Picking the best cherries: Analysing the use of macro-finance variables in predicting monthly realized volatility

Cornelis Vletter (320297)
Erasmus University Rotterdam
MSc Econometrics and Management Science
Quantitative Finance

August 2014

Abstract

This paper analyses the use of macroeconomic and finance variables in predicting monthly realized volatility in four different asset classes: Equities, commodities, foreign exchange rates, and bonds. The predictability is analysed with four different estimation technique classes: Penalized regressions, dynamic factor models, forecast combinations, and bootstrap aggregation. The results, evaluated both statistically and economically, reveal there is predictive content in the macro-finance variables. However, both the estimation technique and subset of variables which are most relevant appear to be asset class specific.

Keywords: realized volatility, forecasting, penalized regressions, dynamic factor models, forecast combinations, bagging

Supervisor: prof. dr. D.J.C. van Dijk
Co-reader: B.J.L. Keijsers MSc

Contents

List of Tables

List of Figures

1 Introduction

Modelling and predicting financial volatility remains to interest and puzzle many researchers. Ever since the introduction of the Autoregressive Conditional Heteroskedasticity (ARCH) model by Engle (1982), the academic literature on modelling and forecasting volatility grew exponentially. However, its analysis is not straightforward as volatility is not directly observable; it is latent. To approximate volatility, classical models generally rely on strong assumptions regarding the movement of volatility and the distributional properties of the underlying returns (see amongst others Bollerslev, 1986; Nelson, 1991; Zakoian, 1991). Recently, the use of alternative techniques to estimate volatility have become more popular. In particular, realized volatility has gained the attention of many (see amongst others Andersen et al., 2006, 2003; Fleming et al., 2003; McAleer and Medeiros, 2008). Realized volatility (RV) provides a consistent non-parametric estimate of return variation over a given historical period, typically estimated on a daily basis using intra-day returns. However, RV can be estimated at any frequency as long as returns at a higher frequency are available (though limitations could arise at higher frequencies due to the micro-market structure). It thus provides a measure of the realized return variation, without having to rely on a specific model. Hence, it becomes *observable* and therefore can be used in more standard estimation techniques.

An interesting analysis of RV is by Christiansen et al. (2012), who analyse the predictive content of financial and macroeconomic variables for monthly realized volatility. Specifically, they employ a set of 38 macro-finance (i.e. macroeconomic and finance) variables to predict the monthly realized volatility of four different asset classes: equities, commodities, foreign exchange rates, and bonds. Using Bayesian estimation techniques they attempt to find the variables that are best in predicting RV. They ultimately find there is added value in these variables to predict volatility for the different classes. More specifically, they find the strongest predictive ability in the variables that are associated with time-varying risk premia, leverage or financial distress. Interestingly, they find the best performance in the variables that are directly related to the asset classes (i.e. the finance variables). The macroeconomic variables are only weakly represented in their best prediction models. Furthermore, their research reveals that variable selection is a fruitful exercise. By selecting the most important predictor variables by means of Bayesian Model Averaging, they find that their forecast models beat autoregressive benchmarks. Their multi-asset focus, combined with a relatively large set of potential predictors is an interesting case and provides new insight in the predictability of RV by macroeconomic variables. However, the idea of linking macroeconomic variables to RV is not new. Mele (2007) suggests that macroeconomic variables have the potential to proxy for 'time-varying risk premia', which makes it reasonable to expect that they have at least some value in predicting RV. The relation between macroeconomic data and financial volatility has been investigated further in the past, but often in a more limited way or in a different context. Schewert (1989) investigates the relation between stock market volatility and several macroeconomic variables, but all in a separate linear regression model. He thereby ignores possible interaction between different variables. Further, Jones et al. (1998) examine the reaction of daily bond market volatility to macroeconomic news releases, using GARCH-type models. In a more recent paper, Beltratti and

Morana (2006) further investigate the relation between macroeconomic data and stock market volatility. They highlight that the relation of macroeconomic variables and volatility is subject to structural breaks. This makes it increasingly difficult to find a strong and consistent relation between macroeconomic variables and volatility. However, taking these breaks into account they find long-run relationships between stock market volatility and macroeconomic variables. Interestingly they find that the causality works both ways, but the evidence of causality running from macroeconomic data to stock market volatility is stronger. Pierdzioch et al. (2008) analyse the relation between monthly stock market volatility and a set of macroeconomic variables, using data from Germany. As they use an expanding window, including all information that is available at t (i.e. today), they implicitly account for structural breaks in volatility. Further, they use a recursive modelling approach to determine the optimal subset of macroeconomic variables to use for forecasting each period. As realized volatility also aims to proxy for risk, it provides an almost direct relation with macroeconomic variables. However, which variables are most useful for forecasting can change each period. Using variable selection techniques to find this optimal subset, such as the recursive approach by Pierdzioch et al. (2008), therefore seems crucial for further research on this topic.

The primary goal of this paper is to shed more light on the relation between macro-finance variables and RV, taking into account the findings of above mentioned papers. This paper further develops the framework established in Christiansen et al. (2012), using a broad set of different estimation and evaluation techniques. More specifically, the extension lies in applying a range of different 'frequentist' estimation techniques, appropriate for analysing large sets of possible explanatory variables, to the same set of variables and different asset classes. The methods aim to select the variables that contribute most to explaining RV. However, as opposed to Bayesian techniques they do not rely on prior information. The estimation techniques applied here are primarily drawn from Kim and Swanson (2014a) and can be subdivided in four different classes: penalized regressions, dynamic factor models, forecast combinations and bootstrap aggregation. The motivation of using a wide range of different techniques is twofold. Firstly, all techniques provide a different perspective on forecasting using a set of large predictors. Secondly, as they all are capable of dealing with many predictors it is beforehand difficult to say which model is best for forecasting RV. Using many different techniques provides an implicit robustness check; is the performance of one model simply due to luck or is it consistent for all the different estimation techniques?

Furthermore, the resulting forecasts are evaluated extensively, both from a statistical and economic perspective. The statistical evaluation focuses on the accuracy of the forecasts and their ability to predict the right direction of realized volatility. The economic evaluation entails an analysis of the forecasting performance via two direct applications. Firstly, the forecasts are used to construct monthly Value-at-Risk forecasts. Secondly, the forecasts are used in a mean-variance strategy to create monthly portfolios.

Summarizing, the paper provides an extensive analysis of the predictability of realized volatility and, in particular, the added value of macro-finance variables. The analysis is executed on different estimation techniques that can be subdivided in four different groups, which all

approach the set of predictors differently. The forecasts resulting from the different estimation techniques are evaluated both statistically and economically, to provide a complete overview. The main research question which follows from the above is: "Are macroeconomic and finance related variables able predict monthly realized volatility?" This question is accommodated by the following sub questions:

1. Which of the used estimation techniques turn out to be most appropriate? What is the best explanation for this? Is there cross-asset consistency?
2. If the macroeconomic variables are able to predict realized volatility, which variables are most relevant for each asset class? Is there cross-asset consistency?
3. What is the best application of realized volatility forecasts?

The analysis reveals that monthly realized volatility is better predictable when using (a subset) of the macro-finance variables, compared to a simple autoregressive model. However, the extent to which predictability of realized volatility is possible differs per asset class. The added value of the variables is most clearly visible for predicting the realized volatility of the equities class. Further, which technique is most appropriate for predicting realized volatility also is different for each class. The evaluation of the Value-at-Risk forecasts and mean-variance portfolio returns reveal that the models perform quite well in general for these two applications. However, it seems that the largest part of this good performance is due to the autoregressive component that is included in each model. Though, the techniques based on forecast combinations appear to perform best on an overall basis, as its forecasts are significantly more accurate than the autoregressive model. Furthermore, which variables are most helpful also is different per asset class. Only limited consistency is found in the variable selection procedure and forecasting performance. For equities, not surprisingly, equity market related variables appear most important. In addition, monthly inflation is variable that is selected often to forecast with. The most selected variables for the commodities class show strong overlap with the equities class, as it also selects predominantly equity market and macroeconomic variables. The foreign exchange rates and bonds class also show some overlap, primarily macroeconomic variables. There appears to be some consistency, as inflation is for all classes an important variable. To shed further light on this issue several extensions are possible, primarily in the choice of estimation techniques and data to which these methods are applied to.

The paper is structured in 7 sections, which includes this introduction as Section 1. Section 2 gives a thorough analysis of the data and a further substantiation of the potential predictor variables. Subsequently, Section 3 provides an individual discussion of the different estimation techniques applied to the data. Further, Section 4 presents the techniques with which the forecasts will be evaluated. The results corresponding to these evaluation techniques are presented and discussed in Section 5. The penultimate Section 6 entails a discussion of the paper and provides possible further extensions. Lastly, Section 7 presents the overall conclusion and answers the main hypotheses of the paper.

2 Data

The dataset comprises the 'short' sample of Christiansen et al. (2012).[1] The sample runs from January 1983 to December 2010, which amounts to a total of 336 monthly observations. It consists of 38 potential predictor variables and four realized volatility series. The RV series each represent an asset class: equities, commodities, foreign exchange rates (henceforth referred to as 'FX-aggregate') and bonds. The equities class is approximated with daily returns of the S&P 500. For bonds it is computed over the daily returns (based on closing prices) of the 10-year Treasury note futures contracts traded on the Chicago Board of Trade (CBOT). Christiansen et al. (2012) point out that the advantage of using futures contracts is that they are very liquid, which allows for a straightforward daily return calculation. Furthermore, the commodities class is represented by the Standard & Poor's GSCI commodity index and the foreign exchange returns are approximated by a basket of exchange rates of different currencies *vis-à-vis* the US Dollar (see Christiansen et al., 2012, for more detailed information). The underlying daily return series for the first three classes are obtained from Datastream, for the purpose of the economic evaluation. The FX-aggregate returns are not included, as its exact construction is unclear. Lastly, the monthly risk-free rate is approximated by the monthly 3-month T-Bill rate, as provided via the database of the Federal Reserve.[2] The 3-month T-Bill rate is available for the full length of the dataset and not discontinued at a certain point. It is therefore favoured over the the 1-month or 4-week T-Bill rate, as they are both discontinued in 2001.

The remainder of this section is structured as follows: Firstly, the anatomy of monthly realized volatility is explained and discussed; Secondly, a preliminary analysis of the realized volatility will be presented (e.g. summary statistics, correlation between the different classes); Lastly, the 38 predictor variables will be further analysed, in particular for their possible relation with realized volatility.

2.1 Realized Volatility

The *monthly* realized volatility of month t for asset class k is computed as the sum of squared *daily* returns of that month:

$$\mathrm{RV}_{k;t} = \sqrt{\sum_{d=1}^{D_t} r_{k,t;d}^2}, \tag{1}$$

where d is the trading day of a month and D_t the last trading day of month t. The daily returns (r) are computed as log returns.[3] Assuming that the *average* daily return is zero, RV is in fact the sum of squared deviations from its mean. Based on this assumption, RV is closely related to the monthly sample standard deviation of daily returns. As standard deviation is a measure to estimate uncertainty, the link between risk and realized volatility becomes clear. As realized volatility is measured over past observations and thus is backward looking, it can best be seen as the (perceived) risk of month t.

[1] Available at the **Journal of Applied Econometrics Data Archive**.
[2] Available at the **Official webstie of the Federal Reserve**.
[3] $r_d \equiv \log(P_d/P_{d-1})$, where P is the price of the asset.

Andersen et al. (2006) point out that logarithmic time series of RV typically are much closer to being homoskedastic and (unconditionally) normally distributed (i.e. the RV is close to being log-normally distributed). As can be observed from Table 1 below, this also holds for the realized volatility series. The 'raw' series represent the untransformed series of (1) and the 'log' series represent the log transformed RV series. All of the raw series are heavily skewed and exhibit strong excess kurtosis, which leads to a rejection of the null hypothesis of the Jarque-Bera test on normality. Compared to its raw series, the distributional properties of log RV (henceforth referred to as 'LRV' or 'y') are more favourable and closer to being normally distributed. The LRV will therefore be used for further analysis.

Although the log transformation is favoured, it can be problematic as the ultimate goal is to forecast the raw series. Using log transformed series as dependent variable in, for example a linear regression model, leads to log forecasts. Hence, the predicted value at $t+k$ should be converted back to its 'raw' value. The naive approach is now to simply take the exponent of this predicted value and assume that:

$$\widehat{\mathrm{RV}}_{t+k|t} = \exp\left(\widehat{\mathrm{LRV}}_{t+k|t}\right) \tag{2}$$

However, the k-step prediction of LRV is in fact the conditional expectation of LRV at $t+k$, using all information available at t ($\mathcal{I}_t$). As the exponent is a convex function, the transformation is not straightforward. The inverse of the expectation of a convex function introduces Jensen's Inequality (1906), which states the following:

$$\mathbb{E}[f(X_{t+k})|\mathcal{I}_t] \geq f\left(\mathbb{E}_t[X_{t+k}|\mathcal{I}_t]\right) \quad \text{or}$$

$$\mathbb{E}[f(X_{t+k})|\mathcal{I}_t] = f\left(\mathbb{E}_t[X_{t+k}|\mathcal{I}_t]\right) + \nu_{t+k}, \tag{3}$$

where ν_{t+k} represents the error term not captured by the naive transformation of (2) and $f(\cdot)$ is a convex function, such as the exponent. The most obvious solution to find ν_{t+k} is to use the relation between the normal and lognormal distribution, which provides an accurate transformation. However, this requires two strong assumptions. The first is that the (log) series have to be normally distributed. As the LRV series for equities is non-normal distributed, it is unclear whether this assumption is appropriate. Secondly, the transformation requires an estimate of the conditional volatility of the dependent variable. This estimate is easily obtained when using OLS, but is quite tedious and often unknown for more advanced techniques. On top of that, Andersen et al. (2006) argues that for short periods the correction term can be considered negligible. As only one-step forecasts are considered here, albeit monthly, the naive transition of (2) will be used to obtain the RV forecasts.

2.2 Preliminary analysis

2.2.1 Realized volatility series

Table 1 below presents an overview of the characteristics of the dataset in five panels. It can be observed from Panel A that the LRV for different asset classes show similarities, yet all of

them hold unique characteristics. Most important is that RV for equities and commodities are non-normal, as the Jarque-Bera test statistic rejects the null hypothesis at a 5% significance level. The stationarity of the data is informally tested by testing whether the AR(1)-coefficient (ϕ_1) is significantly different from 1, which would lead to conclude that the time-series has no unit-root. This appears to be the case and is confirmed by the Augmented-Dickey Fuller test, which rejects the null-hypothesis of a unit root in all four cases.

Panels C to E provide insight in how the different classes relate to each other, in other words, how 'integrated' they are. Panel C provides the standard cross-correlation values between the four classes. The equity class correlates moderately with all classes and, quite remarkably, most with the commodities class. Other than between the bonds and FX-aggregate class, the correlations are negligible. Panel D shows the correlations between the RV of class i at t and the RV of class j at $t-1$. The diagonal contains the first-order autocorrelation for each class, which corresponds with the AR(1) coefficients of Panel B. The 'auto-correlation' between classes can be relatively high (e.g. equities and commodities, bonds and FX-aggregate), but also negligible (e.g. FX-aggregate and commodities, bonds and commodities). This relation is further investigated in Panel E, where the one-month lagged series of class j is added to the AR(1) model of class i:

$$y_{i,t+1} = c + \phi_{1,i} y_{i,t} + \phi_{2,j} y_{j,t} + \varepsilon_{t+1}, \tag{4}$$

which aims to reveal whether the lagged value of class j has additional predictive information for class i, on top of the lagged value of class i. The results are disappointing, as only the t-values between stocks and commodities is significant at a 5% significance level. In further analysis on the predictability of RV, the lags of other asset classes are disregarded.

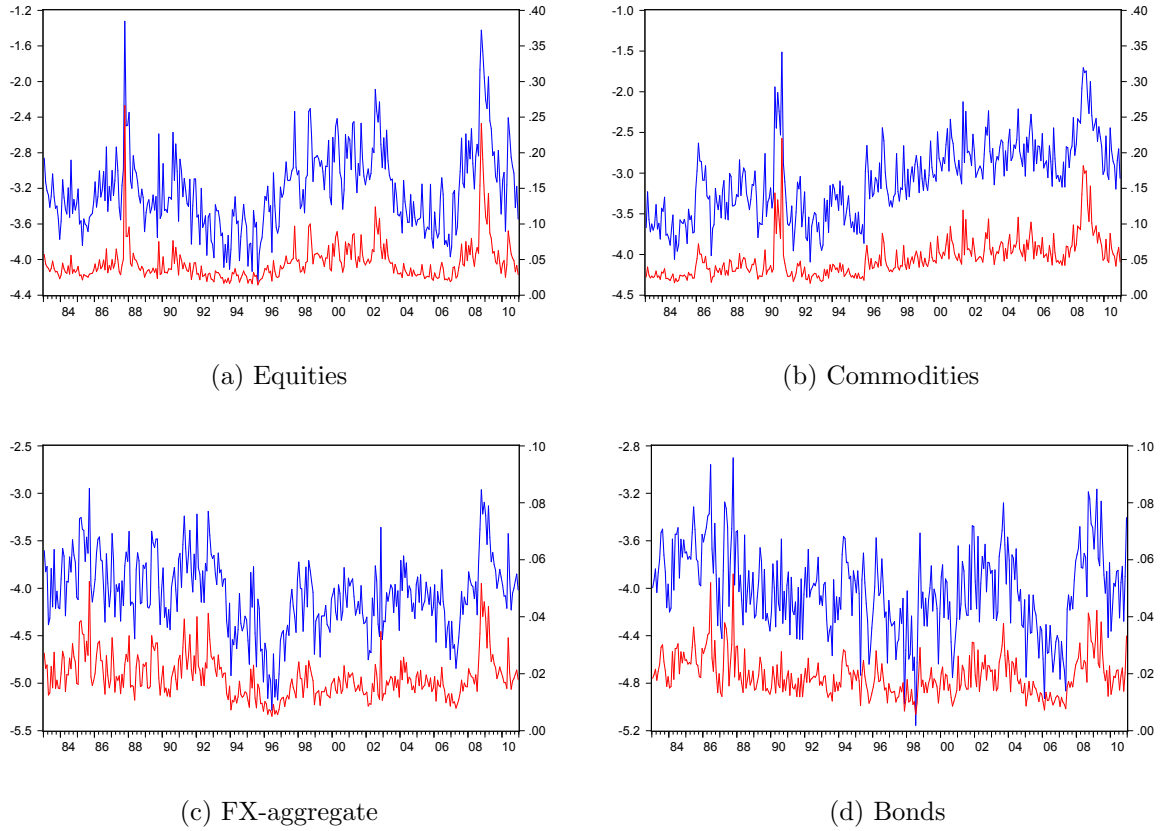
Table 1: Descriptive statistics of the realized volatility series

	Equities		Commodities		FX-aggregate		Bonds	
	Log	Raw	Log	Raw	Log	Raw	Log	Raw
A. Distribution								
Mean	-3.215	0.045	-3.080	0.052	-4.059	0.019	-4.014	0.019
Standard Deviation	0.452	0.028	0.466	0.028	0.394	0.008	0.356	0.007
Skewness	0.814	3.737	0.349	2.156	-0.111	1.264	-0.005	1.220
Kurtosis	4.436	24.404	3.032	10.357	3.355	5.513	3.114	5.795
Jarque-Bera p-value	0.000	0.000	0.033	0.000	0.293	0.000	0.912	0.000
B. Stationarity								
AR(1) coefficient ($\hat{\phi}_1$)	0.712		0.771		0.657		0.582	
t-statistic ($\hat{\phi}_1 - 1$)	-7.483		-6.600		-8.335		-9.325	
R^2	0.507		0.597		0.433		0.335	
Augm. Dickey-Fuller p-value	0.000		0.004		0.000		0.002	
C. Cross-correlation								
Equities								
Commodities	0.480							
FX-aggregate	0.250		0.025					
Bonds	0.325		0.044		0.338			
D. Lagged cross-correlation								
Equities	0.712		0.383		0.206		0.244	
Commodities	0.403		0.773		0.009		-0.011	
FX-aggregate	0.169		-0.037		0.658		0.257	
Bonds	0.190		-0.009		0.278		0.579	
E. Extended AR(1) model: $y_{t+1,i} = c + \phi_{1,i} y_{t,i} + \phi_2 y_{t,j} + \varepsilon_{t+1}$								
t-statistic ($\hat{\phi}_{2,E}$)	–		0.373		1.031		1.250	
t-statistic ($\hat{\phi}_{2,C}$)	1.812		–		-0.186		-0.813	
t-statistic ($\hat{\phi}_{2,F}$)	-0.259		-1.631		–		1.462	
t-statistic ($\hat{\phi}_{2,B}$)	-1.243		-1.227		1.429		–	

Notes: (i) The 'Raw' columns refer to the realized volatility as defined in (1), and the 'Log' columns represent the log transformed Realized volatilities. (ii) The Augm. Dickey-Fuller test is applied on the 'level' series, including a trend, intercept and a number of lags. (iii) The lagged cross-correlation refers to the correlation between the RV series class i and the one month lagged series of class j.

To provide a better idea of the nature of the series, Figure 1 below shows a graphical illustration of both the log (blue) and raw (red) RV of the four asset classes. The raw series appear to be quite stable. However, there are several strong peaks in the series for all four classes. These appear to correspond with periods of large shocks (1987, black Monday) and crises (2008, financial crisis). The log transformation magnifies the behaviour of the RV, both on long and short term. The volatility appears to follow a 'wave-pattern' over the whole window, especially for equities, which signals periods of low and high volatility. The short term pattern in volatility seems to be quite volatile, as range in which the volatility moves is relatively wide. In short, all series display a changing 'behaviour' over time and also appear to act differently between asset classes.

Figure 1: Graphical illustration of the time-seris of (Log) Realized Volatility



(a) Equities

(b) Commodities

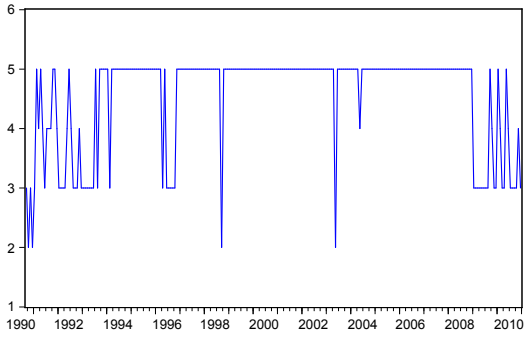
(c) FX-aggregate

(d) Bonds

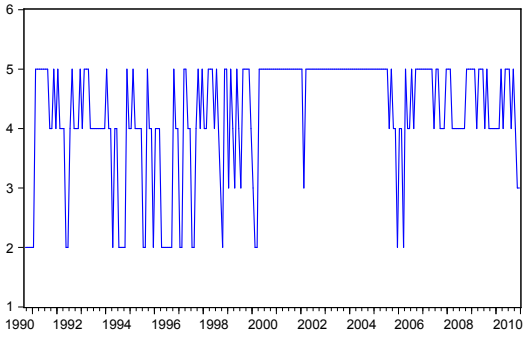
Notes: (i) The blue (red) line and left (right) vertical axis correspond to the log (raw) RV series. (ii) The horizontal axis presents the time frame in years.

The added value of the macro-finance variables will be reviewed with respect to an AR model, which makes it important to account for the dynamic behaviour of the persistence of the RV series. Fixing the number of lags would potentially lead to a suboptimal fit, which on its turn will negatively influence the quality of the analysis. To account for this problem, the number of lags that are used each estimation window is optimized separately for each asset class. The autoregressive model, up to five lags, which yields the lowest BIC value is selected as the baseline model for that sample. The BIC is favoured over other criteria as it assumed to reflect the desired trade-off between model fit and parsimony. The BIC penalizes larger models more strongly than, for example, the Akaike Information Criterion (AIC). This will generally lead to smaller models. The optimization of the number of lags is shown in Figure 2 below. All four asset classes display substantial variation in the optimal number of lags, especially in the commodities and bonds class. Higher order autoregressive models appear to be favoured. Lastly, the optimal number of lags per asset class each period appears to be unique to that class. For example, in the period of 1998 to 2002 the equities class consistently favours the AR(5) model while the optimal number of lags for the bonds class is relatively low during that period. Allowing for different lags each period appears to be the right choice, as there is little consistency of the optimal number of lags both within and between classes.

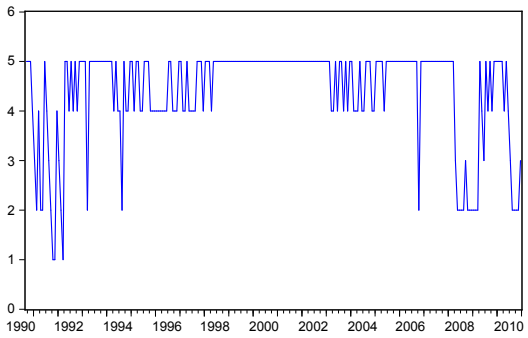
Figure 2: Dynamic behaviour of the optimal autoregressive model



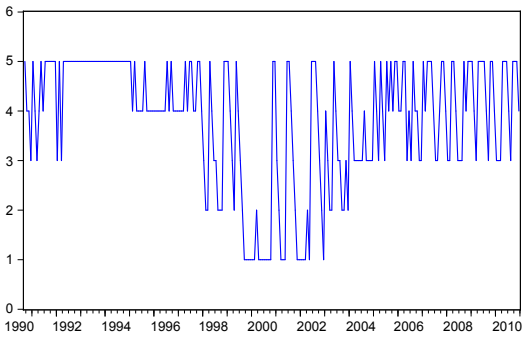
(a) Equities



(b) Commodities



(c) FX-aggregate



(d) Bonds

Notes: (i) The optimal number of lags each period is determined by the autoregressive model which has the lowest BIC value. (ii) The vertical axis represents the number of lags included in the optimal AR model in each period, which can be up to five lags. (iii) The horizontal axis presents the time frame in years.

2.2.2 Macro-finance variables

Table 2 below presents the summary statistics of the complete set of possible predictor variables. The dataset is subdivided in five different groups, consisting of variables that either target a specific asset class (A – C) or variables that proxy for risk not directly related to a particular asset class (D and E):

A. *Equity Market Variables and Risk Factors.* Contains primarily variables that are drawn from the literature on predicting equity returns. Though a good variable in terms of predicting returns is not necessarily useful for predicting realized volatility, there is a strong link. Besides the fact that the RV is based on returns, equity return models often include variables that proxy risk. Possibly the most well-known variable for predicting equity returns, the Fama-French model (Fama and French, 1992), relies solely on risk proxies (i.e. MKT, SMB and HML). Though analysed from a cross-sectional point of view, the concept could hold over time as well. Variables such as the D-P and E-P (see, amongst others Brandt, 2009; Campbell and Shiller, 1988) can be argued according to the same logic. Furthermore, the set holds a short term reversal factor and the return on the MSCI World Index, to

capture the leverage effect. This effect states that negative returns are often associated with higher volatility (Black, 1976), changing patterns in (lagged) returns could therefore correspond with similar patterns in realized volatility. Lastly, the S&P500 monthly turnover is included, as it is claimed to proxy for market consensus and corresponding uncertainty on future values (Scheinkman and Xiong, 2003).

B. *Interest Rates, Spreads and Bond Market Factors.* Holds variables most closely related to bond returns. The T-Bill rate represents the short term return on (effectively riskless) government bonds. It appears to be a useful predictor for predicting equity returns (Ang and Bekaert, 2007). Furthermore, the group contains the long term bond return and term spread. Increased uncertainty or a change in investor sentiment might lead to a shift towards long-term bonds, causing changes in its price (and underlying interest rates). In other words, strong changes in bond returns might signal increased future uncertainty/volatility. The term spread, the difference in the yields between short- and long term bonds, has also been documented as a useful variable, both for equities and bonds (Fama and French, 1993). Lastly, the Cochrane-Piazzesi factor (2005) is included as it should hold information which is not included by common term structure models, let alone a straightforward term spread variable.

C. *FX Variables and Risk Factors.* Includes the average forward discount and two other foreign exchange rate specific variables based on the work of Lustig et al. (2011). The forward discount measures the difference between domestic current and future spot exchange rates to a certain currency. A differential signals that future spot rates might appreciate or depreciate, depending on the sign of the difference. The average forward discount here is the average of the exchange rate of the US dollar against various major currencies, thereby reflecting the overall expected position of the US dollar from a global market point of view. The factors by Lustig et al. (2011) are derived from the first two factors of principal components analysis on a broad set of currency portfolio returns. The first factor is identified as the dollar risk factor, which can be interpreted as the average excess return on currencies. The second factor, referred to as the carry risk factor, represents the difference in high and low interest rate currencies.

D. *Liquidity and Credit Risk Variables.* A change in credit risk is considered as a fourth possible signal for increased market uncertainty, represented here by the default spread. Measured as the yield spread between BAA and AAA bonds, it is argued in Christiansen et al. (2012) that it tends to be positively related to leverage, the ratio of equity over debt. An increase in credit risk will make borrowing more expensive (i.e. higher interest rate), which could lead to a shift towards equity. According valuation models (see amongst others Merton, 1974) an increased leverage should on its turn affect volatility. Furthermore, the TED spread (difference between three-month LIBOR and T-Bill rate) and FX Average Bid-ask spread are used as proxies for the funding liquidity in interbank markets and FX market liquitidy, respectively. Additionally, the liquidity factor of Pastor and Stambaugh (2003) is included to proxy for stock market liquidity. The relation between liquidity and realized volatility is not obvious. High liquidity signals a state of the market where assets are readily available.

This easy access might suppress the bid-ask spread, which could lead to more 'stable' prices and corresponding lower realized volatility.

E. *Macroeconomic Variables.* The final and largest group contains a selection of general macroeconomic variables, representing the state of the general economy. Given that these variables reflect the state of the economy, they could attribute to predicting realized volatility as well. As already briefly mentioned in the introduction, Mele (2007) points out that macroeconomic variables potentially proxy for (time-varying) risk premia. This would directly link the macroeconomic variables to realized volatility, though different variables could proxy for different risk premia.

As some groups of variables are more closely related to specific asset classes, it is not expected that exactly the same set of variables will be selected as most valuable predictors in each class. However, as revealed by Fama and French (1993), their could be some form of predictor 'integration'. The correlation between different asset classes can be substantial, as shown in Table 1, which makes it reasonable to think that at least some variables work for multiple asset classes. Furthermore, it is implicitly assumed that the causal relation between the macro-finance variables and realized volatility is that the former explains the latter. The paper holds the perspective that the macro-finance variables potentially hold information relevant for the future movement of the realized volatility, and not the other way around.

Table 2: Descriptive Statistics of the full set of potential predictor variables

Variable	Abbr.	Mean	Std.	Skew.	Kurt.	JB.
A. Equity Market Variables and Risk Factors						
Dividend-to-Price Ratio (Log) (*)	D-P	-0.003	0.046	0.776	6.437	0.000
Earnings-to-Price Ratio (Log)	E-P	-3.023	0.428	-1.312	6.509	0.000
US Market Excess Return	MKT	0.006	0.046	-0.909	5.756	0.000
Size Factor	SMB	0.001	0.032	0.816	11.458	0.000
Value Factor	HML	0.004	0.032	0.049	5.529	0.000
Short Term Reversal Factor	STR	0.004	0.034	0.176	8.347	0.000
S&P500 Turnover	TUR	0.013	0.165	-0.065	3.377	0.330
Return MSCI World	MSC	0.007	0.043	-1.197	6.424	0.000
B. Interest Rates, Spreads and Bond Market Factors						
T-Bill Rate (Level) (*)	T-B	0.000	0.002	-0.947	5.118	0.000
Rel. T-Bill Rate	RTB	-0.002	0.008	-0.256	2.756	0.105
Long Term Bond Return	LTR	0.008	0.030	0.201	4.804	0.000
Rel. Bond Rate	RBR	-0.002	0.006	-0.345	4.537	0.000
Term Spread (*)	T-S	0.000	0.003	0.343	3.669	0.002
Cochrane Piazzesi Factor	C-P	1.222	1.559	0.411	3.337	0.004
C. FX Variables and Risk Factors						
Dollar Risk Factor	DOL	0.121	2.189	-0.349	4.028	0.000
Carry Trade Factor	C-T	0.055	2.580	-0.688	4.371	0.000
Average Forward Discount	AFD	0.177	0.186	0.883	7.870	0.000
D. Liquidity and Credit Risk Variables						
Default Spread	DEF	0.010	0.004	2.525	12.644	0.000
FX Average Bid-ask Spread	BAS	0.129	0.050	1.949	7.629	0.000
Pastor-Stambaugh Liquidity Factor	PS	-0.026	0.068	-1.759	10.456	0.000
TED Spread	TED	0.007	0.005	1.800	8.767	0.000
E. Macroeconomic Variables						
Inflation Rate, Monthly	INM	0.002	0.003	-1.380	11.274	0.000
Inflation Rate, Yearly	INA	0.029	0.013	-0.479	4.401	0.000
Industrial Production Growth, Monthly	IPM	0.002	0.007	-1.375	10.376	0.000
Industrial Production Growth, Yearly	IPG	0.022	0.043	-1.600	7.455	0.000
Housing Starts	H-S	-0.023	0.248	-0.040	4.565	0.000
M1 Growth, Monthly	M1M	0.004	0.008	1.511	13.751	0.000
M1 Growth, Yearly	M1A	0.048	0.050	0.293	2.307	0.003
Orders, Monthly	ORM	0.001	0.018	-0.044	3.086	0.899
Orders, Yearly	ORA	0.012	0.069	-1.509	8.491	0.000
Return CRB Spot	CRB	0.002	0.027	-1.761	17.660	0.000
Capacity Utilization	CAP	0.000	0.007	-1.143	9.120	0.000
Employment Growth	EMP	0.001	0.002	-0.364	7.392	0.000
Consumer Sentiment	SEN	0.000	0.047	0.069	5.653	0.000
Consumer Confidence	CON	0.000	0.083	-0.291	9.908	0.000
Diffusion Index	DIF	8.669	16.929	-0.635	3.563	0.000
Chicago PM Business Barometer	PMB	55.173	7.323	-0.379	3.390	0.006
ISM PMI	PMI	52.102	5.351	-0.399	3.789	0.000

Notes: (i) Variables with a (*) are non-stationary according to the Augmented Dickey-Fuller (ADF) test, including a constant, trend and a number of lags. To correct for the non-stationarity, they are transformed to their first differences (i.e. $y_{d,t} = y_t - y_{t-1}$). (ii) Variables with a (Log) are transformed to logarithmic variables, i.e. $X(\log) \equiv \log(X)$. (iii) The JB column holds the p-values of the Jarque-Bera test for normality.

3 Estimation techniques

The large set of possible predictor variables (henceforth denoted by X) provides an almost infinite amount of different model specifications. Evaluating every possible combination of predictor variables in a linear predictive regression model implies the estimation of 2^{38} (over 274 billion) different models. Besides the computational intensity of this approach, Stock and Watson (2006) reveal that there are many pitfalls on the road when analysing a large set of regressors. The main problem comes forth from the relation between the number of regressors (K) and observations (N), as the forecast error variance of a model estimated with OLS is proportional to K/N. Clearly, if K is large relative to N, this error become non-negligible. In fact, it could occur that a model without any regressors has a better forecasting performance than a model with all regressors. In other words, it is better to disregard all the information in the possible predictor variables. Thus, parsimony is key when trying to 'improve' upon a model without any predictors. However, this poses the question which predictors to select.

To address this issue, a wide range of 'robust' estimation techniques are applied, primarily based on Kim and Swanson (2014a), which all will be explained into detail in the sections below. The motivation of using a wide range of different techniques is twofold. On the one hand, all techniques provide a different perspective on forecasting using a set of large predictors. On the other hand, using many different techniques provides an implicit robustness check; is the performance of one model simply due to luck or is it consistent for all the different estimation techniques? Given that many of the methods that are used in this paper rely on input parameters or 'hyper-parameters' which have to set in advance, using different methods will also account implicitly for the sensitivity to these parameters. This is especially important for the purposes of this paper, as it aims to identify the most important macroeconomic determinants for predicting realized volatility. The techniques used in this article can roughly be divided in four classes, namely: penalized regressions, dynamic factor models, forecast combinations and bootstrap aggregation.

3.1 General set-up

Though the possible macro-finance predictors are expected to hold information about future values of LRV, their actual relevance should be reviewed in light of their added value to a simple model which solely contains a number of lagged values of LRV. As shown in the previous section, the LRV series are highly persistent, but this persistence varies over time. In other words, consecutive values of LRV are strongly related and are very useful in forecasting future values. For this reason, each model starts in each period with an optimized number of lags, as explained in the previous section:

$$y_t = \beta_0 + \beta_{\mathcal{L}} y_{\mathcal{L},t-1} + \varepsilon_t, \tag{5}$$

where $y_{\mathcal{L}}$ is a matrix that can contain up to five month lagged values of y and $\beta_{\mathcal{L}}$ the corresponding coefficient vector, estimated using OLS. Finding the optimal subset of predictor variables can therefore be seen as a search for a subset variables which improve upon the autoregressive

fit.

To analyse the above, a rolling window of 90 observations is used as estimation window, which translates to 7.5 years of monthly observations. The choice of the number of observations is the result of a trade-off between statistical reliability and economic validity. Considering the number of possible predictors that can be included, using less than 90 observations could severely endanger the reliability of the methods. From a pure statistical point of view, using more than 90 observations would actually be more favourable. However, from an economic point of view this is considered undesirable. By using data for a certain period, it is implicitly assumed that the parameters are constant during that period. As the stability of the relation between RV and the economic variables is unclear, it makes sense to limit the estimation sample to most recent observations. As the aim is to forecast only one month ahead, the 'current' relation between RV and the explanatory variables is most likely found in these observations.

3.2 Penalized regressions

Penalized regressions form the first class of estimation techniques with which the predictability of RV is analysed. The class is represented in this article by four variations: Ridge regressions, the least absolute shrinkage selection operator (lasso), elastic net (EN) and least angle regressions (known as LARS). Penalized regressions in general can be defined as an adjusted OLS minimization problem. Rather than only minimizing the sum of squares, as with OLS, it also takes into account an additional term, often labelled as the 'penalty' term:

$$\hat{\beta}^{pr} = \underset{\beta}{\operatorname{argmin}} ||y - X\beta||_2^2 + \lambda ||\beta||_p, \tag{6}$$

where $||x||_p = (\sum_i |x_i|^p)^{1/p}$, λ a positive scalar that controls the strength of the penalty, and β a vector of length K. For any $\lambda > 0$, this will generally lead to estimates different from the OLS solution ($\hat{\beta}^{ols}$). As the minimization in fact poses a restriction on the magnitude of the betas, this will generally lead to betas closer to zero. In other words, the penalty function 'shrinks' the betas towards zero. The underlying idea of this shrinkage can perhaps best be explained with use of the decomposition of the mean squared error (MSE) of the estimated betas into the squared bias and variance of $\hat{\beta}$:

$$\text{MSE}(\hat{\beta}) = \mathbb{E}\left[(\hat{\beta} - \beta)^2\right]$$

$$\text{MSE}(\hat{\beta}) = \left[\mathbb{E}(\hat{\beta} - \beta)\right]^2 + \left[\mathbb{E}\left(\hat{\beta} - \mathbb{E}(\hat{\beta})\right)\right]^2$$

$$\text{MSE}(\hat{\beta}) = \text{Bias}(\hat{\beta})^2 + \text{Var}(\hat{\beta}) \tag{7}$$

As the betas estimated by OLS are unbiased, its MSE is equal to its variance. Since the resulting betas from a penalized regression move away from the OLS solution by directing them towards smaller values, they will likely introduce some bias into equation (7) above. However, if this *increase* in bias is at least offset by a *decrease* in the variance, the penalized regression has improved upon simple OLS in terms of the MSE. However, the aim is to increase the

forecasting performance and not to minimise the MSE. The penalized regressions will therefore only improve on OLS if this *bias-variance trade-off* also leads to an increased forecasting performance. Suppressing the size of β via the shrinkage parameter makes it less vulnerable to outlying observations and other sample specific effects. This on its turn could improve the forecasting performance of y, the actual goal of this paper. Stock and Watson (2006) point out that OLS with many predictors typically perform badly in terms of (forecasting) accuracy and (i.e. it will lead to estimates of the parameters which are difficult to relate to underlying theory). Letting OLS decide on the parameter values, without imposing restrictions, potentially leads to a model fit which is specific to the sample it is based on. Guiding the parameter estimates to smaller values, in fact tries to prevent that the parameter estimates become too sample specific. As X contains a lot variation in terms of origin and relation to RV, it provides reason to think that penalized regressions will lead to improved forecasts, at least compared to the model which simply includes all predictors. The aim is, however, to improve on the autoregressive benchmark.

In the application of all forms of penalized regressions used in this paper, both y and X are centred around their means. This allows to exclude an intercept term in the minimization function. As the intercept often approximates the mean of y, of which the mean is zero when centred, including an intercept will not influence the estimates of β. More importantly, variables in X are scaled to have unit sample variance. Not scaling (or standardizing) the possible predictor variables would lead to a bias towards variables with very large values compared to y (i.e. this leads to small betas by nature). Standardizing the variables solves this issue. Note that y is not scaled to have unit variance. Standardizing y will also affect the coefficient estimates. Subsequently, this might lead to a different optimal shrinkage parameter. As the shrinkage parameter is optimized in this paper, which will be explained later in this section, standardizing y is not considered to be important. Whether y is standardized or not, a shrinkage parameter will be selected which leads to the 'optimal' solution.

3.2.1 Ridge regressions, lasso and elastic net

Two well-known forms of penalized regressions are the ridge regression (Hoerl and Kennard, 1970) and lasso (Tibshirani, 1996), which use $p = 2$ and $p = 1$ as norm in the penalty term, respectively:

$$\hat{\beta}^{ridge} = \underset{\beta}{\operatorname{argmin}} ||y - X\beta||_2^2 + \lambda ||\beta||_2^2 \tag{8}$$

$$\hat{\beta}^{lasso} = \underset{\beta}{\operatorname{argmin}} ||y - X\beta||_2^2 + \lambda ||\beta||_1 \tag{9}$$

The methods are closely related, as they only differ in the penalty function they employ: A 'squared' norm in the case of ridge regressions and an 'absolute' norm for the lasso. The effect of the shrinkage parameter and the relation of the estimated coefficients of the ridge regression and lasso with the estimated coefficients according to OLS, becomes clear in an orthogonal setting (i.e. the columns in X are uncorrelated) (amongst others, see Kim and Swanson, 2014b). In

this setting, the above optimization functions can be rewritten to:

$$\hat{\beta}^{ridge} = \frac{\hat{\beta}^{ols}}{1+\lambda_1} \quad \text{and} \quad \hat{\beta}^{lasso} = \left(\left|\hat{\beta}^{ols}\right| - \lambda_2/2\right)_{pos} \text{sign}\left(\hat{\beta}^{ols}\right), \tag{10}$$

where $|x|$ is the absolute value of x, sign(x) is equal to ± 1 and $(x)_{pos}$ is the positive part of x (i.e. all negative elements are removed). From above equation it is visible that the OLS coefficients in a ridge regression are linearly scaled by the shrinkage parameter. The conversion of the OLS coefficients to lasso estimates is more complex. Only those coefficients of which the absolute value is larger than half the shrinkage parameter are retained. The lasso therefore has the capability to set the coefficients at exactly zero, while the ridge regression has not.[4] This generally leads to more parsimonious models for the lasso. How the different penalties operate and why they differ can further be explained by rewriting the general penalized regression expression of (6) to a constrained optimization problem:

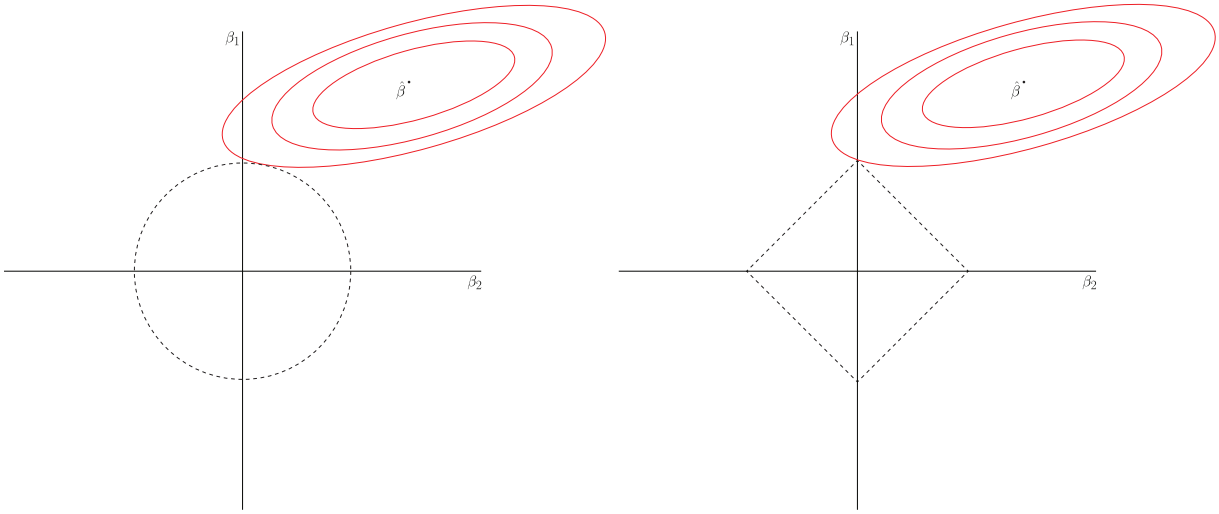
$$\hat{\beta}^{pr} = \underset{\beta}{\text{argmin}} ||y - X\beta||_2^2, \qquad \text{subject to } ||\beta||_p \leq t, \tag{11}$$

with t as parameter that controls the strength of the shrinkage. The interaction between the minimization of the sum of squares and the restriction is illustrated in Figure 3 below. The example limits itself to two regressors, in order to be able to illustrate the effects clearly. The red ovals and dotted lines in Figure represent the confidence ellipse of the estimated betas ($\hat{\beta}$) and the constraints, respectively. The shape of the constraint is determined by the combinations of values for β_1 and β_2 of which the p-norm is equal to t.

In the case of simple OLS, the confidence ellipse of $\hat{\beta}$ will position itself at a point where it minimizes the sum of squares. However, in the case of penalized regressions it also has to satisfy the restriction. Graphically this means that the confidence ellipse has to find a point where it is tangent with the given restriction. This limits the size of the estimates of beta, as the restriction only allows a limited number of combinations via t. The smaller t is, the smaller the shape will be and also the estimates of beta. It is therefore obvious that t controls the strength of the shrinkage. From Figure 3a it is visible that for $p = 2$ in the ridge regression, the restriction becomes a circle. The tangent point can therefore only be close to the y-axis, where β_2 is zero, but can never lie exactly on the y-axis without intersecting the boundary line of the restriction. The absolute norm of the lasso ($p = 1$, Figure 3b) leads to a tilted rectangle, which does provide a tangent point which lies exactly on one of the axes. The lasso, which uses the absolute norm, therefore implicitly executes variable selection. As a result, the lasso generally leads to a more parsimonious model than the ridge regression (Tibshirani, 1996).

[4] Coefficients are considered to be zero if they are 'effectively' zero. That is, if coefficients are smaller than 0.001 they are set at zero. Given the fact that the variables are standardized to have a zero mean and unit variance, the impact of a variable is considered negligible when the coefficient is that small.

Figure 3: A graphical illustration of the penalty functions of the ridge regression and lasso



(a) The squared L2-norm

(b) The 'absolute' L1 norm

Notes: (i) The figure presents the working of the penalty functions of the ridge regression (a) and lasso (b) of a model that consists of two indepdent variables. (ii) The dotted lines represent the penalty function, which' shape is determined by connecting all possible combinations of β_1 and β_2 that lead to a norm equal to t. (iii) The red ovals represent the confidence ellipses of the beta estimates.

Still, Zou and Hastie (2005) reveal that the lasso is an inappropriate method in two cases. Firstly, when K is larger than N, the lasso selects at most N variables, given the nature of the optimization problem. Secondly, in case there is a group of variables which have high pairwise correlations the lasso tends to randomly select only one variable from that group. This variable is not necessarily the 'best predictor' and can therefore affect the predictive power of the resulting model. It is empirically observed that the ridge regression can outperform the lasso in terms of prediction as a result of these two cases (Tibshirani, 1996). As they strongly limit the use of lasso, Zou and Hastie (2005) developed the naive elastic net (NEN) regularization:

$$\hat{\beta}^{nen} = \underset{\beta}{\operatorname{argmin}} ||y - X\beta||_2^2 + \lambda_1 ||\beta||_2^2 + \lambda_2 ||\beta||_1 \tag{12}$$

This is equivalent to solving the residual sum of squares, subject to a constraint that combines the ridge regression and lasso penalty:

$$\hat{\beta}^{nen} = \underset{\beta}{\operatorname{argmin}} ||y - X\beta||_2^2, \qquad \text{subject to } (1-\alpha)||\beta||_1 + \alpha||\beta||_2^2 \leq t, \tag{13}$$

where $\alpha = \frac{\lambda_2}{\lambda_1 + \lambda_2}$. Thus, NEN uses a combination of the penalty terms of the ridge regression and lasso, which they refer to as the 'elastic net penalty'. The strength of each penalty is determined by $\alpha \in [0, 1)$. Correspondingly, NEN has the characteristics of both the ridge regression and lasso when $\alpha > 0$. The two main problems of the lasso are solved by the NEN in the following manner. The solution to the first problem, the limitation of variables to select if N is smaller than K, is solved as follows. It is best explained when using the 'augmented' representation of the NEN, which is equivalent to solving (12). The augmented representation

first augments X and y to:

$$X^*_{(N+K)\times K} = (1+\lambda_2)^{-1/2}\begin{pmatrix} X \\ \sqrt{\lambda_2}I \end{pmatrix}, \quad y^* =_{N+K}= \begin{pmatrix} y \\ \mathbf{0} \end{pmatrix},$$

where X and y, originally of length N, are extended to length $N+K$ through a unit matrix I_K of size K and zero vector $\mathbf{0}_K$ of length K, respectively. Subsequently, by defining $\gamma = \lambda_1/(1+\lambda_2)$ and $\beta^* = \sqrt{1+\lambda_2}\beta$ the solution to the NEN on augmented data can be expressed as:

$$L(\gamma, \beta^*) = \underset{\beta^*}{\operatorname{argmin}} ||y^* - X^*\beta^*||_2^2 + \gamma||\beta^*||_1 \tag{14}$$

The sample size of X^* is now $N+K$, which implies that the NEN procedure can select all K predictors. Hence, the first problem is solved. This problem is, however, no issue in this paper as the number of observations in the estimation sample is always larger than the number of predictors. The second problem is related to correlated predictor variables. Zou and Hastie (2005) argue that the lasso is unable to deal with variables in the full set of predictors which have high pairwise correlations. As a result, the lasso will randomly select one of those correlated variables, which is not necessarily the most relevant one. The elastic net, however, does not select one of these variables, but selects the whole group of these correlated variables at once. This is what Zou and Hastie (2005) refer to as the 'grouping effect'. As the set of predictors used in this paper all to a greater or lesser extent proxy for risk, it could be that one more variables are highly correlated. Tables 12, 13 and 14 show the correlations between the 38 potential predictor variables, measured over the full sample. It appears there are variables which have relatively high correlations with each other, though only a small fraction of the variables have pairwise correlations higher than 0.7. As the correlations are measured over the full sample, it is difficult to say how the correlations behave in different sub-samples. It could be that the correlations are strongly affected by certain periods of extreme observations. Still, the inability of the lasso to deal with this scenario make the use of the elastic net important. Not only because it is able to deal with correlated variables, but also because it can act as a ridge regression-lasso switching model. Using (12) to optimize the shrinkage parameters, without restricting that both should be larger than zero, allows the NEN to be equal to the ridge regression or lasso. Restricting the parameters to be larger than zero would mean that the NEN always has the properties of both the lasso and ridge regression, but does not necessarily imply that its forecasting performance is better than the ridge regression or lasso. Dropping this restriction therefore provides a great amount of flexibility, which ultimately could improve the performance of the forecasts.

Lastly, Zou and Hastie (2005) point out that there is double shrinkage in the NEN, which eventually may lead to extra bias without reducing the variance. The naive elastic net problem is solved by moving through two dimensions. This double shrinkage becomes especially clear in the expression of $\hat{\beta}^{nen}$ in an orthogonal setting:

$$\hat{\beta}^{nen} = \frac{\left(\left|\hat{\beta}^{ols}\right| - \lambda_2/2\right)_{pos}}{1+\lambda_1}\text{sign}\left(\hat{\beta}^{ols}\right) \tag{15}$$

The solution of the naive elastic net in this setting is a combination of the solutions of the

ridge regression and lasso, as presented in (10). It is directly visible that $\hat{\beta}^{ols}$ is subject to the shrinkage of both λ_1 and λ_2. In the optimization of the shrinkage parameters, the function has to evaluate all different combinations of λ_1 and λ_2. For each fixed λ_1 one has to find the optimal λ_2, thereby shrinking the parameters twice. They point out that naive elastic net does not perform that well empirically, as a result of the double shrinkage. With the elastic net (EN) they propose a straightforward correction for this, which removes the double shrinkage and retains all the good properties of the lasso and ridge regression:

$$\hat{\beta}^{en} = (1 + \lambda_1)\hat{\beta}^{nen} \tag{16}$$

The predictor variables in X are evaluated for their added value to the autoregressive model. The coefficients of this autoregressive model are not subject to the shrinkage. As the shrinkage could lead to estimates of β that are very small or even exactly zero, shrinkage on the coefficient estimates of the AR parameters could lead to the situation that a higher order lag is included, but a lower order lag is excluded. This implies that the best forecast comes from ignoring the most recent information and relying on more historical information instead. This is highly undesired. The minimization argument in (6) is therefore adjusted accordingly:

$$\hat{\beta}^* = \underset{\beta}{\operatorname{argmin}} ||y_t - y_{\mathcal{L},t-1}\beta_{\mathcal{L}} - X_{t-1}\beta_X||_2^2 + \lambda||\beta_X||_p, \tag{17}$$

where $y_{\mathcal{L},t-1}$ is the matrix of lags of y_t, and $\beta = \binom{\beta_{\mathcal{L}}}{\beta_X}$. Furthermore, a constant is excluded from the minimization procedure as y is centred around its mean.

Finding the optimal setting of a certain model forms an important aspect in all of the above described methods. To 'tune' the parameters and select the optimal number of variables to forecast with, cross-validation is employed (for example, see Hastie et al., 2001). The use of cross-validation is favoured over other validation methods, such as picking the optimal model which has the lowest value of a certain information criterion, is due the nature of the methods applied here. As explained in the beginning of this section, the penalized regressions are based around the mean-variance trade-off. If a shrinkage parameter is greater than zero, the solution will move away from the typical OLS solution. Hence, its bias increased and its variance potentially decreased. Put differently, the in-sample fit may be worse, but its out-of-sample forecasting performance may have increased compared to the model estimated with OLS. As the model fit forms an important component of information criteria, there might be a tendency towards the OLS solution. Based on a standard linear regression model, $y_t = X_{t-1}\beta + \varepsilon_t$ with $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$, information criteria (IC) in general can be described by the following equation (Zhang et al., 2010):

$$\text{IC} = \log(\hat{\sigma}^2) + \kappa_N df(\hat{\beta}), \tag{18}$$

where $\hat{\sigma}_\lambda^2$ is the maximum likelihood estimator (MLE) of σ^2, κ_N is a penalty function for number of parameters in β and/or number of observations, and $df(\hat{\beta})$ represents the degrees of freedom of the estimate of β. The IC can be broken down in two parts: the goodness-of-fit and penalty function. The first part, represented by $\log(\hat{\sigma}^2)$, determines how well the estimated model fits

to the data. The lower the MLE of σ^2, the error variance, the better the fit. The second part, $\kappa_N df(\hat{\beta})$, is the penalty function. The higher the degrees of freedom of $\hat{\beta}$ or the size of the estimation sample (N), the higher the penalty. As the degrees of freedom corresponds to the number of parameters in the model (i.e. variables), the IC penalizes complex models more strongly. Generally speaking, IC aim to balance the model fit and its complexity. An argument for choosing cross validation over information criteria, is the assumption of normality to find the MLE of the error variance. This assumption generally holds for OLS, but it not necessarily has to hold for the penalized regression estimates. The possibility exists that the OLS solution is always favoured (i.e. $\lambda = 0$) as it has by nature the best fit (i.e. it minimizes the residual sum of squares). Though, the information criteria also incorporate the complexity of the model (i.e. number of parameters) by giving a penalty to larger models. The decrease of the penalty may absorb the increase in the estimated error variance, which could result in an optimal shrinkage larger than zero, but this not necessarily has to be the case. The method of cross-validation does not rely on the underlying fitting procedure of a method and thus is independent of the distribution of errors.[5] It determines the quality of a model on its ability to forecast observations of a sample, when these observations excluded from the sample.

Technically, CV starts by setting a grid of possible tuning parameters, say $(\lambda_1, \ldots, \lambda_p)$, and subdivides the training data in N equal parts. The estimation method is then executed, given a certain fixed parameter λ_p, on all training data but the n-th part. The estimated model is used to predict the n-th part, which was excluded in the estimation. This is repeated N-times, until all parts have been excluded (and predicted) once. The next step is to collect the predictions and calculate the CV prediction error. This is then repeated for each possible value of the tuning parameter and selects the tuning parameter with the lowest CV *estimation* error. The N-fold CV-error is defined by:

$$\mathrm{CV}(\lambda) = \frac{1}{N} \sum_{n=1}^{N} E_n(\lambda), \quad \text{with} \tag{19}$$

$$E_n(\lambda) = \sum_{i \in n} \left(y_i - x_i \hat{\beta}^{n^c}(\lambda) \right)^2$$

The $\hat{\beta}^{n^c}$ in above equation is the estimated β over all but the n-th part of the data. A complication of the use of CV here, is that the order in time-series is of crucial importance. The procedure of CV as described above is to simply remove a part of the data, without explicitly taking the time-series structure into account. CV leads to 'predicting' data, using a model that has been estimated on data that comes from a later time period. To account for this, CV is modified in the following manner. For a sample of K observations, the procedure starts with a set of the first T observations, the estimation sample, which is considered to be the minimum number of observations necessary to estimate a model. This leaves $K - T$ observations left to 'cross-validate' the parameters with. In the first step, the model is estimated on these first T observations and then forecasts observations $T + 1$ as one-step ahead forecasts. In each of the

[5] Note that the selection of the optimal number of lags is executed with use of the BIC. As all the autoregressive models are estimated with OLS, the assumption of the normal distribution of the errors is appropriate.

following steps, the observation that has been predicted in the previous step is added to the starting set of T observations. After the last step, the estimation sample consists of $N-1$ observations and $K-T$ one-step ahead forecasts are obtained. The cross-validation error can now be calculated via (19), similar to standard cross-validation. This process is then repeated for each of the shrinkage parameters and the parameter which leads to the lowest CV *forecast* error is selected as the optimal parameter. In this paper, which uses a moving window of 90 observations, T is set at 60.

The optimization of the shrinkage parameters of the ridge regression, lasso and NEN is limited to the following grid: (0, 1, 2, 3, 5, 7, 10, 20, 30, 50, 70, 100). The grid reflects the trade-off between the computational burden and a sufficient amount of possibilities to optimize over. The grid contains extremes 0 and 100, which represent the cases where no shrinkage is imposed and where all coefficients are (very close to) zero. Furthermore, the values in between reflect limited, medium and heavy shrinkage. Lastly, note that the (N)EN requires two-dimensional CV, as the function contains two parameters. This grid allows the NEN to take the form of a ridge regression, lasso or unrestricted OLS, as either one or both of the parameters can be zero.[6] After finding the optimal shrinkage parameters, the one-step forecast for y_{t+1} is determined as:

$$\hat{y}_{t+1|t} = \hat{\mu}_{y,t} + y_{\mathcal{L},t}\hat{\beta}_{\mathcal{L}} + X_t\hat{\beta}^*_X, \tag{20}$$

where $\hat{\beta}^*_X$ represents the betas as a result of the optimal shrinkage parameter(s) and $\hat{\mu}_{y,t}$ represents the mean of y in the estimation period, which has to be added back as y is centred around its mean.

3.2.2 Least Angle Regressions

LARS as introduced by Efron et al. (2004) is a stepwise estimation algorithm that closely resembles the penalized regressions introduced above. The underlying idea of the LARS algorithm is to build a model step-by-step, by adding one variable at a time. The procedure starts by setting the coefficients of all possible K predictor variables equal zero, thereby having a first estimate of the fit as $\hat{\mu}_0 = X'\hat{\beta} = 0$. Subsequently, the predictor that correlates most with dependent variable y is added to the 'active set' of variables. The coefficient of this predictor is then increased in the direction of its correlation with y, until another predictor has as much correlation with the remaining residual:

$$\hat{c}\left(x_1, (y - \hat{\mu}_{(1)})\right) = \hat{c}\left(x_2, (y - \hat{\mu}_{(1)})\right), \tag{21}$$

where $\hat{\mu}_{(1)} = x_1'\hat{\beta}_1$. This equicorrelation requirement simultaneously determines the next predictor to be added to the active set as well as how to update the current 'fit'. The process continues in the same way for all predictors and is stopped when all have entered the active set. Efron et al. (2004) describe that the lasso only differs from LARS in a sign restriction it enforces on the variables in the active set after each step. The lasso requires that the sign of any non-zero coefficient $\hat{\beta}_j$, which is part of $\hat{\mu} = X\hat{\beta}$, must be equal to the sign s_j of the current

[6] In the case that either λ_1 or λ_2 is zero, then $\hat{\beta}^{nen}$ is not additionally scaled, as there is no double shrinkage.

correlation $\hat{c}_{(j)} = x_j'(y - \hat{\mu})$. If this restriction is not satisfied, the corresponding variable x_j is removed from the active set and disregarded in subsequent steps of the algorithm. LARS and lasso move parallel to each other until the sign restriction of a certain coefficient is violated, say step k. From step k and forward the LARS and lasso will not be similar and each will lead to its own unique solution.

The LARS algorithm by Efron et al. (2004) is followed here and slightly adjusted to fit time-series data (see also Kim and Swanson, 2014b). The process is initialized by estimating the autoregressive model on (5) on a demeaned y. The corresponding residuals (R) are retained and the possible predictors in X are standardized to have a mean of zero and unit variance. The objective is to forecast these residuals and improve upon the autoregressive fit. In each of the following steps i, the variable that correlates most with the current estimate, $\hat{\mu}_{(i)}$, is added to the active set (i.e. set of variables that is incorporated in the current estimate of μ). The estimate of μ is updated accordingly in each step, until the set of predictors is either exhausted or stopped at step N, if the number of observations (N) is smaller than the number of predictors. The exact algorithm is described below:

LARS Algorithm

1. For $i = 1, \ldots, M$, where $M = \min(K, N)$:

   (a) Calculate for all predictors their 'current' correlation with the remaining residual after subtracting $\hat{\mu}_{(i-1)}$ as:

   $$\hat{c}_{(i)} = X'(R - \hat{\mu}_{(i-1)}) \quad \text{with} \quad \hat{C}_{(i)} = \max_j |\hat{c}_{(i)}| \quad \text{for } j \in \mathcal{G}_{(i-1)},$$

   To initialize the process set $\mu_{(0)} = \hat{R}$ and add the predictor that corresponds to $\hat{C}_{(1)}$ to the active set $\mathcal{G}$. Note that the active set contains all variables that have been ranked so far.

   (b) Determine the 'active' matrix for all $X_j \in \mathcal{G}_{i-1}$ as:

   $$\mathcal{X}_{(i)} = (s_{(i),j} X_j) \quad \text{where} \quad s_{(i),j} \equiv \text{sign}(\hat{c}_{(i),j}),$$

   and correspondingly:

   $$\mathcal{D}_{(i)} = \mathcal{X}_{(i)}'\mathcal{X}_{(i)} \quad \text{and} \quad \mathcal{A}_{(i)} = \sqrt{\mathbf{1}_{(i)}'\mathcal{D}_{(i)}^{-1}\mathbf{1}_{(i)}},$$

   where $\mathbf{1}_{\mathcal{G}_{(i)}}$ is a vector ones of a length equal to the number of variables in $\mathcal{G}$.

   (c) Now calculate the equiangular vector $u_{(i)}$ as:

   $$u_{(i)} = \mathcal{X}_{(i)} w_{(i)} \quad \text{where} \quad w_{(i)} = \mathcal{A}_{(i)} \mathcal{D}_{(i)}^{-1} \mathbf{1}_{\mathcal{G}_{(i)}}$$

   (d) Finally update $\hat{\mu}_{(i-1)}$ to $\hat{\mu}_{(i)}$ as:

   $$\hat{\mu}_{(i)} = \hat{\mu}_{(i-1)} + \hat{\gamma}_{(i)} u_{(i)}, \quad \text{with} \tag{22}$$

$$\hat{\gamma}_i = \min_{j \in \mathcal{G}^c_{(i-1)}}{}^{+} \left(\frac{\hat{C}_{(i)} - \hat{c}_{(i),j}}{\mathcal{A}_{(i)} - a_{(i),j}}, \frac{\hat{C}_{(i)} + \hat{c}_{(i),j}}{\mathcal{A}_{(i)} + a_{(i),j}} \right) \quad \text{and} \quad a_{(i),j} = X_j u_{(i)}$$

The fit is updated with the *smallest positive* solution of $\hat{\gamma}$. The corresponding predictor in the inactive set ($\mathcal{G}^c$) is added to the active set.

2. After completion of the algorithm a full ranking of the K predictors is provided, ranked according to their added value to the autoregressive fit. More importantly, the algorithm provides the 'optimal' fit of residuals after each step, $\hat{\mu}_{(i)}$, and can be used for forecasting purposes. The forecast can be established as:

$$\hat{y}_{t+1|t} = \hat{\mu}_{y,t} + L_t \hat{\beta}_L + \hat{\mu}^*(X_t), \tag{23}$$

where $\hat{\mu}_y$ is the mean of y in the estimation sample, $L_t\hat{\beta}_L$ is the autoregressive fit and $\hat{\mu}^*(X_t)$ the optimal LARS estimator. To determine the optimal LARS estimator, tenfold-CV is applied after each step i to a regression model with the AR terms and predictors in the active set. The estimate of μ that corresponds to the step which leads to the lowest CV forecast error is used to forecast with.

3.3 Dynamic Factor Models

Rather than applying penalized regressions, a different method is to 'filter' out most important information in a large set of predictor variables by using dynamic factor models (DFM) combined with principal component analysis (PCA). DFM is based on the theorem that the co-movement of a set of K variables (combined in X) can be described by a number of common unobserved factors (F). Instead of adding the predictors directly into a regression model, DFM first subtracts these common factors from X and uses them as predictor variables. The method aims to find a more parsimonious model by including only a limited number of the extracted factors.

As the factors are unobserved, PCA is applied to X to subtract the factors. PCA applied here is based on the spectral decomposition of the covariance matrix (A) of predictor variables X, which are all stationary and standardized to have zero mean and unit variance. The covariance matrix A can be decomposed as $A = F\Lambda F'$, which eventually leads to *principal component representation* of X:[7]

$$X = FP', \tag{24}$$

where F is the $N \times K$ matrix with all the uncorrelated, unobserved factors (or principal components). Furthermore, P is the $K \times K$ matrix containing the eigenvectors corresponding to the K factors in F. The factors in F are ordered in the amount of variance they explain in X, where the first factor explains most variance. The proportion of variance explained by each factor can be determined by scaling the eigenvalue of each factor, λ_b, by the sum of all the eigenvalues ($\lambda_b / \sum_{i=1}^{K} \lambda_i$). A dynamic factor model can now be constructed by extending the

[7] The algebra to come to (24) is skipped here, but can be found in, for example, Alexander (2008).

model in (5) by adding a subset of the estimated factors, denoted by $\hat{F}_j$:

$$y_t = \beta_0 + \beta'_{\mathcal{L}} y_{\mathcal{L},t-1} + \beta'_F \hat{F}_{j,t-1} + \varepsilon_t, \tag{25}$$

which Stock and Watson (2006) denote as the static representation of DFM.

The issue that remains is to decide on the number of factors to maintain, as the aim of the model is to describe the K economic predictors by a (strongly) reduced set of factors (r). A limited number of r factors should be selected to come to a feasible specification. However, leaving out variables inevitably leads to a loss of information. The issue therefore becomes a trade-off between the number of factors maintained and the variance explained, which are adversely related. Picking too many factors may negatively impact the forecasting power, but picking too few factors could lead to a loss of useful information.

To select r, the following steps are undertaken. In the first step all the factors F are estimated by means of PCA. The second step entails estimating a model which consists of a set of lags of the dependent variable and the first factor, as specified in (25). In each of the following steps one factor is added to the model, until all possible factors have been added. For each model, the CV forecast error is estimated. The factors that are part of the model that yields the lowest error are used for forecasting.

3.3.1 Pre-selecting variables

A possible disadvantage of PCA is that it focuses on capturing the common variation in X, but does not take into account the relation of the variables with the dependent variable. Therefore it could be that PCA is successful in capturing the common variation, but that the corresponding DFM has limited predictive quality. Bai and Ng (2008) find that by pre-selecting the variables in X for their usefulness in forecasting the dependent variable, the forecasting quality of the DFM can be improved. Furthermore, they apply a 'soft' thresholding procedure to pre-select variables (see also Çakmakli and van Dijk, 2010). The soft thresholding makes use of the LARS algorithm, as explained in the previous section. This algorithm acts as a selection technique when stopped after a certain amount of k-steps. The algorithm is stopped at the step which leads to the optimal LARS estimator $\hat{\mu}^*(X_t)$. The variables that are in the active set at that point are assumed to be the optimal subset. In addition to using the LARS as soft-thresholding technique, the lasso and elastic net can also be used for variable-pre selection. These methods also provide a limited set of variables, as variables can be set exactly zero in these methods. This leads to four different variations of the DFM model: unrestricted and with variable pre-selection of the three above described methods.

3.4 Forecast combinations

The third perspective from which the predictability of RV will be assessed is with use of forecast combinations (Rapach et al., 2010). The application of forecast combinations is essentially a two-step approach. The first step is to estimate a simple linear regression model as in (5), consisting of a constant, a set of lags of the dependent variable and one of the predictor variables

in X. This leads to K different sets of forecasts for the dependent variable, each generated by a different explanatory variable. The second step is to combine these forecasts to one forecast, based on a certain weight (ω_i) assigned to each forecast:

$$\hat{y}_{C,t+n|t} = \omega_1 \hat{y}_{1,t+n|t} + \cdots + \omega_K \hat{y}_{K,t+n|t} = \sum_{i=1}^{k} \omega_i \hat{y}_{i,t+n|t} \tag{26}$$

The use of this set-up is motivated by the assumption that different variables can capture different components (i.e. different parts of the variation) of the dependent variable, which is not captured when different variables are combined in one model. This is especially the case with as many predictors as analysed this paper. Combining them all in one model is infeasible, but combining their forecasts could lead to a less volatile and more accurate forecast (Rapach et al., 2010). Simply combining all variables in one model ignores the relevance of each variable to the predictor variable. Weakly related variables potentially only distort the estimates of other, more strongly related variables. Furthermore, adding many variables which all should proxy for risk may give rise to multicollinearity. The big difference with the previous two methods is that the former two methods first combine variables and then estimate a model, whereas this method first estimates a model for each variable separately and then combines the estimates to one forecast.

The combined forecast in (26) is determined by two crucial factors: The weighting function and the set of input variables. The remainder of this section will therefore discuss the possible weighting functions and how to select the set of variables. Three different functions are applied in this paper; equal weights, weights based on the in-sample fit of the models (based on the Discount MSPE of Stock and Watson, 2004), and minimization of the in-sample error. The determination of the weights in the first method are straightforward, as each model is assigned a weight of $1/K$. The second method determines the weight assigned to model i as the fraction of the sum of squared errors (SSE) to the total SSE:

$$\omega_{i,t} = \frac{SSE_{i,t}^{-1}}{\sum_{i=1}^{k} SSE_{i,t}^{-1}} \quad \text{with} \quad SSE_{i,t} = \sum_{j=m}^{t-1} \theta^{t-1-s} \varepsilon_{i,j}^2 \tag{27}$$

Models which have a lower SSE get assigned a higher weight. In addition, the models gives a larger weight to more 'recent' errors via θ. Smaller errors in more recent observations possibly signal that the model performs well in the current situation, which justifies that it should get a higher weight regardless of the fact that model already includes several lags. The value of θ is set at 0.99, which leads to the fact that the 90-th observation only gets a weight of 40%.

The weights of the third method are determined by minimizing the following function:

$$\omega = \underset{\omega}{\operatorname{argmin}} \sum_{j=1}^{t-1} \theta^{t-1-s} \left(y_j - \sum_{i=1}^{k} \omega_i \hat{y}_{i,j} \right)^2, \quad \text{s.t.} \quad \sum_{i=1}^{k} \omega_i = 1 \quad \text{and} \quad \omega_i > 0 \quad \forall i \tag{28}$$

The weights are determined as the combination which leads to the lowest in-sample error. The weights are subject to the restriction that they have to be positive and add up to one. Again,

more weight is given to more recent observations. This method allows the models to interact on a observation-by-observation basis to determine the overall lowest possible SSE. The key difference between this method and the former, is that the former simply gives more weight to the models that have a lower error and this method assigns weights according to which combination would have led to the lowest error. In other words, the last method takes into account the correlation between the different individual forecasts.

Similar to the dynamic factor models, as explained in the previous section, variable selection could be applied to narrow down the total set of variables to a set which contains only the most important variables. The weights, especially for the third method, could be heavily influenced by the number of predictors included. The resulting 'active' variables from lasso, elastic net and LARS are therefore applied in the same manner as explained before.

3.4.1 Forecast combinations with multivariate models

The concept explained above can be extended to combining multivariate models which include p predictors of X instead of 'univariate' models which include only one of the predictors in X. However, this extension leads to an exponential increase in the number of possible models. Including two predictors already leads to 703 possible models, while three predictors allow for over 8,000 models. Including more than two predictors therefore strongly limits the number of methods that can be used to determine the weights for each model. Especially the third method in this paper becomes computationally very burdensome when trying to combine a large number of models. However, there could be value in combining models that include more than one of the macro-finance predictors. The negative relation between the performance of a model and the number of included predictors might only hold for very large models, while models with a limited number of predictors might still outperform univariate models. Different variables combined into one model might interact well and produce a better forecast than when they are estimated separately.

A creative solution is required to make optimal use of the available predictors. One solution that is considered in this paper, is to divide the predictors in different categories based on their origin. The set of predictors can be subdivided into five different groups (G), as can be seen in Table 2: Equity & risk, bond, foreign exchange, liquidity & credit risk, and macroeconomic related variables. Each of those groups provides a different perspective on risk and therefore possibly also predict realized volatility differently. For example, the equity market group includes lagged market returns and the Fama-French risk factors, while the macroeconomic group includes variables such as inflation and industrial production growth. These entirely different variables potentially capture a different share of the variation of realized volatility. However, variables from the same group are likely to capture much of the same variation in realized volatility. Including many 'univariate' models based on variables that stem from the same group might not be helpful. Rather, it could be fruitful to estimate *one* forecast model for each of the five groups.

To establish a model for each group, PCA is applied to each of the groups of predictors. Subsequently, the first factor is used to establish the model for each group. The first factor

holds most common variation of the underlying variables and regarded to be an appropriate 'representative' of the group. Each model has the same structure and is estimated as:

$$y_t = \beta_0 + \beta'_{\mathcal{L}} y_{\mathcal{L},t-1} + \beta_G \hat{F}_G + \varepsilon_t \tag{29}$$

The resulting forecasts, denoted by $\hat{y}_{G,t+1|t}$, are then combined in one forecast:

$$\hat{y}_{GC,t+1|t} = \sum_{g=1}^{G} \omega_g \hat{y}_{g,t+1|t}, \tag{30}$$

where the weights for each model are determined by the above discussed methods. Alternatively, the first factors can also be combined directly in one linear regression model, thereby avoiding forecast combination procedure:

$$y_t = \beta_0 + \beta'_{\mathcal{L}} y_{\mathcal{L},t-1} + \sum_{g=1}^{G} \beta_g \hat{F}_g + \varepsilon_t \tag{31}$$

This 'first factor regression' model will likely deal differently with the correlation between the estimated factors compared to the forecast combinations, which will make it interesting to see how they relate to each other.

3.5 Bootstrap Aggregation

The last method considered in this paper is bootstrap aggregation (bagging), which is characterized as a "device for reducing prediction error in learning algorithms" (Breiman, 1996). As the name already suggests, the method involves drawing bootstrap samples from a certain set of training data. For each bootstrap sample b a certain model is estimated (hence, the learning algorithm), which is then used to forecast a certain value. The final forecast is determined as the average prediction over the different bootstrap samples, which could lead to a reduction in the estimation error. As the estimated parameters (and selected variables) of a certain model could be sample specific, simply forecasting with that model could lead to disappointing results. Bagging provides a solution to this problem, as it averages over forecasts that stem from different bootstrap samples. This method was recently investigated in a time-series context to forecast realized volatility by Hillebrand and Medeiros (2010). They indeed find that their models benefit from bagging, as it leads to increased forecast accuracy. The different bootstrap samples basically act as a robustness check on the original sample. If the estimated model is not strongly influenced by a subset of observations in that sample (i.e. every combination of X and y carries the same relation), the estimated models of different bootstrap samples will likely be close to the estimated model on the original sample. However, if there are observations which significantly negatively influence the estimation, bagging could be useful. Combining the forecasts based on random samples via bootstrapping, will average out the effect of extreme observations. If B is large enough, the average bagging forecast is robust to the effect of extreme observations in the estimation sample. This on its turn could lead to an increase in forecasting accuracy or a decrease in the variance of the forecasts. As the optimal number of B is not clear

beforehand, four different sample sizes will be provided in this paper: $B = (1, 50, 100, 200)$. Bagging on one sample, which is equivalent to estimating the model once on the original sample, is included to illustrate the effect of the number of bootstrap samples. The technical details of bagging can best be explained by converting it to a number of steps which have to be followed to obtain the bagging predictor, which is shown below:

Bagging Algorithm

1. For $b = 1, \ldots, B$:

   (a) Draw a sample from $\{y\}_{s=t-m}^{t}$ and $\{X\}_{s=t-m}^{t}$ *with* re-sampling. To keep the possible dependence between y and X intact, the full rows of $C = (y, X)$ will be used to construct samples. Denote this sample as $y_{(b)}$ and $X_{(b)}$.

   (b) Estimate for each predictor k:

   $$y_{(b),t} = \beta_0 + \beta'_{\mathcal{L}} y_{(b)\mathcal{L},t-1} + \beta_{(b),k} X_{(b),k,t-1} + \varepsilon_t,$$

   and retain only those variables for which $|t(\hat{\beta}^*_{(b),k})| > c$, which will be denoted as subset $\mathcal{X}_{(b)}$. The value of c is in set at 1.96, which is the t-value for a two-sided t-test with a 5% significance level.

   (c) Estimate a regression based on this subset $\mathcal{X}_{(b)}$, a constant, and a number of lagged dependent variables:

   $$y_{(b),t} = \beta_0 + \beta'_{\mathcal{L}} y_{(b)\mathcal{L},t-1} + \beta_x \mathcal{X}_{(b),t-1} + \varepsilon_t,$$

   (d) Subsequently, determine the forecast as:

   $$\hat{y}^*_{(b),t+1|t} = \hat{\beta}_0 + \hat{\beta}'_{\mathcal{L}} y_{(b)\mathcal{L},t} + \hat{\beta}_x \mathcal{X}_{(b),t}$$

2. The bagging forecast is computed as the average over the B samples:

$$\tilde{y}_{t+1|t} = \frac{1}{B} \sum_{b=1}^{B} \hat{y}^*_{(b),t+1|t} \tag{32}$$

4 Out-of-Sample evaluation

The models presented in the previous section are evaluated both for their statistical and economic quality of the forecasts they produce. The statistical quality of the models is analysed from two perspectives; precision and directional accuracy (i.e. how good are the forecasts in predicting the correct direction). The precision is assessed by means of the mean squared prediction error (MSPE). The significance of the MSPEs, compared to the 'optimized' AR model, is determined by the test of Clark and West (2007). An adjusted variation of the test by Breen et al. (1989) is used to assess the directional accuracy of the models. Furthermore, the economic quality is tested by means of two different applications. Firstly, the forecasts are used to predict the monthly Value-at-Risk (VaR). The VaR predictions are evaluated by means of the Conditional Coverage test by Christoffersen (2011). Secondly, the forecasts are applied to a myopic mean-variance investment strategy. The returns resulting from this strategy are used to determine Sharpe Ratios. The significance of these Sharpe Ratios, compared to a simple Buy-and-Hold strategy, is determined by the test of Opdyke (2007). All tests are further substantiated and explained below.

4.1 Statistical criteria

The primary metric to evaluate the accuracy in this paper is the Mean Squared Prediction Error (MSPE). The MSPE provides an indication of the forecasting accuracy in the form of the average squared forecast error, which is determined as:

$$\text{MSPE} = P^{-1} \sum_{t=T}^{T+P-1} \left(y_{t+1} - \hat{y}_{t+1|t}\right)^2, \tag{33}$$

where y_{t+1} is the actual value of LRV at $t+1$ and $\hat{y}_{t+1|t}$ the forecasted value. The significance of the difference between the MSPE of model x and the optimized AR model, which serves as a benchmark, is determined by the MSPE-adjusted test statistic of Clark and West (2007). The test uses the null-hypothesis that the difference between two MSPEs is not significant, or in other words, they do not differ significantly from each other. The added value of this test over, for example, the test by Diebold and Mariano (1995), is that it remains to have its power when applied to nested models. Since all the models start from the same autoregressive model, this cannot be ignored. The MSPE-adjusted statistic is a regression based method, where the adjusted-MSPE (f) is regressed on a constant. The corresponding one-sided p-value of this coefficient determines the significance of the difference between two models. The adjusted-MSPE is defined as:

$$f_{t+1} = (y_{t+1} - \hat{y}_{b,t+1|t})^2 - \left[(y_{t+1} - \hat{y}_{x,t+1|t})^2 - (\hat{y}_{b,t+1|t} - \hat{y}_{x,t+1|t})^2\right], \tag{34}$$

where $\hat{y}_{b,t+1|t}$ is the forecast of the benchmark (b) model and $\hat{y}_{x,t+1|t}$ the forecast of model x.

As outlined above, the test by Breen et al. (1989) instead is used to determine the directional accuracy of the forecasts. This method is favoured over, for example, the PT-statistic by Pesaran

and Timmerman (1992), due to the nature of volatility. The existing literature on volatility tells us that it is likely that volatility exhibits clustering, which will work through in tests that focus on the sign of forecasts of volatility. Ignoring this, by using the PT-Statistic, would possibly lead to the wrong conclusions (i.e. it lacks power). Breen et al. (1989) provide a straightforward solution to this by testing the directional accuracy in a regression framework and correcting for possible heteroskedasticity and/or autocorrelation (i.e. Newey-West standard errors) subsequently. The test revolves around the following regression:

$$I_{s_{t+1}} = a + bI_{\hat{s}_{t+1|t}} + \eta_{t+1}, \tag{35}$$

where $I_{s_{t+1}}$ is a binary variable that is equal to 1 or 0 if s is positive or negative, respectively. Since RV can only be positive (and the log RV is predominantly, if not entirely, negative), the test is based on the *first differences* of y. Subsequently, $I_{\hat{s}_{t+1|t}}$ is adjusted accordingly and reflects the direction of the forecast at $t+1$ compared to the actual value at t: $s_{t+1} = \hat{x}_{t+1|t} - x_t$. In the case of a significant coefficient (both signs possible, though a negative sign would be an interesting case), the test signals that volatility can be *timed*. In other words, the model is able to predict the correct direction of the volatility on average.

As both the dependent and independent variable are indicator variables, the model is quite peculiar and its interpretation is not straightforward. The model should be interpreted as follows. Firstly, the model also includes a constant, which can be interpreted as the average value of the dependent variable. In this case it is the average direction of the true value of LRV. The coefficient value $\hat{b}$ therefore says that if there is a predicted upward movement, thus a one, then the fitted value is the sum of constant and the coefficient value. In the case a forecast model is not at all able to predict the direction, the value of the coefficient will be very small. A predicted direction says little about the actual direction in this case, so it will not add much to the 'average' direction of the constant term. Hence, a small value of $\hat{b}$. On the other hand, a model that is able to predict the right direction will have a higher value of $\hat{b}$. Thus, the coefficient $\hat{b}$ reflects the capability of the independent variable (the predicted value of y) to predict the correct value of the dependent variable (the true value of y).

4.2 Economic criteria

Though the statistical criteria provide extensive insight regarding the predictive quality of model, a model is only as good as its economic performance. A good model from a statistical point of view does not necessarily imply that it returns a high profit, albeit likely that they go hand-in-hand. The economic quality is assessed by means of monthly Value-at-Risk (VaR) forecasts and as the return on a myopic investment strategy (see, amongst others, Çakmakli and van Dijk, 2010). For both applications, the LRV forecasts are converted to RV forecasts using (2).

The α%-Value-at-Risk (VaR_α) is a measure that provides a certain loss L that occurs with a probability of α%. From a statistical point of view, the VaR_α can be defined as

$$\text{VaR}_{\alpha,t} = \mu_t + \sigma_t D^{-1}(\alpha), \tag{36}$$

where μ is the mean return, σ the standard deviation, D a certain distribution and α the α%-quantile. Clearly, the metric relies on the underlying distribution of the asset returns. This makes it difficult to analyse the ability of volatility forecasts to predict the VaR, as the returns not necessarily follow a certain distribution. The focus in this paper is to forecast the one-month VaR_α%, using the volatility forecasts generated by the models, the mean return by a historical mean (HM) model (i.e. a linear regression model with only a constant), and a normal distribution to approximate D.

$$\mathbb{E}_t[\text{VaR}_{\alpha,t+1}] = \mathbb{E}_t[r_{k,t+1}] + \mathbb{E}_t[\sigma_{k,t+1}]\Phi^{-1}(\alpha) \tag{37}$$

Though the choice of a proper distribution is difficult and the normal distribution could be questioned, the aim is to compare the different models and not to evaluate the VaR forecasts on an isolated basis. Though, to analyse the influence of the normal distribution, the analysis is also executed on a Student's t-distribution with 5 degrees of freedom. The quality of the VaR forecasts will be assessed by how many times the VaR forecast is surpassed by the actual return. This will be formally tested with use of the conditional coverage test by Christoffersen (2011), which simultaneously tests for the proportion of exceedances and volatiliy clustering:

$$\text{LR}_{cc} = \frac{\pi_{\text{exp}}^{n_1}(1-\pi_{\text{exp}})^{n_0}}{\pi_{01}^{n_{01}}(1-\pi_{00})^{n_{00}}\pi_{11}^{n_{11}}(1-\pi_{11})^{n_{10}}}, \quad \text{where} \quad -2\ln \text{LR}_{cc} \sim \chi(2), \tag{38}$$

where a 1 represents an actual return that is lower than the predicted VaR and 0 when it is not. A double digit, such as 11, means that an exceedance is followed by another exceedance. The n represents the number of times it occurs and π its corresponding fraction to the total number of predicted observations. In addition, π_{exp} corresponds to the expected number of exceedances, which corresponds to α. The conditional coverage test can be split up in the unconditional coverage test and the independence test:

$$\text{LR}_{cc} = \text{LR}_{ucc} \times \text{LR}_{ind}, \quad \text{where} \tag{39}$$

$$\text{LR}_{ucc} = \frac{\pi_{\text{exp}}^{n_1}(1-\pi_{\text{exp}})^{n_0}}{\pi_{obs}^{n_1}(1-\pi_{obs})^{n_0}} \quad \text{and} \quad \text{LR}_{ind} = \frac{\pi_{obs}^{n_1}(1-\pi_{obs})^{n_0}}{\pi_{01}^{n_{01}}(1-\pi_{01})^{n_{00}}\pi_{11}^{n_{11}}(1-\pi_{11})^{n_{10}}},$$

where both $-2\ln \text{LR}_{ucc}$ and $-2\ln \text{LR}_{ind}$ are asymptotically chi-squared distributed with 1 degree of freedom. The unconditional coverage test focuses explicitly on the number of exceedances, while the independence test evaluates the independence or clustering of the exceedances.

The investment strategy considered here is from the perspective of a myopic, mean-variance investor with a one-month horizon. Furthermore, the investor's portfolio consists of risk-free T-bills (r_f) and one of the assets (r_k) analysed. The portfolio return (r_p) at $t+1$ is therefore given by:

$$r_{p,k;t+1} = r_{f,t+1} + w_{t+1}r_{k,t+1} \tag{40}$$

The weight allocated to the asset is determined by optimizing the following objective function,

where γ represents the risk-averseness of the investor

$$\max_{w_{t+1}} \quad \mathbb{E}_t[r_{p,t+1}] - \frac{\gamma}{2}\mathbb{V}_t[r_{p,t+1}], \tag{41}$$

where $\mathbb{E}_t[\cdot]$ and $\mathbb{V}_t[\cdot]$ are the conditional mean and variance at t, respectively. Assuming that $r_{f,t}$ reflects the risk-free rate between t and $t+1$, allows the optimal weight can to be determined as:

$$w^*_{t+1} = \frac{\mathbb{E}_t[r_{k,t+1}]}{\gamma \mathbb{V}_t[r_{k,t+1}]} \tag{42}$$

To prevent unrealistic weights, the range of w^*_{t+1} is limited to $[-1, 1]$. This setting prohibits extreme positions (i.e. more than %100 short in the asset). To determine the optimal weight given to the asset at $t+1$, forecasts of both the expected return and volatility are required. The forecast of volatility is given by one of the forecast models as discussed in Section 3 and the mean is predicted with the HM model. This leads to a monthly portfolio return equal to:

$$r_{p,k;t+1} = r_{f,t+1} + w^*_{t+1} r_{k,t+1} \tag{43}$$

The profitability of these monthly returns is assessed by means of their Sharpe Ratio (SR):

$$\widehat{SR}_{p,k} = \frac{\mathbb{E}[r_{p,k;t} - r_{f,t}]}{\hat{\sigma}[r_{p,k;t} - r_{f,t}]}, \tag{44}$$

which is equal to the ratio of the average monthly return in excess of the risk-free rate, divided by its standard deviation. The Sharpe Ratio is a measure to put the profitability (i.e. average monthly return) in perspective to its variation. The Sharpe Ratio reveals whether a higher profitability is solely due to increased variance (i.e. the risk-return trade-off), or whether it actually performs better than other portfolios. The Sharpe ratio in fact 'standardizes' the average return and provides a neutral measure to compare portfolios. To evaluate the Sharpe ratios of the myopic portfolios they are compared to a simple Buy-and-Hold (BAH) strategy. The BAH strategy buys 50% of the asset and 50% of the risk-free asset on the first month of the forecast period and holds it throughout this period. It never updates or changes its position. Thus, the monthly return of the BAH strategy is to the average of the monthly actual return and risk-free rate. A good portfolio strategy should at least be able to beat a simple passive strategy, which makes the BAH strategy a suitable benchmark.

The performance of the portfolios with respect to the BAH strategy is formally tested by the test of Opdyke (2007). The test derives the underlying distribution of the difference between two Sharpe Ratios ($\widehat{SR}_d$), to determine whether they differ significantly. Opposed to other tests, such as the more commonly used test by Jobson and Korkie (1981), it does not assume that the underlying returns are normally distributed and estimates the variance using the delta method. As the variance of $\widehat{SR}_d$ reduces to the expression found by Jobson and Korkie (1981) if returns are normally distributed, the variance of Opdyke (2007) can be seen as a more generalized expression. The distribution of $\widehat{SR}_d$ is asymptotically normal, having a zero mean and Var_d as variance:

$$\sqrt{T}\left(\widehat{\text{SR}}_a - \widehat{\text{SR}}_b\right) = \sqrt{T}\left(\widehat{\text{SR}}_d\right) \overset{a}{\sim} \mathcal{N}\left(0, \text{Var}_d\right), \quad \text{where} \tag{45}$$

$$\text{Var}_d = 2 + \frac{\text{SR}_a^2\,(\kappa_a - 1)}{4} - \text{SR}_a s_a + \frac{\text{SR}_b^2\,(\kappa_b - 1)}{4} - \text{SR}_b s_b$$
$$-2\left(\rho_{a,b} + \frac{\text{SR}_a \text{SR}_b}{4}\left(\frac{\mu_{2a,2b}}{\sigma_a^2\sigma_b^2} - 1\right) - \frac{1}{2}\left(\text{SR}_a \frac{\mu_{1b,2a}}{\sigma_b\sigma_a^2} + \text{SR}_b \frac{\mu_{1a,2b}}{\sigma_a\sigma_b^2}\right)\right)$$

where s, κ, σ and ρ stand for skewness, kurtosis, standard deviation and correlation, respectively. Furthermore, $\mu_{nj,mk} = \text{E}\left[(j - \text{E}(j))^n\,(k - \text{E}(k))^m\right]$. The asymptotic normality of $\widehat{SR}_d$ implies that the resulting sample mean and variance of $\widehat{SR}_d$ can be used for a standard z-test for hypothesis testing.

4.3 Benchmark models

Benchmark models take an important place in the evaluation of the models. Next to the optimized AR model, which has a central role throughout the paper, several other benchmarks are presented. The use of additional benchmarks is motivated as a means to analyse the estimated models more in-depth. To evaluate whether optimizing the AR model is the right choice, the benchmarks also include to AR(1) to AR(5) models. Furthermore, an important benchmark to add for the penalized regression methods is the model which simply adds all variables to the optimized AR model and estimates its coefficients by means OLS. This model, denoted as the ARX model, implies the model which sets $\lambda = 0$ each period. The historical mean and random walk (RW) model without drift (i.e. next month's forecast is this month's volatility) are also included as benchmarks. Their performance should reveal whether they can actually outperform the ARX model, as stated in the beginning of this paper.

5 Empirical results

5.1 Statistical performance

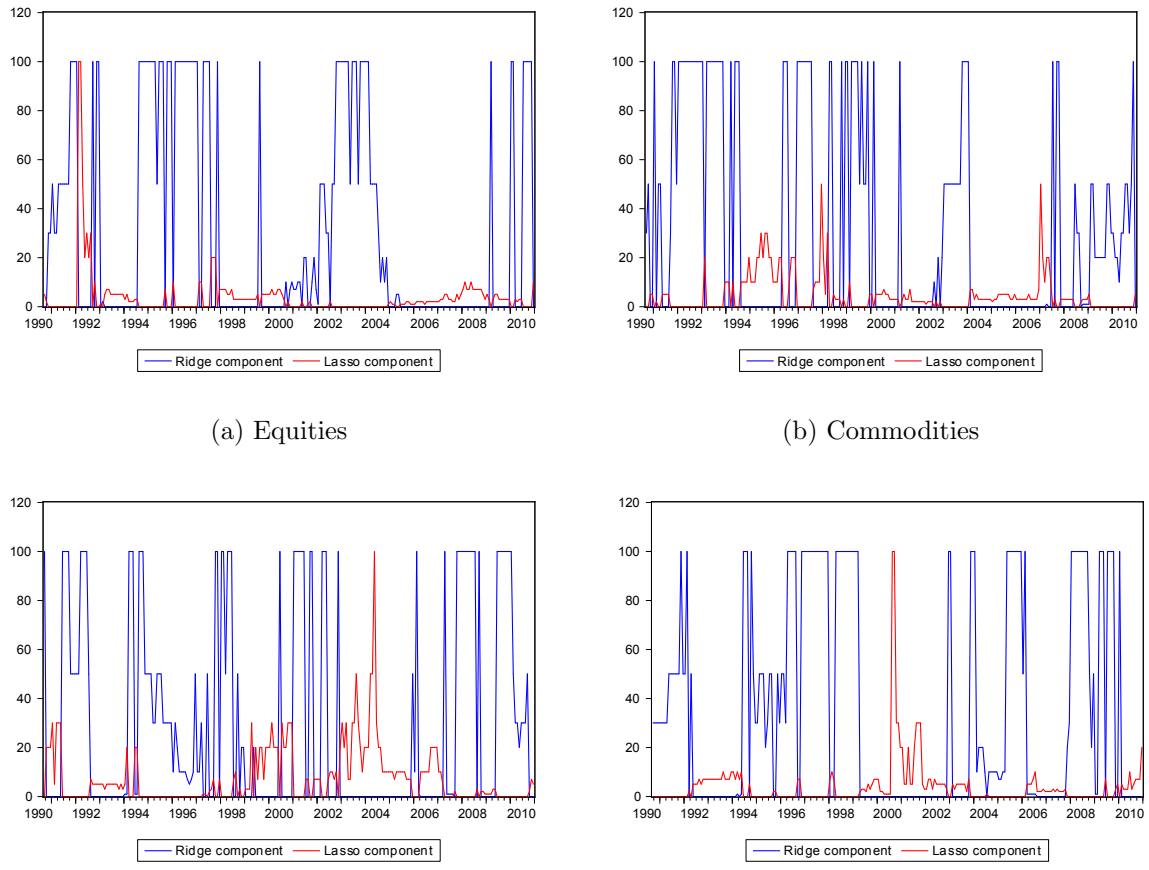
Table 4 below displays the MSPE of the forecasts, both in absolute terms (MSPE_a) and relative (MSPE_r) to the optimized autoregressive benchmark model (AR_{opt}). Thus, a model outperforms the AR_{opt} model in terms of accuracy if the MSPE_r is smaller than one. The AR_{opt} model appears to perform quite well, as it is amongst the best performing benchmark models. However, its average absolute MSPE is highest in the equities class, but appears to be quite substantial in all models. Taking the root of the MSPE_a (RMSPE) gives a better idea by how much the average forecast is off compared to the true value. The RMSPE of the AR_{opt} models for all asset classes is in the range of 0.28 to 0.3, which is indeed quite substantial given that the average absolute value of log RV is between 3 and 4. In addition, the model fails to perform better than one or more of the fixed lag models. For all classes, the AR(3) model performs best and is significantly more accurate than the optimized AR model in all classes. However, at the time the forecast is made it is difficult to find the best specification of the AR model. The fact that the AR(3) model will perform best over time is not known at that point. Optimizing the lags is therefore intuitively a sensible thing to do. The AR_{opt} model is, however, amongst the best performing model and is a hard-to-beat benchmark for the other models. Only a limited number of models have a MSPE_r smaller than one and if they do, its only by a marginal amount. Below the forecasts of each of the estimation technique classes will be discussed separately, before a general conclusion is provided.

The penalized regression techniques fail to beat the benchmark in all asset classes. Their performance is therefore quite disappointing, as the methods are all well-established variable selection and estimation techniques. The ridge regressions performs best on an overall basis, but is outperformed by the LARS in the bonds class. However, the LARS performs worst in the equities class. Besides the bad performance in general, the consistency in performance of the models between asset classes appears to be limited. The penalized regressions all outperform the ARX_{opt} model, which is the model that simply combines all predictors in one model without imposing any shrinkage. The essence of the penalized regressions is to improve the accuracy of this model, in which it succeeds. Nevertheless, adding non of the 38 variables seems to be the better choice. Though the performance is disappointing, it is not necessarily surprising as an optimal in-sample forecasting performance does not guarantee a good out-of-sample performance. Moreover, shrinking the complete set of variables all together may not be optimal. The dimension of the set of predictors may require pre-emptive elimination of variables which surely do not hold any predictive content of LRV.

The performance of the elastic net consistently lies between the performance of the ridge regressions and the lasso. As the elastic is able to switch between the ridge regression and lasso, regardless of the possibility of combining the two, it is in the line of expectation that the EN outperforms both. However, this is based on the assumption that an in-sample optimal choice holds out-of-sample. As argued above, this does not necessarily has to be the case. Figure 4 below provides more insight in the interaction between the ridge regression component (λ_1) and

the lasso component (λ_2) in the elastic net. In only a limited number the ridge (regression) and lasso component are both larger than zero. In most cases the elastic net switches between the two. It is striking to see that the optimization often finds that relatively large values of λ_1 are optimal, while the optimal values of λ_2 are relatively small. Most likely this is due to the fact that the lasso component is able to set coefficients at exactly zero. The benefit of having a more parsimonious model appears to have its limits. For λ_2 a value of 100 would mean that practically all coefficients are set zero, which apparently is not optimal in-sample.

Figure 4: The time-varying behaviour of the ridge regression and lasso shrinkage parameters in the elastic net



(a) Equities

(b) Commodities

(c) FX-aggregate

(d) Bonds

Notes: (i) The Figure shows the optimal values of the shrinkage parameter of the ridge regression and lasso component in the elastic net, respectively. (ii) The blue and red line represent the value of the ridge regression component (λ_1) and the lasso component (λ_2), respectively.

The second class of models, the dynamic factor models, also perform quite poorly. They all fail to beat the AR_{opt} model. It seems that the common variation in the predictors, as extracted by means of principal component analysis, is not able capture and forecast the variation in log realized volatility. This could be the result of two causes. Firstly, the common risk of the predictors that is captured in the factors is not related to realized volatility. PCA can still be successful in subtracting the common variation in the variables, but the factors are of no

use. Secondly, the common variation of the predictors is too low and the corresponding factors therefore have little meaning. As the predictors originate from different segments of the financial market, this is most likely to be the case. This is confirmed by Table 3 below, which provides a further analysis of the performance of the PCA in subtracting the common variation in the variables. The average number of factors (ANF) is quite low, but surprisingly is higher when the variables are pre-selected. This implies that, although the set of variables is limited to the 'most important' ones, the shared variation has decreased. As a result, this requires more factors to capture (proportionally) the same amount of common variation. The average percentage of common variation captured (% Expl.) is also higher than PCA on all variables as it is roughly 60% compared 40% in the unrestricted case. The fact that PCA needs on average more than four factors to capture 50 to 60% of the common variation, this reveals that common variation is fairly low. Pre-selecting variables results in an increase in the captured common variation, but also requires on average more factors. Most important variables do not necessarily share a lot of common variation, which will result in information-rich factors. This appears to be the case, as the pre-selection of variables does not improve the forecasting quality of the DFM model. This leads to concluding that the use of DFM models appears to be limited.

Table 3: Dissection of the dynamic factor models: Average percentage of common variation captured and number of required factors

	Equities		Commodities		FX-aggregate		Bonds	
	% Expl.	ANF	% Expl.	ANF	% Expl.	ANF	% Expl.	ANF
Dynamic factor models								
Unrestricted	46.1	4.8	41.3	3.8	37.9	3.3	38.3	3.7
Lasso	64.4	5.4	67.5	6.0	63.4	5.3	49.3	3.3
Elastic Net	60.6	6.1	54.8	4.7	50.9	3.8	44.4	3.2
LARS	64.4	5.4	67.5	6.0	63.4	5.3	49.3	3.3

Notes: (i) The table presents an overview of the average share of common variation explained (% Expl.) by the average number of optimal factors (ANF) to use for the DFM forecast model. (ii) The four cases represent the unrestricted variant, which includes the full set of 38 predictors, and the sets of variables that come out after applying one of the three penalization methods to the predictors.

The best performance is found in the forecast combinations class, similar to what has been found by Christiansen et al. (2012). Combining the forecast of separate models which add only one of the predictor variables to the AR_{opt} model leads to a significant outperformance of both the AR_{opt} and AR(3) model. Only the forecast combinations according to the second weight type perform poorly. Perhaps this weighting scheme is more prone to sample specific effects. Letting the weights be determined by the combination that leads to the lowest in-sample error, not necessarily implies the best out-of-sample combination. As already mentioned, the equal weighted combinations and combinations according to weight type perform strongly and have the lowest MSPEs of all models. The performance is consistent for the different asset classes, albeit weaker compared to the equity class. Unlike the dynamic factor models, pre-selecting variables does improve the predictive quality of forecast combinations. Narrowing down a set of predictor variables using penalized regressions and then apply forecast combinations appears to be the most successful strategy to produce forecasts with the highest accuracy. The fact that it does work for forecast combinations is most likely due to the construction of forecast

combinations. Each predictor is allowed to hold its own, unique predictive content and not forced into a factor or combined in one model.

Interestingly, the PCA-based combinations appear to be a strong competitor of the regular forecast combinations, as their MSPEs are very similar. Subtracting the common variation of each five different subgroups of variables (see Table 2) and combining the corresponding first principal components, works better than directly applying PCA to all predictor variables. It therefore appears that the weak performance of the dynamic factor model is not due to the inability of PCA to capture the common variation, but more due to a lack of common variation of all variables. This common variation is however found within each subgroup, which explains its good forecasting performance. Surprisingly, adding the same 'first principal components' for each subgroup to the AR_{opt} model leads to a disappointing performance. The use of forecast combinations therefore appears to be crucial for a good forecasting performance.

Lastly, the bagging forecasts appear not to be able to outperform the AR_{opt} model. Bagging on only one sample represents the strategy of simply removing all predictors which are not significant, which not surprisingly leads to a poor forecasting performance. Bagging on only one sample is simply equal to selecting only those predictors that are significant in a linear regression model, which is a superficial method the select the most important variables. The value of bagging lies, however, lies in repeating this procedure many times to 'filter' out sample specific selections. This appears to be the case, as bagging on 100 samples leads to a strong improvement of the predictive quality. More than 100 samples, such as 150 or 200, seems to have no real impact on the forecasting performance of bagging. Though bagging on 100 samples seems sufficient, it still lacks predictive quality compared to the AR_{opt} model. Its main flaw therefore appears to lie in the variable selection component, which could be replaced by more sophisticated techniques such as those provided in this paper. This is, however, left for future research.

Table 4: Mean Squared Prediction Errors

	Equities		Commodities		FX-aggregate		Bonds	
	$MSPE_a$	$MSPE_r$	$MSPE_a$	$MSPE_r$	$MSPE_a$	$MSPE_r$	$MSPE_a$	$MSPE_r$
1. Benchmark models								
Historical mean	0.257	2.672	0.159	1.899	0.162	1.951	0.127	1.566
Random Walk	0.106	1.105	0.101	1.201	0.097	1.176	0.105	1.286
AR_{opt}	0.096	1.000	0.084	1.000	0.083	1.000	0.081	1.000
AR(1)	0.102	1.061	0.090	1.069	0.089	1.080	0.087	1.071
AR(2)	0.097	1.005	0.08**	0.959	0.082**	0.987	0.083	1.025
AR(3)	0.095**	0.992	0.08**	0.961	0.081**	0.977	0.08**	0.981
AR(4)	0.098	1.019	0.082*	0.975	0.081**	0.972	0.082	1.002
AR(5)	0.097	1.004	0.083	0.989	0.084	1.012	0.083	1.020
ARX_{opt}	0.277	2.878	0.211	2.521	0.197	2.380	0.239	2.937
2. Penalized regressions								
Ridge Regressions	0.099	1.030	0.092	1.098	0.088	1.063	0.089	1.090
Lasso	0.103	1.070	0.098	1.174	0.092	1.105	0.093	1.139
Elastic Net	0.105	1.095	0.088	1.052	0.090	1.084	0.088	1.085
LARS	0.105	1.095	0.093	1.109	0.089	1.070	0.087	1.071
3. Dynamic factor models								
Unrestricted	0.114	1.184	0.090	1.073	0.095	1.145	0.087	1.069
Lasso	0.148	1.541	0.118	1.411	0.118	1.426	0.087	1.074
Elastic Net	0.129	1.344	0.107	1.283	0.090	1.088	0.091	1.115
LARS	0.155	1.609	0.104	1.244	0.121	1.457	0.103	1.261
4. Forecast combinations								
a. Equal weights								
Unrestricted	0.096	0.999	0.082**	0.975	0.082*	0.991	0.082	1.008
Lasso	0.094*	0.980	0.083*	0.986	0.083	0.999	0.082	1.010
Elastic Net	0.095	0.987	0.082*	0.977	0.082	0.986	0.083	1.016
LARS	0.095	0.990	0.084	1.002	0.083	1.004	0.082	1.006
b. Weight type 1								
Unrestricted	0.095*	0.989	0.083*	0.994	0.083	0.997	0.081*	0.994
Lasso	0.093*	0.972	0.084	1.006	0.083	1.002	0.081	0.995
Elastic Net	0.094*	0.979	0.083	0.996	0.082	0.991	0.081	1.002
LARS	0.095	0.984	0.085	1.020	0.083	1.008	0.081	0.996
c. Weight type 2								
Unrestricted	0.103	1.073	0.083*	0.996	0.090	1.086	0.087	1.066
Lasso	0.104	1.086	0.086	1.031	0.091	1.102	0.083	1.018
Elastic Net	0.100	1.039	0.084	1.007	0.088	1.060	0.085	1.048
LARS	0.098	1.024	0.087	1.035	0.092	1.114	0.088	1.082
d. PCA based								
Unrestricted	0.095*	0.993	0.082*	0.976	0.082*	0.990	0.082	1.010
SSE weights	0.095**	0.984	0.083	0.995	0.082	0.995	0.081	0.996
SSE minimalized	0.095**	0.989	0.081**	0.970	0.083	1.001	0.083	1.022
First factor regression	0.105	1.094	0.091	1.081	0.090	1.081	0.089	1.099
5. Bootstrap aggregation								
#Bootstrap samples = 1	0.109	1.133	0.094	1.120	0.098	1.189	0.093	1.138
#Bootstrap samples = 100	0.104	1.084	0.090	1.073	0.096	1.159	0.087	1.065
#Bootstrap samples = 150	0.105	1.095	0.090	1.075	0.095	1.144	0.087	1.070
#Bootstrap samples = 200	0.106	1.098	0.090	1.073	0.096	1.155	0.086	1.062

Notes: (i) The table displays the Mean Squared Error, both in absolute sense ($MSPE_a$) and relative to the AR_{opt} model ($MSPE_{opt}$. The ratios for AR_{opt} are therefore all are equal to 1. (ii) The significance of the MSPE is only estimated if the ratio is smaller than one. A * or ** implies significance at a 10% or 5% significance level, respectively.

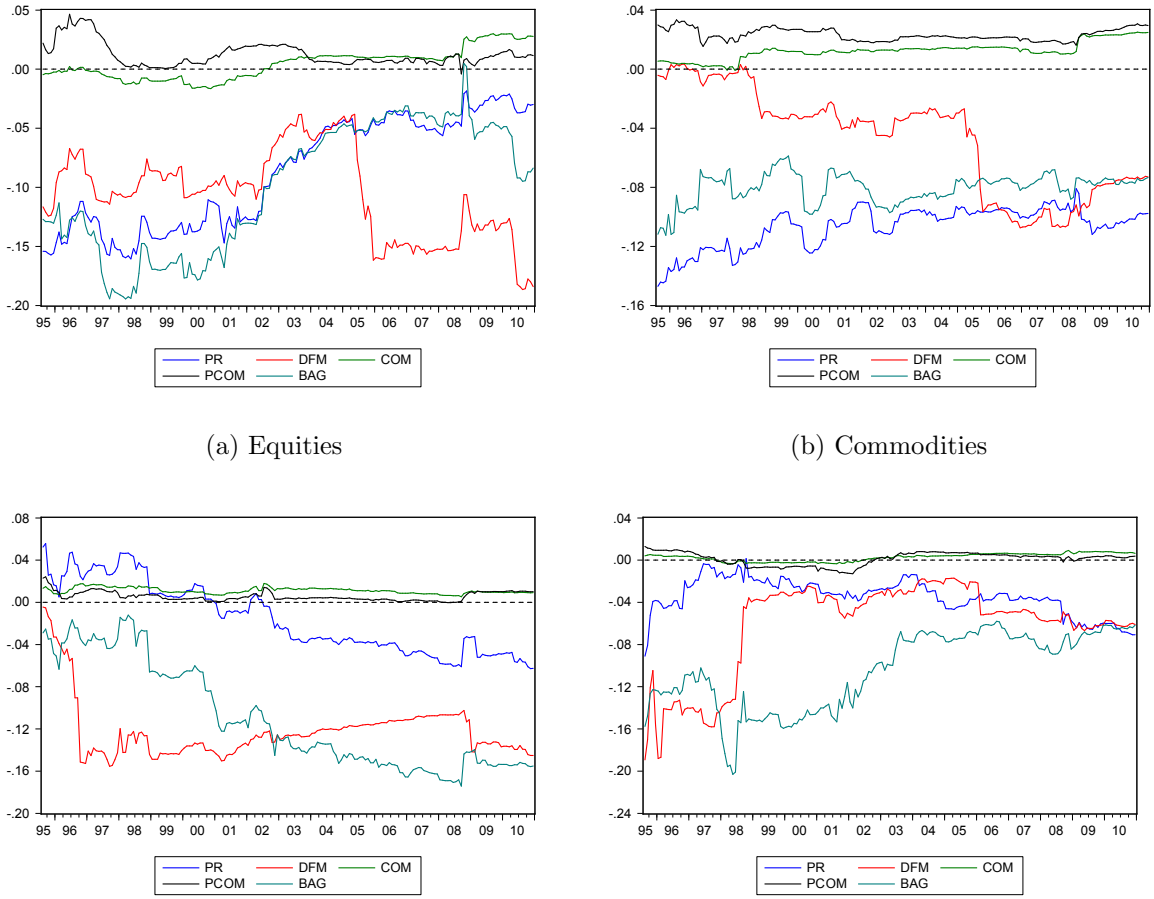
In addition to the accuracy statistics over the full sample, Figure 5 shows how the MSPE of the best model in each 'model class' for each asset class evolves over time. More specifically, there are in total 5 models displayed; (1) penalized regressions (PR), (2) dynamic factor models (DFM), (3) forecast combinations of subgroup a, b or c (COM), (4) PCA-based forecast combinations (PCOM) and (5) bagging (BAG). The movement of the of the MSPE over time is measured relative to the MSPE of the AR(3) model, using the Out-of-Sample R^2 (R^2_{oos}) of Campbell and Thompson (2008):

$$R^2_{oos} = 1 - \frac{\sum_{t=T}^{T+P-1}(y_{t+1} - \hat{y}_{x,t+1|t})^2}{\sum_{t=T}^{T+P-1}(y_{t+1} - \hat{y}_{b,t+1|t})^2} = 1 - \frac{\text{MSPE}_a(\hat{y}_x)}{\text{MSPE}_a(\hat{y}_b)} \tag{46}$$

The Out-of-Sample R^2 is initialized with the first 60 observations of the forecast sample and then expanded with one observation each month until all observations are included. The AR_{opt} is outperformed by a model if the corresponding Out-of-Sample R^2 is larger than zero.

The time-varying R^2_{oos} reveals that there is quite some variation in the predictive quality for some of the models. However, the forecast combinations, the best performing overall models, behave quite stable over time. The strong performance of the PCA-based forecast combinations appears to be primarily due to a good performance in the beginning of the forecast sample ('95 – '96). Quite interesting is the behaviour of the penalized regressions and bagging forecasts for the equities class, which are strong upward sloping for the majority of the forecast sample after hitting its lowest point early in the sample. Though both remain to be below zero during its rise, it must be taken into account that the window for estimating the R^2_{oos} is expanding. A strong rise therefore implies a (relatively) strong performance of the models in period of the rise. Their upward movement signals that these models perform better in more recent periods. A similar pattern is not observed in the other asset classes, which implies the good performance is limited to the equities class. Equally remarkable is the behaviour of the dynamic factor models in the bonds class. Its shows a very strong and persistent outperformance of the AR_{opt} model for the majority of the forecasting sample, but dies out in the most recent period. Though DFM appears not to work for other classes, just like the performance of bagging seems specific to the equities class, it does work for the bonds class. Lastly, the predictor variables have most difficulties in predicting the RV of commodities and FX-aggregate class, as the majority of the models are consistently beaten by the AR_{opt} model. Based on these figures and Table 4, it appears that the macro-finance variables work best for the equities class.

Figure 5: Time-varying Out-of-Sample R^2 for the best models of each asset class



(a) Equities

(b) Commodities

(c) FX-aggregate

(d) Bonds

Notes: (i) The figures show the moving out-of-sample R^2 for the best model of each of the following five classes: (1) penalized regressions (PR, blue line), (2) dynamic factor models (DFM, red line), (3) forecast combinations of subgroup a, b or c (COM, green line), (4) PCA based forecast combinations (PCOM, black line) and (5) bagging (BAG, green/blue line).

Table 5 shows the estimated coefficients $\hat{b}$ of (35), both in absolute terms and relative to the coefficient of the AR_{opt} model ($\hat{b}/\hat{b}_a$), and lastly their corresponding standard errors (std. err.) between brackets. The models appear to be quite successful in timing the volatility, as almost all coefficients are significant. As monthly realized volatility is highly persistent, which is for a large part already captured by the AR_{opt} model, this comes as no surprise. The AR models have the highest values of $\hat{b}$ for all asset classes but the equities class, which signals that a simple AR model is sufficient to predict the direction of next month's volatility and more importantly, the macro-finance variables do not increase the predictability. However, for equities class the results are consistent with the MSPEs of the models. The forecast combinations are as a class best in predicting the correct direction, but the ridge regression is surprisingly the best direction predictor on an overall basis. The general conclusion is that the models are able to predict the correct direction as the almost all coefficients are significant and differences are relatively small.

Overall, there appears to be information in the macro-finance variables which help to improve the forecasting power of the AR_{opt} model. Of all models, the forecast combinations appear to work quite well. In particular, the pre-selection of variables in combination with forecast combinations seems to perform consistently over all different asset classes. Identifying the most important variables and combining their forecasts separately, which keeps their unique predictive content intact is a successful strategy to beat the AR_{opt} model.

Table 5: The Directional Accuracy Test

	Equities			Comm.			FX-aggr.			Bonds		
	$\hat{b}$	$\hat{b}/\hat{b}_a$	std. err.	$\hat{b}$	$\hat{b}/\hat{b}_a$	std. err.	$\hat{b}$	$\hat{b}/\hat{b}_a$	std. err.	$\hat{b}$	$\hat{b}/\hat{b}_a$	std. err.
1. Benchmark models												
Historical mean	0.218	0.872	(0.056)	0.296	0.814	(0.056)	0.348	1.269	(0.053)	0.367	1.026	(0.057)
Random Walk	–	–	–	–	–	–	–	–	–	–	–	–
AR_{opt}	0.249	1.000	(0.053)	0.364	1.000	(0.053)	0.275	1.000	(0.055)	0.358	1.000	(0.055)
AR(1)	0.218	0.872	(0.056)	0.296	0.814	(0.056)	0.307	1.118	(0.054)	0.350	0.978	(0.055)
AR(2)	0.245	0.982	(0.053)	0.330	0.907	(0.062)	0.341	1.240	(0.054)	0.366	1.023	(0.06)
AR(3)	0.196	0.787	(0.052)	0.372	1.022	(0.058)	0.300	1.093	(0.06)	0.440	1.229	(0.056)
AR(4)	0.196	0.787	(0.055)	0.438	1.203	(0.055)	0.332	1.209	(0.058)	0.391	1.093	(0.059)
AR(5)	0.266	1.067	(0.053)	0.373	1.026	(0.051)	0.282	1.027	(0.053)	0.351	0.979	(0.062)
ARX_{opt}	0.145	0.580	(0.061)	0.134	0.368	(0.065)	0.174	0.634	(0.066)	0.188	0.525	(0.062)
2. Penalized regressions												
Ridge Regressions	0.286	1.146	(0.063)	0.298	0.819	(0.057)	0.324	1.180	(0.058)	0.312	0.872	(0.054)
Lasso	0.229	0.919	(0.064)	0.321	0.884	(0.053)	0.284	1.033	(0.057)	0.320	0.894	(0.057)
Elastic Net	0.260	1.042	(0.054)	0.363	0.997	(0.055)	0.274	0.997	(0.059)	0.337	0.941	(0.053)
LARS	0.056	0.223	(0.059)	0.192	0.528	(0.061)	0.200	0.729	(0.055)	0.341	0.953	(0.054)
3. Dynamic factor models												
Unrestricted	0.252	1.008	(0.066)	0.317	0.872	(0.051)	0.300	1.091	(0.062)	0.382	1.068	(0.062)
Lasso	0.302	1.209	(0.059)	0.323	0.888	(0.054)	0.161	0.585	(0.062)	0.383	1.069	(0.057)
Elastic Net	0.284	1.137	(0.063)	0.264	0.725	(0.053)	0.350	1.273	(0.062)	0.392	1.094	(0.062)
LARS	0.252	1.008	(0.064)	0.293	0.806	(0.057)	0.217	0.791	(0.053)	0.377	1.053	(0.053)
4. Forecast combinations												
a. Equal weights												
Unrestricted	0.259	1.037	(0.056)	0.324	0.890	(0.055)	0.299	1.089	(0.057)	0.383	1.069	(0.061)
Lasso	0.265	1.061	(0.055)	0.324	0.890	(0.051)	0.299	1.088	(0.057)	0.358	1.000	(0.06)
Elastic Net	0.249	1.000	(0.055)	0.323	0.888	(0.056)	0.282	1.028	(0.057)	0.374	1.045	(0.059)
LARS	0.259	1.037	(0.058)	0.330	0.908	(0.051)	0.290	1.058	(0.058)	0.367	1.026	(0.06)
b. Weight type 1												
Unrestricted	0.247	0.990	(0.053)	0.348	0.957	(0.052)	0.332	1.209	(0.057)	0.368	1.028	(0.059)
Lasso	0.266	1.067	(0.056)	0.382	1.050	(0.054)	0.307	1.119	(0.058)	0.351	0.982	(0.059)
Elastic Net	0.248	0.995	(0.054)	0.365	1.002	(0.051)	0.324	1.180	(0.059)	0.359	1.002	(0.059)
LARS	0.257	1.031	(0.057)	0.355	0.977	(0.049)	0.291	1.059	(0.058)	0.369	1.031	(0.06)
c. Weight type 2												
Unrestricted	0.188	0.753	(0.056)	0.333	0.914	(0.058)	0.241	0.878	(0.057)	0.382	1.068	(0.053)
Lasso	0.190	0.760	(0.058)	0.331	0.909	(0.055)	0.187	0.680	(0.052)	0.357	0.999	(0.056)
Elastic Net	0.197	0.791	(0.055)	0.331	0.909	(0.059)	0.201	0.734	(0.05)	0.399	1.114	(0.054)
LARS	0.195	0.783	(0.059)	0.305	0.837	(0.056)	0.210	0.765	(0.052)	0.308	0.860	(0.058)
d. PCA based												
Unrestricted	0.275	1.104	(0.056)	0.348	0.957	(0.054)	0.307	1.119	(0.058)	0.366	1.022	(0.061)
SSE weights	0.266	1.067	(0.055)	0.365	1.002	(0.054)	0.333	1.211	(0.058)	0.351	0.979	(0.061)
SSE minimalized	0.265	1.061	(0.058)	0.346	0.952	(0.057)	0.291	1.059	(0.056)	0.366	1.023	(0.06)
First factor regression	0.246	0.986	(0.059)	0.280	0.769	(0.055)	0.299	1.089	(0.055)	0.312	0.872	(0.062)
5. Bootstrap aggregation												
Number of samples = 1	0.277	1.111	(0.059)	0.325	0.893	(0.054)	0.237	0.863	(0.052)	0.294	0.822	(0.06)
Number of samples = 100	0.277	1.111	(0.06)	0.365	1.002	(0.05)	0.194	0.705	(0.049)	0.368	1.029	(0.058)
Number of samples = 150	0.311	1.247	(0.064)	0.349	0.959	(0.051)	0.234	0.851	(0.05)	0.338	0.944	(0.06)
Number of samples = 200	0.247	0.990	(0.062)	0.290	0.797	(0.053)	0.217	0.791	(0.049)	0.353	0.987	(0.059)

Notes: (i) The table shows the estimated coefficients ($\hat{b}$), coefficients relative to the AR_{opt} model ($\hat{b}/\hat{b}_a$) and corresponding standard errors (std. err.) of b in (35). The reported standard errors are Newey-West standard errors, which correct for potential autocorrelation and/or heteroskedasticity. (ii) The significance of a coefficient can be determined as the ratio of $\hat{b}$ over its standard error, which is significant if it is larger than 1.96 (critical value for the 5% significance level). (iii) The coefficients and standard errors are not reported for the Random Walk model, as it is the one-month lagged value of the actual returns. Due to the persistence in realized volatility, the random walk model moves exactly the same as the actual returns.

5.2 Economic performance

The first part of the economic evaluation aims to reveal the ability of the models to predict the monthly Value-at-Risk (VaR). Table 6 below shows the results for the 95% (1-α) VaR using a normal distribution. As explained in the previous section, the results have been subdivided in unconditional coverage (UCC), independence (IND) and conditional coverage (CCV). In addition, the table shows the percentage of violations (π_1) of each model. The VaR performance of each asset class will be evaluated separately below.

For the equity class, the models seem to perform quite well in general. Most models have a proportion of violations below 6% and are close to the expected 5% violations. Furthermore, all models do not appear to suffer from volatility clustering, leading to consecutive violations of the VaR, as the independence test is not rejected for any of the tested models besides the historical mean model. The unconditional coverage, testing whether the number of violations is significantly different from the expected violations, also shows promising results. As both tests yield positive results, it is not surprising that this is also the case for the conditional coverage test.

Though the models work quite well in general, there are subtle differences between them. Firstly, the bagging forecasts seem to outperform all the other models in terms violations. With only a mere 4% of violations, it remains even below the expected 5%. At first sight this seems odd, as their forecasting performance is quite poor from a statistical point of view. However, lack of accuracy can go hand in hand with a 'good' VaR forecast performance. A consistent overestimation of the volatility will also lead to a consistent overestimation of the monthly VaR. This on its turn will cause relatively few violations, as it is set to low in general. A strong deviation from the expected 5% violations, either below or above, therefore also signals a lack of forecasting accuracy. From that perspective, there are several models that perform quite strong. In particular, the ridge regression, elastic net and forecast combinations with equal weights. Those models all have a number violations close to to expectations.

As strong as the performance of the VaR forecasts is for equities, as weak it is for the commodities class. The forecasts consistently underestimate the actual VaR, leading to an average number of violations of roughly 8 to 9%. Hence, it is not surprising that the UCC test is rejected for almost all models. The bagging forecasts based on 100 samples come closest to the expected number of deviations and is the only model for which the null hypothesis of the UCC test is not rejected, though its number of violations is still quite substantial. Similar to what has been observed for the equities class, the models in the commodities class also do not seem to suffer from volatility clustering. This leads to conclude that although the forecasts violate the thresholds quite frequently, they do not happen more frequent in certain periods. This conclusion is also expressed in the CCV test, which is largely determined by the strong negative performance of the UCC test. Only the bagging forecasts appear to be 'able' to capture the monthly VaR of commodities, albeit likely that this is due to the consistent overestimation of the volatility.

The bonds class, however, shows again more promising results. The proportion of violations

are for all models roughly between 5% and 7%. Quite unexpectedly, this does not lead to a rejection of the UCC test for any of the models. Though the number of violations appear to be quite different from the expected 5%, they are not significantly different. Furthermore, as for the other asset classes, the models all seem able to incorporate volatility clustering (for as far it is present). More interestingly is that most models perform quite similarly, some even display the exact same performance. A possible explanation for this is similar to the 'overestimation issue' of the bagging forecasts. If there are two models of which one predicts higher values of volatility in all cases, this will likely show up in an analysis on forecast accuracy. However, it could occur that the violations of the forecasted VaR for the two models are exactly similar. Consequently, this leads to a similar performance in terms of violations.

At first sight, it can be concluded that the forecasts are very well able to forecast the monthly 95% Value-at-Risk for equities and bonds. However, the forecasts seem unable to capture the actual VaR for the commodities class. Several complications make it difficult to draw conclusions on a general level. As explained in Section 2, a naive transformation of the log volatility forecasts is used to convert them back to 'normal' volatility forecasts. Though the error might be small, it cannot be ruled out that it influences the outcome of the economic evaluation. Further, the results are based on a normal distribution. The choice of the normal distribution could explain the large differences in performance. It might just be that the normal distribution is completely inappropriate for the commodities class, while it approximates the distribution of equities and bonds quite well. An evaluation of the VaR forecasts based on the Student's t-distribution should reveal whether this distribution is more appropriate for the commodities class.

Table 18 shows the results of the 95% VaR forecasts using a Student's t-distribution with 5 degrees of freedom. This distribution is known to have fatter tails than the normal distribution, which therefore implicitly assumes there are more observations with 'extreme' values compared to the mean of the distribution. This on its turn leads to higher critical values for the same critical p-values, compared to the normal distribution. Ultimate this leads to a higher VaR, thereby making it more difficult for the actual return to violate the VaR threshold. This is also what is observed, as the proportion of violations is substantially lower for all models in all asset classes. Though this is unnecessary for the equities and bonds class, for which the normal distribution seems appropriate, it appears that a t-distribution fits better to the forecasts of the commodities' realized volatility. The choice of a distribution is therefore crucial in making VaR forecasts as this can strongly influence the performance.

So far the performance of the forecasts as a VaR estimate have only been analysed for an α of 95%. Different levels of α give a different threshold, which may lead to new insights. Setting α at 99%, for example, will reveal whether the violations of the 95% VaR are primarily extreme violations or that the violations are only of a moderate size. As the 99% VaR sets the threshold lower than the 95% VaR, only 1% violations are expected. If more than one 1% violations occur, which leads to a rejection of the UCC test, this implies that the violations are quite extreme. In other words, if a violation occurs a high loss is likely. Similar reasoning can be applied to comparing the 90% VaR and the 95% VaR. An adequate forecast of the 95% VaR

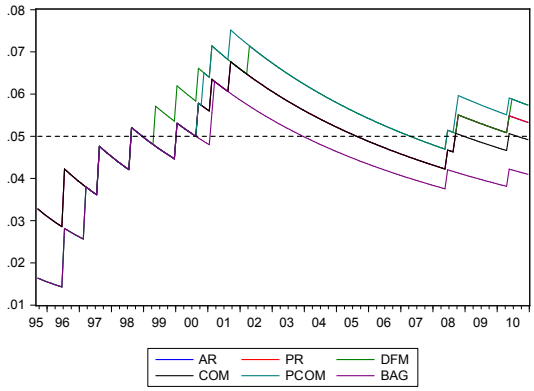
does not necessarily imply that the 90% VaR will perform equally well. If there are primarily small violations this will not be traced by the 95% VaR threshold. A significant violation of the 90% VaR but not of the 95% VaR, signals that violations do exist but that they are on average relatively small. Tables 15 and 16 provide the results for the 90% and 99% VaR based on a normal distribution, respectively. The 90% VaR gives the same picture as the 95% VaR does: It is accurate for equities and bonds, but a poor estimate for commodities. For equities and bonds, this implies that the violations are well distributed and can be accurately captured by the forecasts. The 99% VaR also appears to be accurate for the equities class, but performs poorly for both the commodities and bonds class. As the 95% VaR is accurate for the bonds class, this suggests that the violations of the 95% VaR do not occur more than expected, but are on average quite large.

The 90% and 99% VaR performance based on a t-distribution with 5 degrees of freedom are displayed in Table 17 and 19, respectively. The 90% VaR shows, similar to the results for the normal distribution based VaR, the same picture as the 95% VaR. It works best for the commodities class, though the proportion of violations still is substantial for some models. For the equities and commodities classes the forecasts overestimate the 90% VaR, as the proportion of violations is on average quite small. The 99% VaR appears to work quite well for all classes, though there is a slight overestimation of the 99% VaR by the forecasts of the equities an commodities class. Overall, the 95% VaR based on a normal distribution represents the performance of the forecasts quite well for the equities and bonds class and the 95% based on a t-distribution for the commodities class.

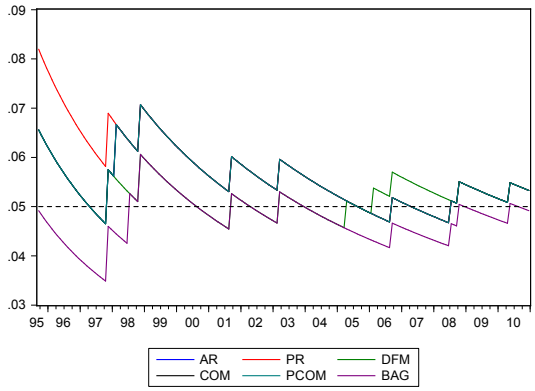
Still, the evaluation over the full sample gives only a one sided view on the performance of the model to act as a 95% Value-at-Risk estimate. Plotting the percentage of violations over time will provide further insight in the performance of the forecasts as VaR estimate. Figure 6 reveals the evolution of the fraction violations using an expanding window for the best model of each model class (i.e. model with the lowest UCC value in each class). A vertical increase in the line implies that there has occurred a violation of the 95% VaR. A downward trend indicates a period in which no new violation has occurred, which will lead to a decline of the fraction of violations as time passes by. Most strikingly, all models appear to perform similarly or even exactly the same (not all lines are visible, which implies that a model exacly mimics the pattern of another model). Though, between asset classes the is different. For the equities class most violations are quite early in the out-of-sample period. The fraction of violations builds up to a peak around 2002. After 2002 it declines slowly, with no violation up until the start of 2008. This is, however, not surprising as markets were struck by the financial crisis back then. The commodities class displays a more stable evolution of the percentage of violations, as the violations are equally spread over the sample. Lastly, the bonds class seems to move in the opposite direction of the equities class. The fraction of violations shrinks to a minimum in 1999 and then goes into a period where violations occur quite frequently. In the most recent period, 2005 to 2010, there are hardly any violations. Overall, the models appear to be well suited to predict the 95% Value-at-Risk of the returns for all asset classes. Though, the models do not seem to be able to perform better (or even differentiate) from the AR_{opt} model, which leads to

question what the added value is of the macroeconomic variables.

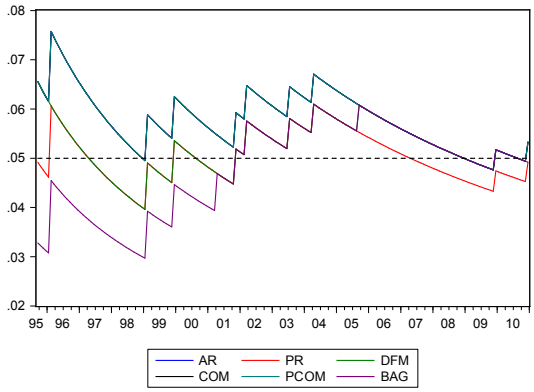
Figure 6: Percentage of violations over time of the 95% Value-at-Risk forecasts of the best performing models



(a) Equities



(b) Commodities



(c) Bonds

Notes: (i) The figure shows the 95% Value-at-Risk violations as a fraction of the sample. The sample is based on an expanding window, that is initialized with the first 60 observations (i.e. 5 years). (ii) Each model class is represented by the model which produces the 95% Value-at-Risk forecasts, which is based on the conditional coverage test value. The classes can be subdivided in penalized regressions (PR), dynamic factor models (DFM), forecast combinations (COM), PCA-based forecast combinations (PCOM) and Bagging (BAG). In addition, the figures display the performance of the optimized autoregressive model (AR). (iii) The 95% Value-at-Risk forecasts are based on a normal distribution for the equities and bonds class, and based on a t-distribution with 5 degrees of freedom for commodities.

Table 6: Results of the 95% Value-at-Risk forecasts using a normal distribution

	Equities				Commodities				Bonds			
	π_1 (%)	*UCC*	*IND*	*CCV*	π_1 (%)	*UCC*	*IND*	*CCV*	π_1 (%)	*UCC*	*IND*	*CCV*
1. Benchmark models												
Historical mean	9.053	6.850	7.021	13.871	12.346	19.949	1.638	21.587	6.173	0.657	1.286	1.943
Random Walk	3.704	0.941	0.971	1.912	9.053	6.850	0.549	7.399	7.407	2.599	0.066	2.665
AR_{opt}	5.350	0.061	1.911	1.972	9.465	8.174	0.347	8.521	5.350	0.061	0.193	0.254
AR(1)	4.938	0.002	0.262	0.264	11.111	14.396	1.472	15.868	6.173	0.657	0.023	0.680
AR(2)	6.173	0.657	1.106	1.763	9.465	8.174	0.347	8.521	6.173	0.657	0.023	0.680
AR(3)	5.761	0.283	1.475	1.758	9.465	8.174	0.347	8.521	6.173	0.657	0.023	0.680
AR(4)	5.350	0.061	1.911	1.972	9.053	6.850	0.549	7.399	5.350	0.061	0.193	0.254
AR(5)	5.350	0.061	1.911	1.972	9.465	8.174	0.347	8.521	5.350	0.061	0.193	0.254
ARX_{opt}	7.407	2.599	1.891	4.490	9.053	6.850	1.999	8.849	5.761	0.283	0.086	0.370
2. Penalized regressions												
Ridge Regressions	4.938	0.002	0.262	0.264	8.230	4.506	0.086	4.592	4.527	0.118	0.947	1.065
Lasso	5.350	0.061	1.470	1.532	8.230	4.506	0.086	4.592	4.938	0.002	0.347	0.349
Elastic Net	5.350	0.061	0.133	0.194	9.465	8.174	1.566	9.740	6.584	1.173	0.947	2.120
LARS	5.350	0.061	0.133	0.194	8.642	5.626	2.493	8.119	6.173	0.657	1.839	2.496
3. Dynamic factor models												
Unrestricted	5.350	0.061	0.133	0.194	9.053	6.850	0.549	7.399	4.115	0.425	0.771	1.196
Lasso	5.761	0.283	0.049	0.332	9.465	8.174	1.566	9.740	5.350	0.061	0.193	0.254
Elastic Net	4.938	0.002	1.354	1.356	10.700	12.708	3.777	16.485	4.938	0.002	0.347	0.349
LARS	6.173	0.657	0.947	1.604	8.642	5.626	0.022	5.648	5.761	0.283	0.086	0.370
4. Forecast combinations												
a. Equal weights												
Unrestricted	5.350	0.061	1.911	1.972	9.465	8.174	0.347	8.521	5.350	0.061	0.193	0.254
Lasso	4.938	0.002	0.262	0.264	9.465	8.174	0.347	8.521	5.350	0.061	0.193	0.254
Elastic Net	5.350	0.061	1.911	1.972	9.465	8.174	0.347	8.521	5.350	0.061	0.193	0.254
LARS	4.527	0.118	0.440	0.559	9.053	6.850	0.000	6.850	5.350	0.061	0.193	0.254
b. Weight type 1												
Unrestricted	5.350	0.061	1.911	1.972	9.465	8.174	0.347	8.521	5.350	0.061	0.193	0.254
Lasso	5.350	0.061	1.911	1.972	9.053	6.850	0.549	7.399	5.350	0.061	0.193	0.254
Elastic Net	5.350	0.061	1.911	1.972	9.465	8.174	0.347	8.521	5.350	0.061	0.193	0.254
LARS	4.938	0.002	0.262	0.264	9.053	6.850	0.000	6.850	5.350	0.061	0.193	0.254
c. Weight type 2												
Unrestricted	5.350	0.061	1.911	1.972	9.053	6.850	0.549	7.399	5.350	0.061	0.193	0.254
Lasso	5.350	0.061	1.911	1.972	9.465	8.174	0.347	8.521	5.350	0.061	0.193	0.254
Elastic Net	5.350	0.061	1.911	1.972	9.465	8.174	0.347	8.521	5.350	0.061	0.193	0.254
LARS	4.527	0.118	0.440	0.559	9.465	8.174	0.018	8.192	5.761	0.283	0.086	0.370
d. PCA based												
Unrestricted	5.350	0.061	1.911	1.972	9.465	8.174	0.347	8.521	5.350	0.061	0.193	0.254
SSE weights	5.350	0.061	1.911	1.972	9.053	6.850	0.549	7.399	5.350	0.061	0.193	0.254
SSE minimalized	5.761	0.283	0.049	0.332	9.877	9.594	1.192	10.786	5.350	0.061	0.193	0.254
First factor	6.173	0.657	0.007	0.664	9.053	6.850	0.549	7.399	5.350	0.061	0.193	0.254
5. Bootstrap aggregation												
#Bootstrap samples = 1	4.115	0.425	0.674	1.099	8.230	4.506	0.342	4.849	4.938	0.002	1.247	1.249
#Bootstrap samples = 100	4.115	0.425	0.859	1.284	7.819	3.495	0.192	3.688	4.938	0.002	1.141	1.143
#Bootstrap samples = 150	4.115	0.425	0.859	1.284	8.230	4.506	0.086	4.592	4.938	0.002	1.141	1.143
#Bootstrap samples = 200	4.115	0.425	0.859	1.284	8.230	4.506	1.107	5.613	4.938	0.002	1.141	1.143

Notes: (i) Abbreviations UCC, IND and CCV stand for Unconditional Coverage, Independence and Conditional Coverage test, respectively. (ii) The displayed values correspond to the Chi-Squared distribution with one degree of freedom for the UCC and IND test, and 2 degrees of freedom for the CCV test. The 5% critical value of the Chi-squared distribution with one or two degrees of freedom is equal to 3.84 and 5.99, respectively. (iii) $\pi_1(\%)$ represents the percentage of violations of the VaR threshold.

The second part of the economic evaluation entails the analysis of the mean-variance investment strategy for a myopic investor with a one-month horizon. Table 7 below shows the results for the investor with a risk aversion coefficient (γ) of 2. The Table is subdivided in average monthly portfolios return in excess of the risk-free rate (approximately 0.37% or 37 basis points (bps) per month), Sharpe Ratios on monthly returns, the t-value resulting from the test based on (45), and the average weight given to the asset. The results of the investment strategy with a risk aversion of 5 and 10 are presented in Table 20 and 21, respectively. As the level of risk aversion does not appear to strongly influence general pattern the results, the focus will lie on the results of the strategy with a γ of 2.

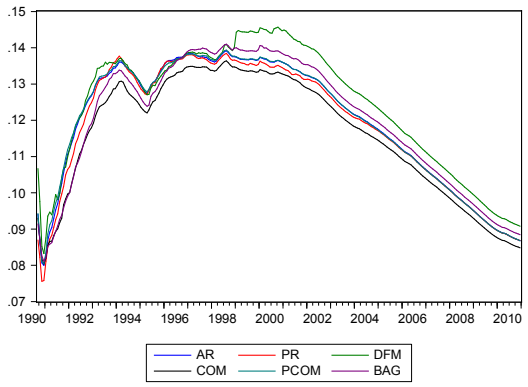
Starting again with the equities class, it is striking to see that the forecasts show a lot of resemblance. All models seem only marginally able to 'beat' the risk-free rate, as the average excess return is roughly only 6 bps and the Sharpe ratios are remarkably low. However, they all to perform better than a simple '50-50' buy-and-hold strategy given the low Sharpe ratio of the strategy compared to the portfolio returns based on the forecast models. As the BAH strategy has the highest average monthly excess return but has the lowest Sharpe ratio, this signals that the main improvement of the mean-variance portfolios lies in a reduction of the return variance. This reduction in variance is most likely due to the relatively small weight that is invested in the asset, as the average weight is roughly 9%. This implies that the risk-free asset gets on average a relatively large weight, which will lead to portfolio returns close to the return of the risk-free asset. As the average returns and Sharpe ratios are measured over the returns in excess of the risk-free rate, this explains why these values are relatively low. As it seems unlikely that the full out-of-sample period was subject to strong volatility, which would explain a flee towards the risk-free asset, the most logical explanation is that the volatility forecasts on average overestimate the actual volatility. A similar pattern was observed for the Value-at-Risk estimates, where it seemed that the VaR threshold was often set too low.

What is observed for the equities class also holds for the commodities class. The average monthly excess returns and Sharpe ratios are even lower, though they are positive, which makes it almost not worth investing in the commodities asset by means of a mean-variance portfolio strategy. The best performance of the mean-variance portfolios is found in the bonds class. Due to the average negative return obtained by the BAH strategy, it is relatively easy for the mean-variance portfolios to outperform this strategy. However, actively managing the weight assigned to the asset does pay off as the average returns of the portfolios are slightly positive. Subsequently, the Sharpe ratios of the portfolios are almost all positive and significant. The average weight assigned to the asset is, however, remarkably low. Only investing a small amount in the bonds asset already leads to a better average return than simply investing in the risk-free asset.

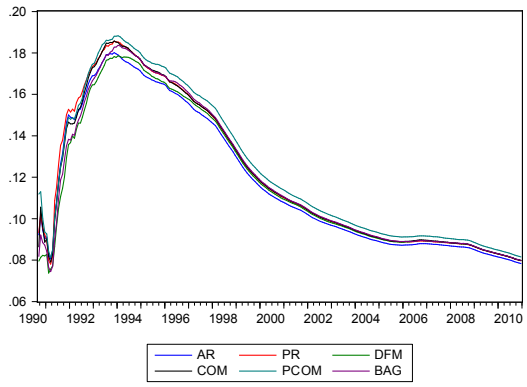
As the weights assigned to the assets are surprisingly low, it becomes especially interesting to analyse the behaviour of the weights over time. Figure 7 below shows the average weight assigned to the asset for the best model of each model class (i.e. model with the highest Sharpe ratio in each class) using an expanding window. The models within each asset class all follow more or less the same pattern and also show strong similarities between asset classes. For the

equities class, the average weight slowly increases to a peak at roughly 14% in 2001 after which it quite steeply declines to the average weight of approximately 9% of the full sample. This reveals that the relatively low weight of the full sample is primarily due to the decline of the weight in the most recent period. Though the decline stretches over a quite long period, it could be linked to the recent crisis. An increase in market volatility directly influences the weight given to the asset in a mean-variance portfolio. As already mentioned, the same patterns are observed in the commodities and bonds classes. The peak in the commodities is more extreme and comes even earlier in the dataset, but its decline is more gentle. The bonds class also reaches its peak quite early in the dataset and also declines afterwards, but there is more variation in the speed of the decline. Overall, the low average weights of the full sample can be explained by the decrease in the later, more recent part of the out-of-sample period. This hints at increased volatility in more recent periods, which inevitably leads to a shift towards to the risk-free asset.

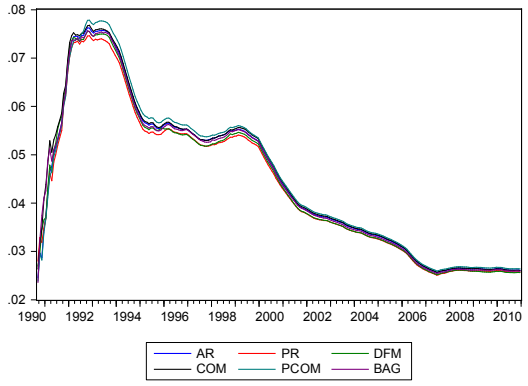
Figure 7: Evolution of the portfolio weights over time of the mean-variance portfolio strategy based on the best models



(a) Equities



(b) Commodities

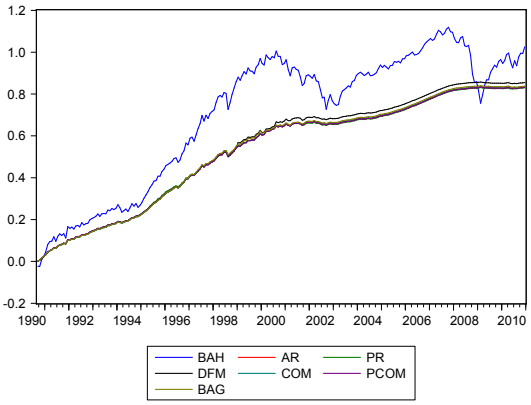


(c) Bonds

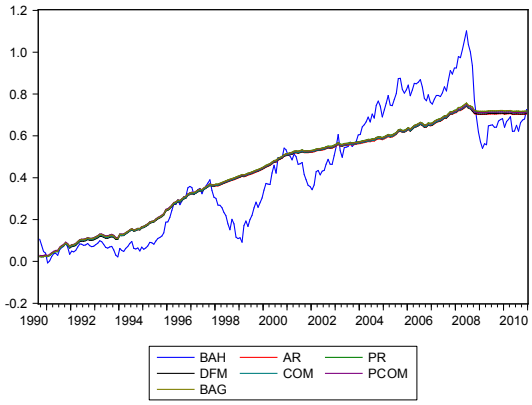
Notes: (i) The figure shows the average sample weight assigned to the asset in the mean-variance portfolio strategy, based on the best LRV forecast model of each model class. The sample is based on an expanding window. (ii) The classes can be subdivided in penalized regressions (PR), dynamic factor models (DFM), forecast combinations (COM), PCA-based forecast combinations (PCOM) and Bagging (BAG). In addition, the figures display the performance of the optimized autoregressive model (AR).

Similar to the above discussed statistical and economic evaluation metrics, the portfolio returns can also be analysed over time to provide further insight in the performance of the mean-variance portfolios. Figure 8 below displays the cumulative returns, estimated as the sum of monthly excess returns, of the Buy-and-Hold strategy and the best performing portfolio of each model class (i.e. model that has the portfolio with the highest Sharpe ratio). The figures of each asset class all highlight the big difference in the return volatility of an investor who manages his portfolio according to the BAH strategy and the investor who applies the RV forecasts to invest according to the mean-variance portfolio. As expected, the cumulative returns of the mean-variance portfolios increase gradually over time, displaying little variance. In contrast, the BAH strategy is strongly influenced by the volatile behaviour of the asset returns. This immediately reveals why the portfolios almost all have a higher Sharpe ratio than the BAH strategy. The average monthly return of the BAH strategy may be higher, but it comes at the price of increased volatility. As the cumulative returns of BAH and mean-variance portfolios in the equities and commodities class are fairly close to each other, it appears to pay-off to actively manage the portfolio. This is especially the case for the bonds class, as the mean-variance portfolios perform significantly better than the BAH-strategy. The mean-variance strategy thus seem to be a strategy worth pursuing, though the returns are not corrected for any transaction costs. This correction may eat up a large share of the profit, but this is particularly the case when the portfolio is managed frequently. As monthly returns and thus monthly revisions are analysed here, the results may still hold even without a correction for transaction costs.

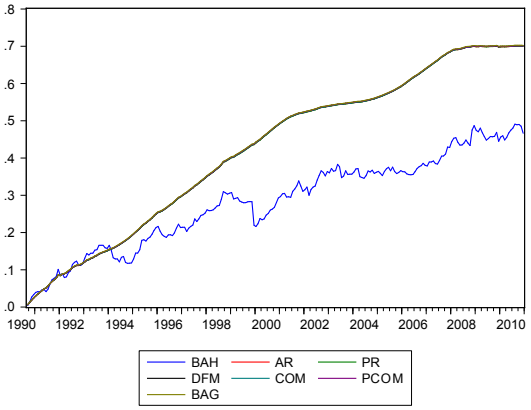
Figure 8: Cumulative Returns over time of the mean-variance portfolios based on the best models



(a) Equities



(b) Commodities



(c) Bonds

Notes: (i) The figure shows the evolution of the cumulative returns of the mean-variance portfolios, based on the best LRV forecast model of each model class. The sample is based on an expanding window. (ii) The classes can be subdivided in penalized regressions (PR), dynamic factor models (DFM), forecast combinations (COM), PCA-based forecast combinations (PCOM) and Bagging (BAG). In addition, the figures display the performance of the optimized autoregressive model (AR) and the Buy-and-Hold strategy (BAH) .

Table 7: Performance of the mean-variance portfolio with $\gamma = 2$

	Equities				Commodities				Bonds			
	$\bar{r}$ (bps)	SR	t-val.	$\bar{w}$ (%)	$\bar{r}$ (bps)	SR	t-val.	$\bar{w}$ (%)	$\bar{r}$ (bps)	SR	t-val.	$\bar{w}$ (%)
Buy-and-Hold Strategy	13.74	0.06	–	–	1.46	0.00	–	–	-9.18	-0.09	–	–
1. Benchmark models												
Historical mean	6.49	0.12	1.18	9.19	0.43	0.01	0.08	8.04	0.18	0.02	2.43	2.56
Random Walk	6.12	0.15	1.70	9.04	0.84	0.02	0.30	8.16	0.57	0.06	2.90	2.77
AR_{opt}	5.85	0.14	1.56	8.68	0.56	0.01	0.19	7.84	0.39	0.05	2.68	2.61
AR(1)	5.91	0.14	1.49	8.67	0.49	0.01	0.15	7.98	0.27	0.03	2.47	2.55
AR(2)	5.68	0.13	1.47	8.69	0.58	0.01	0.20	7.93	0.30	0.04	2.47	2.59
AR(3)	5.77	0.14	1.49	8.75	0.60	0.01	0.21	7.91	0.35	0.04	2.60	2.60
AR(4)	5.81	0.14	1.54	8.68	0.50	0.01	0.16	7.89	0.38	0.04	2.67	2.59
AR(5)	5.86	0.14	1.57	8.68	0.55	0.01	0.18	7.86	0.40	0.05	2.72	2.61
ARX_{opt}	5.19	0.12	1.14	9.03	1.11	0.02	0.41	8.19	-0.16	-0.01	1.28	3.30
2. Penalized regressions												
Ridge Regressions	5.56	0.13	1.36	8.67	1.14	0.03	0.48	7.97	0.40	0.05	2.68	2.63
Lasso	5.63	0.13	1.39	8.68	0.86	0.02	0.35	7.85	0.39	0.05	2.71	2.57
Elastic Net	5.64	0.13	1.37	8.80	1.25	0.03	0.54	7.95	0.34	0.04	2.56	2.53
LARS	5.73	0.14	1.49	8.68	0.77	0.02	0.29	7.86	0.44	0.05	2.70	2.62
3. Dynamic factor models												
Unrestricted	5.80	0.14	1.49	8.66	0.55	0.01	0.18	7.90	0.43	0.05	2.72	2.56
Lasso	6.76	0.14	1.66	9.08	-0.04	0.00	-0.13	8.01	0.36	0.04	2.59	2.57
Elastic Net	7.36	0.15	1.91	8.90	-0.29	-0.01	-0.26	8.00	0.38	0.04	2.60	2.58
LARS	5.23	0.11	0.90	8.90	0.67	0.02	0.23	7.99	0.36	0.04	2.66	2.68
4. Forecast combinations												
a. Equal weights												
Unrestricted	5.84	0.14	1.57	8.67	0.74	0.02	0.28	7.84	0.39	0.05	2.67	2.61
Lasso	5.79	0.14	1.55	8.69	0.75	0.02	0.29	7.85	0.39	0.05	2.68	2.61
Elastic Net	5.77	0.14	1.53	8.67	0.74	0.02	0.28	7.85	0.39	0.05	2.68	2.61
LARS	5.78	0.14	1.53	8.68	0.84	0.02	0.33	7.84	0.40	0.05	2.68	2.61
b. Weight type 1												
Unrestricted	5.83	0.14	1.56	8.67	0.64	0.01	0.23	7.84	0.39	0.05	2.67	2.60
Lasso	5.80	0.14	1.54	8.69	0.64	0.01	0.23	7.85	0.39	0.05	2.67	2.60
Elastic Net	5.76	0.14	1.52	8.67	0.64	0.01	0.23	7.86	0.39	0.05	2.67	2.61
LARS	5.77	0.14	1.51	8.68	0.75	0.02	0.28	7.85	0.39	0.05	2.67	2.60
c. Weight type 2												
Unrestricted	5.83	0.14	1.65	8.44	0.61	0.01	0.21	7.91	0.36	0.04	2.62	2.57
Lasso	5.93	0.15	1.69	8.48	0.62	0.01	0.21	7.96	0.42	0.05	2.74	2.58
Elastic Net	5.78	0.14	1.63	8.48	0.69	0.02	0.25	8.00	0.37	0.04	2.64	2.59
LARS	5.79	0.14	1.61	8.52	0.90	0.02	0.36	7.97	0.42	0.05	2.71	2.61
d. PCA based												
Unrestricted	5.81	0.14	1.56	8.68	0.74	0.02	0.28	7.86	0.40	0.05	2.68	2.61
SSE weights	5.83	0.14	1.55	8.68	0.64	0.01	0.22	7.87	0.40	0.05	2.68	2.60
SSE minimalized	5.76	0.14	1.52	8.71	0.69	0.02	0.25	7.96	0.42	0.05	2.74	2.64
First factor	5.72	0.13	1.43	8.81	0.91	0.02	0.34	8.15	0.40	0.05	2.68	2.59
5. Bootstrap aggregation												
#Bootstrap samples = 1	6.03	0.14	1.56	8.84	0.91	0.02	0.36	7.95	0.48	0.06	2.85	2.59
#Bootstrap samples = 100	5.62	0.13	1.43	8.74	1.10	0.02	0.46	7.99	0.46	0.05	2.85	2.56
#Bootstrap samples = 150	5.68	0.13	1.46	8.75	1.17	0.03	0.49	7.99	0.45	0.05	2.83	2.55
#Bootstrap samples = 200	5.62	0.13	1.42	8.74	1.19	0.03	0.50	7.98	0.45	0.05	2.84	2.55

Notes: (i) Abbreviations $\bar{r}$, SR, t-val. and $\bar{w}$ stand for average monthly return in excess of the risk-free rate (expressed in percentages), monthly Sharpe Ratio, the t-statistic for the comparison of the Sharpe Ratios, and average weight (expressed in percentages) assigned to the asset, respectively. (ii) The t-value is significant at a 5% significant level if its absolute value is larger than 1.96.

5.3 Which variables matter

Table 8 shows the proportion of times a variable is used in a forecasting model for all three variable selection techniques: lasso, elastic net and LARS. It appears that lasso and LARS quite severely penalize larger models as they select only 12 variables on average. The impact of the 'ridge component' in elastic net is therefore clearly visible, as it prevents variables to become exactly zero and selects 25 variables on average. As the ridge regression is amongst the best performing models of the penalized regressions class, especially for equities, it seems that the advantage of having a more parsimonious model has its limits.

Though the selection techniques have their own unique characteristics they are, of course, closely related. This is also expressed in the ranking of variables, based on the proportion of times they were selected. Focusing on the five highest ranked variables for each technique and asset class there appear to be similarities both between techniques and asset classes. Starting with the equity class, all three techniques select the equity market variables and the monthly inflation rate the most. This consistency between different techniques leads to think that equity market variables such as the dividend-to-price ratio and return of the underlying index, combined with monthly inflation, work best to predict the realized volatility. The other macroeconomic variables are considered less useful. The average proportion for the elastic net is quite high, which is not really surprising, given that the model on average selects roughly 28 variables.

The best five variables for the commodities class show quite some resemblance compared to the equities class, as the majority of the best ranked variables are either from the equity market or macroeconomic variables group. This resemblance is most likely due to the fact that the realized volatility of equities and commodities are most correlated with each other, as shown in Table 1. Quite interesting to see is that the annual inflation rate (INFA) is selected more often than the monthly inflation rate, as is the case in the equities class. It could be that the realized volatility of commodities behaves more stable, though this is not directly visible in Figure 1 or the standard deviation in Table 1, which as a result makes it better predictable by the annual inflation rate.

The average forward discount appears to be the most important variable in predicting the realized volatility of the foreign exchange class. As this variable proxies the general appreciation or depreciation of the US dollar compared to other major currencies, it might be best able to capture the volatility as well in this market. Similar to the leverage effect as observed for equity returns, a strong change in the position of the US dollar against the world might be associated by an increase of volatility in the foreign exchange market. Furthermore, the macroeconomic variables also play an important role. As exchange rates are directly influenced by the state of the economy, in particular by inflation, this is not surprising.

More remarkable is the overlap of best predictors of the foreign exchange class and the bonds class. However, their realized volatilities are most correlated with each other as well. The interest rates, spread and other bond market factors appear not be very helpful in predicting the realized volatility for bonds. As these variables are almost never amongst the highest ranked variables, it is difficult to assess their added value. At first sight, it appears to be limited.

The liquidity and credit risk variables, however, seem to be more useful for the bonds class. A (sudden) change in credit risk perhaps may also lead to increased uncertainty and thereby leading to increased volatility.

In general, the macroeconomic variables can be considered as the most relevant or important variables to predict realized volatility for all asset classes, inflation in particular. This leads to suspect that current inflation level is a good proxy what to expect for next month's volatility. As inflation directly affects the value of money, it also directly affects the spending behaviour of individuals. An increase in inflation will make money worth less, thereby making (financial) products relatively more costly. Periods of high inflation could therefore result in a decrease in expenditures, as people will be tempted to postpone their investments. This also affects the financial market, as people may prefer to hold their money and wait for a period of deflation. A decrease in market activity leads to a decrease in bid and ask orders, which makes the markets less 'liquid'. This implies that a single market order of reasonable size will more likely affect the stock price. Hence, the market volatility will increase. As the reaction of market participants may be lagged to some extent, the current value of inflation may give a good indication of near future market volatility.

Table 8: Average weights assigned to each predictor variable according to the lasso, Elastic Net and LARS selection procedures

	Equities			Commodities			FX-aggr.			Bonds		
	LAS	EN	LAR	LAS	EN	LAR	LAS	EN	LAR	LAS	EN	LAR
Average no. of selected variables												
	12.8	25.3	11.9	12.2	23.9	12.8	12.1	22.5	11.2	12.1	22.5	10.2
A. Equity Market Variables and Risk Factors												
D-P	59.4	76.6	46.3	23.0	56.1	18.4	43.0	55.3	44.3	43.0	55.3	25.0
E-P	56.1	83.6	52.9	36.9	68.0	27.9	27.0	65.2	32.0	27.0	65.2	18.9
MKT	75.8	83.2	52.5	18.4	46.7	9.4	19.7	50.0	17.2	19.7	50.0	5.3
SMB	22.1	66.0	35.7	52.9	79.1	66.4	33.2	58.2	35.7	33.2	58.2	26.2
HML	42.2	64.8	41.0	11.5	49.6	23.4	24.2	60.7	27.5	24.2	60.7	17.2
STR	66.8	77.5	73.8	43.0	66.8	41.8	25.0	56.1	28.3	25.0	56.1	34.0
TURN	30.7	67.2	28.3	54.5	83.2	57.8	22.5	54.1	35.7	22.5	54.1	68.9
MSCI	30.7	66.0	10.7	25.8	53.7	28.3	20.9	53.3	9.4	20.9	53.3	22.5
B. Interest Rates, Spreads and Bond Market Factors												
T-B	14.8	59.0	18.9	23.4	53.3	13.1	18.4	56.1	13.9	18.4	56.1	28.7
RTB	25.0	63.9	22.5	29.5	54.9	16.0	11.1	54.1	22.5	11.1	54.1	23.0
LTR	41.8	65.2	35.7	15.6	46.3	12.7	12.3	53.7	11.9	12.3	53.7	32.4
RBR	29.5	59.0	26.2	41.0	70.5	28.3	22.5	56.1	12.7	22.5	56.1	33.2
T-S	35.7	52.0	38.5	11.5	50.0	21.7	9.8	51.2	12.3	9.8	51.2	26.2
C-P	40.6	63.1	38.5	48.8	70.9	48.8	44.7	61.9	38.9	44.7	61.9	23.8
C. FX Variables and Risk Factors												
DOL	41.4	61.1	40.6	47.5	73.4	54.5	23.0	54.9	15.2	23.0	54.9	57.8
C-T	26.2	53.7	29.9	37.7	64.3	40.2	34.4	54.9	24.2	34.4	54.9	18.4
AFD	48.4	76.2	36.5	40.6	74.2	27.9	58.2	77.0	54.9	58.2	77.0	31.1
D. Liquidity and Credit Risk Variables												
DEF	29.5	67.2	21.7	27.9	66.8	18.9	31.1	57.0	36.5	31.1	57.0	11.9
BAS	29.9	66.0	22.1	32.4	59.0	18.4	33.6	62.3	27.5	33.6	62.3	45.5
PS	43.0	68.4	52.0	51.2	67.2	45.5	51.6	69.7	54.9	51.6	69.7	27.0
TED	29.1	63.5	23.4	42.2	69.7	29.5	32.8	57.8	27.9	32.8	57.8	35.2
E. Macroeconomic Variables												
INFM	67.6	81.6	54.9	33.2	53.3	38.5	48.8	69.7	66.8	48.8	69.7	38.1
INFA	29.5	59.4	22.5	58.6	83.2	50.4	29.1	54.5	23.4	29.1	54.5	21.3
IPM	42.6	70.9	41.0	23.0	50.8	15.6	63.5	73.8	56.6	63.5	73.8	29.5
IPGA	22.1	66.0	14.3	33.6	64.8	33.2	23.8	54.5	11.1	23.8	54.5	35.7
H-S	18.4	58.6	32.0	50.4	70.9	56.6	46.3	66.8	59.0	46.3	66.8	25.8
M1M	31.6	62.3	36.1	40.6	77.9	33.2	32.4	64.3	45.1	32.4	64.3	29.9
M1A	44.3	84.0	32.4	42.2	73.0	9.4	25.8	56.1	21.7	25.8	56.1	31.1
ORDM	40.6	68.9	46.7	40.6	70.5	48.4	16.4	55.3	25.0	16.4	55.3	48.4
ORDA	12.3	57.8	9.0	39.3	62.3	36.9	18.4	52.9	16.0	18.4	52.9	1.2
CRB	44.3	65.2	34.0	50.0	69.7	38.9	50.4	68.9	59.0	50.4	68.9	24.6
CAP	31.6	62.3	14.3	16.0	56.1	6.6	45.1	64.3	37.3	45.1	64.3	39.8
EMPL	46.7	77.5	27.5	23.8	59.8	22.1	41.8	62.3	48.4	41.8	62.3	18.0
SENT	40.6	70.1	44.7	30.3	56.1	21.3	24.6	57.4	29.1	24.6	57.4	34.4
CONF	19.3	57.0	25.4	26.6	49.6	26.6	48.8	59.0	36.9	48.8	59.0	33.2
DIFF	22.5	60.7	19.7	35.2	61.5	36.5	22.5	56.6	22.1	22.5	56.6	38.5
PMBB	23.4	66.0	21.7	14.8	50.0	9.0	37.7	55.3	33.6	37.7	55.3	44.3
PMI	12.7	57.0	6.1	19.7	53.3	6.1	35.2	61.9	28.7	35.2	61.9	11.9

Notes: (i) The lasso, elastic net and LARS are abbreviated as LAS, EN and LAR, respectively. (ii) All weights are expressed in percentages. (iii) The description and the full names of the variables can be found in Table 2.

As the best out-of-sample performance is found in the forecast combinations class of weight type 1 (i.e. SSE weighted) and PCA-based combinations (both weight type 1 and 2), it is worth investigating the weight distribution of these models. Tables 9 and 11 below show the weight distribution for weight type 1 and the PCA-based, respectively. The reported weights represent the average weight given to a forecast model which only includes variable x_i, which is part of the set of variables of the full set (unrestricted case, UR) or if it is part of the active set found by the lasso (LS), elastic net (EN) or LARS (LR).

What stands out from both Tables is that the weights according to type 1 based on the full set of variables, is roughly the same as the weights based on equal weighed combinations. As there are 38 variables in the full set, the variables in the equal weighted combinations all get a weight of 1/38 or 2.6%. This implies that the SSE of the individual models are on average similar. Individual predictors appear not to be able to significantly reduce the SSE of the AR_{opt} model, which would lead to a different, unequal distribution of the weights. This means that the average weights assigned to the variables based on one of the three variable selection techniques, is directly related to the number of times the variables are selected by those techniques. Tables

8 and 9 are therefore directly related to each other. The subset of variables selected by one of the three penalized regression methods, is most likely an equal weighted combination of those variables. This explains why the performance of the forecasts based on equal weighted combinations and of weight type 1 are so similar throughout the paper. However, this is not completely true, as the forecast combinations according to weight type 1 are slightly better than the corresponding equally weighted combinations. The PCA-based combinations confirm this idea, as the factors all get an average weight of 20%, which is equal to 1/5.

Table 9: Average weights assigned to the variables in the forecast combinations based on weight type 1

	Equities				Commodities				FX-aggr.				Bonds			
	UR	LS	EN	LR	UR	LS	EN	LR	UR	LS	EN	LR	UR	LS	EN	LR
A. Equity Market Variables and Risk Factors																
D-P	2.6	7.1	4.3	11.2	2.6	7.6	3.8	7.3	2.6	7.9	3.5	7.8	2.6	7.9	3.5	7.2
E-P	2.6	7.3	4.6	10.2	2.6	7.8	4.3	6.7	2.6	10.2	6.5	11.2	2.6	10.2	6.4	5.9
MKT	2.6	7.4	4.4	8.1	2.6	8.5	3.1	4.5	2.6	8.5	2.9	8.0	2.6	8.4	2.9	6.0
SMB	2.6	7.2	3.9	6.6	2.6	7.4	5.1	13.3	2.6	7.5	3.9	6.8	2.6	7.5	3.9	6.3
HML	2.6	7.5	4.3	8.4	2.6	7.5	3.2	4.9	2.6	10.1	4.2	8.3	2.6	10.1	4.2	7.2
STR	2.6	7.5	4.2	9.8	2.6	8.3	4.2	9.5	2.6	7.2	3.9	6.5	2.6	7.2	3.9	7.2
TURN	2.6	6.3	3.9	6.0	2.6	7.9	5.2	10.1	2.6	7.8	3.7	7.1	2.6	7.8	3.7	10.7
MSCI	2.6	7.7	4.0	4.9	2.6	7.6	3.3	5.4	2.6	8.6	3.8	5.6	2.6	8.6	3.8	10.7
B. Interest Rates, Spreads and Bond Market Factors																
T-B	2.6	6.7	3.5	6.8	2.6	6.6	3.4	5.1	2.6	10.3	4.0	6.1	2.6	10.3	4.0	7.7
RTB	2.6	7.5	4.1	6.3	2.6	7.0	3.6	5.9	2.6	9.5	3.6	9.7	2.6	9.4	3.6	10.5
LTR	2.6	7.2	3.6	7.9	2.6	7.1	2.8	5.3	2.6	7.3	3.9	4.7	2.6	7.3	3.9	10.1
RBR	2.6	7.8	3.6	13.5	2.6	7.1	4.6	8.4	2.6	9.1	4.6	5.2	2.6	9.1	4.6	11.1
T-S	2.6	7.6	3.1	6.9	2.6	7.0	3.2	4.8	2.6	7.4	3.4	5.0	2.6	7.4	3.4	6.7
C-P	2.6	7.7	3.8	7.8	2.6	9.2	4.7	11.0	2.6	7.9	4.2	7.8	2.6	7.9	4.2	6.0
C. FX Variables and Risk Factors																
DOL	2.6	7.7	3.7	8.3	2.6	8.7	4.9	11.2	2.6	8.1	3.4	6.4	2.6	8.1	3.4	11.2
C-T	2.6	7.2	3.2	6.5	2.6	7.3	4.0	7.2	2.6	7.9	3.7	6.5	2.6	7.9	3.7	7.0
AFD	2.6	7.2	4.2	9.5	2.6	7.7	4.6	9.4	2.6	8.0	6.0	10.1	2.6	8.0	6.0	8.6
D. Liquidity and Credit Risk Variables																
DEF	2.6	7.2	4.1	8.9	2.6	7.6	4.5	6.0	2.6	10.5	4.9	8.4	2.6	10.6	4.9	6.3
BAS	2.6	6.6	3.9	7.1	2.6	6.8	4.2	10.2	2.6	8.5	4.7	6.8	2.6	8.5	4.7	9.8
PS	2.6	6.7	3.8	7.7	2.6	7.3	4.1	8.0	2.6	8.2	6.2	11.3	2.6	8.2	6.2	8.5
TED	2.6	7.3	4.1	8.4	2.6	7.3	4.6	8.9	2.6	7.2	3.8	7.0	2.6	7.2	3.8	9.2
E. Macroeconomic Variables																
INFM	2.6	7.4	4.7	8.3	2.6	8.3	3.5	10.1	2.6	7.6	5.8	9.3	2.6	7.6	5.8	8.0
INFA	2.6	7.6	3.6	7.0	2.6	7.5	5.5	9.3	2.6	7.7	3.6	6.3	2.6	7.7	3.6	8.4
IPM	2.6	7.4	4.2	8.9	2.6	6.3	3.4	5.1	2.6	7.8	6.1	9.2	2.6	7.8	6.1	8.3
IPGA	2.6	8.0	4.0	5.4	2.6	7.9	4.0	6.6	2.6	9.9	4.0	12.8	2.6	9.9	4.0	12.2
H-S	2.6	6.8	3.6	6.8	2.6	9.1	4.5	14.6	2.6	8.2	5.0	10.4	2.6	8.2	5.0	6.4
M1M	2.6	8.2	4.1	7.8	2.6	7.0	4.8	7.5	2.6	7.2	4.7	7.7	2.6	7.2	4.7	6.0
M1A	2.6	7.0	4.8	7.3	2.6	7.3	4.8	5.3	2.6	8.0	3.6	5.9	2.6	8.0	3.6	9.8
ORDM	2.6	7.1	4.0	7.6	2.6	7.8	4.5	10.1	2.6	9.4	4.0	6.0	2.6	9.4	4.0	12.1
ORDA	2.6	9.3	3.4	5.4	2.6	7.4	4.0	6.9	2.6	9.1	3.4	7.0	2.6	9.1	3.4	3.7
CRB	2.6	6.6	3.7	7.8	2.6	7.9	4.8	9.5	2.6	7.8	5.1	9.0	2.6	7.8	5.1	10.8
CAP	2.6	7.1	3.7	8.5	2.6	6.7	3.8	4.7	2.6	7.4	5.0	7.6	2.6	7.4	5.0	11.2
EMPL	2.6	7.4	4.3	9.0	2.6	9.3	4.3	6.5	2.6	7.3	4.9	8.0	2.6	7.3	4.9	6.7
SENT	2.6	6.7	3.9	7.2	2.6	8.0	3.7	8.3	2.6	9.0	4.5	9.0	2.6	9.0	4.5	6.5
CONF	2.6	7.2	3.4	5.7	2.6	7.6	3.1	7.0	2.6	7.5	4.1	7.5	2.6	7.5	4.1	9.1
DIFF	2.6	6.9	3.4	5.7	2.6	7.6	4.0	8.1	2.6	7.8	4.3	6.9	2.6	7.8	4.3	9.2
PMBB	2.6	7.3	3.9	11.5	2.6	8.0	3.2	4.7	2.6	9.0	3.7	8.9	2.6	9.0	3.7	8.8
PMI	2.6	9.1	3.6	6.0	2.6	8.5	3.4	6.7	2.6	11.1	4.3	10.7	2.6	11.1	4.3	5.7

Notes: (i) The 38 variables are represented by their abbreviation, which is introduced in Table 2. (ii) The weights are determined by the first weighting type introduced in this paper. This weighting type determines the weights based on the proportion of SSE of each individual model, compared to the total SSE. (iii) The average weights are based on the average weight they have when they are part of the active set of variables. The active set of variables is either the full, unrestricted set (UR) or determined by one of the three variable selection techniques: lasso (LS), elastic net (EN) and LARS (LR).

More variation is expected is the behaviour of the weights according to the second weighting method. This method minimizes the total SSE of each estimation sample by finding the optimal combination of different available forecast models. This likely to be better able to take correlations into account, but it can backfire as well. This weighting scheme can also lead to an optimization of the in-sample error, which can lead to very extreme weights. These extreme weights do not necessarily have to optimal to forecast with. This appears to be the case, as

found in the analysis on the forecast accuracy of the models. Table 10 displays the average weights according to the second weighting method, which are indeed quite extreme. Many of the variables are disregarded completely, while others get assigned a weight of 100%. The weighting scheme appears to be highly unstable, which makes it difficult to analyse the distribution of the weights. Given that the best results come from the equal weighted combinations and weight type 1, it can be concluded that an optimization of the in-sample total SSE is not optimal and seems to have a limited application in forecasting the RV.

Table 10: Average weights assigned to the variables in the forecast combinations based on weight type 2

	Equities				Commodities				FX-aggr.				Bonds			
	UR	LS	EN	LR	UR	LS	EN	LR	UR	LS	EN	LR	UR	LS	EN	LR
A. Equity Market Variables and Risk Factors																
D-P	7.3	40.0	22.6	68.2	0.0	76.6	89.0	38.2	0.0	95.0	94.3	91.0	0.0	94.2	77.6	100
E-P	96.5	96.0	96.3	55.1	93.3	93.5	94.3	29.8	100	100	100	100	100	100	100	0.0
MKT	0.0	0.0	0.0	43.1	0.0	56.0	0.0	0.0	0.0	100	0.0	0.0	0.0	100	0.0	0.0
SMB	0.0	100	0.0	62.9	1.3	36.3	35.3	83.5	0.0	90.3	86.8	100	0.0	89.2	85.9	0.0
HML	0.0	100	0.0	64.2	0.0	0.0	0.0	85.1	0.0	96.3	88.0	87.6	0.0	98.4	0.0	21.9
STR	0.0	100	100	80.1	0.0	100	0.0	97.1	0.0	0.0	0.0	95.2	0.0	0.0	0.0	30.4
TURN	0.0	0.0	0.0	76.3	0.0	0.0	0.0	73.1	0.0	0.0	0.0	92.0	0.0	0.0	0.0	80.1
MSCI	0.0	0.0	0.0	86.0	0.0	2.3	0.0	76.6	0.0	100	0.0	0.0	0.0	100	0.0	81.0
B. Interest Rates, Spreads and Bond Market Factors																
T-B	0.0	100	100	96.8	0.0	87.1	0.0	62.9	0.0	100	100	75.8	0.0	100	100	48.5
RTB	0.0	89.9	94.8	65.2	0.0	79.6	100	82.5	0.0	82.0	100	100	0.0	99.6	100	81.7
LTR	0.0	0.0	0.0	77.9	0.0	0.0	0.0	60.0	0.0	0.0	0.0	100	0.0	0.0	0.0	77.8
RBR	0.0	95.8	96.8	98.1	0.0	96.0	95.5	68.3	0.0	75.0	100	76.0	0.0	72.5	100	69.1
T-S	0.0	88.4	0.0	86.1	0.0	54.6	100	70.0	0.0	80.4	100	55.9	0.0	75.6	100	55.6
C-P	2.7	8.5	5.0	74.9	8.3	12.4	6.9	77.6	0.0	37.5	41.7	90.5	0.0	43.3	43.2	79.4
C. FX Variables and Risk Factors																
DOL	3.8	9.1	3.8	79.2	3.4	14.1	5.4	59.6	0.0	28.3	23.6	100	0.0	18.3	24.5	44.0
C-T	0.6	13.2	3.0	57.4	1.3	16.3	6.5	53.2	0.0	32.4	19.3	75.0	0.0	35.9	24.2	11.3
AFD	5.3	5.5	4.7	67.0	12.1	5.7	13.5	47.1	0.0	0.0	100	83.0	0.0	0.0	100	86.9
D. Liquidity and Credit Risk Variables																
DEF	0.0	0.0	0.0	83.3	0.0	0.0	0.0	76.0	0.0	0.0	100	87.9	0.0	33.2	100	46.8
BAS	0.0	0.5	0.0	68.3	0.0	0.2	0.0	58.4	0.0	100	0.0	80.6	0.0	100	0.0	83.2
PS	0.0	93.8	99.8	64.9	0.0	88.9	98.3	64.3	0.0	88.2	94.1	90.0	0.0	84.8	91.5	77.6
TED	0.0	0.0	0.0	99.0	0.0	100	0.0	88.4	0.0	0.0	0.0	76.0	0.0	0.0	0.0	48.8
E. Macroeconomic Variables																
INFM	0.0	1.3	0.0	92.5	0.0	84.9	0.0	84.9	0.0	100	81.1	91.9	0.0	100	77.9	75.2
INFA	0.0	0.0	0.0	86.4	0.0	0.0	100	87.7	0.0	0.0	100	53.5	0.0	0.0	100	78.6
IPM	0.0	0.0	0.0	61.0	0.0	0.0	0.0	67.0	0.0	82.6	85.5	96.1	0.0	62.4	69.4	0.0
IPGA	0.0	0.0	0.0	58.0	0.0	0.0	0.0	50.9	0.0	0.0	0.0	91.2	0.0	0.0	0.0	57.6
H-S	0.0	100	100	25.3	0.0	95.9	100	85.5	0.0	87.5	99.1	99.5	0.0	81.6	99.2	88.7
M1M	0.0	94.3	92.9	100	0.0	12.1	3.3	78.1	0.0	100	100	72.6	0.0	100	0.0	91.0
M1A	0.0	1.0	0.0	35.9	5.5	12.2	4.1	66.6	0.0	0.0	0.0	47.2	0.0	0.0	0.0	78.5
ORDM	0.0	87.9	79.7	76.3	0.0	58.5	100	55.9	0.0	80.9	68.0	90.8	0.0	64.0	63.8	56.9
ORDA	0.0	100	0.0	0.0	0.0	70.4	70.9	77.2	0.0	100	0.0	100	0.0	66.8	0.0	0.0
CRB	0.0	0.0	0.0	24.3	0.0	93.3	100	70.4	0.0	100	97.1	78.2	0.0	94.6	93.5	70.4
CAP	0.0	81.9	0.0	38.5	0.0	0.0	85.9	100	0.0	83.4	80.0	82.4	0.0	38.7	76.0	78.9
EMPL	0.0	99.5	0.0	99.8	0.0	100	100	33.1	0.0	74.9	100	69.8	0.0	70.1	100	86.1
SENT	0.0	84.2	0.0	20.1	0.0	10.1	0.0	63.2	0.0	71.0	100	61.4	0.0	88.8	100	83.0
CONF	0.0	100	65.7	28.5	0.0	21.1	0.0	75.4	0.0	94.1	100	100	0.0	92.9	100	94.3
DIFF	0.0	0.0	0.0	5.8	0.0	0.0	0.0	41.6	0.0	0.0	0.0	74.0	0.0	0.0	0.0	86.8
PMBB	0.0	0.0	0.0	0.0	0.0	0.0	0.0	59.4	0.0	0.0	0.0	73.1	0.0	0.0	0.0	96.9
PMI	0.0	0.0	0.0	1.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	88.5	0.0	0.0	0.0	0.0

Notes: Notes: (i) The 38 variables are represented by their abbreviation, which is introduced in Table second weighting type introduced in this paper. This weighting type determines the weights based on the lowest SSE that is possible by combining the individual models. (iii) The average weights are based on the average weight they have when they are part of the active set of variables. The active set of variables is either the full, unrestricted set (UR) or determined by one of the three variable selection techniques: lasso (LS), elastic net (EN) and LARS (LR).

Table 11: Average weights assigned to the factors of the PCA-based forecast combinations

	Equities		Commodities		FX-aggr.		Bonds	
	Type 1	Type 2	Type 1	Type 2	Type 1	Type 2	Type 1	Type 2
Group A	20.0	25.2	20.0	25.8	20.0	27.3	20.0	27.7
Group B	20.0	33.2	20.0	27.7	20.0	36.7	20.0	34.6
Group C	20.0	28.7	20.0	27.7	20.0	31.7	20.0	31.3
Group D	20.0	26.7	20.0	25.1	20.0	31.7	20.0	32.1
Group E	20.0	36.2	20.0	30.7	20.0	33.6	20.0	33.9

Notes: Notes: (i) The table presents the average weight assigned to the different subgroups of variables, based on weighting type 1 or 2. (ii) The first weighting type determines the weights based on the proportion of SSE of each individual model, compared to the total SSE. The second weighing type determines the weights based on the lowest SSE that is possible by combining the individual models. (iii) The groups represent the equity market variables (A), bond market variables (B), FX variables (C), Liquidity and credit risk variables (D), and macroeconomic variables (E).

6 Discussion and further extensions

The primary goal of the paper was to assess the value of macroeconomic variables in predicting monthly realized volatility, using a large variety of different estimation techniques. The analysis revealed that monthly realized volatility is predictable, though strong variation is observed on various levels. Most useful techniques appear to be the forecast combinations, as also found by Christiansen et al. (2012), but they fail to behave consistently over the different asset classes. As the extent to which realized volatility is predictable changes over time, this makes it increasingly difficult for the models to perform consistently. To provide further insight in this topic, various suggestions for possible extensions are presented below.

Already a broad set of different estimation and variable selection techniques are used to assess the use of macroeconomic variables in explaining and predicting realized volatility. The goal of this was to analyse the issue from several perspectives and to provide an implicit robustness check. As little consistency both within and between methods is observed, the question arises what alternative methods are available. Many alternatives are available, an almost countless amount, which all treat a large number of predictors differently. Kim and Swanson (2014a), amongst others, provide several alternatives to the methods applied in this paper. Potential extensions or improvements of the analysis therefore lie primarily in the methodology.

More specifically, most fruitful extensions lie primarily in the area of variable selection. This forms the central component of the paper and works through in all different methods. Further analysing the possibilities of variable selection, will reveal more insight in the predictability of realized volatility and the added value of macroeconomic information. Moreover, the bootstrap aggregation procedure was applied in its most elementary form. A fairly straightforward was applied to select the variables used in the forecasting model, which could be simply replaced by more sophisticated techniques.

Though methodology is an important part, most fundamental is the data they are applied to. The paper covers the four major asset markets, but they could all be approximated by different assets. Also, the analysis could be narrowed down to more specific, individual assets, rather than analysing it from an index point of view. Applying the methodology to, for example, different individual stocks, provides an interesting perspective that could add value to this paper. As individual effects possibly cancel out in an index, the analysis here might not apply to individual assets. Further, to be able to obtain reliable estimates, this paper uses a moving window of 7.5 years. From an economic point of view, this period is quite long. Switching to weekly or daily realized volatility will make it possible to analyse short periods of time, using therefore more recent observations, which could increase the forecasting performance.

A last point that could be further investigated is the actual application of the realized volatility forecasts. The perspective of this paper is quite broad, as the focus lies on predicting realized volatility on a monthly basis for four large indices of different assets. The forecasts are therefore perhaps most useful for economic decision making. The predictions give an idea of the expected risk in financial markets and the accommodated uncertainty, which could help decision makers such as central banks to set, for example, the risk-free rate. Taking on, for

example, a more individual perspective by using individual assets might also change the use of realized volatility forecasts.

7 Conclusion

This paper presents an analysis of the use of 38 different macro-finance (macroeconomic and finance) variables to predict monthly realized volatility in four different asset classes: Equities, commodities, foreign exchange rates and bonds. A variety of different estimation- and variable selection techniques are used which can be roughly subdivided in four groups: penalized regressions, dynamic factor models, forecast combinations and bootstrap aggregation. All approach the set of possible predictor variables differently, thereby providing different perspectives on the predictability. The added value of the the macro-finance variables is analysed with respect to an optimized autoregressive model, which is the best elementary model possible without using any exogenous predictors. The goal of the paper is thus to analyse whether macro-finance variables add value to this autoregressive model model.

The forecast combination models appear to be most successful in predicting realized volatility. However, all models lack consistency between different asset classes. This highlights the difficulty of finding one forecast model or estimation technique to forecast realized volatility of different asset classes. Overall, the realized volatility of equities seems to be best predictable. The analysis also reveals, quite surprisingly, a weak performance of dynamic factor models. This leads to conclude that the common variation (or information) in macroeconomic variables is not able to capture the variation in realized volatility.

The forecasts of all models are quite well able to predict the monthly 95% Value-at-Risk. This signals that the realized volatility forecasts might be useful for economic decision making, as they provide a good estimate of next month 'worst case scenario'. The realized volatility forecasts also seem usable in portfolio context, as almost all models are able to outperform a simple buy-and-hold strategy and the risk-free rate. In general, the economic evaluation is not able to strongly differentiate between different models. This makes it difficult to assess which estimation technique performs best, though the penalized regression and forecast combination based forecasts belong consistently amongst the best performing models.

Lastly, and perhaps most importantly, the paper provides an analysis of which variables get selected most by the variable selection techniques. For equities, not surprisingly, equity market related variables appear most important. In addition, monthly inflation is selected often to forecast with. The most selected variables for the commodities class show strong overlap with the equities class, as it also selects predominantly equity market and macroeconomic variables. The latter two asset classes, foreign exchange rates and bonds also share many observations, primarily economic variables. There appears to be some consistency, as inflation is for all classes an important variable.

The overall conclusion of the analysis is that monthly realized volatility is predictable. More importantly, the use of macro-finance variables further improve the forecasts, as including them leads to outperforming a simple autoregressive model. Which variables are most helpful differs for each asset class as only limited consistency is found in the variable selection procedure and forecasting performance. To shed further light on this issue several extensions are possible, primarily in the choice of estimation techniques and data to which these methods are applied

to.

A Additional Tables

A.1 Correlation matrix

Table 12: Correlation matrix part 1/3

	E-P	MKT	SMB	HML	STR	TUR	MSC	T-B	RTB	LTR	RBR	T-S	C-P
D-P	0.53	-0.02	-0.07	0.00	-0.02	-0.02	0.01	0.53	-0.09	0.07	-0.13	0.30	0.49
E-P	–	-0.03	-0.10	0.01	0.02	0.00	0.03	0.69	0.28	0.06	0.06	-0.14	0.20
MKT		–	0.22	-0.31	0.28	-0.02	0.91	0.01	0.02	0.04	-0.15	-0.04	-0.02
SMB			–	-0.34	0.09	-0.08	0.14	-0.13	-0.09	-0.19	0.00	0.14	0.05
HML				–	-0.04	-0.06	-0.20	0.03	0.05	0.04	0.04	0.04	0.13
STR					–	0.04	0.23	0.00	-0.04	0.00	-0.03	0.06	0.06
TUR						–	-0.04	0.03	0.01	0.04	-0.04	-0.06	-0.03
MSC							–	0.08	0.06	0.00	-0.15	-0.06	0.00
T-B								–	0.21	0.06	0.04	-0.39	0.26
RTB									–	-0.02	0.46	-0.32	-0.32
LTR										–	-0.38	-0.12	-0.09
RBR											–	0.12	0.03
T-S												–	0.64

Notes: The table shows the correlation between the 38 possible predictor variables, as presented in Table 2.

Table 13: Correlation matrix part 2/3

	DOL	C-T	AFD	DEF	BAS	PS	TED	INM	INA	IPM	IPG	H-S	M1M
D-P	0.00	-0.13	0.09	0.43	0.19	0.07	0.41	0.09	0.35	0.03	-0.02	0.03	0.26
E-P	-0.08	-0.12	-0.28	-0.19	0.14	0.14	0.27	0.18	0.65	0.23	0.61	0.28	-0.05
MKT	0.16	0.27	-0.09	-0.08	0.04	0.27	-0.19	-0.01	-0.12	-0.06	-0.01	-0.00	-0.06
SMB	-0.02	0.11	0.02	0.03	-0.05	0.06	-0.08	-0.02	-0.04	-0.05	-0.12	0.02	0.02
HML	0.00	-0.08	0.01	-0.08	-0.06	-0.01	-0.08	0.08	0.05	0.09	0.04	0.13	0.06
STR	0.04	0.16	-0.07	-0.07	0.03	-0.01	-0.11	0.08	-0.01	0.00	0.06	0.08	-0.04
TUR	-0.04	-0.16	-0.03	-0.03	0.01	-0.20	0.01	0.15	0.03	-0.08	0.01	0.13	-0.01
MSC	0.12	0.27	-0.12	-0.07	0.07	0.31	-0.17	-0.02	-0.11	-0.03	0.04	0.05	-0.06
T-B	-0.10	-0.10	-0.32	-0.08	0.38	0.11	0.39	0.22	0.58	0.16	0.40	0.20	-0.10
RTB	-0.04	-0.04	-0.32	-0.37	-0.17	0.06	-0.00	0.13	0.10	0.25	0.52	0.13	-0.34
LTR	0.07	-0.17	0.11	0.07	0.07	-0.06	0.02	-0.24	0.02	-0.12	0.02	-0.12	0.04
RBR	-0.07	0.03	-0.30	-0.25	-0.31	0.07	0.12	0.18	0.18	0.26	0.33	0.07	-0.23
T-S	0.05	0.05	0.48	0.28	-0.01	0.00	-0.08	-0.06	-0.18	0.06	-0.14	0.18	0.33
C-P	0.04	0.00	0.25	0.03	0.32	0.11	-0.04	0.14	0.14	0.14	0.07	0.36	0.20
DOL	–	0.07	0.12	-0.02	0.00	-0.05	-0.15	0.08	-0.12	-0.02	-0.06	-0.05	0.03
C-T		–	-0.02	0.02	-0.08	0.17	-0.22	0.06	-0.10	0.00	-0.07	0.07	-0.10
AFD			–	0.32	0.23	-0.15	-0.19	-0.13	-0.27	0.00	-0.22	0.10	0.29
DEF				–	0.07	-0.18	0.45	-0.24	-0.14	-0.31	-0.56	-0.31	0.37
BAS					–	0.00	0.15	-0.05	-0.08	0.12	0.12	0.30	0.14
PS						–	-0.19	0.02	0.03	0.08	0.04	0.06	-0.14
TED							–	-0.05	0.34	-0.16	0.02	-0.26	0.13
INM								–	0.30	0.10	0.15	0.20	-0.17
INA									–	-0.05	0.29	0.04	-0.12
IPM										–	0.44	0.35	-0.22
IPG											–	0.40	-0.26
H-S												–	-0.04

Notes: The table shows the correlation between the 38 possible predictor variables, as presented in Table 2

Table 14: Correlation matrix part 3/3

	M1A	ORM	ORA	CRB	CAP	EMP	SEN	CON	DIFF	PMB	PMI
D-P	0.50	0.05	0.13	-0.04	0.09	0.16	0.02	0.00	0.07	0.08	0.01
E-P	-0.15	0.09	0.61	-0.03	0.21	0.58	-0.05	-0.02	0.32	0.47	0.36
MKT	0.02	0.05	0.03	0.21	-0.06	0.03	0.22	0.25	0.03	0.02	0.02
SMB	0.05	0.00	-0.09	0.02	-0.01	-0.09	0.22	0.15	-0.04	-0.05	-0.03
HML	0.04	0.09	0.08	0.14	0.11	0.10	0.06	0.14	0.16	0.10	0.11
STR	-0.02	0.07	0.06	0.12	0.01	0.07	0.12	-0.02	0.03	0.06	0.04
TUR	-0.02	0.07	0.03	-0.06	-0.11	0.02	-0.05	-0.14	0.01	-0.02	-0.02
MSC	0.05	0.07	0.08	0.19	-0.04	0.11	0.22	0.24	0.11	0.07	0.07
T-B	-0.14	0.09	0.42	-0.08	0.05	0.51	-0.01	0.01	0.17	0.16	0.09
RTB	-0.25	0.04	0.39	0.13	0.22	0.44	0.02	0.10	0.43	0.61	0.57
LTR	-0.02	-0.05	-0.04	-0.17	-0.13	-0.04	0.02	-0.08	-0.06	-0.07	-0.11
RBR	-0.14	0.03	0.31	0.13	0.25	0.24	-0.04	0.06	0.28	0.44	0.49
T-S	0.65	0.06	0.04	0.03	0.18	-0.10	0.06	0.06	0.23	0.15	0.20
C-P	0.47	0.14	0.30	-0.00	0.14	0.20	0.08	0.09	0.28	0.13	0.15
DOL	0.11	0.09	-0.00	0.29	0.01	-0.01	-0.10	-0.04	0.01	0.01	0.02
C-T	-0.05	0.09	-0.05	0.18	0.04	-0.09	0.14	0.14	-0.04	-0.06	-0.02
AFD	0.42	0.04	-0.09	-0.08	0.06	-0.20	0.01	-0.05	0.01	-0.10	-0.08
DEF	0.50	-0.14	-0.48	-0.10	-0.18	-0.43	0.04	-0.04	-0.29	-0.36	-0.35
BAS	0.17	0.11	0.16	-0.19	-0.00	0.28	0.03	0.02	0.12	-0.02	-0.04
PS	0.02	0.09	0.11	0.08	0.07	0.10	0.07	0.19	0.10	0.09	0.08
TED	0.07	-0.13	-0.05	-0.21	-0.16	0.05	-0.05	-0.05	-0.06	0.04	-0.02
INM	-0.04	0.10	0.18	0.24	0.10	0.21	-0.20	-0.20	0.12	0.17	0.17
INA	-0.26	-0.05	0.29	-0.15	-0.06	0.24	-0.10	-0.11	-0.06	0.13	0.01
IPM	-0.01	0.38	0.48	0.12	0.96	0.51	-0.04	0.08	0.49	0.41	0.49
IPG	-0.35	0.12	0.86	0.03	0.33	0.73	-0.04	0.02	0.51	0.71	0.63
H-S	0.07	0.24	0.59	0.08	0.31	0.48	0.05	0.10	0.52	0.37	0.43
M1M	0.50	-0.04	-0.14	-0.05	-0.15	-0.18	0.10	0.02	-0.08	-0.16	-0.13
M1A	–	0.07	-0.09	0.07	0.13	-0.11	0.08	0.08	0.15	0.01	0.08
ORM		–	0.30	0.15	0.37	0.23	-0.03	0.04	0.24	0.11	0.18
ORA			–	0.09	0.41	0.73	-0.00	0.06	0.63	0.72	0.70
CRB				–	0.17	0.04	0.14	0.24	0.18	0.14	0.22
CAP					–	0.41	-0.04	0.08	0.49	0.41	0.52
EMP						–	0.04	0.14	0.62	0.61	0.60
SEN							–	0.60	0.15	0.06	0.06
CON								–	0.22	0.13	0.16
DIFF									–	0.74	0.80
PMB										–	0.89

Notes: The table shows the correlation between the 38 possible predictor variables, as presented in Table 2

A.2 Additional Results

A.2.1 Value-at-Risk

Table 15: Results of the 90% Value-at-Risk forecasts using a normal distribution

	Equities				Commodities				Bonds			
	π_1 (%)	*UCC*	*IND*	*CCV*	π_1 (%)	*UCC*	*IND*	*CCV*	π_1 (%)	*UCC*	*IND*	*CCV*
1. Benchmark models												
Historical mean	13.580	3.149	1.496	4.645	17.695	13.329	6.996	20.325	11.934	0.957	0.046	1.003
Random Walk	10.700	0.129	0.607	0.736	13.992	3.877	1.303	5.180	11.523	0.600	0.561	1.161
AR_{opt}	10.288	0.022	2.124	2.146	14.403	4.673	2.125	6.798	10.700	0.129	0.048	0.177
AR(1)	10.288	0.022	2.367	2.390	15.638	7.457	2.013	9.470	10.700	0.129	0.048	0.177
AR(2)	10.288	0.022	2.124	2.146	14.403	4.673	2.125	6.798	11.111	0.323	0.005	0.328
AR(3)	10.700	0.129	1.676	1.805	14.403	4.673	2.125	6.798	11.111	0.323	0.005	0.328
AR(4)	10.700	0.129	1.676	1.805	14.815	5.536	1.672	7.209	11.111	0.323	0.005	0.328
AR(5)	10.288	0.022	2.124	2.146	14.403	4.673	2.125	6.798	10.700	0.129	0.048	0.177
ARX_{opt}	11.934	0.957	0.104	1.062	11.934	0.957	6.155	7.113	9.877	0.004	1.060	1.064
2. Penalized regressions												
Ridge Regressions	10.700	0.129	1.676	1.805	14.815	5.536	1.672	7.209	9.877	0.004	0.041	0.045
Lasso	11.111	0.323	1.286	1.609	13.992	3.877	2.636	6.513	9.877	0.004	0.265	0.269
Elastic Net	11.111	0.323	0.302	0.625	15.226	6.465	0.439	6.904	10.288	0.022	0.116	0.139
LARS	10.288	0.022	2.367	2.390	13.992	3.877	1.303	5.180	10.700	0.129	0.229	0.358
3. Dynamic factor models												
Unrestricted	11.111	0.323	2.860	3.183	13.580	3.149	0.642	3.791	9.877	0.004	0.041	0.045
Lasso	9.053	0.249	0.443	0.692	14.403	4.673	2.125	6.798	9.877	0.004	0.041	0.045
Elastic Net	10.288	0.022	0.048	0.070	15.638	7.457	2.013	9.470	9.877	0.004	0.041	0.045
LARS	9.465	0.079	0.265	0.344	15.226	6.465	1.276	7.741	11.934	0.957	0.046	1.003
4. Forecast combinations												
a. Equal weights												
Unrestricted	10.700	0.129	1.676	1.805	13.992	3.877	2.636	6.513	10.700	0.129	0.048	0.177
Lasso	10.700	0.129	1.676	1.805	14.403	4.673	2.125	6.798	10.700	0.129	0.048	0.177
Elastic Net	10.700	0.129	1.676	1.805	13.992	3.877	2.636	6.513	10.700	0.129	0.048	0.177
LARS	10.700	0.129	1.676	1.805	14.403	4.673	2.125	6.798	10.700	0.129	0.048	0.177
b. Weight type 1												
Unrestricted	10.288	0.022	2.124	2.146	14.815	5.536	1.672	7.209	10.700	0.129	0.048	0.177
Lasso	10.288	0.022	2.124	2.146	14.815	5.536	1.672	7.209	10.700	0.129	0.048	0.177
Elastic Net	10.288	0.022	2.124	2.146	14.815	5.536	1.672	7.209	10.700	0.129	0.048	0.177
LARS	10.288	0.022	2.124	2.146	14.815	5.536	1.672	7.209	10.700	0.129	0.048	0.177
c. Weight type 2												
Unrestricted	10.288	0.022	0.735	0.757	14.403	4.673	2.125	6.798	11.523	0.600	0.005	0.605
Lasso	10.288	0.022	0.735	0.757	13.992	3.877	2.636	6.513	11.523	0.600	0.005	0.605
Elastic Net	10.288	0.022	0.735	0.757	14.403	4.673	2.125	6.798	11.523	0.600	0.005	0.605
LARS	10.288	0.022	2.124	2.146	15.226	6.465	1.276	7.741	11.111	0.323	0.005	0.328
d. PCA based												
Unrestricted	10.700	0.129	1.676	1.805	13.580	3.149	1.700	4.849	10.700	0.129	0.048	0.177
SSE weights	10.700	0.129	1.676	1.805	13.992	3.877	1.303	5.180	10.700	0.129	0.048	0.177
SSE minimalized	11.111	0.323	1.286	1.609	13.992	3.877	1.303	5.180	10.700	0.129	0.048	0.177
First factor	9.877	0.004	0.116	0.120	13.992	3.877	1.303	5.180	10.700	0.129	0.048	0.177
5. Bootstrap aggregation												
#Bootstrap samples = 1	10.288	0.022	2.124	2.146	14.403	4.673	2.125	6.798	9.053	0.249	0.006	0.255
#Bootstrap samples = 100	11.523	0.600	0.951	1.551	14.403	4.673	2.125	6.798	9.053	0.249	0.006	0.255
#Bootstrap samples = 150	11.523	0.600	0.951	1.551	14.403	4.673	2.125	6.798	9.053	0.249	0.006	0.255
#Bootstrap samples = 200	11.523	0.600	0.951	1.551	14.403	4.673	2.125	6.798	9.053	0.249	0.006	0.255

Notes: (i) Abbreviations UCC, IND and CCV stand for Unconditional Coverage, Independence and Conditional Coverage test, respectively. (ii) The displayed values correspond to the Chi-Squared distribution with one degree of freedom for the UCC and IND test, and 2 degrees of freedom for the CCV test. The 5% critical value of the Chi-squared distribution with one or two degrees of freedom is equal to 3.84 and 5.99, respectively. (iii) $\pi_1(\%)$ represents the percentage of violations of the VaR threshold.

Table 16: Results of the 99% Value-at-Risk forecasts using a normal distribution

	Equities				Commodities				Bonds			
	π_1 (%)	*UCC*	*IND*	*CCV*	π_1 (%)	*UCC*	*IND*	*CCV*	π_1 (%)	*UCC*	*IND*	*CCV*
1. Benchmark models												
Historical mean	3.292	8.055	1.342	9.397	3.704	10.609	0.971	11.580	2.881	5.760	0.415	6.175
Random Walk	0.823	0.082	0.033	0.115	2.881	5.760	0.415	6.175	2.881	5.760	0.415	6.175
AR_{opt}	0.412	1.093	0.008	1.101	2.469	3.760	0.304	4.063	2.469	3.760	0.304	4.063
AR(1)	0.412	1.093	0.008	1.101	2.469	3.760	0.304	4.063	2.469	3.760	0.304	4.063
AR(2)	0.412	1.093	0.008	1.101	2.469	3.760	0.304	4.063	3.292	8.055	0.545	8.600
AR(3)	0.412	1.093	0.008	1.101	2.469	3.760	0.304	4.063	2.469	3.760	0.304	4.063
AR(4)	0.412	1.093	0.008	1.101	2.469	3.760	0.304	4.063	2.469	3.760	0.304	4.063
AR(5)	0.412	1.093	0.008	1.101	2.469	3.760	0.304	4.063	2.469	3.760	0.304	4.063
ARX_{opt}	2.058	2.103	0.210	2.313	2.881	5.760	0.415	6.175	2.881	5.760	0.355	6.115
2. Penalized regressions												
Ridge Regressions	0.823	0.082	0.033	0.115	2.881	5.760	0.415	6.175	2.469	3.760	0.304	4.063
Lasso	1.235	0.126	0.075	0.201	3.292	8.055	0.545	8.600	2.469	3.760	0.304	4.063
Elastic Net	1.235	0.126	0.075	0.201	3.292	8.055	0.545	8.600	2.469	3.760	0.304	4.063
LARS	0.412	1.093	0.008	1.101	2.469	3.760	0.304	4.063	2.881	5.760	0.415	6.175
3. Dynamic factor models												
Unrestricted	1.235	0.126	0.075	0.201	4.115	13.395	0.859	14.253	2.469	3.760	0.304	4.063
Lasso	2.058	2.103	0.210	2.313	3.704	10.609	0.692	11.302	2.881	5.760	0.415	6.175
Elastic Net	0.823	0.082	0.033	0.115	3.292	8.055	0.545	8.600	2.881	5.760	0.415	6.175
LARS	1.235	0.126	0.075	0.201	3.292	8.055	1.342	9.397	2.469	3.760	0.304	4.063
4. Forecast combinations												
a. Equal weights												
Unrestricted	0.412	1.093	0.008	1.101	2.881	5.760	0.415	6.175	2.469	3.760	0.304	4.063
Lasso	0.412	1.093	0.008	1.101	2.881	5.760	0.415	6.175	2.469	3.760	0.304	4.063
Elastic Net	0.412	1.093	0.008	1.101	2.881	5.760	0.415	6.175	2.881	5.760	0.415	6.175
LARS	0.412	1.093	0.008	1.101	2.881	5.760	0.415	6.175	2.469	3.760	0.304	4.063
b. Weight type 1												
Unrestricted	0.412	1.093	0.008	1.101	2.469	3.760	0.304	4.063	2.469	3.760	0.304	4.063
Lasso	0.412	1.093	0.008	1.101	2.469	3.760	0.304	4.063	2.469	3.760	0.304	4.063
Elastic Net	0.412	1.093	0.008	1.101	2.469	3.760	0.304	4.063	2.881	5.760	0.415	6.175
LARS	0.412	1.093	0.008	1.101	2.881	5.760	0.415	6.175	2.469	3.760	0.304	4.063
c. Weight type 2												
Unrestricted	0.412	1.093	0.008	1.101	2.469	3.760	0.304	4.063	2.881	5.760	0.415	6.175
Lasso	0.412	1.093	0.008	1.101	3.292	8.055	0.545	8.600	2.469	3.760	0.304	4.063
Elastic Net	0.412	1.093	0.008	1.101	2.058	2.103	0.210	2.313	2.881	5.760	0.415	6.175
LARS	0.412	1.093	0.008	1.101	2.469	3.760	0.304	4.063	2.881	5.760	0.415	6.175
d. PCA based												
Unrestricted	0.412	1.093	0.008	1.101	3.292	8.055	0.545	8.600	2.469	3.760	0.304	4.063
SSE weights	0.412	1.093	0.008	1.101	3.292	8.055	0.545	8.600	2.469	3.760	0.304	4.063
SSE minimalized	0.412	1.093	0.008	1.101	2.881	5.760	0.415	6.175	2.469	3.760	0.304	4.063
First factor	1.646	0.857	0.134	0.991	3.704	10.609	0.971	11.580	2.469	3.760	0.304	4.063
5. Bootstrap aggregation												
#Bootstrap samples = 1	1.235	0.126	0.075	0.201	3.292	8.055	0.545	8.600	2.058	2.103	0.210	2.313
#Bootstrap samples = 100	1.646	0.857	0.134	0.991	3.704	10.609	0.692	11.302	2.469	3.760	0.304	4.063
#Bootstrap samples = 150	1.235	0.126	0.075	0.201	3.292	8.055	0.545	8.600	2.469	3.760	0.304	4.063
#Bootstrap samples = 200	1.235	0.126	0.075	0.201	3.292	8.055	0.545	8.600	2.469	3.760	0.304	4.063

Notes: (i) Abbreviations UCC, IND and CCV stand for Unconditional Coverage, Independence and Conditional Coverage test, respectively. (ii) The displayed values correspond to the Chi-Squared distribution with one degree of freedom for the UCC and IND test, and 2 degrees of freedom for the CCV test. The 5% critical value of the Chi-squared distribution with one or two degrees of freedom is equal to 3.84 and 5.99, respectively. (iii) $\pi_1(\%)$ represents the percentage of violations of the VaR threshold.

Table 17: Results of the 90% Value-at-Risk forecasts using a t distribution with 5 degrees of freedom

	Equities				Commodities				Bonds			
	π_1 (%)	*UCC*	*IND*	*CCV*	π_1 (%)	*UCC*	*IND*	*CCV*	π_1 (%)	*UCC*	*IND*	*CCV*
1. Benchmark models												
Historical mean	10.288	0.022	4.121	4.143	13.580	3.149	1.700	4.849	8.230	0.894	1.283	2.177
Random Walk	6.584	3.539	0.798	4.337	11.934	0.957	2.081	3.038	8.642	0.519	0.421	0.940
AR_{opt}	8.230	0.894	1.107	2.001	11.934	0.957	2.081	3.038	7.407	1.976	0.066	2.042
AR(1)	7.819	1.378	1.469	2.847	12.757	1.905	1.252	3.157	6.996	2.694	0.015	2.709
AR(2)	9.053	0.249	4.186	4.435	11.934	0.957	2.081	3.038	7.819	1.378	0.264	1.642
AR(3)	9.465	0.079	3.511	3.590	11.934	0.957	2.081	3.038	7.407	1.976	0.440	2.416
AR(4)	9.053	0.249	4.186	4.435	12.346	1.393	3.241	4.634	7.407	1.976	0.066	2.042
AR(5)	8.230	0.894	1.107	2.001	11.111	0.323	1.472	1.795	7.407	1.976	0.066	2.042
ARX_{opt}	9.053	0.249	1.999	2.248	9.877	0.004	2.907	2.911	7.819	1.378	0.151	1.529
2. Penalized regressions												
Ridge Regressions	6.996	2.694	2.377	5.071	11.111	0.323	3.148	3.471	7.819	1.378	0.264	1.642
Lasso	8.230	0.894	1.107	2.001	10.700	0.129	1.890	2.020	7.407	1.976	0.440	2.416
Elastic Net	7.407	1.976	0.345	2.321	11.934	0.957	2.081	3.038	8.230	0.894	0.134	1.028
LARS	7.819	1.378	3.685	5.063	13.169	2.491	2.153	4.644	8.230	0.894	0.270	1.164
3. Dynamic factor models												
Unrestricted	7.407	1.976	1.891	3.867	11.111	0.323	1.472	1.795	7.819	1.378	0.151	1.529
Lasso	6.996	2.694	0.440	3.134	12.346	1.393	1.638	3.031	6.996	2.694	0.015	2.709
Elastic Net	6.173	4.517	0.000	4.517	13.169	2.491	3.844	6.335	7.819	1.378	0.151	1.529
LARS	6.173	4.517	0.947	5.464	12.757	1.905	2.666	4.571	7.819	1.378	0.151	1.529
4. Forecast combinations												
a. Equal weights												
Unrestricted	8.642	0.519	0.802	1.321	11.523	0.600	1.110	1.710	7.407	1.976	0.066	2.042
Lasso	8.642	0.519	0.802	1.321	11.523	0.600	1.110	1.710	7.407	1.976	0.066	2.042
Elastic Net	8.642	0.519	0.802	1.321	11.523	0.600	1.110	1.710	7.407	1.976	0.066	2.042
LARS	8.642	0.519	0.802	1.321	11.523	0.600	1.110	1.710	7.819	1.378	0.151	1.529
b. Weight type 1												
Unrestricted	8.230	0.894	1.107	2.001	11.523	0.600	1.110	1.710	7.407	1.976	0.066	2.042
Lasso	8.230	0.894	1.107	2.001	11.934	0.957	2.081	3.038	6.996	2.694	0.015	2.709
Elastic Net	8.230	0.894	1.107	2.001	11.934	0.957	2.081	3.038	7.407	1.976	0.066	2.042
LARS	8.230	0.894	3.054	3.948	11.934	0.957	2.081	3.038	7.819	1.378	0.151	1.529
c. Weight type 2												
Unrestricted	7.819	1.378	1.469	2.847	11.523	0.600	1.110	1.710	7.819	1.378	0.151	1.529
Lasso	7.819	1.378	1.469	2.847	11.934	0.957	2.081	3.038	7.819	1.378	0.151	1.529
Elastic Net	7.819	1.378	1.469	2.847	12.346	1.393	3.241	4.634	7.819	1.378	0.151	1.529
LARS	7.407	1.976	0.345	2.321	11.111	0.323	0.391	0.714	8.642	0.519	0.049	0.568
d. PCA based												
Unrestricted	8.230	0.894	0.086	0.980	11.523	0.600	1.110	1.710	7.407	1.976	0.066	2.042
SSE weights	7.819	1.378	0.192	1.570	11.934	0.957	2.081	3.038	6.996	2.694	0.015	2.709
SSE minimalized	8.642	0.519	0.802	1.321	11.523	0.600	1.110	1.710	8.230	0.894	0.270	1.164
First factor	7.407	1.976	0.105	2.081	11.523	0.600	1.110	1.710	7.407	1.976	0.440	2.416
5. Bootstrap aggregation												
#Bootstrap samples = 1	6.996	2.694	2.377	5.071	12.346	1.393	1.638	3.031	7.819	1.378	0.151	1.529
#Bootstrap samples = 100	6.173	4.517	0.947	5.464	11.111	0.323	1.472	1.795	6.584	3.539	0.000	3.539
#Bootstrap samples = 150	6.584	3.539	0.667	4.206	11.111	0.323	3.148	3.471	6.584	3.539	0.000	3.539
#Bootstrap samples = 200	6.584	3.539	0.667	4.206	11.523	0.600	2.584	3.183	6.584	3.539	0.000	3.539

Notes: (i) Abbreviations UCC, IND and CCV stand for Unconditional Coverage, Independence and Conditional Coverage test, respectively. (ii) The displayed values correspond to the Chi-Squared distribution with one degree of freedom for the UCC and IND test, and 2 degrees of freedom for the CCV test. The 5% critical value of the Chi-squared distribution with one or two degrees of freedom is equal to 3.84 and 5.99, respectively. (iii) $\pi_1(\%)$ represents the percentage of violations of the VaR threshold.

Table 18: Results of the 95% Value-at-Risk forecasts using a t distribution with 5 degrees of freedom

	Equities				Commodities				Bonds			
	π_1 (%)	*UCC*	*IND*	*CCV*	π_1 (%)	*UCC*	*IND*	*CCV*	π_1 (%)	*UCC*	*IND*	*CCV*
1. Benchmark models												
Historical mean	4.938	0.002	0.262	0.264	8.230	4.506	5.765	10.271	3.292	1.688	0.476	2.164
Random Walk	2.469	3.996	0.304	4.299	5.350	0.061	0.133	0.194	4.115	0.425	0.859	1.284
AR_{opt}	2.058	5.640	0.210	5.850	5.350	0.061	0.133	0.194	3.292	1.688	0.545	2.233
AR(1)	2.469	3.996	0.304	4.299	4.938	0.002	0.262	0.264	3.292	1.688	0.545	2.233
AR(2)	2.058	5.640	0.210	5.850	4.938	0.002	0.262	0.264	4.115	0.425	0.771	1.196
AR(3)	2.058	5.640	0.210	5.850	4.938	0.002	0.262	0.264	4.115	0.425	0.859	1.284
AR(4)	1.646	7.696	0.134	7.830	5.350	0.061	0.133	0.194	3.704	0.941	0.692	1.633
AR(5)	1.646	7.696	0.134	7.830	5.350	0.061	0.133	0.194	3.704	0.941	0.692	1.633
ARX_{opt}	2.881	2.694	0.415	3.109	5.761	0.283	1.475	1.758	3.704	0.941	0.614	1.555
2. Penalized regressions												
Ridge Regressions	1.646	7.696	0.134	7.830	4.527	0.118	1.043	1.162	3.704	0.941	0.614	1.555
Lasso	2.058	5.640	0.210	5.850	4.527	0.118	0.440	0.559	3.704	0.941	0.692	1.633
Elastic Net	2.469	3.996	0.304	4.299	4.527	0.118	1.043	1.162	3.292	1.688	0.545	2.233
LARS	2.058	5.640	0.210	5.850	5.350	0.061	0.133	0.194	3.704	0.941	0.692	1.633
3. Dynamic factor models												
Unrestricted	2.058	5.640	0.210	5.850	4.938	0.002	0.262	0.264	3.704	0.941	0.692	1.633
Lasso	3.704	0.941	0.692	1.633	5.350	0.061	1.470	1.532	2.881	2.694	0.415	3.109
Elastic Net	2.469	3.996	0.304	4.299	6.584	1.173	2.934	4.107	3.292	1.688	0.545	2.233
LARS	3.292	1.688	1.342	3.030	5.761	0.283	1.475	1.758	3.704	0.941	0.692	1.633
4. Forecast combinations												
a. Equal weights												
Unrestricted	1.646	7.696	0.134	7.830	4.938	0.002	0.262	0.264	3.292	1.688	0.545	2.233
Lasso	2.058	5.640	0.210	5.850	4.938	0.002	0.262	0.264	2.881	2.694	0.415	3.109
Elastic Net	2.058	5.640	0.210	5.850	4.938	0.002	0.262	0.264	3.292	1.688	0.545	2.233
LARS	2.058	5.640	0.210	5.850	4.938	0.002	0.262	0.264	3.292	1.688	0.545	2.233
b. Weight type 1												
Unrestricted	1.646	7.696	0.134	7.830	5.350	0.061	0.133	0.194	3.292	1.688	0.545	2.233
Lasso	2.058	5.640	0.210	5.850	4.938	0.002	0.262	0.264	2.881	2.694	0.415	3.109
Elastic Net	2.058	5.640	0.210	5.850	4.938	0.002	0.262	0.264	3.292	1.688	0.545	2.233
LARS	2.058	5.640	0.210	5.850	4.938	0.002	0.262	0.264	3.292	1.688	0.545	2.233
c. Weight type 2												
Unrestricted	1.646	7.696	0.134	7.830	4.938	0.002	0.262	0.264	3.704	0.941	0.692	1.633
Lasso	1.646	7.696	0.134	7.830	5.350	0.061	0.133	0.194	2.881	2.694	0.415	3.109
Elastic Net	1.235	10.266	0.075	10.341	4.938	0.002	0.262	0.264	3.704	0.941	0.692	1.633
LARS	1.235	10.266	0.075	10.341	5.350	0.061	0.133	0.194	3.704	0.941	0.692	1.633
d. PCA based												
Unrestricted	1.646	7.696	0.134	7.830	5.350	0.061	0.133	0.194	3.292	1.688	0.545	2.233
SSE weights	1.646	7.696	0.134	7.830	5.350	0.061	0.133	0.194	3.292	1.688	0.545	2.233
SSE minimalized	1.646	7.696	0.134	7.830	5.350	0.061	0.133	0.194	3.292	1.688	0.545	2.233
First factor	2.881	2.694	0.415	3.109	6.173	0.657	1.106	1.763	3.704	0.941	0.692	1.633
5. Bootstrap aggregation												
#Bootstrap samples = 1	1.646	7.696	0.134	7.830	5.350	0.061	1.470	1.532	3.704	0.941	0.692	1.633
#Bootstrap samples = 100	2.058	5.640	0.210	5.850	4.938	0.002	1.247	1.249	2.881	2.694	0.415	3.109
#Bootstrap samples = 150	2.469	3.996	0.304	4.299	4.938	0.002	1.247	1.249	2.881	2.694	0.415	3.109
#Bootstrap samples = 200	2.469	3.996	0.304	4.299	4.938	0.002	1.247	1.249	3.292	1.688	0.545	2.233

Notes: (i) Abbreviations UCC, IND and CCV stand for Unconditional Coverage, Independence and Conditional Coverage test, respectively. (ii) The displayed values correspond to the Chi-Squared distribution with one degree of freedom for the UCC and IND test, and 2 degrees of freedom for the CCV test. The 5% critical value of the Chi-squared distribution with one or two degrees of freedom is equal to 3.84 and 5.99, respectively. (iii) $\pi_1(\%)$ represents the percentage of violations of the VaR threshold.

Table 19: Results of the 99% Value-at-Risk forecasts using a t distribution with 5 degrees of freedom

	Equities				Commodities				Bonds			
	π_1 (%)	*UCC*	*IND*	*CCV*	π_1 (%)	*UCC*	*IND*	*CCV*	π_1 (%)	*UCC*	*IND*	*CCV*
1. Benchmark models												
Historical mean	0.823	0.082	0.033	0.115	0.823	0.082	0.033	0.115	0.823	0.082	0.033	0.115
Random Walk	0.412	1.093	0.008	1.101	0.000	4.884	0.000	0.000	0.823	0.082	0.033	0.115
AR_{opt}	0.412	1.093	0.008	1.101	0.412	1.093	0.008	1.101	0.823	0.082	0.033	0.115
AR(1)	0.412	1.093	0.008	1.101	0.412	1.093	0.008	1.101	0.823	0.082	0.033	0.115
AR(2)	0.412	1.093	0.008	1.101	0.412	1.093	0.008	1.101	0.412	1.093	0.008	1.101
AR(3)	0.412	1.093	0.008	1.101	0.412	1.093	0.008	1.101	0.823	0.082	0.033	0.115
AR(4)	0.412	1.093	0.008	1.101	0.412	1.093	0.008	1.101	0.823	0.082	0.033	0.115
AR(5)	0.412	1.093	0.008	1.101	0.412	1.093	0.008	1.101	0.823	0.082	0.033	0.115
ARX_{opt}	0.412	1.093	0.008	1.101	1.646	0.857	0.134	0.991	1.235	0.126	0.075	0.201
2. Penalized regressions												
Ridge Regressions	0.412	1.093	0.008	1.101	0.412	1.093	0.008	1.101	0.823	0.082	0.033	0.115
Lasso	0.412	1.093	0.008	1.101	0.412	1.093	0.008	1.101	0.412	1.093	0.008	1.101
Elastic Net	0.412	1.093	0.008	1.101	0.412	1.093	0.008	1.101	0.412	1.093	0.008	1.101
LARS	0.412	1.093	0.008	1.101	0.000	4.884	0.000	0.000	0.823	0.082	0.033	0.115
3. Dynamic factor models												
Unrestricted	0.412	1.093	0.008	1.101	0.412	1.093	0.008	1.101	0.823	0.082	0.033	0.115
Lasso	0.823	0.082	0.033	0.115	0.823	0.082	0.033	0.115	0.823	0.082	0.033	0.115
Elastic Net	0.412	1.093	0.008	1.101	0.412	1.093	0.008	1.101	0.823	0.082	0.033	0.115
LARS	0.823	0.082	0.033	0.115	0.000	4.884	0.000	0.000	0.823	0.082	0.033	0.115
4. Forecast combinations												
a. Equal weights												
Unrestricted	0.412	1.093	0.008	1.101	0.412	1.093	0.008	1.101	0.823	0.082	0.033	0.115
Lasso	0.412	1.093	0.008	1.101	0.412	1.093	0.008	1.101	0.823	0.082	0.033	0.115
Elastic Net	0.412	1.093	0.008	1.101	0.412	1.093	0.008	1.101	0.823	0.082	0.033	0.115
LARS	0.412	1.093	0.008	1.101	0.412	1.093	0.008	1.101	0.823	0.082	0.033	0.115
b. Weight type 1												
Unrestricted	0.412	1.093	0.008	1.101	0.412	1.093	0.008	1.101	0.823	0.082	0.033	0.115
Lasso	0.412	1.093	0.008	1.101	0.412	1.093	0.008	1.101	0.823	0.082	0.033	0.115
Elastic Net	0.412	1.093	0.008	1.101	0.412	1.093	0.008	1.101	0.823	0.082	0.033	0.115
LARS	0.412	1.093	0.008	1.101	0.412	1.093	0.008	1.101	0.823	0.082	0.033	0.115
c. Weight type 2												
Unrestricted	0.412	1.093	0.008	1.101	0.412	1.093	0.008	1.101	1.235	0.126	0.075	0.201
Lasso	0.412	1.093	0.008	1.101	0.412	1.093	0.008	1.101	1.235	0.126	0.075	0.201
Elastic Net	0.412	1.093	0.008	1.101	0.412	1.093	0.008	1.101	0.823	0.082	0.033	0.115
LARS	0.412	1.093	0.008	1.101	0.000	4.884	0.000	0.000	0.823	0.082	0.033	0.115
d. PCA based												
Unrestricted	0.412	1.093	0.008	1.101	0.412	1.093	0.008	1.101	0.823	0.082	0.033	0.115
SSE weights	0.412	1.093	0.008	1.101	0.412	1.093	0.008	1.101	0.823	0.082	0.033	0.115
SSE minimalized	0.412	1.093	0.008	1.101	0.412	1.093	0.008	1.101	0.823	0.082	0.033	0.115
First factor	0.412	1.093	0.008	1.101	0.412	1.093	0.008	1.101	0.412	1.093	0.008	1.101
5. Bootstrap aggregation												
#Bootstrap samples = 1	0.412	1.093	0.008	1.101	0.412	1.093	0.008	1.101	0.412	1.093	0.008	1.101
#Bootstrap samples = 100	0.412	1.093	0.008	1.101	0.412	1.093	0.008	1.101	0.412	1.093	0.008	1.101
#Bootstrap samples = 150	0.412	1.093	0.008	1.101	0.412	1.093	0.008	1.101	0.412	1.093	0.008	1.101
#Bootstrap samples = 200	0.412	1.093	0.008	1.101	0.412	1.093	0.008	1.101	0.412	1.093	0.008	1.101

Notes: (i) Abbreviations UCC, IND and CCV stand for Unconditional Coverage, Independence and Conditional Coverage test, respectively. (ii) The displayed values correspond to the Chi-Squared distribution with one degree of freedom for the UCC and IND test, and 2 degrees of freedom for the CCV test. The 5% critical value of the Chi-squared distribution with one or two degrees of freedom is equal to 3.84 and 5.99, respectively. (iii) $\pi_1(\%)$ represents the percentage of violations of the VaR threshold.

A.2.2 Mean-variance portfolio returns

Table 20: Performance of the mean-variance portfolio with $\gamma = 5$

	Equities				Commodities				Bonds			
	$\bar{r}$ (bps)	SR	t-val.	$\bar{w}$ (%)	$\bar{r}$ (bps)	SR	t-val.	$\bar{w}$ (%)	$\bar{r}$ (bps)	SR	t-val.	$\bar{w}$ (%)
Buy-and-Hold Strategy	13.74	0.06	–	–	1.46	0.0046	–	–	-9.18	-0.09	–	–
1. Benchmark models												
Historical mean	2.60	0.12	1.18	3.67	0.17	0.01	0.08	3.22	0.07	0.02	2.43	1.02
Random Walk	2.45	0.15	1.70	3.62	0.34	0.02	0.30	3.26	0.23	0.06	2.90	1.11
AR_{opt}	2.34	0.14	1.56	3.47	0.22	0.01	0.19	3.13	0.16	0.05	2.68	1.04
AR(1)	2.36	0.14	1.49	3.47	0.19	0.01	0.15	3.19	0.11	0.03	2.47	1.02
AR(2)	2.27	0.13	1.47	3.48	0.23	0.01	0.20	3.17	0.12	0.04	2.47	1.04
AR(3)	2.31	0.14	1.49	3.50	0.24	0.01	0.21	3.16	0.14	0.04	2.60	1.04
AR(4)	2.33	0.14	1.54	3.47	0.20	0.01	0.16	3.16	0.15	0.04	2.67	1.04
AR(5)	2.34	0.14	1.57	3.47	0.22	0.01	0.18	3.14	0.16	0.05	2.72	1.04
ARX_{opt}	2.08	0.12	1.14	3.61	0.45	0.02	0.41	3.27	-0.07	-0.01	1.28	1.32
2. Penalized regressions												
Ridge Regressions	2.22	0.13	1.36	3.47	0.46	0.03	0.48	3.19	0.16	0.05	2.68	1.05
Lasso	2.25	0.13	1.39	3.47	0.35	0.02	0.35	3.14	0.16	0.05	2.71	1.03
Elastic Net	2.26	0.13	1.37	3.52	0.50	0.03	0.54	3.18	0.13	0.04	2.56	1.01
LARS	2.29	0.14	1.49	3.47	0.31	0.02	0.29	3.14	0.18	0.05	2.70	1.05
3. Dynamic factor models												
Unrestricted	2.32	0.14	1.49	3.46	0.22	0.01	0.18	3.16	0.17	0.05	2.72	1.02
Lasso	2.70	0.14	1.66	3.63	-0.02	0.00	-0.13	3.20	0.15	0.04	2.59	1.03
Elastic Net	2.95	0.15	1.91	3.56	-0.12	-0.01	-0.26	3.20	0.15	0.04	2.60	1.03
LARS	2.09	0.11	0.90	3.56	0.27	0.02	0.23	3.19	0.14	0.04	2.66	1.07
4. Forecast combinations												
a. Equal weights												
Unrestricted	2.33	0.14	1.57	3.47	0.30	0.02	0.28	3.13	0.16	0.05	2.67	1.04
Lasso	2.32	0.14	1.55	3.48	0.30	0.02	0.29	3.14	0.16	0.05	2.68	1.04
Elastic Net	2.31	0.14	1.53	3.47	0.30	0.02	0.28	3.14	0.16	0.05	2.68	1.05
LARS	2.31	0.14	1.53	3.47	0.34	0.02	0.33	3.14	0.16	0.05	2.68	1.04
b. Weight type 1												
Unrestricted	2.33	0.14	1.56	3.47	0.26	0.01	0.23	3.14	0.16	0.05	2.67	1.04
Lasso	2.32	0.14	1.54	3.48	0.26	0.01	0.23	3.14	0.15	0.05	2.67	1.04
Elastic Net	2.31	0.14	1.52	3.47	0.25	0.01	0.23	3.14	0.16	0.05	2.67	1.04
LARS	2.31	0.14	1.51	3.47	0.30	0.02	0.28	3.14	0.16	0.05	2.67	1.04
c. Weight type 2												
Unrestricted	2.33	0.14	1.65	3.38	0.24	0.01	0.21	3.16	0.14	0.04	2.62	1.03
Lasso	2.37	0.15	1.69	3.39	0.25	0.01	0.21	3.18	0.17	0.05	2.74	1.03
Elastic Net	2.31	0.14	1.63	3.39	0.27	0.02	0.25	3.20	0.15	0.04	2.64	1.04
LARS	2.31	0.14	1.61	3.41	0.36	0.02	0.36	3.19	0.17	0.05	2.71	1.04
d. PCA based												
Unrestricted	2.33	0.14	1.56	3.47	0.29	0.02	0.28	3.14	0.16	0.05	2.68	1.04
SSE weights	2.33	0.14	1.55	3.47	0.25	0.01	0.22	3.15	0.16	0.05	2.68	1.04
SSE minimalized	2.30	0.14	1.52	3.48	0.27	0.02	0.25	3.18	0.17	0.05	2.74	1.06
First factor	2.29	0.13	1.43	3.52	0.36	0.02	0.34	3.26	0.16	0.05	2.68	1.03
5. Bootstrap aggregation												
#Bootstrap samples = 1	2.41	0.14	1.56	3.54	0.36	0.02	0.36	3.18	0.19	0.06	2.85	1.04
#Bootstrap samples = 100	2.25	0.13	1.43	3.49	0.44	0.02	0.46	3.20	0.18	0.05	2.85	1.02
#Bootstrap samples = 150	2.27	0.13	1.46	3.50	0.47	0.03	0.49	3.20	0.18	0.05	2.83	1.02
#Bootstrap samples = 200	2.25	0.13	1.42	3.49	0.48	0.03	0.50	3.19	0.18	0.05	2.84	1.02

Notes: (i) Abbreviations $\bar{r}$, SR, t-val. and $\bar{w}$ stand for average monthly return in excess of the risk-free rate (expressed in percentages), monthly Sharpe Ratio, the t-statistic for the comparison of the Sharpe Ratios, and average weight (expressed in percentages) assigned to the asset, respectively. (ii) The t-value is significant at a 5% significant level if its absolute value is larger than 1.96.

Table 21: Performance of the mean-variance portfolio with $\gamma = 10$

	Equities				Commodities				Bonds			
	$\bar{r}$ (bps)	SR	t-val.	$\bar{w}$ (%)	$\bar{r}$ (bps)	SR	t-val.	$\bar{w}$ (%)	$\bar{r}$ (bps)	SR	t-val.	$\bar{w}$ (%)
Buy-and-Hold Strategy	13.74	0.06	–	–	1.46	0.0046	–	–	-9.18	-0.09	–	–
1. Benchmark models												
Historical mean	1.30	0.12	1.18	1.84	0.09	0.01	0.08	1.61	0.04	0.02	2.43	0.51
Random Walk	1.22	0.15	1.70	1.81	0.17	0.02	0.30	1.63	0.11	0.06	2.90	0.55
AR_{opt}	1.17	0.14	1.56	1.74	0.11	0.01	0.19	1.57	0.08	0.05	2.68	0.52
AR(1)	1.18	0.14	1.49	1.73	0.10	0.01	0.15	1.60	0.05	0.03	2.47	0.51
AR(2)	1.14	0.13	1.47	1.74	0.12	0.01	0.20	1.59	0.06	0.04	2.47	0.52
AR(3)	1.15	0.14	1.49	1.75	0.12	0.01	0.21	1.58	0.07	0.04	2.60	0.52
AR(4)	1.16	0.14	1.54	1.74	0.10	0.01	0.16	1.58	0.08	0.04	2.67	0.52
AR(5)	1.17	0.14	1.57	1.74	0.11	0.01	0.18	1.57	0.08	0.05	2.72	0.52
ARX_{opt}	1.04	0.12	1.14	1.81	0.22	0.02	0.41	1.64	-0.03	-0.01	1.28	0.66
2. Penalized regressions												
Ridge Regressions	1.11	0.13	1.36	1.73	0.23	0.03	0.48	1.59	0.08	0.05	2.68	0.53
Lasso	1.13	0.13	1.39	1.74	0.17	0.02	0.35	1.57	0.08	0.05	2.71	0.51
Elastic Net	1.13	0.13	1.37	1.76	0.25	0.03	0.54	1.59	0.07	0.04	2.56	0.51
LARS	1.15	0.14	1.49	1.74	0.15	0.02	0.29	1.57	0.09	0.05	2.70	0.52
3. Dynamic factor models												
Unrestricted	1.16	0.14	1.49	1.73	0.11	0.01	0.18	1.58	0.09	0.05	2.72	0.51
Lasso	1.35	0.14	1.66	1.82	-0.01	0.00	-0.13	1.60	0.07	0.04	2.59	0.51
Elastic Net	1.47	0.15	1.91	1.78	-0.06	-0.01	-0.26	1.60	0.08	0.04	2.60	0.52
LARS	1.05	0.11	0.90	1.78	0.13	0.02	0.23	1.60	0.07	0.04	2.66	0.54
4. Forecast combinations												
a. Equal weights												
Unrestricted	1.17	0.14	1.57	1.73	0.15	0.02	0.28	1.57	0.08	0.05	2.67	0.52
Lasso	1.16	0.14	1.55	1.74	0.15	0.02	0.29	1.57	0.08	0.05	2.68	0.52
Elastic Net	1.15	0.14	1.53	1.73	0.15	0.02	0.28	1.57	0.08	0.05	2.68	0.52
LARS	1.16	0.14	1.53	1.74	0.17	0.02	0.33	1.57	0.08	0.05	2.68	0.52
b. Weight type 1												
Unrestricted	1.17	0.14	1.56	1.73	0.13	0.01	0.23	1.57	0.08	0.05	2.67	0.52
Lasso	1.16	0.14	1.54	1.74	0.13	0.01	0.23	1.57	0.08	0.05	2.67	0.52
Elastic Net	1.15	0.14	1.52	1.73	0.13	0.01	0.23	1.57	0.08	0.05	2.67	0.52
LARS	1.15	0.14	1.51	1.74	0.15	0.02	0.28	1.57	0.08	0.05	2.67	0.52
c. Weight type 2												
Unrestricted	1.17	0.14	1.65	1.69	0.12	0.01	0.21	1.58	0.07	0.04	2.62	0.51
Lasso	1.19	0.15	1.69	1.70	0.12	0.01	0.21	1.59	0.08	0.05	2.74	0.52
Elastic Net	1.16	0.14	1.63	1.70	0.14	0.02	0.25	1.60	0.07	0.04	2.64	0.52
LARS	1.16	0.14	1.61	1.70	0.18	0.02	0.36	1.59	0.08	0.05	2.71	0.52
d. PCA based												
Unrestricted	1.16	0.14	1.56	1.74	0.15	0.02	0.28	1.57	0.08	0.05	2.68	0.52
SSE weights	1.17	0.14	1.55	1.74	0.13	0.01	0.22	1.57	0.08	0.05	2.68	0.52
SSE minimalized	1.15	0.14	1.52	1.74	0.14	0.02	0.25	1.59	0.08	0.05	2.74	0.53
First factor	1.14	0.13	1.43	1.76	0.18	0.02	0.34	1.63	0.08	0.05	2.68	0.52
5. Bootstrap aggregation												
#Bootstrap samples = 1	1.21	0.14	1.56	1.77	0.18	0.02	0.36	1.59	0.10	0.06	2.85	0.52
#Bootstrap samples = 100	1.12	0.13	1.43	1.75	0.22	0.02	0.46	1.60	0.09	0.05	2.85	0.51
#Bootstrap samples = 150	1.14	0.13	1.46	1.75	0.23	0.03	0.49	1.60	0.09	0.05	2.83	0.51
#Bootstrap samples = 200	1.12	0.13	1.42	1.75	0.24	0.03	0.50	1.60	0.09	0.05	2.84	0.51

Notes: (i) Abbreviations $\bar{r}$, SR, t-val. and $\bar{w}$ stand for average monthly return in excess of the risk-free rate (expressed in percentages), monthly Sharpe Ratio, the t-statistic for the comparison of the Sharpe Ratios, and average weight (expressed in percentages) assigned to the asset, respectively. (ii) The t-value is significant at a 5% significant level if its absolute value is larger than 1.96.

References

Alexander, C., 2008, *Market Risk Analysis, Quantitative Methods in Finance*, Market Risk Analysis (Wiley).

Andersen, T., T. Bollerslev, P. Christoffersen, and F. Diebold, 2006, Volatility and correlation forecasting, *Handbook of Economic Forecasting* 778–878.

Andersen, T., T. Bollerslev, F. Diebold, and P. Labys, 2003, Modeling and forecasting realized volatility, *Econometrica* 71, 579–625.

Ang, A., and G. Bekaert, 2007, Stock return predictability: Is it there?, *Review of Financial Studies* 20, 651–707.

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

Beltratti, A., and C. Morana, 2006, Breaks and persistency: macroeconomic causes of stock market volatility, *Journal of Econometrics* 131, 151–177.

Black, F., 1976, Studies in stock price volatility changes, *Proceedings of the American Statistical Association, Business and Economic Statistics Section* 177–181.

Bollerslev, T., 1986, Generalized autoregressive conditional heteroskedasticity, *Journal of Financial Econometrics* 31, 307–327.

Brandt, M., 2009, Portfolio choice problems, *Handbook of Financial Econometrics* 1, 269–336.

Breen, W., L. Glosten, and R. Jagannathan, 1989, Economic sginificance of predictable variations in stock returns, *Journal of Finance* 44, 1177–1189.

Breiman, L., 1996, Bagging predictors, *Machine Learning* 24, 123–140.

Campbell, J., and R. Shiller, 1988, The dividend-price ratio and expectations of future dividends and discount factors, *The Review of Financial Studies* 1, 195–228.

Campbell, J., and S. Thompson, 2008, Predicting the equity premium out of sample: Can anything beat the historical average?, *Review of Financial Studies* 21, 1509–1531.

Çakmakli, C., and D. van Dijk, 2010, Getting the most ouf of macroeconomic information for predicting stock returns and volatility, *Working paper* .

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.

Christoffersen, P., 2011, *Elements of Financial Risk Management* (Elsevier Science).

Clark, T., and K. West, 2007, Approximately normal tests for equal predictive accuracy in nested models, *Journal of Economics* 138, 291–311.

Cochrane, J., and M. Piazzesi, 2005, Bond risk premia, *American Economic Review* 95, 138–160.

Diebold, F.X., and R.S. Mariano, 1995, Comparing predictive accuracy, *Journal of Business and Economic Statistics* 13, 253–263.

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

Engle, R., 1982, Autorregressive conditional heteroskedasticity with estimates of united kingdom ination, *Econometrica* 50, 987–1008.

Fama, E., and K. French, 1992, The cross-section of expected stock returns, *Journal of Finance* 47, 427–465.

Fama, E., and K. French, 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3–56.

Fleming, J., C. Kirby, and B. Ostdiek, 2003, The economic value of volatility timing using realized volatility, *Journal of Financial Economics* 67, 473–509.

Hastie, T., R. Tibshirani, and J. Friedman, 2001, *The Elements of Statistical Learning; Data Mining, Inference and Prediction* (New York: Springer).

Hillebrand, E., and M. Medeiros, 2010, The benefits of bagging for forecast models of realised volatility, *Econometric Reviews* 29.

Hoerl, A., and R. Kennard, 1970, Ridge regression: biased estimation for nonorthogonal problems, *Technometrics* 12, 382–417.

Jensen, J., 1906, Sur les fonctions convexes et les ingalits entre les valeurs moyennes, *Acta Math.* 30, 175–193.

Jobson, J., and B. Korkie, 1981, Performance hypothesis testing with the sharpe and treynor measures, *Journal of Finance* 36, 889–908.

Jones, C., O. Lamont, and R. Lumsdaine, 1998, Macroeconomic news and bond market volatility, *Journal of Financial Econometrics* 47, 315–337.

Kim, H., and N. Swanson, 2014a, Forecasting financial and macroeconomic variables using data reduction methods: New empirical evidence, *Journal of Econometrics* 178, 352–367.

Kim, H., and N. Swanson, 2014b, Mining big data using parsimonious factor and shrinkage methods, *Working paper* .

Lustig, H., N. Roussanov, and A. Verdelhan, 2011, Common risk factors in currency markets, *Review of Financial Studies* 24, 3731–3777.

McAleer, M., and M. Medeiros, 2008, Realized volatility: A review, *Econometric Reviews* 27, 10–45.

Mele, A., 2007, Asymmetric stock market volatility and the cyclical behaviour of expected returns, *Journal of Financial Economics* 86, 681–718.

Merton, R., 1974, On the pricing of corporate debt: The risk structure of interest rates, *Journal of Finance* 29, 449–470.

Nelson, D., 1991, Conditional heteroskedasticity in asset returns: A new approach, *Econometrica* 59, 347–370.

Opdyke, J., 2007, Comparing sharpe ratios: So where are the p-values?, *Asset Management* 8, 308–336.

Pastor, L., and R. Stambaugh, 2003, Liquidity risk and expected stock returns, *Journal of Political Economy* 111, 642–685.

Pesaran, M., and A. Timmerman, 1992, A simple nonparametric test of predictive performance, *Journal of Business and Economic Statistics* 10, 461–465.

Pierdzioch, C., J. Döpke, and D. Hartmann, 2008, Forecasting stock market volatility with macroeconomic variables in real time, *Journal of Economics and Business* 60, 256–276.

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

Scheinkman, J., and W. Xiong, 2003, Overconfidence and speculative bubbles, *Journal of Political Economy* 111, 1183–1219.

Schewert, G., 1989, Why does stock market volatility change over time?, *Journal of Finance* 44, 1115–1153.

Stock, J., and M. Watson, 2004, Combination forecasts of output growth in a seven-country data set, *Journal of Forecasting* 23, 405–430.

Stock, J., and M. Watson, 2006, *Handbook of Economic Forecasting* (Elsevier Science).

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

Zakoian, J.M., 1991, Threshold heteroskedastic models, *Working paper* .

Zhang, Y., Li R., and C. Tsai, 2010, Regularization parameter selections via generalized information criterion, *Journal of the American Statistical Association* 105, 312–323.

Zou, H., and T. Hastie, 2005, Regularization and variable selection via the elastic net, *Journal of Royal Satistical Society* Series B 67, 301–320.