ERASMUS UNIVERSITY ROTTERDAM

ERASMUS SCHOOL OF ECONOMICS

BACHELOR THESIS ECONOMETRICS AND OPERATIONAL RESEARCH

July 28, 2020

Forecasting tail risks

Author:

Adrienne DEN OUDEN

Student ID number: 481145

Supervisor:

B. VAN OS MSC

Second Assessor:

Dr. Y. SUN

Abstract

This paper tries to find a reliable early warning system for tail risk. To do this we forecast Value at Risks of real and financial variables. The forecasts are made through autoregressive models which are estimated through linear regression, lineair quantile regression and quantile regression forests. Furthermore we investigate the impact of extending the models through the addition of factors. These factors are extracted from a large data set through principal components analysis and partial least squares regression. Last, we combine the indivual model forecasts through equally weighted pools to find out its possibly added value.

Contents

1 Introduction

This research is based on the paper 'Forecasting Tail Risks' by De Nicolò & Lucchetta (2017). The informal definition of tail risk is the possibility of large loss due to an extremely rare event. An example of such an event is the financial crisis in 2008. As the financial crisis has had a tremendous impact, it would be useful to create an Early Warning System (EWS) for similar future events. For this reason De Nicolò & Lucchetta (2017) try to come up with such a model. The difficulty lies in creating an EWS that is reliable. The feature that makes an early warning system reliable is its out-of-sample forecasting performance. For this research we implement the same methods and models as De Nicolò & Lucchetta (2017) and extend them to try to find a model that generates reliable early warnings.

A standard risk measure that has been widely used is Value at Risk (VaR). De Nicolò & Lucchetta (2017) propose autoregressive models and factor-augmented autoregressive models to forecast the Value at Risks of indicators of real activity and financial stress. Just like De Nicolò & Lucchetta (2017) we use the methods linear regression and quantile linear regression to generate VaR forecasts. As an extension, this research adds quantile regression forests to those methods.

We consider several autoregressive models that differ in the sense that the forecasts are done iterated or direct, whether a rolling or an expanding window is applied and whether the models are extended through the addition of factors. De Nicolò & Lucchetta (2017) extract these factors from a large dataset using Principal Components Analysis (PCA). As an extension to this, this research also uses Partial Least Squares (PLS) regression to extract factors from the data and find out if there is a significant difference between the methods.

Besides the individual model forecasts, we also analyse the impact of Equally Weighted Pools (EWPs) which generate Value at Risks forecasts that are averages of the Value at Risks forecasts generated by individual models.

Through comparing all models we aim on becoming acquainted with: which estimation method is preferable (linear regression, quantile linear regression or quantile regression forests), whether the addition of factors to the autoregressive models gives a significant improvement, and whether EWP forecasts are superior to those generated by individual models. Through this acquired knowledge we try to answer the research question:

"Can we find a model or several models that can deliver reliable early warning signals of tail real and financial risk?"

Using a data set of monthly US data ranging from January 1973 to December 2014 we find a few main results. First, factor-augmented models can give forecasts that are more accurate than the models without factors, however, this is variable dependent. Furthermore, adding factors seems to improve the quantile models to a larger extent than the models estimated through linear regression. Using PCA or PLS does in general not give significantly different forecasts. However, in combination with quantile models, PLS seems to be preferred over PCA. Second, combining the models through an EWP equals, or even improves the predictive ability of the individual models in the pool. Third, the linear regression method is preferable for most cases compared to the two quantile estimation methods. However, for one particular financial variable we find that quantile forest regression strongly outperforms the other methods.

The remainder of the article is organized as follows. Section 2 discusses previously written literature on this topic. Section 3 describes the data that has been used. Section 4 discusses the models, methods and forecasting evaluation tools. section 6 concludes, and Section 7 gives some final comments and discusses possible future research. The Appendix provides supporting information in the form of extra Tables and Figures.

2 Literature review

This research is mainly based on the article of De Nicolò & Lucchetta (2017) 'Forecasting tail risks'. They propose several methods for forecasting tail risks, this section explains the reasoning behind choosing these particular methods and extensions to these methods based on previously published literature.

However, theoretical aspects of tail risks have been discussed by Acemoglu et al. (2017), they do not forecast them. De Nicolò & Lucchetta (2017) propose linear regression and quantile linear regression for the estimation of autoregressive models and factor-augmented autoregressive models for forecasting tail risks. Autoregressive models are among the most commonly used in the field of macroeconomic time series introduced by Akaike (1969). The key idea of an autoregressive model is to regress the response variable on its own previous observations. However, adding factors extracted from a large data set might improve the autoregressive models as additional economic information could be relevant to modeling the dynamics of the tail risk indicators. Stock & Watson (2006) find that factor models with many predictors have a superior predictive ability. De Nicolò & Lucchetta (2017) make those factors through the well-known PCA method. Although the idea of PCA had already been acknowleged in traditional statistics, it has become more popular thanks to Stock & Watson (2002). Another method, however, for factor extraction is PLS regression. Groen & Kapetanios (2016) find that when the factor structure in the data gets weak PLS outperfroms PCA. In empirical research on a large set of monthly U.S. macroeconomic data set they find that PLS usually gives a better out-of-sample performance.

De Nicolò & Lucchetta (2017) measure the tail risks by means of Value-at-Risk (VaR). This is by far the most popular measure of downside risk. Taken from van Os & van Dijk (2020) VaR is the maximum value that will *not* be exceeded over a given time period with a probability $(1-\alpha)$. The other way around it is also the minimum value that could occur over a given time period with a specified probability α. Using the autoregressive models the one-day VaR at $(1-\alpha)$ can be estimated based on normal density. A crucial assumption for this way of estimating VaR is that the values of the variable of interest have the same distribution. As Value at Risk is a quantile α of the distribution of the variable of interest, it can also be estimated in 'one step' through quantile linear regression of the autoregressive model and factor-augmented autoregressive model. Quantile autoregressive models do not require the knowledge of the underlying distribution which could be a potential advantage (Komunjer, 2013). An alternative to quantile linear regression is quantile regression forest. This is a quantile version of the random forests that were introduced by Breiman (2001). This is a machine learning tool that has gained great popularity throughout the years. Quantile regression forests have been designed to take the advantages of quantile linear regression and random forests to predict quantiles. To my knowledge there is no generally known literature yet on using quantile forests for estimating Value at Risk of a macroeconomic time series. The use of EWPs of the individual models is based on Geweke & Amisano (2011) who find that simple pooled forecasts are superior for first moments. We are interested if this result also extends to the quantiles of a predicted distribution.

3 Data

The data used in this research is retrieved from the database of the 'Journal of Applied Econometrics'. It is monthly U.S. data that ranges from January 1973 until December 2014 (504 months). The datafile consists of a total of 164 macroeconomic series which are divided into nine groups. Except for 36 macroeconomic series that are retrieved from DataStream and one variable that is retrieved from the FRED Chicago website, the rest of the series are retrieved from the FRED-MD database. The series were retrieved from these platforms including all revisions known at that date. After obtaining the data we transform it, transformation codes and their explanation can be found in the Appendix (Table A1).

3.1 Tail risk measures

From the 164 macroeconomic variablese, we use five macroeconomic variables (IPG, EMG, CDI, BDI and DNFCI) to measure tail risk. The tail risk measures are VaRs of these five variables. As stated in Section 2, VaR is the minimum value that could occur over a given time period with a specified probability. We only consider the probability level 5% ($\alpha = 0.05$). Descriptive statistics and transformation of these five variables can be found in Table A1 in the Appendix.

IPG and EMG are indicators of real activity. The VaRs of those variables will be used as measures of tail real risk. The variable IPG is the log change in the industrial production index. The industrial production index is a monthly economic indicator "measuring real output in the manufacturing, mining, electric and gas industries (*Industrial Production Index (IPG)*, n.d.)". EMG is the log change in total employment. A defining characteristic of the log change is that it represents the percentage change.

CDI, BDI and DNFCI are indicators of financial stress. The VaRs of those variables will be used as measures of tail financial risk. 'DI' in CDI ad BDI stands for distance-to-insolvency (DI). In short, DI is the ratio of the firms' leverage over the asset volatility. A negative percentage corresponds with the firm being in debt. The DI ratio gives the decrease in asset value that would make the firm insolvent, measured in units of the firms asset standard deviation. CDI and BDI are portfolio versions of DI which have been introduced by Atkeson et al. (2017). CDI and BDI are respectively the DI of a value-weighted portfolio of the corporate sector and the banking sector. Summarizing, CDI and BDI can be interpreted as the lower bound of the chance on insolvency of respectively the entire corporate and banking sector.

DNFCI is the negative first difference in the National Financial Condition Index (NFCI). The NFCI has been produced by the Federal Reserve Bank of Chicago ans is a weekly updated index on "financial conditions in money markets, debt and equity markets and the traditional and "shadow" banking systems (*National Financial Conditions Index (NFCI)*, n.d.)". The NFCI serie ranges from January 1973 until December 2013 for a total of 492 observations. As DNFCI is NFCI's first difference this variable ranges from February 1973 until December 2013 for a total of 491 observations.

4 Methodology

De Nicolò & Lucchetta (2017) propose several models and methods for forecasting the five tail risk measures which are elaborated on in this section. An important part of the research is to perform every estimation and forecast pseudo real time. This implies that on time i we can only make use of the data known until time i. Furthermore, we use two window-based forecasting schemes: a rolling window of 120 months and an expanding window starting from 120 months. We compute the forecasts at a 3-month, 6-month, and 12-month horizon. An overview of all models is given in Table A2 in the Appendix. This Table also reports the EWPs, which are the average forecasts of the forecasts generated by the individual models in the pool. Finally, this section elaborates on scores and tests used for forecast evaluation.

4.1 Autoregressive models

This section will elaborate on all implemented autoregressive models estimated through linear regression. We consider, autoregressive (AR) models (4.1.1) which deliver direct and iterated forecasts, factor-augmented autoregressive (FAAR) models (4.1.2), which generate direct forecasts and factor-augmented vector autoregressive (FAVAR) models (4.1.3) which create iterated forecasts.

4.1.1 AR models

In an AR model, the dependent variable is described by a regression model that includes only its own lagged observations. An AR model of order p (AR(p)) can be written as:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \cdots + \beta_p y_{t-p} + \varepsilon_t. \tag{1}$$

In which β_0 is a constant, $\beta_1, ..., \beta_{t-p}$ are (unknown) parameters and ε_t is white noise with zero mean and constant variance σ^2. From this equation you could state that the observation y_t is related to the p previous observations, however this is a bit misleading as there is actually a dependence between y_t and all its past observations. The forecasts created by this model are made in an iterated or in a direct way. We also make use of two different forecasting windows, this gives a total of four individual AR models, all of them have five lags ($p = 5$). (See Table A2).

Direct Based on Marcellino et al. (2006), direct forecasts are made using a horizon-specific estimated model, where the dependent variable is the multi-period ahead value being forecasted. The model generating direct forecasts is given by the following equation:

$$y_{t+fh} = \beta_0 + \beta_1 y_t + \beta_2 y_{t-1} + \beta_3 y_{t-2} + \beta_4 y_{t-3} + \beta_5 y_{t-4} + \varepsilon_{t+fh}. \tag{2}$$

In which fh represents the forecasting horizon ($fh = 3, 6$ or 12). The variance is estimated in-sample.

Example: First window $(1 : 120)$ **and** 3**-month forecasting horizon** To give an illustration we consider the following example for a 3-month forecasting horizon and the first estimation window $(1 : 120)$. For this window we regress the vector $(y_8, \dots, y_{120})'$ on a matrix consisting of a column of ones and five vectors with its lagged valued $(y_5, \dots, y_{117})'$ until $(y_1, \dots, y_{112})'$. Such that, for example, y_{50} will be regressed on a constant, $y_{47}, y_{46}, y_{45}, y_{44}$ and y_{43}. In this way we estimate the model in Equation 2 and retrieve estimated parameters $b_0, \dots, b_5$ for $\beta_0, \dots, \beta_5$. To get the forecasted value $\hat{y}_{123}$ we fill in the values y_{120}, ..., y_{116} in the estimated model:

$$\hat{y}_{123|120} = b_0 + b_1 y_{120} + b_2 y_{119} + b_3 y_{118} + b_4 y_{117} + b_5 y_{116} \tag{3}$$

We retrieve the forecasted value $\hat{\sigma}_{123|120}$ through taking the root mean squared error of the estimated model. This is the root of the variance of the residuals. Subsequently we generate the VaR_α as:

$$VaR_\alpha(y_{t+fh|t}) = \hat{y}_{t+fh|t} + \hat{\sigma}_{t+fh|t} F^{-1}(\alpha). \tag{4}$$

In which $F^{-1}(\alpha)$ is the inverse Gaussian cumulative distribution function (cdf) which is equal to -1.645 for $\alpha = 0.05$. Furthermore fh is the forecasting horizon and t is the last observation of the current window.

Shift variables Note that, as the shift variables IPG, EMG and DNFCI represent monthly changes, we need to adjust them to get their realized values. Instead of regressing the original variable on its lagged values, we regress the realized variable on the lagged values of the original variable. The adjusted versions of the original variable differ for each forecasting horizon and can be computed as:

$$y_{realized,t} = y_t + y_{t-1} + \cdots + y_{t-fh+1}. \tag{5}$$

Note that for DNFCI we take the negative sum as this variable is a negative first difference. Thus for the shift variables, the model generating direct forecasts can be written as:

$$y_{realized,t+fh} = \beta_0 + \beta_1 y_t + \beta_{2t-1} + \beta_3 y_{t-2} + \beta_4 y_{t-3} + \beta_5 y_{t-4} + \varepsilon_{t+fh}. \tag{6}$$

As we regress the realized variable, the model also automatically returns a 'realized' version of the variance.

Iterated In contrast to direct forecasts, iterated forecasts are made using a one-period ahead model (Equation 1) and subsequently iterated forward for the desired number of periods (Marcellino et al., 2006). For all iterated forecasts we introduce a GARCH(1,1) specification to account for time-varying volatility.

Example: First window $(1:120)$ **and 3-month forecasting horizon** To illustrate, we take the first window $(1:120)$ and a 3-month forecasting horizon as an example. We estimate a model by regressing the vector $(y_6,\ldots,y_{120})'$ on a matrix containing a vector of ones and five vectors containing the lagged values of y: $(y_5,\ldots,y_{119})'$ until $(y_1,\ldots,y_{116})'$. We retrieve a model with estimated parameters $b_0,b_1,\ldots,b_5$ for $\beta_0,\ldots,\beta_5$ in Equation 1 (for $p=5$). To get the value $\hat{y}_{123}$, we first need to estimate y_{122} and y_{121} through the estimated model and then plug in those values to get $\hat{y}_{123}$:

$$\hat{y}_{123|120} = a + b_1\hat{y}_{122|120} + b_2\hat{y}_{121|120} + b_3 y_{120} + b_4 y_{119} + b_5 y_{118}. \tag{7}$$

Thus for every estimation window we (re-)estimate the model, and predict fh values for y instead of only one directly as done with direct forecasting. The estimated variance is retrieved through a GARCH(1,1) model:

$$\sigma_t^2 = \omega + \alpha\sigma_{t-1}^2 + \beta\varepsilon_{t-1}^2. \tag{8}$$

For forecasting the volatility through the GARCH(1,1) model we make use of the following equations:

$$\begin{aligned}\hat{\sigma}_{t+1|t}^2 &= \omega + \alpha\sigma_t^2 + \beta\varepsilon_t^2,\\ \hat{\sigma}_{t+fh|t} &= \omega + (\alpha+\beta)\hat{\sigma}_{t+fh-1|t}.\end{aligned} \tag{9}$$

We compute the VaR_α as in Equation 4.

Shift Variables Again we need to take note of the fact that the variables IPG, EMG and DNFCI represent monthly changes. For the iterated forecasts, we use the original variables as response variable, but we still need to make a slight adjustion. To get the forecast of the realized variable instead of the original variable, we add up all one-period ahead forecasts and get the multi-period forecasts:

$$\begin{aligned}\hat{y}_{realized,t+fh|t} &= \hat{y}_{t+fh|t} + \hat{y}_{t+fh-1|t} + \cdots + \hat{y}_{t+1|t}\\ \hat{\sigma}_{realized,t+fh}^2 &= \hat{\sigma}_{t+fh}^2 + \hat{\sigma}_{t+fh-1}^2 + \cdots + \hat{\sigma}_{t+1}^2.\end{aligned} \tag{10}$$

Note that, to get $\hat{y}_{realized,t+fh|t}$ for DNFCI, we take the negative sum of the one-period ahead forecasts as this variable is the negative first difference. To get $\hat{\sigma}_{adjusted,t+fh}^2$ for DNFCI we take the positive sum as in Equation 10, due to $Variance(A-B) = Variance(A) + Variance(B)$, assuming no (or little) correlation. We compute the VaR_α as

$$VaR_\alpha(y_{t+fh}) = \hat{y}_{adjusted,t+fh|t} + \hat{\sigma}_{realized,t+fh|t}F^{-1}(\alpha). \tag{11}$$

4.1.2 FAAR models

The factor-augmented autoregressive (FAAR) model generates direct forecasts, which means the same 'rules' apply to the shift variables IPG, EMG and DNFCI as described in the 'Direct' section in 4.1.1. A FAAR model of order p can be specified as:

$$y_{t+fh} = \alpha + \beta_1 y_t + \cdots + \beta_p y_{t-p+1} + \gamma_{1,1}f_{1,t} + \cdots + \gamma_{1,p}f_{1,t-p+1} + \cdots + \gamma_{k,1}f_{k,t} + \cdots + \gamma_{k,p}f_{k,t-p+1} + \varepsilon_{t+fh}. \tag{12}$$

In which $f_{i,t}$ is the value of factor i at time t, α is a constant and $\beta_1, \ldots, \gamma_{k,p}$ are (unknown) parameters. Or shorter as:

$$y_{t+fh} = \alpha + \beta_p(L)y_t + \gamma_{1,p}(L)f_{1,t} + \cdots + \gamma_{k,p}(L)f_{k,t} + \varepsilon_t, \tag{13}$$

or

$$y_{t+fh} = \alpha + \beta_p(L)y_t + \Gamma_p(L)F_t + \varepsilon_t. \tag{14}$$

In which L is the lag operator and $\beta_p(L) = (\beta_1 + \beta_2 L + \beta_3 L^2 + \cdots + \beta_{p-1}L^p)$, $\Gamma_p(L) = \begin{bmatrix}\gamma_{1,p}(L) & \ldots & \gamma_{k,p}(L)\end{bmatrix}$ and $F_t = \begin{bmatrix}f_{1,t} & \ldots & f_{k,t}\end{bmatrix}'$.

4.1.3 FAVAR models

In this section we use multivariate time series model instead of a univariate time series model as in section 4.1.1 and 4.1.2. There are a few reasons for using a multivariate time series model according to Wang (2020). We might be able to provide a better forecast with additional information on closely related variables. We are able to examine the dynamic relationship between several variables and furthermore, ignoring relationships with other variables may give rise to complications in univariate modelling. For this reasons we consider factor-augmented vector autoregressive (FAVAR) models. As the FAVAR model generates iterated forecasts we need to apply the same 'rules' for the shift variables IPG, EMG and DNFCI as described in the 'Iterated' section in 4.1.1. A FAVAR model of order p is set up as follows:

$$\begin{bmatrix} F_{t+1} \\ y_{t+1} \end{bmatrix} = A^{(1)} \begin{bmatrix} F_t \\ y_t \end{bmatrix} + \cdots + A^{(d)} \begin{bmatrix} F_{t-p+1} \\ y_{t-p+1} \end{bmatrix} + \begin{bmatrix} \eta_{t+1} \\ \varepsilon_{t+1} \end{bmatrix}. \tag{15}$$

Where the volatility of y again follows a GARCH(1,1) process (Equation 8 and 9). Note again that the iterated volatility and y forecasts of the shift variables IPG, EMG and DNFCI need to be summed as in Equation 10 to get the multi-period forecasts. The VaR_α for these variables is then again computed as Equation 11.

4.2 Quantile projections

This section elaborates on quantile autoregressive (QAR) models (4.2.1) and on factor-augmented (FAQAR) models (4.2.2) that generate forecasts using quantile linear regression.

4.2.1 QAR models

In contrast to the autoregressive models described in previous section 4.1, which are estimated through least squares regression, QARs are estimated using quantile linear regression. Least squares regression, on the one hand, wants to model the conditional mean of the response variable, and quantile regression, on the other hand, models the conditional α-th quantile of the response variable for some value of $\alpha \in (0, 1)$. In this research we are interested in forecasting the VaR_α for a value of $\alpha = 0.05$, which is the 0.05-th quantile. The α-th conditional quantile function of y_{t+fh} can be written as:

$$Q(\alpha|y_t, \ldots, y_{t-p+1}) = \beta_0(\alpha) + \beta_1(\alpha)y_t + \cdots + \beta_p(\alpha)y_{t-p+1}, \tag{16}$$

based on Koenker & Xiao (2006). Note that the AR models delivering direct forecasts can be seen as a special case of the QAR model by setting β_i for $i = 1, \ldots, p$ to a constant. Note too, that for the shift variables IPG, EMG and DNFCI we thus need to apply the rules as described in the 'Direct' section in 4.1.1.

The parameter vector $\beta(\alpha)$ minimizes

$$\frac{1}{T - fh} \sum_{t=1}^{T-fh} \rho_\alpha(y_{t+fh} - \beta_0(\alpha) - \beta_1(\alpha)y_t - \cdots - \beta_p(\alpha)y_{t-p+1}), \tag{17}$$

where fh is the forecasting horizon of interest and $\rho_\alpha(u)$ defines a loss function:

$$\rho_\alpha(u) = \begin{cases} u\alpha & u \geq 0 \\ u(1-\alpha) & u < 0. \end{cases} \tag{18}$$

The forecast generated by the QAR model of order p is thus given by

$$Q_\alpha(y_{t+fh}) = VaR_\alpha(y_{t+fh}) = b_0(\alpha) + b_1(\alpha)y_t + \cdots + b_p y_{t-p+1}. \tag{19}$$

In which $b_0(\alpha), \ldots, b_p(\alpha)$ are estimated parameters of $\beta_0, \ldots, \beta_p$ in Equation 17.

4.2.2 FAQAR models

Adding factors to QAR models gives factor-augmented quantile autoregressive (FAQAR) models. We get these models by taking $(1, y_t, \ldots, y_{t-p+1}, f_{1,t}, \ldots, f_{k,t-p+1})$ as regressor variables in Equation 16 and 17 instead of $(1, y_t, \ldots, y_{t-p+1})$. The loss function is still defined as in Equation 18. For the FAQAR models we again need to take into account that the sift variables IPG, EMG and DNFCI need to be adjusted as described in the 'Direct' section in 4.1.1.

4.3 Factor extraction methods

This section describes how the two factor extraction used in this research work. The essence of most factor methods is that a few factors, summarizing a large set of data, are used in forecasting equations.

4.3.1 Principal Component Analysis

The most well-known method for factor extaction is Principal Component Analysis (PCA). For this research, PCA finds a linear combinations of the 164 variables that are uncorrelated and have maximum variance. PCA can be applied on the correlation matrix or on the covariance matrix. De Nicolò & Lucchetta (2017) choose to use the correlation matrix. To be able to compare results we do so too. Let

$$\hat{V} = \frac{1}{T} \sum_{t=i}^{T} (X_t - \bar{X})(X_t - \bar{X})' \tag{20}$$

denote the $(N \times N)$ sample covariance matrix of $X_t = (X_{1t}, X_{2t}, \ldots, X_{Nt})'$. Where $\bar{X} = \frac{1}{T}\sum_{t=1}^{T} X_t$, note that for our data $N = 164$ and $T = 504$. Then the i-th principal component is the linear combination $f_i t = e_i' X_t$ that maximizes Variance$(f_{it}) = e_i' \hat{V} e_i$ subject to the constraint $e_i' e_i = 1$ and Cov$(f_{it}, f_{jt}) = e_i' \hat{V} e_j = 0$ for

$j = 1, \ldots, i-1$. Note that the solution e_i to the maximization problem is an eigenvector of $\hat{V}$. Each eigenvector of matrix $\hat{V}$ corresponds to an eigenvalue (constant) λ_i such that $\hat{V}e_i = \lambda_i e_i$. The sum of the eigenvalues is equal to the total variance of X_t and the fraction $\frac{\lambda_i}{\sum_{j=1}^{N} \lambda_j}$ is the fraction of the total variance in X_t explained by the i-th principal component. If $(\lambda_1, e_i), \ldots, (\lambda_N, e_N)$ are the eigenvalues-eigenvector pairs of $\hat{V}$, then they are ordered such that $\lambda_a \geq \lambda_2 \geq \cdots \geq \lambda_N \geq 0$. This implies that the first principal component explains the largest part in the total variance in X_t after which the second principal component follows and so on.

We make use of this fact when selecting the number of factors to add to the model. We either select five factors to add or the number of factors is determined according to the AH selection criterion (Ahn & Horenstein, 2013). The AH selection criterion selects the number of factors that maximizes the ratio of two sequential eigenvalues arranged in descending order. As we have seen that the factor corresponding to the largest eigenvalue explains the greatest part of the variation in the data.

4.3.2 Partial Least Squares

A more unknown alternative of factor extraction is Partial Least Squares (PLS) regression. Both methods construct new variables, known as factors or components, as linear combinations of a large data set. However, there is a main difference between PCA and PLS. PLS also takes the response variable into account, whereas PCA does not. PLS asssumes that the response variable is directly depending on the set of 164 variables. Linear, orthogonal combinations of the set of 164 variables are made such that the covariance between the response variable and the factors is maximized.

Again we either select five factors to add to the model or we use a similar selection procedure as the AH selection criterion: dividing the percentages of variance explained in the data set of 164 series X and subsequently maximizing this ratio gives us the number of factors to add.

4.4 Quantile regression forests

As an alternative to quantile linear regression which is the method used for estimating the Quantile Autoregressive and factor-augmented Quantile Autoregressive models, we consider quantile regression forest (Meinshausen, 2006). This section will shortly explain this method.

The quantile regression method in section 4.2 is solved by optimizing the parameters such that the the loss function in Equation 18 is minimized. However, the quantile regression forest method does not directly use this loss function but is based on the random forest method (Breiman, 2001). The main idea of random forests is that it grows an ensemble of trees, using independent observations of the predictor variables and the response variable. A defining characteristic of regression trees are its splitpoints. For each tree and node, the random forests use randomness for splitpoint selection, which the name of the method reveals. The main difference between quantile forests and random forests is that random forests, for each node in each tree, only keep the mean of the observations that fall into the node and ignores all other information. In contrast, quantile regression forests does not only keep the wanted quantile but keeps the value of all observations in this node and in this way it assesses the conditional distribution based on all observations. The mathematical notation is as follows (Meinshausen, 2006):

$$F(y|X=x) = P(Y \leq y|X=x) = E(\mathbb{I}(Y \leq y)|X=x)|X=x) \tag{21}$$

is the conditional distribution function of Y given $X = x$, which is subsequently estimated as:

$$\hat{F}(y|X=x) = \sum_{i=1}^{n} w_i(x)\mathbb{I}(Y_i \leq y). \tag{22}$$

Where $w_i(x) = k^{-1}\sum_{t=1}^{k} w_i(x, \theta_t)$, where k is the number of trees, each constructed with an idenpendent and identically distributed vector θ_t.

4.5 Forecast evaluation

To analyse the forecasts we make use of two score functions. We consider the quantile score (QS) and quantile-weighted probability score (QWPS) (Gneiting & Ranjan, 2011). The scoring rules are set up such that a lower score indicates a better performance. Consequently we can rank the forecasting methods and test if they significantly differ from each other by means of a Diebold Mariano (DM) test.

Quantile score To assign a quantile score to forecast $F^{-1}(\alpha)$ for y we use the following equation:

$$QS(F^{-1}(\alpha), y, \alpha) = \begin{cases} 2(1-\alpha)(F^{-1}(\alpha) - y) & y \leq F^{-1}(\alpha) \\ -2\alpha(F^{-1}(\alpha) - y) & y > F^{-1}(\alpha), \end{cases} \tag{23}$$

where the forecast $F^{-1}(\alpha)$ is a VaR_α forecast. The quantile score can be interpreted as the corresponding loss of an under- or overprediction.

Quantile-weighted probability score The quantile score assigns weight $w(\alpha) = 1$ to the all values of asymmetry parameter $\alpha \in (0, 1)$. However, this may not represent realistic assumptions on actual costs corresponding to the loss. In our case, as we are interested in the Value at Risk forecasts (for $\alpha = 0.05$), we will focus on the left tail. Gneiting & Ranjan (2011) propose a quantile weigthed version of the continuous ranked probability score as:

$$QWPS(f, y) = \int_0^1 QS(F^{-1}(\alpha), y)w(\alpha)d\alpha \tag{24}$$

Gneiting & Ranjan (2011) propose a few weight functions $w(\alpha)$. For emphasis on the left tail, as we wish, they propose $w(\alpha) = (1-\alpha)^2$. Note that we cannot compute the QWPS for the VaR forecasts generated through quantile linear regression nor quantile regression forests as these methods require an input of α. To compute the QWPS, we approach the integral in Equation 24, by using the 2-nd until the 99-th value of a vector with 100 evenly spaced values between 0 and 1.

Diebold Mariano test The Diebold Mariano test is used to compare the predictive ability of two forecasts. The DM test statistic is computed as:

$$DM = \frac{\bar{d}}{\sqrt{\hat{\sigma}_d^2/T}}, \tag{25}$$

where $\bar{d}$ represents the sample mean of the difference between two scores $d_t = (QS_{1,t} - QS_{2,t})$ or $d_t = (QWPS_{1,t} - QWPS_{2,t})$ and $\hat{\sigma}_d^2$ represents the variance of d_t. T is the number of forecasts, and is equal to 382, 379 or 373 for the 3-, 6- and 12-month forecasting horizon respectively. Except for the variable DNFCI, the values of T are 371, 368 and 362 respectively. Under the null hypothesis of equal performance, the DM statistic is asymptotically $N(0,1)$ distributed. As this is a two-sided test, we reject the null hypothesis of equal predictive performance at a $(1-\alpha)$ confidence interval if $|DM| > z_{alpha/2}$. For $\alpha = 0.05$, $z_{alpha/2}$ is equal to 1.96.

5 Results

5.1 Results for (factor-augmented) AR models

Table A3 in the Appendix gives the average left tail QWPSs of the four individual AR models, the two EWPs of the AR models, the eight individual factor-augmented AR models[1] and the two EWPs of the factor-augmented AR models. Thus in total it reports the results for sixteen individual AR models and four EWPs. According to pair-wise DM test, the models with the yellow marks are the models with superior predictive ability for that particular variable and forecasting horizon. The other bold values indicate that this particular model does not significantly differ from the best (yellow-marked) model in that column. We consider the AR models and the factor-augmented AR models seperately. From Table A3 there is no ultimate type of estimation window, way of forecasting, number of AR lags or number of factors that stands out to be preferable for all variables and horizons. However, for some variables the way of forecasting (direct or iterated) is dominating. For example for the factor-augmented AR models, we find that for the variables EMG, CDI and BDI the direct forecasts are dominating and for DNFCI iterated forecasts are dominating. This is in line with Marcellino et al. (2006) and Pesaran et al. (2011) who find that direct or iterated forecasts may be preferred depending on the variable and forecasting horizons. Table 1 reports the ratios of the QWPSs of the best individual factor-augmented AR model over the best individual AR model. In this case 'best' means that the forecast generated by this particular model gives the lowest average QWPS. Ratios in bold indicate that a pair-wise DM test of equal predictive ability has been rejected. Illustrating, a bold ratio less than one indicates that the best individual factor-augmented AR model for the particular variable and horizon is strictly better than the best individual AR model.

From Table 1 we can conclude that for the variables CDI and BDI a factor-augmentation of the AR model does not improve their forecasts as the ratios are greater than one. For half of these cases a DM test proves

[1]Throughout the Results section, the name 'factor-augmented AR' models is used as a general name for the FAAR and FAVAR models.

Table 1: Ratios of average left tail QWPSs of best FA(V)AR models over average left tail QWPSs of best AR models

	Horizon		
Variable	**3 months**	**6 months**	**12 months**
IPG	**0.91**	0.91	0.92
EMG	**0.84**	**0.81**	**0.88**
CDI	**1.03**	1.02	1.03
BDI	1.02	**1.04**	**1.05**
DNFCI	0.95	0.95	0.99

Bold ratios indicate a rejection of equal performance at 5%.

the difference between the models is significant.

For the rest of the variables IPG, EMG and DNFCI we find ratios less than one, indicating the best factor-augmented model gives a lower average QWPS than the best AR model without factors. For EMG we even find that for all horizons the best factor-augmented AR model is strictly better than the best AR model without factors. We can thus conclude that adding factors to the AR models improves the model for certain variables and horizons.

To analyse the gain of using EWPs over individual AR models we set up Table 2 which consists of four subtables. Table 2(a) reports ratios of QWPSs of the best AR EWP over the average QWPSs of the best individual AR model. Table 2(b) reports ratios of the average QWPSs of the best FA(V)AR EWP over the average QWPSs of the best individual factor-augmented AR model. Table 2(c) reports ratios of average QWPS of the worst AR EWP over the average QWPS of the worst individual AR model. Table 2(d) reports ratios of the average QWPSs of the worst FA(V)AR EWP over the average QWPSs of the worst individual factor-augmented AR model.

From Table 2(a) and 2(b) we can conclude that an EWP does not necessarily improve the predictive performance compared to the best individual models as most ratios are equal to one or greater than one. However there are only few cases where equal predictive performance is actually rejected. There are even a few cases where ratios are less than one indicating that the best EWP for the corresponding horizon and variable gives even a lower QWPS than the best individual model. For BDI and a 12-month forecasting horizon, the best EWP of factor-augmented AR models (FAAR D) is strictly better than the best individual factor-augmented AR model (FAAR D 5 EW).

From Subtable 2(c) and 2(d) we can conclude that the worst EWP is always strictly better than the worst individual AR or factor-augmented AR model. The only exception applies to variable BDI and a 3-month forecasting horizon in Subtable 2(c). For this case the worst individual AR model (AR D RW) and the worst EWP (AR I) are not proven to be signficantly different.

Thus from Table 2 the worst EWPs strictly outperform the worst individual model. Compared to the best individual models EWPs are not proven to perform significantly different. Considering this finding we have a closer look at the EWPs. Table 3 gives the DM statistics of pair-wise DM tests between the four EWPs of AR and factor-augmented AR models.

Table 2: Ratios of average left tail QWPSs of best (factor-augmented) AR EWP over average left tail QWPSs of best individual (factor-augmented) AR.

	Horizon		
Variable	**3 months**	**6 months**	**12 months**
IPG	1.00	1.02	**1.04**
EMG	1.01	1.01	1.01
CDI	1.03	1.03	1.03
BDI	1.00	1.00	**1.02**
DNFCI	0.98	1.00	**1.04**

(a) Best AR EWP versus best AR

	Horizon		
Variable	**3 months**	**6 months**	**12 months**
IPG	0.99	1.05	**1.06**
EMG	0.98	0.97	0.98
CDI	**1.02**	**1.03**	0.99
BDI	1.01	1.00	**0.96**
DNFCI	0.97	0.97	1.04

(b) Best FA(V)AR EWP versus best FA(V)AR

	Horizon		
Variable	**3 months**	**6 months**	**12 months**
IPG	**0.97**	**0.96**	**0.94**
EMG	**0.96**	**0.94**	**0.91**
CDI	**0.97**	**0.96**	**0.94**
BDI	0.99	**0.98**	**0.95**
DNFCI	**0.87**	**0.93**	**0.91**

(c) Worst AR EWP versus worst AR

	Horizon		
Variable	**3 months**	**6 months**	**12 months**
IPG	**0.93**	**0.89**	**0.90**
EMG	**0.96**	**0.96**	**0.95**
CDI	**0.94**	**0.92**	**0.86**
BDI	**0.92**	**0.88**	**0.84**
DNFCI	**0.79**	**0.84**	**0.81**

(d) Worst FA(V)AR EWP versus worst FA(V)AR

Bold ratios indicate a rejection of equal performance at 5%.

Table 3: Pair-wise DM statistics of left tail QWPSs of EWPs of AR and factor-augmented AR models

	Horizon	**3 months**			**6 months**			**12 months**		
Variable	**Model**	**AR I**	**AR D**	**FAAR D**	**AR I**	**AR D**	**FAAR D**	**AR I**	**AR D**	**FAAR D**
IPG	**AR D**	**5.65**			**5.85**			**3.42**		
	FAAR D	**-2.49**	**-4.41**		-1.54	**-4.42**		-1.71	**-4.63**	
	FAVAR I	-1.17	**-3.55**	**2.32**	-1.66	**-5.14**	-0.13	-0.67	**-3.37**	0.94
EMG	**AR D**	**2.66**			**2.88**			**3.26**		
	FAAR D	**-3.87**	**-4.19**		**-4.04**	**-4.72**		**-3.15**	**-4.60**	
	FAVAR I	0.95	0.62	**5.60**	1.04	0.58	**5.35**	1.61	1.03	**4.30**
CDI	**AR D**	0.83			1.00			0.57		
	FAAR D	1.77	0.80		**2.50**	1.25		-0.04	-0.69	
	FAVAR I	**4.03**	**3.53**	**2.79**	**4.97**	**4.35**	**3.40**	**5.64**	**6.29**	**5.77**
BDI	**AR D**	0.22			-1.88			**-4.01**		
	FAAR D	**2.21**	1.35		0.32	**2.34**		**-2.17**	1.26	
	FAVAR I	**5.72**	**5.81**	**4.94**	**4.53**	**5.84**	**4.59**	**4.90**	**7.56**	**6.60**
DNFCI	**AR D**	**8.50**			**9.38**			**5.22**		
	FAAR D	**3.27**	-1.80		**3.71**	**-3.08**		1.13	**-2.46**	
	FAVAR I	-0.89	**-6.97**	**-6.88**	-0.98	**-6.27**	**-4.84**	-0.26	**-4.46**	-1.60

Bold ratios indicate a rejection of equal performance at 5%

A negative and significant (in bold) DM statistic implies the EWP in the corresponding row performs strictly better than the EWP in the corresponding column. The results differ among the variables, however for each variable we can see quite the same patterns for each forecasting horizon.

For the real variables IPG and EMG we see that for each horizon FAAR D strictly dominates AR D. For both variables the FAAR D model also generates better forecasts than, or no significantly different forecasts than AR I and FAVAR I.

For the financial variables CDI and BDI the forecasts generated by the FAVAR I EWP is clearly outperformed by the other EWPs. For BDI, this is in line with the paper by De Nicolò & Lucchetta (2017). Furthermore, in the paper, CDI and BDI are the only variables for which the FAAR D EWP strictly outperforms the direct EWP without factors (AR D) for every horizon. We find this too.

For the financial variable DNFCI, it seems that the iterated forecasting method with GARCH volatility is dominating. For all forecast horizons the models AR I and FAVAR I deliver strictly better forecasts than AR D. Furthermore, for the 3- and 6-month forecasting horizons AR I and FAVAR I also outperform FAAR D. There is no significant difference between FAVAR I and AR I, however the FAVAR I forecasts have slightly lower average QWPSs.

5.2 Results for (factor-augmented) Quantile Projections

Table A4 in the Appendix reports the average QSs of the two individual QAR models, the six individual factor-augmented QAR models, an EWP of the QAR models and an EWP of the factor-augmented QAR models.

First we consider the impact of adding factors to the individual QAR models. From Table A4 it seems that again, for the variables IPG, EMG and DNFCI the factor-augmented models are preferable and for variable CDI the models without factors are peferable. For BDI there seems to be no clear preference. Table 4 gives the ratios of the QS of the best individual FAQAR model over the QS of the best individual QAR model for each variable and horizon.

Table 4: Ratios of QSs of best QAR models over QSs of best FAQAR models

	Horizon		
Variable	**3 months**	**6 months**	**12 months**
IPG	**0.82**	**0.81**	**0.79**
EMG	**0.88**	**0.74**	**0.85**
CDI	**1.18**	**1.12**	1.09
BDI	**1.20**	0.99	1.14
DNFCI	**0.62**	**0.64**	**0.73**

Bold ratios in the tables above (left and right) indicate a rejection of equal performance at 5%

Table 5: Ratios of average QSs of FAQAR EWPs over average QSs of QAR EWPs

	Horizon		
Variable	**3 months**	**6 months**	**12 months**
IPG	**0.85**	**0.84**	**0.88**
EMG	**0.77**	**0.81**	0.90
CDI	**1.20**	**1.15**	**1.17**
BDI	**1.11**	1.07	0.95
DNFCI	**0.64**	**0.65**	**0.66**

From Table 4 we can confirm that the addition of factors to the individual QAR models indeed strictly improves the forecasting ability for each horizon for the real variables IPG and EMG and the financial variable DNFCI. The extent to which the QAR models are improved through the addition of factors compared to the AR models is even greater. For the other financial variables CDI and BDI, however, we find that the best QAR model without factors is better than, or not significantly different from the best FAQAR models.

Just like for the AR models we find that the best individual models only for a few cases significantly outperform the best EWPs, and again from the other point of view, the worst EWP significantly outperfoms the worst individual model. Ratios supporting this are given in Table A5 in the Appendix.

Ratios of QSs of FAQAR EWPs over QSs of QAR EWPs are given in Table 5. The FAQAR EWP significantly outperforms the QAR EWP for the financial variable DNFCI and the real variables IPG and EMG. The 12-month forecasting horizon for EMG is the only exception. However for the financial variable CDI we find that the QAR EWP is strictly better than the FAQAR EWP. As we have found similar results for the individual models (Table 4) and considering the observations of previous section, we recommend using EWPs. we will from now on particularly focus on the EWPs.

5.3 Comparing (factor-augmented) AR and (factor-augmented) QAR models

Table 6 compares the FAQAR EWPs and the best factor-augmented AR EWPs (FAAR D or FAVAR I) for the corresponding variable and forecasting horizon by dividing their average quantile scores.

Table 6: Ratios of average QSs of FAQAR EWPs over average QSs of best FA(V)AR EWPs

	Horizon		
Variable	**3 months**	**6 months**	**12 months**
IPG	1.00	1.02	1.01
EMG	**1.16**	**1.27**	**1.31**
CDI	1.07	**1.11**	1.00
BDI	1.11	1.05	**1.26**
DNFCI	**1.22**	**1.20**	1.11

Bold ratios indicate a rejection of equal performance at 5%.

All ratios in Table 6 are greater than one, indicating factor-augmented AR EWP forecasts get lower average quantile scores than FAQAR EWP forecasts. However, for the variable IPG, the DM test indicates that there is no significant difference between the FAQAR EWPs and best factor-augmented AR EWP for all horizons. For the variable EMG, on the other hand, there clearly is. For CDI and BDI it seems better to not add factors. For this reason we also compare the best (factor-augmented) AR EWP (AR I, AR D, FAAR D or FAVAR I) to the best quantile projection EWP (QAR or FAQAR), see Table A6. This gives simalar results: (Factor-augmented) AR EWPs in general produce lower quantile scores than (factor-augmented) quantile projection EWPs. Only for CDI there seems to be potential in using quantile projections.

Based on De Nicolò & Lucchetta (2017) we also compare the best (factor-augmented) AR EWPs and the FAQAR EWPs through coverage ratios. In this way we want to find out if the VaR forecasts are appropriate as early warning signals. The coverage ratios are given in Table A7, for the whole sample and for a subsample. The subsample starts from 2007, to see how the coverage ratios behave when a reliable early warning system is needed most (just before a financial crisis). A coverage ratio higher than 0.05 means that the VaR forecast underestimates tail risk, as this means that the VaR forecast is violated more than 5% of the time. A coverage ratio lower than 0.05 means that the VaR forecast overestimates risk, as the VaR forecast is violated less than 5% of the time. Assuming one is risk-adverse, we prefer overestimating risk over underestimating risk which corresponds to preferring a lower coverage ratio than than the target probability (0.05) over a higher coverage ratio than the target.

For all variables and horizons, coverage ratios for the best (factor-augmented) QAR EWP forecasts are higher than for the best (factor-augmented) AR EWP. For the 12-month horizon and the subsample 2007 - 2014 (or 2007 - 2013 for DNFCI) all ratios clearly exceed 0.05. However, the coverage ratios of FAQAR EWP forecast are even double the coverage ratios of the best (factor-augmented) QAR EWP forecasts. This finding confirms that (factor-augmented) AR EWPs are generally preferable over the (factor-augmented) QAR models. Figure 3 until 12 in the Appendix show the two best (factor-augmented) AR EWP and best (factor-augmented) QAR EWP 12-month horizon forecasts of each variable.

5.4 Results for Partial Least Squares

In Table A8 all average QWPSs of PLS factor-augmented AR models and their EWPs are reported. Just like for PCA factor-augmented AR models, the best forecasting method (iterated or direct) seems to be variable depending. For the variables IPG, EMG, CDI and BDI the direct method seems to be preferable and for DNFCI the iterated method. For the individual models these methods seem to be preferable in combination with respectively an expanding window and a rolling window. There seems to be no particular preference for number of factors or number of lags.

Table 7 compares the PCA factor-augmented AR EWPs and the PLS factor-augmented AR EWPs.

From Table 7 it becomes clear that for the real variables IPG and EMG the direct forecasting method is indeed strictly preferable. Both direct PLS EWPs and direct PCA EWPs outperform the iterated EWPs. Between the two direct EWPs is no significant difference. For the financial variables CDI and BDI, the direct PCA EWP significantly outperforms the two iterated EWPs. Furthermore, for CDI the iterated PCA EWP always outperforms the iterated PLS EWP whereas for BDI the iterated PLS EWP always outperforms the iterated PCA EWP. For the financial variable DNFCI, the iterated EWPs are preferable.

Table A9 in the Appendix reports all average QSs of PLS factor-augmented quantile projections. Table 8 compares the PCA factor-augmented QAR models and the PLS factor-augmented QAR models. For CDI the PLS method strictly outperforms the PCA method for the 6-, and 12-month forecasting horizon. For BDI the PLS method strictly outperforms the PCA method for the 3-month horizon. For DNFCI however, the PCA method seems to dominate, which also accounts for IPG for the 3-month forecasting horizon. Thus the preferable factor extraction method with regards to quantile projections,differs among the variables and horizons or does not differ significantly.

For BDI the PLS method is significantly proven to improve the FAQAR EWP the most compared to the PCA method for a 3-month forecasting horizon.

Table 7: Pair-wise DM statistics of left tail QWPSs of PCA factor-augmented AR EWPs and PLS factor-augmented AR EWPs

	Horizon	3 months			6 months			12 months		
Variable	**Model**	**D PCA**	**I PCA**	**D PLS**	**D PCA**	**I PCA**	**D PLS**	**D PCA**	**I PCA**	**D PLS**
IPG	**I PCA**	**2.32**			-0.13			**0.94**		
	D PLS	0.46	-1.65		-0.62	-0.39		1.29	-0.11	
	I PLS	**2.44**	0.82	**2.59**	1.20	**3.03**	**2.13**	**3.31**	**7.83**	**3.41**
EMG	**I PCA**	**5.60**			**5.35**			**4.30**		
	D PLS	1.95	**-4.09**		1.95	**-4.44**		0.99	**-4.06**	
	I PLS	**6.49**	**3.37**	**5.28**	**7.18**	**6.49**	**6.80**	**3.93**	**2.11**	**3.40**
CDI	**I PCA**	**2.79**			**3.40**			**5.77**		
	D PLS	0.80	-1.66		1.62	-1.50		1.78	**-4.35**	
	I PLS	**3.64**	**3.42**	**2.48**	**4.03**	**2.92**	**2.03**	**6.50**	**3.24**	**5.03**
BDI	**I PCA**	**4.94**			**4.59**			**6.60**		
	D PLS	1.75	-1.68		1.41	**-2.59**		**1.30**	**-4.57**	
	I PLS	**4.37**	**-2.20**	1.24	**4.25**	**-2.73**	**2.18**	**5.58**	**-5.23**	**3.83**
DNFCI	**I PCA**	**-6.88**			**-4.84**			-1.60		
	D PLS	**2.05**	**6.76**		**3.02**	**6.88**		1.86	**3.90**	
	I PLS	**-4.14**	**0.45**	**-8.16**	**-3.72**	0.70	**-9.08**	-1.02	**0.04**	**-5.69**

Bold ratios indicate a rejection of equal performance at 5%.

Table 8: Ratios of average QSs of PLS FAQAR EWPs over average QSs of PCA FAQAR EWPs

	Horizon		
Variable	**3 months**	**6 months**	**12 months**
IPG	**1.13**	0.94	1.05
EMG	0.90	0.92	0.94
CDI	0.97	**0.93**	**0.91**
BDI	**0.90**	1.02	1.08
DNFCI	**1.42**	**1.29**	1.15

Bold ratios indicate a rejection of equal performance at 5%.

5.5 Results for Quantile Forests

Table A10 reports all average QSs for two individual quantile forests (QF) models, four individual factor-augmented quantile forest (FAQF) models and two EWPs. Just as for the least squares regression and quantile linear regression methods, we analyse if the addition of factors to the models give any improvement using the ratios in Table 9. From this Table we can conclude that forecasts produced by the FAQF EWP are either not significantly different from, or outperforming the forecasts generated by the QF EWP .

Table 10 reports the ratios of the average QSs of the best quantile forest EWP (QF EWP or FAQF EWP) over the average QSs of the best quantile projection EWP (QAR EWP or FAQAR EWP).

Table 10 shows that for the variable DNFCI the use of Quantile Forests is a great improvement. For the other variables however, there is no significant difference between the forecasts generated by the quantile forest

Table 9: Ratios of average QSs of FAQF EWPs over average QSs of QF EWPs

	Horizon		
Variable	**3 months**	**6 months**	**12 months**
IPG	**0.78**	**0.86**	0.94
EMG	0.92	**0.84**	**0.85**
CDI	1.07	1.04	1.03
BDI	1.01	1.02	0.98
DNFCI	1.02	1.02	1.06

Bold ratios indicate rejected equal predictive performance at 5%.

Table 10: Ratios QSs best (FA)QF EWP over QSs best (FA)QAR EWP

	Horizon		
Variable	**3 months**	**6 months**	**12 months**
IPG	1.05	**1.10**	1.04
EMG	1.09	0.99	0.99
CDI	1.08	1.08	1.06
BDI	1.02	**1.12**	1.01
DNFCI	**0.69**	**0.66**	**0.69**

Table 11: Ratios average QSs PLS FAQF EWPs over average QSs PCA FAQF EWPs

	Horizon		
Variable	**3 months**	**6 months**	**12 months**
IPG	0.94	0.90	**1.07**
EMG	0.99	0.95	1.04
CDI	0.94	0.97	0.99
BDI	0.98	0.99	1.00
DNFCI	**1.17**	**1.13**	1.01

Bold ratios in the two tables above (left and right) indicate rejected equal predictive performance at 5%.

EWPs and quantile projection EWPs. Or the quantile projection EWPs are proved to be significantly better (IPG and BDI: 6-month horizon).

To give an illustration, Figure 1 shows the 12-month FAQAR EWP VaR forecast for DNFCI and Figure 2 shows the 12-month QF EWP VaR forecast. Both are respectively the best quantile projection EWP and quantile forest EWP for DNFCI for a 12-month horizon. Remarkable is that the quantile forest EWP does way better with regards to the period around 2008 when the financial crisis happened. This seems promising as it is for such periods that we particularly need an early warning system to be reliable. For DNFCI, the quantile forest EWP even strictly outperforms the best AR EWP with ratios of their average quantile scores being equal to respectively 0.85, 0.79 and 0.76.

Finally we have combined the two methods of PLS factor-extraction and quantile forests. Table A11 reports all average QSs for four individual PLS factor-augmented QFs and one EWP (FAQF PLS). Table 11 helps us compare the two factor-augmented quantile forest EWPs: FAQF PCA EWP and FAQF PLS EWP. Most of the ratios in he Table are less than one, indicating that the PLS factor-augmented quantile forest EWP forecasts have a lower average quantile score than the PCA factor-augmented quantile forest EWP forecasts. However, this improvement is not proven to be significant for any horizon nor variable.

As our main goal is to generate an early warning system for events like the financial crisis as in 2007-2009, we have a better look at the coverage ratios (VaR violations) of the 12-month forecasts created by the PCA FAQF EWPs and PLS FAQF EWPs for subsample 2007 - 2014. These coverage ratios are given in Table 12.

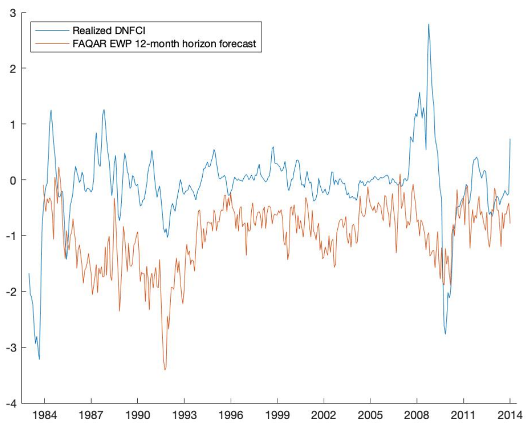

Figure 1: Realized DNFCI and it's FAQAR EWP $\text{VaR}_{0.05}$ forecast for a 12-month horizon

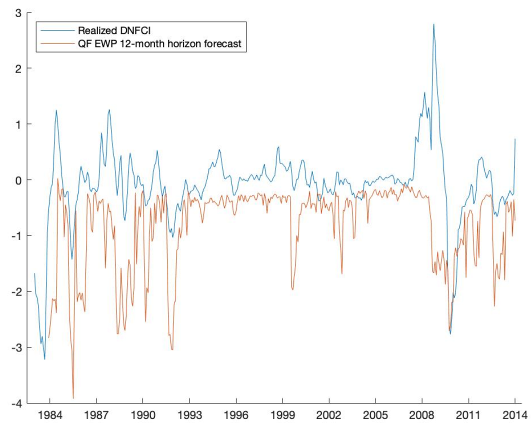

Figure 2: Realized DNFCI and it's QF EWP $\text{VaR}_{0.05}$ forecast for a 12-month horizon

Table 12: Coverage ratios for 12-month horizon PCA FAQF EWP and PLS FAQF EWP VaR forecasts for the subsample 2007M01 - 2014M12

Model | Variable	IPG	EMG	CDI	BDI	DNFCI
FAQF PCA EWP	0.15	0.25	0.15	0.23	0.12
FAQF PLS EWP	0.15	0.20	0.11	0.23	0.06

Although both models still produce VaR forecasts which are violated way more than the target probability, we do see an improvement of using PLS over PCA in this case. Furthermore, both the PCA FAQF EWP as the PLS FAQF EWP give lower coverage ratios than the best quantile projection EWP for a 12-month forecasting horizon for all variables (except one, Table A7).

6 Conclusion

In this research we have tried to find an early warning system for tail risk. Therefore we have analysed Value at Risks forecasts of two real variables (IPG,EMG) and three financial variables (CDI,BDI,DNFCI) created by individual autoregressive and factor-autoregressive models and their equally weighted pools. We have found a few main results. To start we have found that extending the models through factors seem to really improve the autoregressive and quantile autoregressive models for the real variables IPG and EMG and the financial variable DNFCI. Whether the addition of factors improves the model is thus variable dependent. Second, we find that for the factor-augmented autoregressive models the way of forecasting seems to dominate (in line with Marcellino et al. (2006) and Pesaran et al. (2011)) instead of the factor-extraction method. As the factor-augmented quantile autoregressive models are all similar to the direct forecast method we cannot draw the same conclusions for those models. For the factor-augmented quantile autoregressive models, we find that the PLS method is strictly superior to the PCA method for several variables and forecasting horizons. Except for DNFCI, PLS gives better evaluation scores of the Value at Risks forecasts than PCA for at least one forecasting horizon for all variables, extending the emperical results of Groen & Kapetanios (2016). Third, we find that it can be recommended to use equally weighted pool forecasts instead of forecasts generated by only one individual model. As sometimes EWPs are proven to outperform all individual models and the best model is horizon and variable dependent, EWPs are considered more reliable. Comparing the (factor-augmented) autoregressive and (factor-augmented) quantile autoregressive EWPS, we find that the (factor-augmented) autoregressive EWPs are superior, the coverage ratios of the Value at Risks confirm this. Last, we have analysed the possible improvement of the Value at Risk forecasts through to the use of a quantile forest method instead of quantile linear regression for estimating the quantile autoregressive models. We find that for the financial variable DNFCI this method strongly improves the Value at Risk forecasts for every horizon. The quantile forest EWP even outperforms its best (factor-augmented) autoregressive EWP forecast. Again extending the quantile forest models through the addition of factor generates significantly better forecast scores, but only for particular variables. In this case it is improving for the EWP forecasts of the real variables IPG and EMG. For the other variables we do not specifically see a significant improvement. Using PLS instead of PCA for the factor-augmented quantile forests does in general not create significantly different forecasts according to the quantile scores. However, if we have a look at the coverage ratios for when early warnings are needed most, adding PLS seems to be more recommendable than PCA. All in one, we do not find one very best model for all variables. However, we do find some important extending results on De Nicolò & Lucchetta (2017) for the research field of forecasting tail risks.

7 Discussion

Looking back on this research, there are a few points of attention and possibilities for future research. First, this research analyses several models that forecast Value at Risks. However, a disadvantage of Value at Risks is that is does not say anything about the amount of loss when exceeding the Value at Risk threshold. A complementing measure could be Expected Shortfall. An Expected Shortfall of α is defined as the conditional expectation of exceedances when violating the corresponding VaR_α. Li et al. (2020) for example propose a method that jointly forecasts the Value at Risk and Expected Shortfall through a Bayesian model. De Nicolo (2018) actually already considers forecasting combinations of Value at Risks and Expected Shortfalls. Another comment on this research is that the models are evaluated through Quantile Scores and Quantile Weighted Probability Scores (Gneiting & Ranjan, 2011). However, we cannot assign probability scores to the forecasts made through quantile regression. This makes it slightly harder to compare the autoregressive models and the quantile models, as sometimes the best model according to the quantile score does not match the best model according to the quantile weighted probability score. Possible further research could thus be on finding an extended evaluation tool to compare the VaR generated through the density forecasts and quantile forecasts. As the autoregressive models appeared to be best, we could also consider optimizing the weighing scheme for the models in the pools using scoring rules according to Opschoor et al. (2017). As the results for quantile forests seem promising, we could also consider other quantile methods such as in Taylor (2000).

References

Acemoglu, D., Ozdaglar, A., & Tahbaz-Salehi, A. (2017). Microeconomic origins of macroeconomic tail risks. *American Economic Review*, *107*(1), 54–108.

Ahn, S. C., & Horenstein, A. R. (2013). Eigenvalue ratio test for the number of factors. *Econometrica*, *81*(3), 1203–1227.

Akaike, H. (1969). Fitting autoregressive models for prediction. *Annals of the institute of Statistical Mathematics*, *21*(1), 243–247.

Atkeson, A. G., Eisfeldt, A. L., & Weill, P.-O. (2017). Measuring the financial soundness of us firms, 1926–2012. *Research in Economics*, *71*(3), 613–635.

Breiman, L. (2001). Random forests. *Machine learning*, *45*(1), 5–32.

De Nicolo, G. (2018). An early warning system for systemic risks. *Available at SSRN 3132520*.

De Nicolò, G., & Lucchetta, M. (2017). Forecasting tail risks. *Journal of Applied Econometrics*, *32*(1), 159–170.

Geweke, J., & Amisano, G. (2011). Optimal prediction pools. *Journal of Econometrics*, *164*(1), 130–141.

Gneiting, T., & Ranjan, R. (2011). Comparing density forecasts using threshold-and quantile-weighted scoring rules. *Journal of Business & Economic Statistics*, *29*(3), 411–422.

Groen, J. J., & Kapetanios, G. (2016). Revisiting useful approaches to data-rich macroeconomic forecasting. *Computational Statistics & Data Analysis*, *100*, 221–239.

Industrial production index (ipg). (n.d.). Retrieved from `https://fred.stlouisfed.org/series/INDPRO`

Koenker, R., & Xiao, Z. (2006). Quantile autoregression. *Journal of the American Statistical Association*, *101*(475), 980–990.

Komunjer, I. (2013). Quantile prediction. In *Handbook of economic forecasting* (Vol. 2, pp. 961–994). Elsevier.

Li, Z., Tran, M.-N., Wang, C., Gerlach, R., & Gao, J. (2020). A bayesian long short-term memory model for value at risk and expected shortfall joint forecasting. *arXiv preprint arXiv:2001.08374*.

Marcellino, M., Stock, J. H., & Watson, M. W. (2006). A comparison of direct and iterated multistep ar methods for forecasting macroeconomic time series. *Journal of econometrics*, *135*(1-2), 499–526.

Meinshausen, N. (2006). Quantile regression forests. *Journal of Machine Learning Research*, *7*(Jun), 983–999.

National financial conditions index (nfci). (n.d.). Retrieved from `https://www.chicagofed.org/publications/nfci/index`

Opschoor, A., Van Dijk, D., & van der Wel, M. (2017). Combining density forecasts using focused scoring rules. *Journal of Applied Econometrics*, *32*(7), 1298–1313.

Pesaran, M. H., Pick, A., & Timmermann, A. (2011). Variable selection, estimation and inference for multi-period forecasting problems. *Journal of Econometrics*, *164*(1), 173–187.

Stock, J. H., & Watson, M. W. (2002). Forecasting using principal components from a large number of predictors. *Journal of the American statistical association*, *97*(460), 1167–1179.

Stock, J. H., & Watson, M. W. (2006). Forecasting with many predictors. *Handbook of economic forecasting*, *1*, 515–554.

Taylor, J. W. (2000). A quantile regression neural network approach to estimating the conditional density of multiperiod returns. *Journal of Forecasting*, *19*(4), 299–311.

van Os, B., & van Dijk, D. (2020). Quantitative methods for finance slides week 1.

Wang, W. (2020). Advanced econometrics course slides. topic 3: Multivariate time series.

Appendix

Table A1: Descriptive statistics

Variable	Group	Serie	Transformation	Observations	Mean	Std dev	Min	Max
IPG	1	3	(3)	503	0.176	0.731	-4.299	2.068
EMG	2	11	(3)	503	0.114	0.276	-0.852	1.502
CDI	9	2	(1)	504	0.083	0.032	0.014	0.241
BDI	9	7	(2)	504	0.082	0.044	0.008	0.316
DNFCI	9	10	(2)	491	0.002	0.253	-1.389	1.281

The data is retrieved from the ***database*** of the Joumrnal of Applied Econometrics. The data consists of 9 groups of series and should be transformed according to their transformation code. (1) means no transformation, (2) means first difference and (3) stands for the first difference of the natural logarithm. The group, serie and transformation for the five variables we use to forecast their VaR as tail risk measures are stated in the Table above. Together with their descriptive statistics.

Table A2: Model overview

Model number	Model name	Lags	Factors	Estimation window	Estimation	Volatility
			AR models			
1	AR D RW	5	-	Rolling window	Direct	In-sample
2	AR D EW	5	-	Expanding window	Direct	In-sample
3	AR I RW	5	-	Rolling window	Iterated	GARCH(1,1)
4	AR I EW	5	-	Expanding window	Iterated	GARCH(1,1)
			Factor-augmented AR models			
5	FAAR D AH RW	2	AH	Rolling window	Direct	In-sample
6	FAAR D AH EW	2	AH	Expanding window	Direct	In-sample
7	FAAR D 5 RW	2	5	Rolling window	Direct	In-sample
8	FAAR D 5 EW	2	5	Expanding window	Direct	In-sample
9	FAVAR I AH RW	2	AH	Rolling window	Iterated	GARCH(1,1)
10	FAVAR I AH EW	2	AH	Expanding window	Iterated	GARCH(1,1)
11	FAVAR I 5 RW	1	5	Rolling window	Iterated	GARCH(1,1)
12	FAVAR I 5 EW	1	5	Expanding window	Iterated	GARCH(1,1)

			Quantile Projections	
1	QAR RW	5	-	Rolling window
2	QAR EW	5	-	Expanding window
			Factor-augmented Quantile Projections	
3	FAQAR AH RW	2	AH	Rolling window
4	FAQAR AH EW	2	AH	Expanding window
5	FAQAR 5 RW	1	5	Rolling window
6	FAQAR 5 EW	1	5	Expanding window

Equally weighted pools			
EWP AR models		**EWP QP models**	
(1,2)	AR D	**(1,2)**	QAR
(3,4)	AR I	**(3,4,5,6)**	FAQAR
(5,6,7,8)	FAAR D		
(9,10,11,12)	FAVAR I		

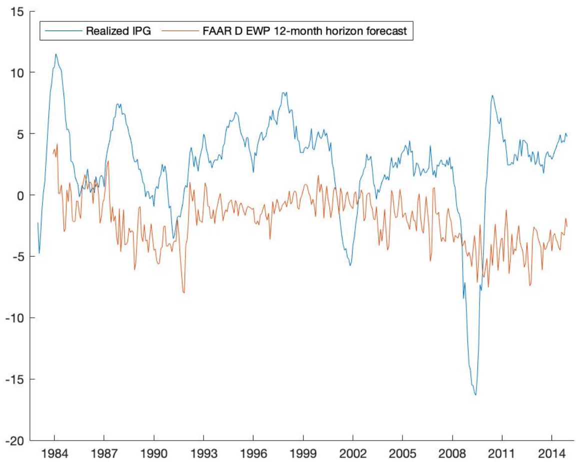

Figure 3: Realized IPG and it's FAAR D EWP $\mathrm{VaR}_{0.05}$ forecast for a 12-month horizon

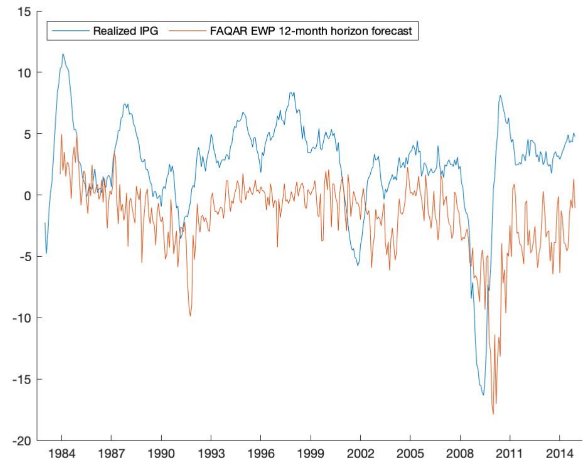

Figure 4: Realized IPG and it's FAQAR EWP $\mathrm{VaR}_{0.05}$ forecast for a 12-month horizon

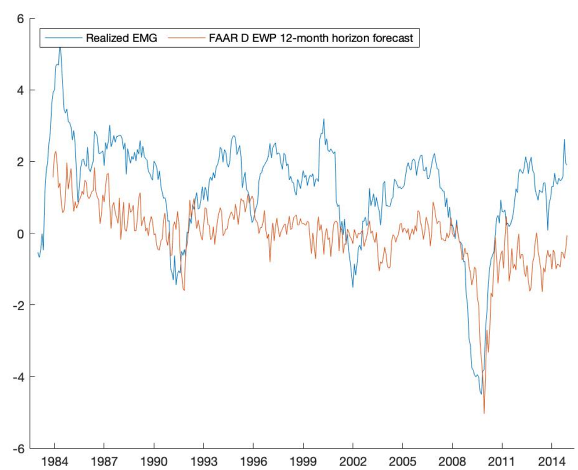

Figure 5: Realized DNFCI and it's FAAR D EWP $\mathrm{VaR}_{0.05}$ forecast for a 12-month horizon

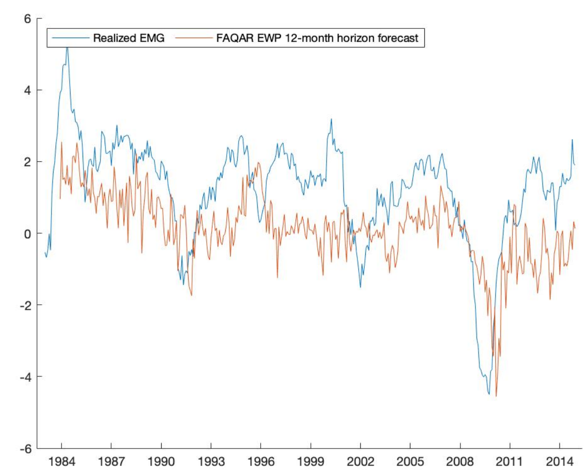

Figure 6: Realized DNFCI and it's QF EWP $\mathrm{VaR}_{0.05}$ forecast for a 12-month horizon

Table A3: Average QWPS for AR models, factor-augmented AR models and their EWPs.

Variable	**IPG**			**EMG**			**CDI**			**BDI**			**DNFCI**		
Model | Horizon	**3m**	**6m**	**12m**	**3m**	**6m**	**12m**	**3m**	**6m**	**12m**	**3m**	**6m**	**12m**	**3m**	**6m**	**12m**
AR D RW	20.87	42.83	82.73	8.06	14.34	28.25	0.488	0.517	0.564	**0.443**	0.496	0.584	5.28	8.45	10.85
AR D EW	21.14	40.87	75.50	**7.65**	**13.09**	**24.18**	**0.465**	**0.483**	**0.513**	**0.442**	**0.479**	**0.539**	6.57	9.42	12.14
AR I RW	19.94	**39.59**	77.70	7.82	**13.46**	25.27	0.480	0.499	0.538	**0.443**	**0.501**	0.583	**4.78**	**6.64**	**8.42**
AR I EW	**19.37**	**37.13**	**69.82**	**7.60**	**13.05**	**24.35**	**0.460**	**0.482**	**0.522**	**0.441**	0.510	0.630	**4.85**	**6.82**	9.37
EWP AR D	20.48	41.00	77.51	7.78	13.50	25.66	**0.474**	**0.495**	0.533	**0.437**	**0.481**	0.551	5.74	8.72	11.05
EWP AR I	**19.33**	**37.76**	72.43	**7.66**	**13.13**	**24.49**	**0.469**	**0.488**	**0.527**	**0.436**	0.498	0.595	**4.67**	**6.62**	8.75
FAAR D AH RW	19.20	39.52	78.47	6.73	11.41	22.99	0.495	0.521	0.576	**0.444**	**0.509**	0.593	5.05	7.34	**9.07**
FAAR D AH EW	**17.79**	**34.90**	**66.32**	6.64	11.41	**22.30**	**0.474**	**0.493**	**0.534**	**0.457**	**0.501**	0.590	6.54	8.84	10.96
FAAR D 5 RW	19.36	40.10	76.14	6.76	11.26	24.45	0.519	0.576	0.562	0.485	0.584	0.637	5.94	8.47	9.73
FAAR D 5 EW	**17.72**	**33.93**	**64.51**	**6.47**	**10.94**	**21.69**	**0.475**	0.502	**0.535**	0.462	**0.511**	0.589	6.90	9.16	11.51
FAVAR I AH RW	19.21	37.57	78.60	8.01	14.35	28.42	0.530	0.576	0.636	0.509	0.591	0.726	**4.61**	**6.64**	**8.97**
FAVAR I AH EW	**18.22**	**34.59**	**67.33**	7.85	13.37	26.19	0.522	0.551	0.586	0.529	0.600	0.712	5.00	**6.48**	**8.34**
FAVAR I 5 RW	20.06	37.74	75.10	8.34	14.51	27.66	0.554	0.622	0.750	0.543	0.635	0.816	**4.91**	6.97	9.60
FAVAR I 5 EW	19.38	36.55	69.88	8.34	14.16	27.21	0.540	0.608	0.762	0.569	0.694	0.907	5.32	7.26	9.85
EWP FAAR D	**17.62**	35.65	68.27	**6.37**	**10.62**	**21.36**	0.482	0.506	**0.527**	**0.448**	**0.500**	**0.565**	5.46	7.67	9.31
EWP FAVAR I	18.61	**35.52**	70.78	8.00	13.89	27.00	0.521	0.571	0.656	0.524	0.612	0.765	**4.46**	**6.30**	**8.65**

The values in yellow indicates this model gets the lowest average QS of that column. Other values in bold in the same column are not significantly different from the best, yellow marked, model according to a DM test at 5% Factor augmented AR models and AR models without factors are considered separately.

Table A4: Average QSs for QAR models, factor-augmented QAR models and their EWPs.

Variable	IPG			EMG			CDI			BDI			DNFCI		
Model | Horizon	**3m**	**6m**	**12m**	**3m**	**6m**	**12m**	**3m**	**6m**	**12m**	**3m**	**6m**	**12m**	**3m**	**6m**	**12m**
QAR RW	0.32	0.72	1.80	0.13	0.27	0.58	**0.0050**	**0.0056**	0.0059	**0.0056**	**0.0059**	0.0082	0.12	0.17	0.22
QAR EW	0.32	0.64	1.29	0.12	0.22	0.47	**0.0050**	**0.0054**	**0.0054**	**0.0051**	**0.0053**	**0.0062**	0.12	0.18	0.26
FAQAR AH RW	**0.26**	**0.59**	1.42	**0.10**	0.23	0.51	0.0066	0.0066	0.0078	0.0064	0.0067	**0.0070**	**0.07**	**0.11**	0.17
FAQAR AH EW	**0.28**	**0.52**	**1.02**	**0.11**	**0.19**	**0.40**	0.0059	0.0063	0.0070	0.0061	0.0069	0.0082	**0.08**	**0.11**	**0.16**
FAQAR 5 RW	0.28	0.65	1.45	0.11	0.27	0.66	0.0065	0.0077	0.0087	0.0065	0.0077	**0.0071**	**0.08**	**0.12**	0.18
FAQAR 5 EW	0.29	**0.58**	1.18	0.10	**0.16**	**0.41**	0.0059	0.0060	**0.0059**	0.0065	**0.0052**	0.0075	0.09	0.14	0.17
EWP QAR	0.29	0.62	1.31	0.12	0.23	0.49	**0.0050**	**0.0054**	0.0056	**0.0052**	**0.0055**	**0.0070**	0.11	0.16	0.21
EWP FAQAR	**0.25**	**0.52**	1.15	**0.09**	0.19	0.44	0.0060	0.0062	0.0066	0.0058	0.0059	**0.0067**	**0.07**	**0.11**	**0.14**

The values in yellow indicates that this model gets the lowest average QS of that column. Other values in bold in the same column are not significantly different from the best, yellow marked, model according to a DM test at 5%. QAR and factor-augmented QAR models are considered together.

Table A5: Best EWPs versus individual models Quantile Projections

(a) Best EWP versus best individual model

	Horizon		
Variable	**3 months**	**6 months**	**12 months**
IPG	0.95	1.00	**1.13**
EMG	0.93	**1.16**	**1.12**
CDI	1.00	1.01	**1.04**
BDI	1.02	1.06	1.08
DNFCI	0.95	0.95	0.88

(b) Worst EWP versus worst individual model

	Horizon		
Variable	**3 months**	**6 months**	**12 months**
IPG	**0.92**	**0.86**	**0.72**
EMG	**0.91**	**0.86**	**0.75**
CDI	**0.91**	**0.81**	**0.76**
BDI	0.90	**0.76**	**0.86**
DNFCI	0.92	0.93	**0.83**

Bold ratios indicate a rejection of equal performance at 5%

Table A6: Ratios average QSs of best (FA)QAR EWP over average QSs of best (FA(V))AR EWP

	Horizon		
Variable	**3 months**	**6 months**	**12 months**
IPG	1.00	1.02	1.01
EMG	**1.16**	**1.27**	**1.31**
CDI	**0.89**	0.98	1.02
BDI	1.00	0.98	**1.26**
DNFCI	**1.22**	**1.20**	1.11

Bold ratios indicate a rejection of equal performance at 5%

Table A7: Coverage ratios of the best (factor-augmented) AR EWP and the best (factor-augmented) QAR EWP for each variable.

	Model	**FA(V)AR EWP**			**FAQAR EWP**		
Variable	**Sample | Horizon**	**3m**	**6m**	**12m**	**3m**	**6m**	**12m**
IPG	1984M01 - 2014M12	0.03	0.03	0.06	0.05	0.08	0.10
	2007M01 - 2014M12	0.05	0.09	0.13	0.07	0.09	0.14
EMG	1984M01 - 2014M12	0.06	0.08	0.12	0.09	0.11	0.17
	2007M01 - 2014M12	0.09	0.15	0.18	0.16	0.17	0.28
CDI	1984M01 - 2014M14	0.06	0.07	0.06	0.06	0.08	0.08
	2007M01 - 2014M12	0.08	0.10	0.08	0.08	0.16	0.16
BDI	1984M01 - 2014M12	0.04	0.04	0.05	0.09	0.09	0.17
	2007M01 - 2014M12	0.06	0.06	0.11	0.13	0.18	0.26
DNFCI	1984M01 - 2013M12	0.03	0.03	0.03	0.02	0.05	0.06
	2007M01 - 2013M12	0.07	0.07	0.07	0.08	0.13	0.19

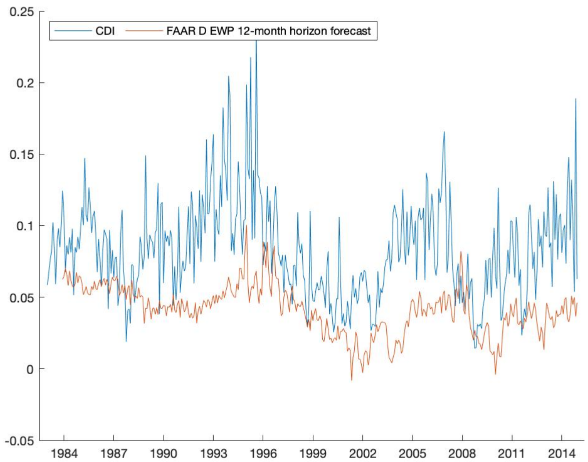

Figure 7: CDI and it's FAAR D EWP $\mathrm{VaR}_{0.05}$ forecast for a 12-month horizon

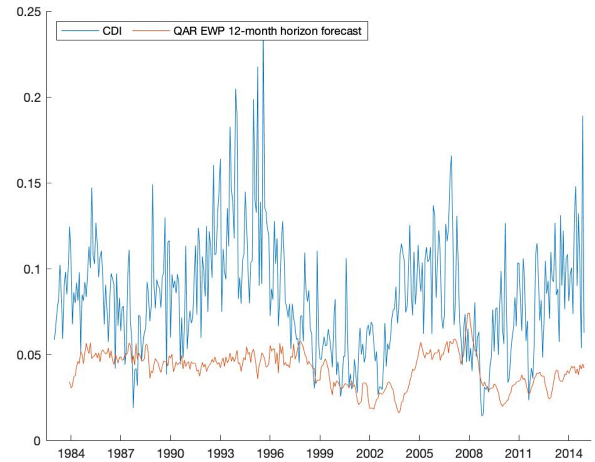

Figure 8: CDI and it's QAR EWP $\mathrm{VaR}_{0.05}$ forecast for a 12-month horizon

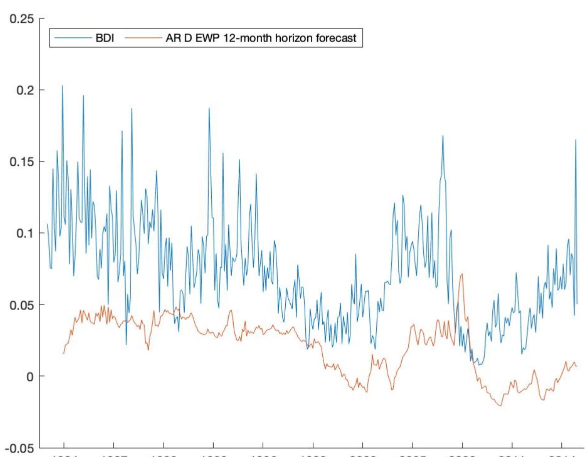

Figure 9: BDI and it's AR D EWP $\mathrm{VaR}_{0.05}$ forecast for a 12-month horizon

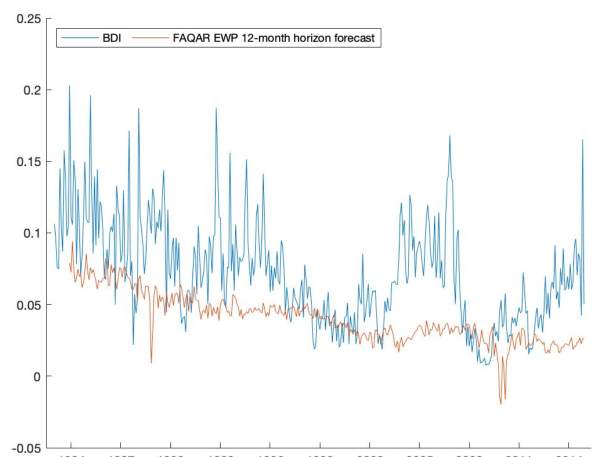

Figure 10: BDI and it's QAR EWP $\mathrm{VaR}_{0.05}$ forecast for a 12-month horizon

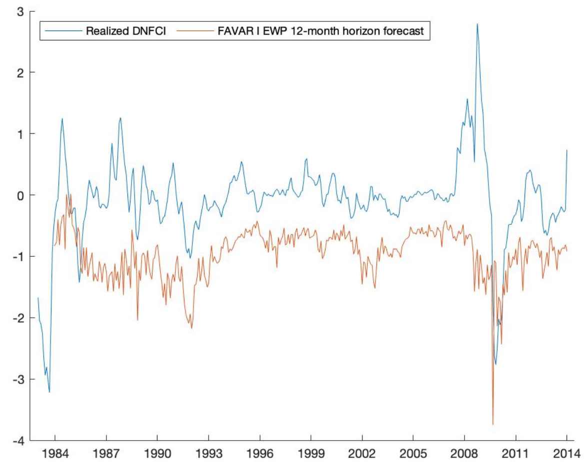

Figure 11: Realized DNFCI and it's FAVAR I EWP $\mathrm{VaR}_{0.05}$ forecast for a 12-month horizon

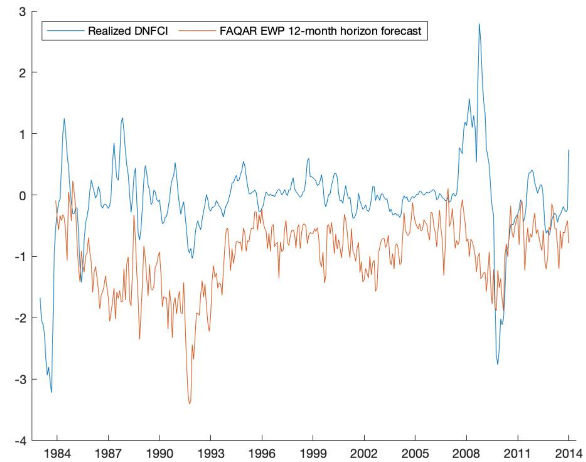

Figure 12: Realized DNFCI and it's FAQAR EWP $\mathrm{VaR}_{0.05}$ forecast for a 12-month horizon

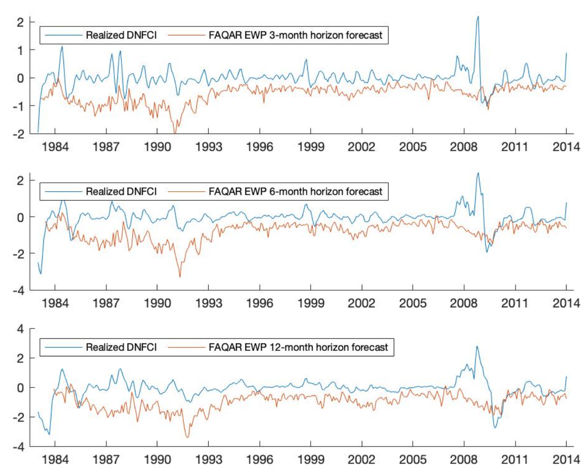

Figure 13: Realized DNFCI and it's 3-, 6- and 12- month FAQAR EWP $\mathrm{VaR}_0.05$ forecast

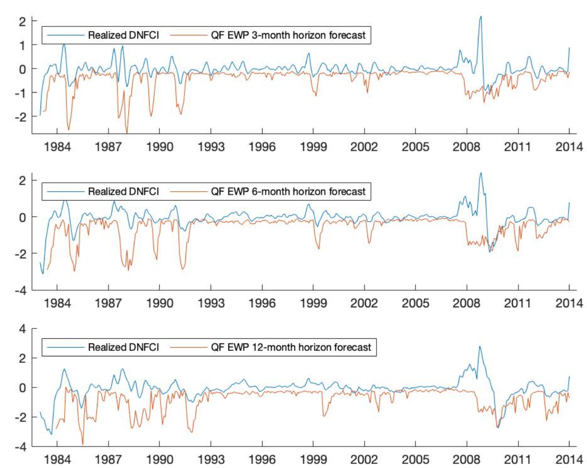

Figure 14: Realized DNFCI and it's 3-, 6- and 12- month QF EWP $\mathrm{VaR}_0.05$ forecast

Table A8: Average left tail QWPSs for PLS factor-augmented AR models and their EWPs

Variable	IPG			EMG			CDI			BDI			DNFCI		
Model \ Horizon	**3m**	**6m**	**12m**	**3m**	**6m**	**12m**	**3m**	**6m**	**12m**	**3m**	**6m**	**12m**	**3m**	**6m**	**12m**
FAAR D AH RW	20.36	43.42	85.39	**6.96**	12.28	25.36	0.531	0.581	0.611	**0.529**	0.586	0.688	5.53	8.16	10.07
FAAR D AH EW	**18.17**	**35.12**	**67.36**	**6.97**	11.93	**21.71**	**0.475**	**0.498**	**0.521**	**0.464**	**0.505**	**0.568**	6.62	9.60	12.20
FAAR D 5 RW	19.53	39.48	79.29	**7.01**	**11.96**	26.81	0.534	0.603	0.607	0.541	0.614	0.657	6.08	8.33	11.01
FAAR D 5 EW	**18.06**	**34.24**	**66.14**	7.03	11.88	**21.89**	**0.474**	**0.502**	**0.522**	**0.468**	**0.511**	**0.578**	6.72	9.48	11.51
FAVAR I AH RW	20.11	39.19	80.49	8.45	15.40	41.11	0.531	0.570	0.610	0.501	0.588	0.679	**4.87**	7.21	**9.79**
FAVAR I AH EW	18.67	**35.50**	74.23	8.09	14.44	29.13	0.523	0.557	0.592	0.514	0.579	0.678	**4.86**	6.85	9.23
FAVAR I 5 RW	20.66	41.05	83.45	8.82	16.11	30.62	0.577	0.640	0.780	0.536	0.628	0.788	**4.68**	**6.25**	**8.31**
FAVAR I 5 EW	**18.68**	37.06	76.82	8.67	15.82	31.66	0.564	0.635	0.792	0.560	0.671	0.871	5.19	7.20	9.25
EWP FAAR D	**17.79**	**35.13**	**70.54**	**6.73**	**11.26**	**22.17**	**0.492**	0.533	0.548	**0.482**	**0.530**	**0.594**	5.86	8.36	10.24
EWP FAVAR I	18.89	37.27	77.94	8.32	15.22	32.78	0.537	0.585	0.673	0.513	0.595	0.729	**4.54**	**6.47**	**8.67**

The values in yellow indicates this model gets the lowest average QWPS of that column. Other values in bold in the same column are not significantly different from the best, yellow marked, model according to a DM test at 5%.

Table A9: Average QSs of PLS factor-augmented QAR models and the EWP.

Variable	IPG			EMG			CDI			BDI			DNFCI		
Model \ Horizon	**3m**	**6m**	**12m**	**3m**	**6m**	**12m**	**3m**	**6m**	**12m**	**3m**	**6m**	**12m**	**3m**	**6m**	**12m**
FAQAR AH RW	0.34	0.68	1.58	0.10	0.27	0.54	0.0062	0.0066	0.0066	0.0062	0.0074	0.0087	**0.11**	**0.15**	**0.20**
FAQAR AH EW	**0.29**	**0.51**	**0.97**	0.10	**0.17**	**0.35**	**0.0056**	**0.0059**	**0.0062**	**0.0053**	**0.0062**	**0.0070**	0.12	0.19	0.25
FAQAR 5 RW	0.33	0.65	1.70	**0.09**	0.23	0.60	0.0069	0.0064	0.0067	0.0061	**0.0066**	0.0085	0.10	**0.16**	**0.19**
FAQAR 5 EW	**0.29**	**0.52**	**1.03**	0.11	**0.19**	0.41	**0.0056**	**0.0058**	**0.0061**	**0.0056**	**0.0058**	**0.0072**	0.12	0.17	**0.18**
EWP FAQAR	**0.28**	**0.49**	1.22	**0.08**	**0.17**	0.41	**0.0058**	**0.0058**	**0.0060**	**0.0052**	**0.0060**	**0.0072**	**0.10**	**0.14**	**0.16**

The values in yellow indicates this model gets the lowest average QS of that column. Other values in bold in the same column are not significantly different from the best, yellow marked, model according to a DM test at 5%.

Table A10: Average QSs for Quantile Forest, factor-augmented Quantile Forest models and their EWPs

Variable	**IPG**			**EMG**			**CDI**			**BDI**			**DNFCI**		
Model \ Horizon	**3m**	**6m**	**12m**	**3m**	**6m**	**12m**	**3m**	**6m**	**12m**	**3m**	**6m**	**12m**	**3m**	**6m**	**12m**
QF RW	0.35	0.84	1.78	0.12	0.24	0.58	**0.0056**	**0.0061**	**0.0061**	**0.0057**	**0.0064**	**0.0072**	0.063	0.116	0.173
QF EW	0.36	0.66	**1.13**	**0.11**	0.25	0.49	**0.0054**	**0.0060**	**0.0063**	**0.0053**	**0.0062**	**0.0069**	0.052	**0.072**	**0.095**
FAQF AH RW	0.33	0.69	1.60	0.12	0.24	0.55	0.0059	0.0065	**0.0064**	**0.0057**	**0.0068**	**0.0072**	**0.055**	0.116	0.179
FAQF AH EW	**0.24**	**0.51**	**1.05**	**0.10**	**0.18**	**0.41**	0.0060	**0.0063**	0.0067	**0.0056**	**0.0064**	0.0074	**0.052**	**0.075**	**0.097**
FAQF 5 RW	0.33	0.74	1.58	**0.12**	0.23	0.55	0.0060	**0.0062**	**0.0064**	**0.0057**	**0.0069**	**0.0069**	0.064	0.128	0.174
FAQF 5 EW	**0.25**	**0.51**	**1.02**	**0.11**	**0.19**	**0.44**	0.0062	**0.0064**	**0.0064**	**0.0057**	**0.0066**	0.0076	0.059	0.091	0.117
EWP QF	0.33	0.66	1.29	**0.11**	0.23	0.51	**0.0054**	**0.0058**	**0.0059**	**0.0053**	**0.0062**	**0.0069**	**0.047**	**0.071**	**0.097**
EWP FAQF	**0.26**	**0.57**	1.21	**0.10**	**0.19**	**0.44**	**0.0057**	**0.0061**	**0.0061**	**0.0054**	**0.0063**	**0.0068**	**0.048**	**0.072**	**0.103**

The values in yellow indicates this model gets the lowest average QS of that column. Other values in bold in the same column are not significantly different from the best, yellow marked, model according to a DM test at 5%. The QF models and factor-augmented QF models are considered together.

Table A11: QSs for PLS factor-augmented Quantile Forest models and the EWP.

Variable	**IPG**			**EMG**			**CDI**			**BDI**			**DNFCI**		
Model \ Horizon	**3m**	**6m**	**12m**	**3m**	**6m**	**12m**	**3m**	**6m**	**12m**	**3m**	**6m**	**12m**	**3m**	**6m**	**12m**
FAQF AH RW	0.34	0.78	1.78	**0.11**	0.22	0.52	**0.0055**	**0.0059**	**0.0062**	0.0059	0.0069	**0.0067**	**0.068**	0.120	0.183
FAQF AH EW	**0.25**	**0.49**	**1.09**	**0.11**	**0.19**	**0.43**	**0.0056**	0.0064	**0.0064**	**0.0056**	**0.0061**	**0.0074**	**0.056**	**0.084**	**0.113**
FAQF 5 RW	0.34	0.79	1.79	0.12	0.23	0.55	0.0059	**0.0062**	**0.0063**	**0.0057**	**0.0063**	**0.0071**	**0.068**	0.124	0.189
FAQF 5 EW	**0.25**	**0.45**	**1.04**	**0.10**	**0.19**	**0.44**	**0.0056**	**0.0062**	0.0065	**0.0055**	0.0073	**0.0074**	0.071	0.102	0.124
EWP FAQF	**0.24**	**0.51**	1.29	**0.10**	**0.18**	**0.45**	**0.0054**	**0.0059**	**0.0061**	**0.0053**	**0.0062**	**0.0067**	**0.057**	**0.082**	**0.105**

The values in yellow indicates this model get the lowest average QS of that column. Other values in bold in the same column are not significantly different from the best, yellow marked, model according to a DM test at 5%