 18 preselected variables is not beaten and for the 10-year moving window it can be seen that the benchmark even has a smaller MSPE. Fortunately the random walk model produces significantly worse forecasts than the linear kernel ridge regressions.

From the Diebold-Mariano statistics can be concluded that the forecasts constructed with a moving window of 10 years are significantly better than those with a moving window of 5 years. A possible explanation might be that the moving window of 10 years contains more information. Preselecting the variables with LARS sometimes gives significantly better results, but not always.

No significant differences in MSPE are found between the tuning parameter estimation method with principal components and the method with the complete matrix with independent variables.

All forecasts are efficient, apart from the model with the 18 preselected variables, the tuning parameter estimation by four principal components and a moving window of 10 years.

For commodities similar results are found, but for the asset class of foreign exchanges, the forecasts of the linear kernel almost never significantly beat the benchmark models of AR(1) and ARX(1).

However, all linear kernel forecasts are unbiased and the random walk model is beaten most of the times. The forecasts of the linear kernel for the asset class bonds are even worse in terms of accuracy.

4.1.2 Quadratic kernel

In this section the polynomial kernel with two degrees is evaluated. Table 4.2 shows the results in terms of unbiasedness, accuracy and efficiency.

Model	Unbiasedness	MSPE	Mincer-Zarnowitz	Out-of-sample R ²		
			p-value	AR(1)	ARX(1)	RW
10y PC	-1.275	0.059	0.067	0.384*	0.374*	0.423*
10y OLS	-1.844	0.060	0.072	0.375*	0.365*	0.414*
10y PC 18var	-1.282	0.060	0.031*	0.373*	-0.067	0.413*
10y OLS 18var	-1.532	0.061	0.034*	0.361*	-0.088	0.401*
5y PC	-0.602	0.068	0.157	0.347*	0.840*	0.336*
5y OLS	-1.212	0.070	0.169	0.331*	0.836*	0.320*
5y PC 18var	-0.622	0.067	0.078	0.358*	0.068	0.348*
5y OLS 18var	-1.027	0.068	0.095	0.349*	0.054	0.338*

Table 4.2 Evaluation of quadratic kernel for the asset class stocks where the shrinkage parameters are estimated with an OLS regression of y on 4 principal components (PC) and a regression of y on the complete set of macroeconomic and financial variables (OLS). Furthermore this is done for a moving window of 5 and 10 years.

Note: In the column of unbiasedness the z-statistics are given, where the null hypothesis is that the forecast is unbiased. The column of Mincer-Zarnowitz shows the p-values of the Wald test using the F-test statistic. The null hypothesis is that the forecast is efficient.

The significance of the out-of-sample R² is determined by the Diebold-Mariano statistics. For each model the corresponding moving window length and the number of variables is chosen in the benchmark for the out-of-sample R² (for example the 10y PC 18var model is compared with the ARX(1) 10y 18 var). If there is an *, it means that there is a significant difference in accuracy

* denotes the rejection of the null hypothesis for a significance level of 5 percent

Results in Table 4.2 are quite similar to the results of the linear kernel. A main difference is that all forecasts of the quadratic kernel are unbiased, but the forecasts constructed with 18 preselected variables and a 10 year moving window are now both inefficient. Again the ARX benchmark is not beaten by the forecasts with 18 preselected variables and forecasts constructed with a 10 year moving window perform better than those with a 5 year moving window. There are no significant differences in the estimation method of the tuning parameters. From the pre-selection of variables can only be concluded that for a moving window of 5 years and the OLS tuning method, pre-selection provides better forecasts.

For the asset class of commodities, there is no significant difference in accuracy for the two moving windows and none of the forecasts with a moving window of 5 year is efficient. Furthermore, almost all forecast with a moving window of 10 years are not unbiased.

No efficient forecasts are also found for the asset class of foreign exchanges and the same holds for bonds with a 5 year moving window.

4.1.3 Gaussian kernel

In this section with the results of the different kernels used, the Gaussian kernel is statistically evaluated. The table below shows the results in terms of unbiasedness, accuracy and efficiency.

Model	Unbiasedness	MSPE	Mincer-Zarnowitz	Out-of-sample R ²		
				AR(1)	ARX(1)	RW
			p-value			
10y PC	-0.843	0.060	0.124	0.380*	0.370*	0.419*
10y OLS	-0.950	0.059	0.075	0.382*	0.371*	0.420*
10y PC 18var	-0.077	0.058	0.060	0.3968*	-0.028	0.464*
10y OLS 18var	0.000	0.060	0.007*	0.380*	-0.056	0.419*
5y PC	0.199	0.063	0.107	0.393*	0.851*	0.383*
5y OLS	-0.02	0.064	0.084	0.388*	0.850*	0.378*
5y PC 18var	0.913	0.064	0.005*	0.386*	0.108	0.376*
5y OLS 18var	1.008	0.065	0.002*	0.380*	0.100	0.370*

Table 4.3 Evaluation of Gaussian q kernel for the asset class stocks where the shrinkage parameters are estimated with an OLS regression of y on 4 principal components (PC) and a regression of y on the complete set of macroeconomic and financial variables (OLS). Furthermore this is done for a moving window of 5 and 10 years.

Note: In the column of unbiasedness the z-statistics are given, where the null hypothesis is that the forecast is unbiased. The column of Mincer-Zarnowitz shows the p-values of the Wald test using the F-test statistic. The null hypothesis is that the forecast is efficient.

The significance of the out-of-sample R² is determined by the Diebold-Mariano statistics. For each model the corresponding moving window length and the number of variables is chosen in the benchmark for the out-of-sample R² (for example the 10y PC 18var model is compared with the ARX(1) 10y 18 var). If there is an *, it means that there is a significant difference in accuracy

* denotes the rejection of the null hypothesis for a significance level of 5 percent

Similarly to the quadratic kernel, all forecasts are unbiased. Again the ARX(1) benchmark cannot be beaten by the models where 18 variables were preselected. Additionally, inefficient predictions are found for three out of four forecasts which were constructed with pre-selection of variables.

For the Gaussian kernel the difference in forecast accuracy for the moving windows is less convincing than for the previous two kernels. Only for the kernel ridge regression with a pre-selection of 18 variables and usage of the principal component estimation method, the Diebold-Mariano statistic concludes that the 10 year moving window is better than the moving window of 5 years. No statistical evidence is found for the difference in the estimation methods of the shrinkage parameter and the pre-selection of variables.

A remarkable observation is that no efficient forecasts are constructed for the other asset classes (commodities, foreign exchanges and bonds). For the two asset classes, foreign exchanges and bonds, the Gaussian kernel never beats the AR(1) benchmark and sometimes it even performs worse in terms of accuracy. Besides, not all forecasts constructed with the Gaussian kernel are unbiased for these two assets.

The kernels are compared while keeping the shrinkage parameter selection method and the number of macroeconomic and financial predictors the same. There aren't any significant differences in accuracy between the different kernels for the asset class stocks, except that the Gaussian kernel

provides more accurate forecasts than the quadratic kernel when the ridge parameter is estimated with a regression of y on all macroeconomic and financial variables and with a 5 year moving window.

For commodities, the linear kernel is the best of all kernels when using the OLS estimation method and a moving window of 5 years.

From the Diebold-Mariano statistics for the asset class stocks can be concluded that the quadratic kernel performs significantly worse than the other two kernels if a moving window of 5 years and the principal components estimation is used.

The Diebold-Mariano statistics of the last asset class bonds show that for a moving window of 5 years and the OLS estimation method the linear kernel performs best. The same holds for a 10 year moving window with the OLS estimation method and a pre-selection of the variables.

Holtrop et al. (2014) concluded that forecast combinations of linear models with 18 variables, chosen with the LARS method, outperform all other models they used in their research. This unfortunately also holds true for the KRR models.

4.2 Multi-step ahead forecasts

In this section, the results of the multi-step ahead forecasts are discussed. The section is divided in three sections related to the different forecast horizons; three months, half a year and a full year ahead. All evaluations are performed on the ‘total’ volatility, which is the sum of the predicted volatilities over the total forecasting horizon. $(\sum_{h=1}^H y_{t+h})$. Besides this, all forecasts are constructed with a moving window of 10 years. The statistical evaluation of the other asset classes can be found in Appendix B.

4.2.1 Three months ahead forecasts

First, the forecasts with a horizon of three months are evaluated. These are like the KRR analyzed in terms of unbiasedness, accuracy and efficiency and the results are given in the table below

Model	Unbiasedness	MSPE	Mincer-Zarnowitz p-value
AR(1)	-0.770	0.838	0.537
FC 6 vars	-0.922	0.783	0.625
FC 12 vars	-1.239	0.801	0.145
ARX 6 vars	-0.915	0.853	0.001*
ARX 12 vars	-1.074	0.905	0.000*
PCA	-0.514	0.816	0.710
PLS	-1.772	0.805	0.095
RW	-0.047	0.946	0.000*

KRR PC Poly1	0.270	0.857	0.000*
KRR PC Poly2	0.831	0.865	0.000*
KRR PC Gauss	-3.670*	1.383	0.000*
KRR OLS Poly1	0.383	0.854	0.000*
KRR OLS Poly2	1.386	1.820	0.000*
KRR OLS Gauss	-0.391	4.360	0.000*

Table 4.4 Evaluation of three months ahead forecasts for the asset class stocks. This is done for auto-regressive model with one lag (AR(1)), forecast combinations (FC) and ARX model with 6 and 12 variables (selected by LARS), Principal Component Analysis (PCA), Partially Least Squares (PLS), Random Walk (RW) and Kernel Ridge Regressions (KRR) with two shrinkage parameter estimations (PC, OLS) and three different kernels (Poly1, Poly2, Gauss)

* denotes the rejection of the null hypothesis for a significance level of 5 percent

Most of the 3-month ahead forecasts are unbiased, except for the predictions made with the Gaussian Kernel and the shrinkage parameter estimation by regressing y on all macroeconomic and financial variables. Additionally, most of the forecasts are according to the Mincer-Zarnowitz regression inefficient. From the Diebold-Mariano statistics presented in Appendix C can be concluded that 'KRR OLS Gauss' provides significantly worse forecasts than all other models. All models give biased (except the Random Walk model) and inefficient forecasts for the asset class of commodities. Nonetheless, in terms of accuracy the quadratic kernel ridge regression model with the principal component tuning parameter estimation method is significantly better for most models except PCA, PLS and some KRR.

For the asset class of foreign exchanges, all forecasts are unbiased except the forecasts of forecast combinations and the ARX model. According to the Diebold-Mariano statistics the quadratic kernel with the OLS tuning parameter estimation method provides the most accurate forecast. However, it is not significantly better than the random walk and the linear kernel with the principal component tuning parameter estimation method.

In terms of unbiasedness the same holds for bonds, however in terms of accuracy the best forecast is now given by partial least squares. Nevertheless, it is not significantly better for two KRR with the OLS tuning parameter estimation method.

4.2.2 Six months ahead forecasts

In this section the forecasts with a horizon of six months are analyzed. These are again evaluated in terms of unbiasedness, accuracy and efficiency with results in Table 4.5

Model	Unbiasedness	MSPE	Mincer-Zarnowitz
AR(1)	-1.497	3.408	p-value 0.329
FC 6 var	-1.674	3.390	0.060
FC 12 vars	-1.873	3.540	0.007*

ARX 6 vars	-1.549	4.106	0.000*
ARX 12 vars	-1.574	4.396	0.000*
PCA	-0.931	3.461	0.245
PLS	-2.280*	3.938	0.001*
RW	-0.553	4.278	0.000*
KRR PC Poly1	0.403	3.709	0.000*
KRR PC Poly2	0.878	3.984	0.000*
KRR PC Gauss	1.647	5.521	0.000*
KRR OLS Poly1	0.953	3.621	0.000*
KRR OLS Poly2	1.874	6.443	0.000*
KRR OLS Gauss	7.668*	33.23	0.000*

Table 4.5 Evaluation of six months ahead forecasts for the asset class stocks. This is done for auto-regressive model with one lag (AR(1)), forecast combinations (FC) and ARX model with 6 and 12 variables (selected by LARS), Principal Component Analysis (PCA), Partially Least Squares (PLS), Random Walk (RW) and Kernel Ridge Regressions (KRR) with two shrinkage parameter estimations (PC, OLS) and three different kernels (Poly1, Poly2, Gauss)

* denotes the rejection of the null hypothesis for a significance level of 5 percent

Almost all 6-month ahead forecasts are inefficient. Mean squared prediction errors are larger than in the 3-month ahead forecast, because of the accumulation of errors when the model is not correctly specified and since the sum is taken over all forecast horizons.

In terms of accuracy, the results are the same as for the three month ahead forecast; the KRR OLS Gauss performs worst.

For the other asset classes, similar results are found. More models provide inefficient forecasts and MSPE's rise. In addition, the random walk model gets beaten by more models.

4.2.3 Twelve months ahead forecasts

Finally, the forecasts with a horizon of twelve months are evaluated.

Model	Unbiasedness	MSPE	Mincer-Zarnowitz
			p-value
AR(1)	-2.232*	14.76	0.015*
FC 6 vars	-2.738*	15.33	0.000*
FC 12 vars	-2.495*	16.62	0.000*
ARX 6 vars	-3.004*	18.95	0.000*
ARX 12 vars	-1.998	23.23	0.000*
PCA	-1.393	15.97	0.001*
PLS	-2.888*	18.68	0.000*
RW	-0.793	20.27	0.000*
KRR PC Poly1	0.260	17.45	0.000*

KRR PC Poly2	0.426	18.00	0.000*
KRR PC Gauss	2.568*	20.00	0.000*
KRR OLS Poly1	1.090	19.75	0.000*
KRR OLS Poly2	2.190*	26.86	0.000*
KRR OLS Gauss	5.931*	70.97	0.000*

Table 4.6 Evaluation of twelve months ahead forecasts for the asset class stocks. This is done for auto-regressive model with one lag (AR(1)), forecast combinations (FC) and ARX model with 6 and 12 variables (selected by LARS), Principal Component Analysis (PCA), Partially Least Squares (PLS), Random Walk (RW) and Kernel Ridge Regressions (KRR) with two shrinkage parameter estimations (PC, OLS) and three different kernels (Poly1, Poly2, Gauss)

* denotes the rejection of the null hypothesis for a significance level of 5 percent

It can be seen that almost all forecasts are biased and the Mincer-Zarnowitz shows that all forecast are inefficient. The simple first-order autoregressive model and the factor-model with principal components have the lowest MSPE's and according to the Diebold-Mariano statistic they beat 7 out of the 14 models.

4.3 Economical evaluation

In this section, the forecasts of the kernel ridge regression will be evaluated economically for the asset class stocks. The results in the table below are for models constructed with a 10 year moving window. For the other asset classes and the moving window of 5 years, see Appendix D.

Stocks	Mean	STD										
100% Market	21.94	67.38										
50% Market	8.89	29.03										
0% Market	3.38	1.95										
			Weights $\in [0,1]$					Weights $\in [-1,2]$				
Model	Mean	STD	$\Delta 50$	$\Delta 100$	$\Delta 0^1$	Mean	STD	$\Delta 50$	$\Delta 100$	$\Delta 0^1$		
Real	9.36	17.73	2007	6240	152	9.74	19.48	1902	6229	-1534		
KRR PC Poly1	9.26	19.13	1883	6190	44	9.78	20.54	1805	6200	-1672		
KRR OLS Poly1	9.30	19.91	1818	6172	9	9.57	20.75	1764	6173	-1932		
KRR PC Poly2	9.21	19.40	1855	6177	43	9.74	20.86	1770	6186	-1731		
KRR OLS Poly2	9.31	20.46	1766	6155	22	9.68	21.50	-2261	6159	-1908		
KRR PC Gauss	9.28	18.90	1905	6200	52	9.59	19.86	1851	6202	-1634		

¹ For Delta 0, the risk aversion rate is set to $\gamma=1$. For $\gamma=8$, the utility function of the buy-and-hold strategy 0% market does not coincide with the utility function of any model.

KRR OLS Gauss	9.51	21.81	-224	6131	34	9.77	22.55	1586	6132	-1913
----------------------	------	-------	------	------	----	------	-------	------	------	-------

Table 4.7 Economic evaluation of the stock return volatility forecasts of the different kernel ridge regressions where the shrinkage parameters are estimated with an OLS regression of y on 4 principal components (PC) and a regression of y on the complete set of macroeconomic and financial variables (OLS) with a moving window of 10 years. Delta50, Delta100 and Delta0 are the performance fees an investor is willing to pay extra to use the models instead of the standard strategies, displayed in basis points. Note: Mean and STD are respectively the average and the standard deviation of the portfolio return. The 'real' model is the economic evaluation where the optimal weights were to be constructed with the real values for the variances. Two weightings schemes are used, where in the second weighting scheme short selling and lending is allowed.

Table 4.7 shows that the 100% market strategy does have a high mean, but it also has a high standard deviation, which is not optimal for the considered investor. The risk-free strategy has a low standard deviation, but consequently a relatively low return. So one can argue that the 50% market strategy would be an optimal balance between risk and return for an investor with a quadratic utility function. For an investor with these preferences, the table shows that using the KRR models provide a larger mean and a lower standard deviation than the buy-and-hold strategies.

In the table can be seen that for Delta 50, the performance fee that the investor is willing to pay is the largest for the Gaussian kernel ridge regression using principal components as tuning parameter estimation method. The other estimation method however shows negative values, which means that the investor is not willing to use the volatility forecasts of this model. In fact the investor would have to be paid to use these volatility forecasts. This could be because the standard deviation of the returns for this model is considerably higher than the rest.

When allowing short selling and lending the best model in terms of economic evaluation stays the same, but now the quadratic kernel with the OLS estimation method performs worst. In general for all kernels, the OLS estimation method is worse for all kernels and market strategies. The second weighting scheme has slightly higher average returns, but this automatically leads to higher standard deviations

5 Conclusion

This paper researched whether the nonlinear method of kernel ridge regressions (KRR) provides better forecasts than linear models and the models from Holtrop et al. (2014) constructed with the Least Angle Regression (LARS) method. Besides this it was examined which models would perform best when forecasting over a longer horizon, using the direct forecasting approach.

These two research topics were investigated with data of the volatilities of four different asset classes, i.e. stocks, bonds, commodities and foreign exchanges.

The kernel ridge regressions were performed with three different kernels, namely the linear kernel, the quadratic kernel and the Gaussian kernel. Furthermore, estimating the necessary tuning parameters was done in two ways. First, by regressing the relevant volatility on the first four principal

components of the data with 38 macroeconomic and financial variables. Second, by regressing it on the complete dataset of macroeconomic and financial variables. Moreover, KRR were done with the complete set of variables and with 18 variables preselected by LARS.

The statistical evaluation showed that the linear benchmarks were often beaten by the KRR.

However, no clear statistical evidence was found which type of kernel would give the best results. In general, the 10 year moving window includes more information and provides better outcomes.

Sometimes the pre-selection of variables gave better result, but this difference was not always significant. Finally, for the methods of selecting the tuning parameters, no explicit conclusion can be given for the performance when this is statistically evaluated.

Unfortunately, the KRR does not give better results than the model with forecast combinations made by Holtrop et al. (2014). However, forecast combinations become computationally far more problematic than KRR, when the number of predictors rises.

When economically evaluating the KRR, it is noticed that the Gaussian kernel using the principal component estimation method for the tuning parameters gives the investor the highest utility, based on her risk-averse behavior.

The analysis of the multi-step ahead forecasts shows that the Gaussian kernel ridge regression performs worse than the other models and that the prediction errors of the simple AR(1) model get relatively smaller compared to the other models, when the forecast horizon gets larger. The macroeconomic and financial variables influence the forecasts badly, since there is insufficient information how these variables develop over time. Better predictions can be made if all these variables are forecasted properly by other models.

KRR can be examined further in future research by looking at higher order kernels or estimating the tuning parameters with a different number of principal components. Additional research can also be done for the multi-step ahead predictions by examining different moving windows.

References

Andersen, T. G., T. Bollerslev, P. F. Christoffersen and F.X. Diebold (2006), Volatility and correlation forecasting, in G. Elliott, C.W.J. Granger and A. Timmermann (eds.) *Handbook of Economic Forecasting* Volume I, pp. 777-878.

Chen, Q., Y. Hong (2010), Predictability of Equity Returns over Different Time Horizons: A Nonparametric Approach, *Manuscript, Cornell University*.

Christiansen, C., M. Schmeling and A. Schrimpf (2012), A comprehensive look at financial volatility prediction by economic variables, *Journal of Applied Econometrics* **27**, 956-977.

Exterkate, P., P.J.F. Groenen, C. Heij, D. van Dijk (2013), Nonlinear Forecasting With Many Predictors Using Kernel Ridge Regression.

Exterkate, P. (2013), Model selection in kernel ridge regression, *Computational Statistics and Data Analysis* (2013), 1-16.

Granger, W. J., S. Poon (2003), Forecasting volatility in financial markets: a review, *Journal of Economic Literature*, pp. 478-539.

Holtrop, N., W. Kers, F. Mourer and M. Verkuijlen (2014), Volatility's Next Top Driver. Forecasting volatility with the use of macroeconomic and financial variables.

Kim, H.H., N.R. Swanson (2013), Forecasting financial and macroeconomic variables using data reduction methods: New empirical evidence. *Journal of Econometrics*.

McNeil, A. and R. Frey (2000), Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach, *Journal of Empirical Finance* **7**, 271-300.

Medeiros, M.C., T. Teräsvirta and G. Rech (2006), Building neural network models for time series: A statistical approach, *Journal of Forecasting* **25**, 49-75.

Minkah, R. (2007), Forecasting volatility, *U.U.D.M. Project Report* **7**, 1-61.

Stock, J.H. and M.W. Watson (1999), A comparison of linear and nonlinear univariate models for forecasting macroeconomic time series. In R.F. Engle and H. White, editors, *Cointegration, Causality and Forecasting. A Festschrift in Honour of Clive W.J. Granger*, pp. 1-44.

Taieb, S.B., G. Bontempi, A. Atiya, A. Sorjamaa (2012), A review and comparison of strategies for multi-step ahead time series forecasting based on the NN5 forecasting competition, *Expert Systems with Applications* **39**, Issue 8, 7067-7083

Teräsvirta, T., D. van Dijk and M.C. Medeiros (2005), Linear models, smooth transition autoregressions, and neural networks for forecasting macroeconomic time series: A re-examination, *International Journal of Forecasting* **21**, 755-774

Appendices

Appendix A Statistical evaluation KRR

Evaluation of the kernels for the asset classes commodities, foreign exchanges and bonds. The shrinkage parameters are estimated with an OLS regression of y on 4 principal components (PC) and a regression of y on the complete set of macroeconomic and financial variables (OLS). Furthermore this is done for a moving window of 5 and 10 years.

Note: In the column of unbiasedness the z-statistics are given, where the null hypothesis is that the forecast is unbiased. The column of Mincer-Zarnowitz shows the p-values of the Wald test using the F-test statistic. The null hypothesis is that the forecast is efficient.

The significance of the out-of-sample R^2 is determined by the Diebold-Mariano statistics. For each model the corresponding moving window length and the number of variables is chosen in the benchmark for the out-of-sample R^2 (for example the 10y PC 18var model is compared with the ARX(1) 10y 18 var). If there is an *, it means that there is a significant difference in accuracy

* denotes the rejection of the null hypothesis for a significance level of 5 percent

COMMODITIES

Linear Kernel

Model	Unbiasedness	MSPE	Mincer-Zarnowitz	Out-of-sample R^2		
				AR(1)	ARX(1)	RW
			p-value			
10y PC	-1.922	0.069	0.014*	0.087	0.420*	0.207*
10y OLS	-2.969*	0.066	0.009*	0.128*	0.446*	0.242*
10y PC 18var	-1.921	0.071	0.005*	0.067	0.058	0.189*
10y OLS 18var	-2.458*	0.066	0.017*	0.127*	0.119	0.241*
5y PC	-0.010	0.074	0.001*	0.000	0.824*	0.154*
5y OLS	-1.717	0.062	0.015*	0.165*	0.853*	0.294*
5y PC 18var	-0.062	0.073	0.001*	0.009	-0.097	0.162*
5y OLS 18var	-1.225	0.063	0.004*	0.149*	0.058*	0.280*

Quadratic Kernel

Model	Unbiasedness	MSPE	Mincer-Zarnowitz	Out-of-sample R^2		
				AR(1)	ARX(1)	RW
			p-value			
10y PC	-2.049*	0.069	0.013*	0.094*	0.425*	0.212*
10y OLS	-2.475*	0.066	0.017*	0.126*	0.445*	0.240*
10y PC 18var	-1.730	0.072	0.003*	0.053	0.044	0.177*
10y OLS 18var	-2.037*	0.066	0.014*	0.117*	0.109	0.233*
5y PC	0.133	0.076	0.000*	-0.030	0.818*	0.129*
5y OLS	-1.610	0.065	0.000*	0.144	0.844*	0.251*
5y PC 18var	-0.061	0.075	0.000*	-0.011	-0.119	0.145*
5y OLS 18var	-0.753	0.069	0.000*	0.059	-0.042	0.204*

Gaussian Kernel

Model	Unbiasedness	MSPE	Mincer-Zarnowitz	Out-of-sample R ²		
			p-value	AR(1)	ARX(1)	RW
10y PC	-1.358	0.065	0.004*	0.140*	0.454*	0.253*
10y OLS	-0.884	0.067	0.001*	0.110	0.435*	0.227*
10y PC 18var	-0.260	0.073	0.000*	0.033	0.024	0.160*
10y OLS 18var	-0.360	0.075	0.000*	0.014	0.005	0.143*
5y PC	0.734	0.070	0.000*	0.047	0.832*	0.194*
5y OLS	0.853	0.072	0.000*	0.018	0.827*	0.170*
5y PC 18var	1.531	0.075	0.000*	-0.017	-0.126	0.140*
5y OLS 18var	1.999*	0.081	0.000*	-0.102	-0.220*	0.068

FOREIGN EXCHANGES

Linear Kernel

Model	Unbiasedness	MSPE	Mincer-Zarnowitz	Out-of-sample R ²		
			p-value	AR(1)	ARX(1)	RW
10y PC	0.461	0.084	0.001*	0.057	0.777*	0.110*
10y OLS	0.128	0.082	0.356	0.083*	0.783*	0.134*
10y PC 18var	0.416	0.084	0.002*	0.057	-0.079	0.110*
10y OLS 18var	0.009	0.082	0.106	0.082	-0.051	0.133*
5y PC	0.133	0.087	0.000*	-0.039	0.788*	0.078*
5y OLS	0.146	0.081	0.031*	0.035	0.803*	0.143*
5y PC 18var	0.095	0.087	0.000*	-0.043	0.043	0.074*
5y OLS 18var	0.174	0.085	0.002*	-0.009	0.075	0.104

Quadratic Kernel

Model	Unbiasedness	MSPE	Mincer-Zarnowitz	Out-of-sample R ²		
			p-value	AR(1)	ARX(1)	RW
10y PC	0.385	0.084	0.001*	0.057	0.777*	0.110*
10y OLS	0.023	0.083	0.107	0.065	0.778*	0.117*
10y PC 18var	0.287	0.084	0.001*	0.056	-0.080	0.109*
10y OLS 18var	-0.232	0.084	0.021*	0.057	-0.079	0.110*
5y PC	0.236	0.093	0.000*	-0.109	0.773*	0.016
5y OLS	0.126	0.084	0.002*	-0.002	0.795*	0.110*
5y PC 18var	0.274	0.094	0.000*	-0.121	-0.028	0.005

5y OLS 18var	0.250	0.088	0.000*	-0.051	0.036	0.067
---------------------	-------	-------	--------	--------	-------	-------

Gaussian Kernel

Model	Unbiasedness	MSPE	Mincer-Zarnowitz	Out-of-sample R ²		
			p-value	AR(1)	ARX(1)	RW
10y PC	1.021	0.081	0.001*	0.087	0.784*	0.138*
10y OLS	0.903	0.082	0.002*	0.082	0.782*	0.133*
10y PC 18var	1.396	0.086	0.000*	0.031	-0.109	0.086*
10y OLS 18var	1.557	0.088	0.000*	0.016	-0.126	0.071*
5y PC	1.178	0.086	0.000*	-0.030	0.790*	0.086*
5y OLS	1.678	0.089	0.000*	-0.062	0.783*	0.057
5y PC 18var	1.555	0.094	0.000*	-0.124	-0.031	0.002
5y OLS 18var	2.342*	0.100	0.000*	-0.194*	-0.095	-0.060

BONDS

Linear Kernel

Model	Unbiasedness	MSPE	Mincer-Zarnowitz	Out-of-sample R ²		
			p-value	AR(1)	ARX(1)	RW
10y PC	0.388	0.084	0.003*	-0.005	0.237*	0.188*
10y OLS	0.506	0.074	0.737	0.113*	0.327*	0.284*
10y PC 18var	0.458	0.084	0.001*	-0.005	-0.228*	0.188*
10y OLS 18var	0.373	0.075	0.375	0.105*	-0.093	0.277*
5y PC	0.459	0.094	0.000*	-0.035	0.868*	0.090*
5y OLS	0.301	0.081	0.025*	0.113*	0.886*	0.220*
5y PC 18var	0.448	0.093	0.000*	-0.014	-0.007	0.108*
5y OLS 18var	0.496	0.083	0.001*	0.086	0.093	0.196*

Quadratic Kernel

Model	Unbiasedness	MSPE	Mincer-Zarnowitz	Out-of-sample R ²		
			p-value	AR(1)	ARX(1)	RW
10y PC	0.442	0.085	0.001*	-0.014	0.230*	0.181*
10y OLS	0.318	0.075	0.516	0.107*	0.322*	0.278*
10y PC 18var	0.425	0.084	0.001*	0.000	-0.222*	0.192*
10y OLS 18var	0.200	0.077	0.134	0.084	-0.119	0.260*
5y PC	0.241	0.092	0.000*	-0.014	0.870*	0.108*

5y OLS	0.450	0.083	0.002*	0.087	0.883*	0.197*
5y PC 18var	0.384	0.092	0.000*	-0.012	-0.005	0.110*
5y OLS 18var	0.390	0.087	0.000*	0.042	0.048	0.157*

Gaussian Kernel

Model	Unbiasedness	MSPE	Mincer-Zarnowitz	Out-of-sample R²		
			p-value	AR(1)	ARX(1)	RW
10y PC	1.537	0.081	0.000*	-0.038	0.270*	0.223*
10y OLS	1.633	0.079	0.001*	0.056	0.283*	0.237*
10y PC 18var	1.921	0.087	0.000*	-0.036	-0.266*	0.163*
10y OLS 18var	2.265*	0.087	0.000*	-0.038	-0.268*	0.162*
5y PC	1.337	0.091	0.000*	-0.001	0.872*	0.119*
5y OLS	2.120*	0.094	0.000*	-0.0296	0.869*	0.097
5y PC 18var	1.517	0.100	0.000*	-0.096	-0.088	0.036
5y OLS 18var	2.295*	0.101	0.000*	-0.103	-0.095	0.030

Appendix B Statistical evaluation multi-step ahead forecasts

Evaluation of (direct) multi step ahead forecasts for the asset classes commodities, foreign exchanges and bonds. This is done for auto-regressive model with one lag (AR(1)), forecast combinations (FC) and ARX model with 6 and 12 variables (selected by LARS), Principal Component Analysis (PCA), Partially Least Squares (PLS), Random Walk (RW) and Kernel Ridge Regressions (KRR) with two shrinkage parameter estimations (PC, OLS) and three different kernels (Poly1, Poly2, Gauss)

* denotes the rejection of the null hypothesis for a significance level of 5 percent.

COMMODITIES

3-months ahead

Model	Unbiasedness	MSPE	Mincer-Zarnowitz
			p-value
AR(1)	-6.029*	0.508	0.000*
FC 6 vars	-10.49*	0.977	0.000*
FC 12 vars	-9.043*	0.830	0.000*
ARX 6 vars	-8.417*	0.901	0.000*
ARX 12 vars	-4.042*	0.730	0.000*
PCA	-6.157*	0.474	0.000*
PLS	-6.667*	0.472	0.000*
RW	-0.486	0.605	0.000*
KRR PC Poly1	-2.503*	0.446	0.000*
KRR PC Poly2	-2.838*	0.444	0.000*
KRR PC Gauss	-5.109*	0.484	0.000*
KRR OLS Poly1	-4.205*	0.472	0.000*
KRR OLS Poly2	-3.708*	0.497	0.000*
KRR OLS Gauss	-6.450*	0.965	0.000*

6-months ahead

Model	Unbiasedness	MSPE	Mincer-Zarnowitz
			p-value
AR(1)	-8.053*	2.136	0.000*
FC 6 vars	12.65*	3.492	0.000*
FC 12 vars	11.28*	2.855	0.000*
ARX 6 vars	-9.623*	3.609	0.000*
ARX 12 vars	-4.868*	2.940	0.000*
PCA	-7.665*	2.125	0.000*
PLS	-8.154*	2.044	0.000*
RW	-0.579	2.614	0.000*

KRR PC Poly1	-3.020*	1.821	0.000*
KRR PC Poly2	-3.370*	1.826	0.000*
KRR PC Gauss	-6.032*	2.069	0.000*
KRR OLS Poly1	-5.138*	1.964	0.000*
KRR OLS Poly2	-4.723*	2.106	0.000*
KRR OLS Gauss	-8.110*	4.904	0.000*

12-months ahead

Model	Unbiasedness	MSPE	Mincer-Zarnowitz
			p-value
AR(1)	-10.38*	3.482	0.000*
FC 6 vars	-15.78*	14.81	0.000*
FC 12 vars	-15.10*	12.54	0.000*
ARX 6 vars	-12.46*	18.04	0.000*
ARX 12 vars	-8.598*	12.14	0.000*
PCA	-8.883*	10.27	0.000*
PLS	-3.449*	3.625	0.000*
RW	-0.744	11.89	0.000*
KRR PC Poly1	-3.154*	8.463	0.000*
KRR PC Poly2	-3.496*	8.449	0.000*
KRR PC Gauss	-2.321*	8.426	0.000*
KRR OLS Poly1	-5.770*	8.366	0.000*
KRR OLS Poly2	-5.502*	8.831	0.000*
KRR OLS Gauss	-4.652*	18.80	0.000*

FOREIGN EXCHANGES

3-months ahead

Model	Unbiasedness	MSPE	Mincer-Zarnowitz
			p-value
AR(1)	1.629	0.730	0.106
FC 6 vars	21.35*	2.980	0.000*
FC 12 vars	18.34*	2.292	0.000*
ARX 6 vars	15.03*	1.818	0.000*
ARX 12 vars	7.883*	1.113	0.000*
PCA	1.715	0.733	0.045*

PLS	1.118	0.728	0.023*
RW	0.041	0.684	0.000*
KRR PC Poly1	1.085	0.593	0.000*
KRR PC Poly2	0.837	0.566	0.000*
KRR PC Gauss	-0.123	0.821	0.000*
KRR OLS Poly1	0.928	0.573	0.001*
KRR OLS Poly2	1.098	0.641	0.000*
KRR OLS Gauss	-3.502*	1.921	0.000*

6-months ahead

Model	Unbiasedness	MSPE	Mincer-Zarnowitz
			p-value
AR(1)	1.690	3.182	0.002*
FC 6 vars	18.58*	10.76	0.000*
FC 12 vars	16.55*	8.833	0.000*
ARX 6 vars	11.76*	8.005	0.000*
ARX 12 vars	8.013*	6.011	0.000*
PCA	1.731	3.346	0.000*
PLS	0.646	3.288	0.000*
RW	0.074	2.71	0.000*
KRR PC Poly1	1.430	2.564	0.000*
KRR PC Poly2	1.052	2.502	0.000*
KRR PC Gauss	0.539	3.740	0.000*
KRR OLS Poly1	1.078	2.705	0.000*
KRR OLS Poly2	0.955	3.176	0.000*
KRR OLS Gauss	-5.285*	14.117	0.000*

12-months ahead

Model	Unbiasedness	MSPE	Mincer-Zarnowitz
			p-value
AR(1)	0.454	13.26	0.000*
FC 6 vars	12.29*	30.66	0.000*
FC 12 vars	10.61*	26.66	0.000*
ARX 6 vars	6.724*	29.84	0.000*
ARX 12 vars	4.701*	29.61	0.000*

PCA	0.657	13.91	0.000*
PLS	-0.683	13.87	0.000*
RW	-0.713	14.89	0.000*
KRR PC Poly1	1.374	15.15	0.000*
KRR PC Poly2	0.723	13.18	0.000*
KRR PC Gauss	0.666	19.31	0.000*
KRR OLS Poly1	0.906	15.38	0.000*
KRR OLS Poly2	0.818	20.33	0.000*
KRR OLS Gauss	-6.028*	66.27	0.000*

BONDS

3-months ahead

Model	Unbiasedness	MSPE	Mincer-Zarnowitz
			p-value
AR(1)	0.858	0.532	0.425
FC 6 vars	17.16*	1.364	0.000*
FC 12 vars	13.87*	1.116	0.000*
ARX 6 vars	13.36*	1.129	0.000*
ARX 12 vars	7.253*	0.930	0.000*
PCA	1.271	0.502	0.430
PLS	0.448	0.452	0.903
RW	-0.062	0.764	0.000*
KRR PC Poly1	0.149	0.579	0.000*
KRR PC Poly2	0.453	0.579	0.000*
KRR PC Gauss	1.559	0.678	0.000*
KRR OLS Poly1	0.445	0.476	0.023*
KRR OLS Poly2	-0.049	0.529	0.000*
KRR OLS Gauss	-0.740	1.948	0.000*

6-months ahead

Model	Unbiasedness	MSPE	Mincer-Zarnowitz
			p-value
AR(1)	0.539	1.937	0.708
FC 6 vars	12.37*	3.445	0.000*
FC 12 vars	9.023*	2.983	0.000*

ARX 6 vars	7.726*	2.953	0.000*
ARX 12 vars	3.014*	2.934	0.000*
PCA	1.266	1.834	0.265
PLS	0.588	1.711	0.191
RW	-0.228	3.228	0.000*
KRR PC Poly1	-0.035	2.406	0.000*
KRR PC Poly2	0.214	2.466	0.000*
KRR PC Gauss	4.876*	3.834	0.000*
KRR OLS Poly1	0.108	2.059	0.000*
KRR OLS Poly2	-0.392	2.448	0.000*
KRR OLS Gauss	1.669	12.23	0.000*

12-months ahead

Model	Unbiasedness	MSPE	Mincer-Zarnowitz
			p-value
AR(1)	0.639	8.206	0.224
FC 6 vars	8.271	11.20	0.000*
FC 12 vars	6.475	11.80	0.000*
ARX 6 vars	4.335	10.93	0.000*
ARX 12 vars	3.446	12.63	0.000*
PCA	1.498	8.190	0.000*
PLS	1.091	8.137	0.000*
RW	0.186	14.49	0.000*
KRR PC Poly1	0.208	11.01	0.000*
KRR PC Poly2	0.455	11.39	0.000*
KRR PC Gauss	3.845	18.03	0.000*
KRR OLS Poly1	0.307	10.25	0.000*
KRR OLS Poly2	-0.117	12.63	0.000*
KRR OLS Gauss	0.059	44.25	0.000*

20	Quadratic	OLS all variables	18	5 years
21	Gaussian	Principal components	38	5 years
22	Gaussian	Principal components	18	5 years
23	Gaussian	OLS all variables	38	5 years
24	Gaussian	OLS all variables	18	5 years