

## FORECASTING TAIL RISKS

GIANNI DE NICOLÒ<sup>a,b,\*</sup> AND MARCELLA LUCCHETTA<sup>c</sup>

<sup>a</sup> *International Monetary Fund, Washington, DC, USA*

<sup>b</sup> *CEifo, Munich, Germany*

<sup>c</sup> *University of Venice, Italy*

### SUMMARY

This paper presents an early warning system as a set of multi-period forecasts of indicators of tail real and financial risks obtained using a large database of monthly US data for the period 1972:1–2014:12. Pseudo-real-time forecasts are generated from: (a) sets of autoregressive and factor-augmented vector autoregressions (VARs), and (b) sets of autoregressive and factor-augmented quantile projections. Our key finding is that forecasts obtained with AR and factor-augmented VAR forecasts significantly underestimate tail risks, while quantile projections deliver fairly accurate forecasts and reliable early warning signals for tail real and financial risks up to a 1-year horizon. Copyright © 2016 John Wiley & Sons, Ltd.

*Received 18 March 2013; Revised 20 October 2015*



*Supporting information may be found in the online version of this article.*

### 1. INTRODUCTION

The 2007–2009 financial crisis and its adverse impact on real activity have spurred renewed efforts in modeling adverse ‘tail’ events in both the financial and real sectors. Bisias *et al.* (2012) provide an extensive survey of the models currently available to measure and track indicators of tail (systemic) financial risk. Yet most of these models focus exclusively on vulnerabilities in the financial system or some of its components, with no assessment of either their impact on real activity, or on how vulnerabilities in the real sector may affect the financial sector. Most importantly, the out-of-sample forecasting power of many of the proposed measures is seldom assessed, making it difficult to gauge their usefulness as early warning signals. Tail real risks are also the focus of an important theoretical literature—briefly reviewed by Acemoglu *et al.* (2015)—which aims at explaining how aggregate tail real risks can arise from a variety of shock configurations at disaggregated levels of an economy. To the best of our knowledge, however, this literature has not tackled the issue of forecasting tail risks. Operationally, reliable early warning signals for tail real and financial risks—where reliability is defined as the ability of a model to issue signals with relatively small percentages of missed realizations of future adverse extreme events—are essential for timely implementation of macroeconomic and macroprudential policies.

Extending our previous work (De Nicolò and Lucchetta, 2012, 2013), this paper develops an *early warning system* (EWS) as a set of multi-period forecasts of indicators of *tail real and financial risks*. Our analysis has three objectives. First, we compare the forecasting performance of *multi-period* forecasts of indicators of tail real and financial risks obtained using two sets of forecasting models: autoregressions and factor-augmented autoregressions, categorized as AR models, and autoregressive and factor-augmented quantile projections (QPs henceforth). We wish to assess the comparative forecasting performance for tail risks of workhorse AR models *vis-à-vis* QPs. As pointed out by Komunjer (2013), a potential advantage of QPs is that they do not require assumptions about the underlying distribution of a variable to be forecast, this potentially permitting to capture any type of asymmetry.

\* Correspondence to: Gianni De Nicolò, International Monetary Fund, Washington, DC, USA. E-mail: gdenicolo@imf.org

Second, we assess the forecasting performance of equally weighted pools of forecasts obtained under AR models and QPs to gauge whether the superiority of simple pooled forecasts for first moments documented in Geweke and Amisano (2011) extends to the tails of a predicted distribution or its quantiles. Lastly, we gauge the out-of sample forecasting accuracy of tail risk predictions under each model using scoring rules which place either heavier weight on the tails of interest (for the AR models) or quantile scores (for both AR models and QPs). In essence, we aim at identifying models that can deliver reliable early warning signals of tail real and financial risks.

We construct measures of tail risks following a standard risk management approach. Tail *real* risks are measured by the value at-risk (VaR) of two standard aggregate macroeconomic variables: industrial production growth and employment growth. Tail *financial* risk in the *corporate* and *banking* sectors are measured by the VaR of a 'portfolio' version of the *distance to insolvency* measure introduced by Atkenson *et al.* (2013), which is based on a large class of theoretical structural models, and is germane to other theory-based indicators of tail financial risk used in recent studies (see, for example, Acharya *et al.*, 2012; Brownlee and Engle, 2011). Tail risks in *financial markets* are measured by the VaR of changes in a financial condition index.

We implement our EWS using a large set of monthly US data for the period 1973:2–2014:12. Estimation and forecasting are conducted using both a moving and an expanding window of data: the moving window estimation is used to account for time variation in parameters and possible structural breaks, while the estimation based on the longer expanding window provides us a forecasting 'hedge' against possible imprecision of parameter estimation under the shorter moving window. For each variable underlying our tail risk measures, we compute multi-period density and VaR forecasts at 3-month, 6-month and 12-month horizons, and compare their accuracy using the quantile weighted probability score (QWPS) for AR models, and the quantile scores (for both AR models and QPs) introduced by Gneiting and Ranjan (2011).

Our analysis delivers three main results. First, factor-augmented models deliver density and VaR forecasts significantly more accurate than, or at least as accurate as, those of AR models and QPs without factors. This result extends the finding of superior predictive ability of factor models with many predictors (see, for example, Stock and Watson, 2006) to density and VaR forecasts at multiple forecasting horizons.

Second, density and VaR forecasts of equally weighted pools (EWP henceforth) of both AR models and QPs are significantly more accurate than, or at least as accurate as, those obtained from each model in the pool. This result extends the finding of superior predictive ability of EWPs documented in Geweke and Amisano (2011, 2012) to tails and quantile forecasts at multiple forecasting horizons as well.

Our third result is perhaps the most important operationally, since it involves an assessment of the ability of our tail risk forecasts to serve as reliable early warning signals. In this regard, we have both bad news and good news for risk-averse policymakers.

The bad news is that the AR models deliver VaR forecasts that *significantly underestimate tail risks*. Furthermore, their accuracy decreases substantially for the subsample that includes the recent financial crisis, this implying that their reliability falls when it is needed most. The failure of this class of models to issue reliable early warning signals for tail risks is due to their inability to capture asymmetric and time-varying changes in the shapes of the tails of distributions. Policymakers may be particularly concerned about this result, since forecasts of this class of model (as well as their DSGE versions) are often used in central banks and international organizations as inputs for stress-testing purposes.

The good news is that forecasts obtained with factor-augmented QPs are significantly superior to those of the AR models and deliver fairly reliable early warning signals for horizons up to 1 year ahead. Importantly, their reliability is broadly preserved for the subsample that includes the recent financial crisis, indicating that their reliability as early warning signals does not drop significantly when it is most needed. In sum, factor-augmented QPs seem to capture those asymmetric changes in the shape of a distribution that may result in significant future changes in its tails.

The remainder of the paper is composed of four sections. Section 2 defines our tail risk measures and details estimation and choice of factors as predictors. Section 3 describes the forecasting models, their estimation and the evaluation of their forecasting accuracy. Section 4 describes the results and Section 5 concludes. An online Appendix (supporting information) details a number of supporting results as well as data and their sources.

## 2. TAIL RISK MEASURES AND FACTORS AS PREDICTORS

### 2.1. Tail Risk Measures

Our tail risk measures are  $\text{VaR}_\alpha$ s of indicators of real activity and financial stress, with probability levels  $\alpha$  set equal to 5% and 10% (i.e.  $\alpha \in \{0.05, 0.10\}$ ). Tail *real* risks are measured by the  $\text{VaR}_\alpha$  of the (log) change in the industrial production index IPG, denoted by  $\text{VaR}_\alpha$  (IPG) (also called *industrial production-at-risk*), and the  $\text{VaR}_\alpha$  of the (log) change in total employment EMG, denoted by  $\text{VaR}_\alpha$  (EMG) (also called *employment-at-risk*).

Tail *financial* risks in the corporate and banking sectors are measured by the  $\text{VaR}_\alpha$  of a portfolio version of the distance-to-insolvency (DI) introduced by Atkeson *et al.* (2013), who show that: (i) DI is a measure of the adequacy of a firm's equity cushion relative to the volatility of the value of its assets based on Leland's (1994) structural model of credit risk; (ii) it is proxied by the reciprocal of its estimated instantaneous equity volatility; and (iii) it tracks closely measures of default risk derived from a wide range of structural models of firm valuation, such as the distance-to-default. In our implementation, we compute the DI of value-weighted portfolios including all firms in the DataStream equity indexes of non-financial firms and banks, denoted by CDI and BDI respectively. These portfolios represent large portions of the corporate and banking sectors, the latter including all banks considered 'systemically important'. Thus a 'portfolio' DI is a lower bound of the probability of insolvency of these two sectors, as profits and losses of each firm in the portfolio are evened out. As in Atkeson *et al.* (2013), a proxy measure of the instantaneous equity return volatility is obtained by monthly averaging daily squared returns. Tail risk in the corporate sector is measured by the  $\text{VaR}_\alpha$  of the DI of the portfolio of non-financial firms, denoted by  $\text{VaR}_\alpha$  (CDI) (also called *corporate sector-at-risk*). Tail risk in the banking sector is measured by the  $\text{VaR}_\alpha$  of the DI of the portfolio of banks, denoted by  $\text{VaR}_\alpha$  (BDI) (also called *banking sector-at-risk*).

Our measure of tail risks in *financial markets* is based on the National Financial Condition Index (NFCI) produced by the Federal Reserve Bank of Chicago. Brave and Butters (2011) document its construction, obtained as a weighted average of more than 100 standardized indicators of risk, credit and leverage in the financial system. They show that it captures well-known periods of financial stress. Financial conditions indexes—initially studied by Hatzius *et al.* (2010) in the aftermath of the 2007–2009 financial crisis, and later produced in several central banks and international organizations—have been typically designed either to measure whether broad financial conditions are loose or tight by historical standards, thus serving as coincident indicators, or to assess whether the financial system experiences historically unusual stress, therefore serving as early warning indicators. As documented in Aramonte *et al.* (2013), however, the existing evidence on whether financial conditions indexes are useful as coincident or early warning indicators is mixed. In our study, we measure tail risk in financial markets by the  $\text{VaR}_\alpha$  of the negative first differences in the NFCI<sup>1</sup> (DNFCI), denoted by  $\text{VaR}_\alpha$  (DNFCI) (also called *financial markets-at-risk*).

Table A1 in the online Appendix reports descriptive statistics of the series underlying our tail risk indicators and their correlation matrix. As may be expected, there is a negative and significant

<sup>1</sup> Positive (negative) values of the NFCI indicate financial conditions that are tighter (looser). We take the negative of NFCI first difference to preserve the negative orientation of all our other tail risk measures.

contemporaneous correlation between indicators of financial risk and real activity. In other words, measured risk in financial institutions and markets is counter-cyclical.

## 2.2. Factors as Predictors

Factors used as predictors in all models are estimated by principal components (PCA factors) obtained as the solution of a maximization problem subject to a factor loading constraint, as detailed in Stock and Watson (2011), and implemented by Stock and Watson (2002, 2012) and Pesaran *et al.* (2011). The dataset from which PCA factors are extracted includes 164 monthly time series of the US economy for the period 1973:1–2014:12 taken from the FRED-MD database constructed by McCracken and Ng (2014) and from DataStream.<sup>2</sup> As shown in Table A2 of the online Appendix, the distribution of the series by group is fairly comprehensive and balanced.<sup>3</sup>

To assess whether tail forecasts differ significantly when a different number of factors are used as predictors, we consider models with the number of factors chosen according to a selection criterion, and models with five factors. As a selection criterion, we use the eigenvalue ratio (ER) criterion of Ahn and Horenstein (2013) (AH henceforth), which selects the number of factors that maximize the ratio of two adjacent eigenvalues arranged in descending order. For the entire time range of our dataset, this criterion delivers three factors which explain 0.34% of the total variation in the data.<sup>4</sup> The choice of a maximum number of five factors is motivated by McCracken and Ng's (2014) evidence on the decrease of explanatory power of factors after the estimated fifth, and the use of five factors by Stock and Watson (2012) as a benchmark. For the entire time range of our dataset, five factors explain 0.43% of the total variation in the data. As shown in Table A3 of the online Appendix, the explanatory power of factors in contemporaneous regressions of the variables underlying our measures of tail risk on factors and AR terms is anything but trivial:  $R^2$ 's with five factors range from 0.28 to 0.76, and the addition of five AR terms in each equation yields  $R^2$ 's ranging from 0.56 to 0.76.

## 3. FORECASTING MODELS

We consider tail forecasts obtained by: (a) a set of direct AR forecasts; (b) a set of iterated AR and factor-augmented VAR forecasts (FAVAR) with linear generalized autoregressive conditional heteroskedasticity (GARCH) volatility; (c) a set of QPs with no factors as well as factor-augmented specified similarly to direct AR forecasts; (d) equally weighted pools (EWPs) of direct and iterated AR and FAVAR models specified with and without factors; (e) equally weighted pools of QP models specified with and without factors.<sup>5</sup>

The rationales underlying our choice of these models are as follows. First, we estimate both direct and iterated AR and FAVAR models since the empirical findings of Marcellino *et al.* (2006) and Pesaran *et al.* (2011) indicate that direct or iterated forecasts may be preferable depending on the series considered and the forecasting horizon. Second, the introduction of a linear GARCH specification in

<sup>2</sup> As our focus is on comparing forecasts obtained from different models, we use data of the 2014 vintage, as in McCracken and Ng (2014). The counting of series in our dataset includes the NFCI series taken from the Federal Reserve Bank of Chicago website.

<sup>3</sup> Descriptions of the series and the relevant transformations used to ensure their stationarity are detailed in the online Appendix.

<sup>4</sup> AH also propose a second criterion (GR), which selects the number of factors that maximizes the rate of growth of adjacent eigenvalues arranged in descending order. Our results using GR are identical to those using ER. Note that the widely used Bai and Ng (2002) criteria typically select between seven and nine factors. In the FRED-MD dataset, which is a subset of our dataset in the chosen range, McCracken and Ng (2014) find eight factors selected according to the Bai and Ng (2002) PCP2 criterion. However, they observe that the incremental explanatory power of PCA factors declines significantly moving from five to eight factors, which may negatively affect forecasts through increased estimation error, as noted by Boivin and Ng (2006).

<sup>5</sup> To the best of our knowledge, this study is the first to consider factor-augmented quantile predictions in the mold of the literature of forecasting with many predictors.

the iterated AR and FAVAR forecasts is instrumental to accounting for time-varying volatility, which is a common feature of the series whose VaR we wish to predict. This specification is also supported by the superiority of iterated forecasts for second moments over constant specifications documented in Ghysels *et al.* (2009), allowing us to assess whether a time-varying variance specification can account for the differential behavior in the tails of real variables noted by Acemoglu *et al.* (2015). Third, we consider EWP forecasts as we wish to gauge whether the superior forecasting power over component models in the context of density forecasts based on equally weighted pools documented by Geweke and Amisano (2011, 2012) carries over to tail forecasts. Fourth, each factor-augmented model is estimated with AH factors and with five factors with the purpose of assessing whether forecasts differ significantly depending on the number of factors. Lastly, we used two window-based forecasting schemes for all models to account for time variation in parameters and possible structural breaks. The first window is a rolling window of 120 months, while the second one is an expanding window starting with the first estimation period of 120 months (1973:2–1984:1), and adding one observation sequentially at each forecasting date.

In all models described next, we denote by  $Y_t \equiv \{IPG_t, EMG_t, BDI_t, CDI_t, DNFCI_t\}$  the set of variables we wish to forecast, with  $y_t$  an element of  $Y_t$ , and with  $F_t$  a vector of PCA factors.

### 3.1. AR Models

#### 3.1.1. Direct AR Forecasts

The model generating direct forecasts is given by

$$y_{t+h} = A + B(L)F_t + C(L)y_t + \varepsilon_{t+h} \tag{1}$$

where estimated factors  $F_t$  and lagged values of  $y_t$  are the predictors of  $y_{t+h}$  at each forecasting horizon ( $h = 3, 6$  and  $12$ ). The  $\text{VaR}_\alpha$  of  $y_{t+h}$  is given by

$$\text{VaR}_\alpha(y_{t+h}) = Q_\alpha(y_{t+h}) = \hat{y}_{t+h} + \hat{\sigma}_t F^{-1}(\alpha) \tag{2}$$

where the ‘hat’ in equation (2) denotes estimated parameters,  $F^{-1}(\alpha)$  is the inverse Gaussian cumulative distribution function (cdf) and  $\hat{\sigma}_t$  is estimated within the forecasting sample.

#### 3.1.2. Iterated Forecasts with Time-Varying Volatility (FAVAR)

For each  $y_t \in Y_t$ , the specification we use is a FAVAR given by

$$\begin{bmatrix} F_t \\ y_t \end{bmatrix} = \begin{pmatrix} A(L) & B(L) \\ a(L) & b(L) \end{pmatrix} \begin{bmatrix} F_{t-1} \\ y_{t-1} \end{bmatrix} + \begin{bmatrix} \eta_t \\ u_{yt} \end{bmatrix} \tag{3}$$

$$\begin{aligned} u_{yt} &= \sigma_{yt} \varepsilon_{yt} \\ \sigma_{yt} &= a + b\sigma_{yt-1} + c|u_{yt-1}| \end{aligned} \tag{4}$$

Note that in equation (3) we do not impose  $B(L)=0$ , which is a restriction that has been applied in many forecasting exercises in the literature (see, for example, Pesaran *et al.*, 2011). Stock and Watson (2005) found that this restriction was rejected in their FAVAR version of an approximate dynamic factor model, but its impact was not quantitatively significant. However, this is not the case for our dataset and the variables underlying our tail risk measures, as documented in Table A4 of the online Appendix.

Equation (4) describes a linear GARCH(1,1) process for  $y_t$ , where  $\varepsilon_{y_t}$  are assumed to be i.i.d. 0-mean random variables distributed with a Gaussian cdf, and  $\sigma_{y_t}$  is the conditional standard deviation. This specification is supported by standard ARCH tests, which reveal that the null of absence of time variation in second moments is rejected for all variables except EMG. Denote the mean and volatility forecasts of  $y$  at horizon  $h \geq 1$  by  $\hat{y}_{t+h}$  and  $\hat{\sigma}_{t+h}$  respectively. Since IPG and EMG are expressed in percentage changes and DNFCI is expressed in first differences, their multi-period mean forecasts at horizon  $h$  are given by  $\hat{y}_{t+h} = \sum_{i=1}^h \hat{y}_{t+i}$ . Following Ghysels *et al.* (2009) and Andersen *et al.* (2010), the multi-period volatility forecasts at horizon  $h$  are proxied by the expected quadratic variation, given by  $\hat{\sigma}_{t+h} = \sum_{i=1}^h \hat{\sigma}_{t+i}$ , where each term of this sum is the forward iteration of the linear GARCH equation. As the variables CDI and BDI are in levels, the relevant forecasts are just the iterated forecast values  $h$  periods ahead. Therefore, the VaR of  $y_{t+h}$  at probability level  $\alpha \in (0, 1)$  is given by

$$\text{VaR}_\alpha(y_{t+h}) = Q_\alpha(y_{t+h}) = \hat{y}_{t+h} + \hat{\sigma}_{t+h} F^{-1}(\alpha) \quad (5)$$

where  $F^{-1}(\alpha)$  is the inverse Gaussian cdf.

### 3.2. Quantile Projections (QPs)

Let  $P(y_{t+h}|X_t)$  denote the distribution of  $y_{t+h}$  conditional on  $X_t$ . For a given  $\alpha \in (0, 1)$ , a conditional quantile projection  $q_\alpha(X)$  satisfies  $P(y_{t+h} \leq q_\alpha(X) | X_t = X) = \alpha$ , where  $q_\alpha(X)$  is monotonically increasing in  $\alpha \in (0, 1)$ . Defining  $X_t = (1, F_t, \dots, y_t, y_{t-1}, \dots)$  and assuming a linear quantile function  $q_\alpha(X) = X' \beta(\alpha)$ , the parameter vector  $\beta(\alpha)$  minimizes  $\frac{1}{t} \sum_{i=0}^{t-1} \rho_\alpha(y_{t+k} - X_t' \beta(\alpha))$ , where  $\rho_\alpha(u) = u(\alpha - I(u < 0))$  denotes a ‘check’ function defined for any scalar  $u$ . The quantile projection forecast of  $y_{t+h}$  is given by

$$\text{VaR}_\alpha(y_{t+h}) = X_t' \hat{\beta}(\alpha) \quad (6)$$

where the ‘hat’ denotes the estimated parameters of the linear quantile function.

### 4.4. Model Specifications and Forecast Evaluation

Table A5 of the On-line Appendix reports the specifications of all models. For each  $y_t \in Y_t$ , we estimate 12 individual and four EWP AR models, and six individual and two EWP QPs. Models differ according to the use of direct or iterated forecasts (marked with I or D respectively), whether the number of factors is determined by the AH criterion or is set equal to five, and whether estimation is carried out using an expanding window or a rolling window (marked EW and RW respectively). The specifications for the volatility and AR lags for each model are reported in the relevant columns of the table.

Following the density forecast literature (see, for example, Corradi and Swanson, 2006), we compared the accuracy of tail forecasts generated by these models using the quantile-weighted probability score (QWPS) proposed by Gneiting and Ranjan (2011). The QWPS allows us to compare the accuracy of density forecasts with reference to particular regions of a distribution, such as its tails, as well as the predictive accuracy of specific quantiles (or VaRs). The (continuous) QWPS, denoted by  $\text{QWPS}(f, y)$ , is defined by

$$\text{QWPS}(f, y) = \int_0^1 QS(F^{-1}(\alpha), y, \alpha) w(\alpha) d\alpha \quad (7)$$

where  $f$  denotes a density forecast,  $y$  is the realization of the forecast variable,  $F$  is the cdf corresponding to the density  $f$ ,  $F^{-1}(\alpha)$  the *predicted* quantile at level  $\alpha \in (0, 1)$ ,  $w(\alpha)$  is a non-negative weighting function on the unit interval and  $QS(F^{-1}(\alpha), y, \alpha)$  is the *score* of quantile  $\alpha$  given by

$$QS(F^{-1}(\alpha), y, \alpha) \equiv 2(I\{y \leq F^{-1}(\alpha)\} - \alpha)(F^{-1}(\alpha) - y) \quad (8)$$

where  $I\{\cdot\}$  is an indicator function. The QWPS and the quantile score have negative orientation, with lower values indicating better performance.<sup>6</sup>

To evaluate the performance of density forecasts of the AR models over the left tail, we constructed a discretized version of left tail QWPS setting  $w(\alpha) = (1 - \alpha)^2$ , as suggested by Gneiting and Ranjan (2011), using a grid of 98 quantiles (from 1 to 99). If for a given  $\alpha \in (0, 1)$  we set  $w(\alpha) = 1$  and 0 otherwise, then the QWPS collapses to the quantile score (7), which we use to compare the predictive ability of  $\text{VaR}_\alpha$  forecasts of different models.

As in Amisano and Giacomini (2007) and Gneiting and Ranjan (2011), we compare the predictive power of density and  $\text{VaR}_\alpha$  forecasts applying Diebold and Mariano (1995) tests (DM henceforth) of equal forecasting performance of two different left tail QWPS or quantile scores. Under standard regularity conditions, the DM statistics of the differences between the scores associated with two forecasts is asymptotically standard normal under the null hypothesis of equality of scores.<sup>7</sup> Given the negative orientation of QWPS and quantile scores, a forecast with score  $m$  is superior (inferior) to a forecast with score  $h$  if the DM statistic is significantly negative (positive).

#### 4. RESULTS

Estimation and forecasting were conducted in pseudo-real time, with factors and parameters of all models re-estimated for each estimation window, yielding a total number of forecasts equal to 369 at a 3-month horizon, 366 at a 6-month horizon and 362 at a 12-month horizon for each model.

##### 5.1. Results for AR Models

Table A6 of the online Appendix reports left tail QWPSs of the 12 individual models and the four EWPs, with marks indicating the models with superior forecasting performance as obtained by pair-wise DM tests. For each variable and forecasting horizon the predictive ability of each subset of individual models does not differ significantly according to the number of factors used, the type of estimation window or the number of AR lags, but their forecasting performance is strongly or weakly dominated by the EWP forecasts for models both without and with factors as predictors. This result complements the evidence reported in Geweke and Amisano (2012) about EWPs as enhancers of forecasting performance of overall density forecasts, since factors also enhance density forecasts heavily weighted on the left tail.<sup>8</sup>

Focusing on EWP forecasts, Table I reports the matrix of DM tests of left tail QWPSs of the EWP models for each variable, estimation window and forecast horizon. Each cell of the matrix contains the DM statistics of the difference test between the left tail QWPS of the row model and the column model.

<sup>6</sup> Note that quantile prediction  $F^{-1}(\alpha)$  is optimal when the ex post loss is  $L(x, y, \alpha) = 2(1 - \alpha)y - x|$  in the case of an over-prediction ( $x \geq y$ ), and  $L(x, y, \alpha) = 2\alpha|y - x|$  in the case of an under-prediction ( $x \leq y$ ).

<sup>7</sup> The DM statistic is computed using heteroskedasticity and autocorrelation robust (HAC) standard errors. Diks *et al.* (2011) implement Monte Carlo experiments testing the power of DM tests for comparisons of scoring rules, finding that their power is adequate when the number of occurrences in the rejection region is greater than 40, which is a threshold satisfied by our forecasts using the expanding window.

<sup>8</sup> All rankings of the forecasting ability of the models based on left tail QWPSs are preserved when we consider the full matrix of DM tests of equal predictive ability of  $\text{VaR}_\alpha$  forecasts using quantile scores at both the 5% and 10% probability level, which we do not report for brevity.

Table I. DM pair-wise tests of left tail QWPS of EWPs of AR models

Forecasting horizon	3 months			6 months			12 months		
	AR-I	AR-D	FAAR-D	AR-I	AR-D	FAAR-D	AR-I	AR-D	FAAR-D
<i>IPG</i>									
AR-D	-0.06			0.45			0.18		
FAAR-D	-1.66	<b>-2.40</b>		-0.48	-1.23		-0.46	-0.90	
FAVARG-I	<b>-2.56</b>	<b>-3.43</b>	<b>-1.71</b>	<b>-2.54</b>	<b>-3.47</b>	<b>-2.33</b>	<b>-2.76</b>	<b>-3.03</b>	<b>-2.62</b>
<i>EMG</i>									
AR-D	<b>-3.48</b>			<b>-3.77</b>			<b>-3.71</b>		
FAAR-D	<b>-4.73</b>	<b>-3.83</b>		<b>-4.84</b>	<b>-3.71</b>		<b>-4.03</b>	<b>-2.25</b>	
FAVARG-I	<b>-4.22</b>	<b>-3.07</b>	0.31	<b>-4.48</b>	<b>-2.59</b>	1.12	<b>-3.74</b>	<b>-1.37</b>	0.17
<i>BDI</i>									
AR-D	<b>-2.92</b>			<b>-4.41</b>			<b>-5.53</b>		
FAAR-D	<b>-2.28</b>	<b>1.88</b>		<b>-4.21</b>	0.90		<b>-5.01</b>	-0.46	
FAVARG-I	<b>3.18</b>	<b>4.37</b>	<b>5.13</b>	<b>2.24</b>	<b>5.53</b>	<b>5.38</b>	1.55	<b>6.56</b>	<b>6.29</b>
<i>CDI</i>									
AR-D	<b>-1.80</b>			-1.67			<b>-1.81</b>		
FAAR-D	-1.21	1.07		-1.55	0.48		<b>-2.56</b>	-1.26	
FAVARG-I	0.92	<b>2.41</b>	<b>2.94</b>	1.14	2.54	2.57	<b>-2.07</b>	0.06	1.51
<i>DNFCI</i>									
AR-D	<b>4.78</b>			<b>5.64</b>			<b>3.37</b>		
FAAR-D	<b>3.71</b>	-0.56		<b>3.19</b>	<b>-2.15</b>		<b>1.95</b>	<b>-2.16</b>	
FAVARG-I	<b>2.22</b>	<b>-7.20</b>	<b>-8.65</b>	<b>1.91</b>	<b>-6.92</b>	<b>-6.36</b>	1.35	<b>-5.31</b>	<b>-5.69</b>

Note: Entries in bold indicate DM statistics corresponding to  $p$ -values  $\leq 0.05$  (if negative) or  $p$ -values  $\geq 0.95$  (if positive) of pair-wise tests of left tail QWPS of model column-left tail QWPS of model row.

A negative (positive) and significant DM statistic indicates that the left tail QWPS of the model row is superior (inferior) to that of the column model. The key result is that factor-augmented models exhibit tail risk forecasts significantly better than, or not significantly different from, those obtained with AR models for all variables and forecasting horizons with only two exceptions.

Specifically, DM tests for the real variables IPG and EMG show that EWP iterated FAVAR forecasts with linear GARCH volatility deliver significantly better tail forecasts than both those of the AR models without factors and direct factor-augmented AR forecasts at any forecasting horizon. Second, factor-augmented direct AR EWP forecasts for the financial variables BDI and CDI deliver significantly better forecasting performance than, or at least as good as, the AR models without factors. Third, for the financial variable DNFCI, AR models without factors deliver a better forecasting performance at the 3- and 6-month forecasting horizon, but they perform worse than iterated FAVAR forecasts at the 12-month horizon.

## 5.2. Results for Quantile Projections

Table A7 in the online Appendix reports the results of the full set of pair-wise DM tests of  $\text{VaR}_{0.05}$  and  $\text{VaR}_{0.10}$  of all QPs using the relevant quantile score. Results are qualitatively similar to, and generally statistically stronger than, those obtained for the AR models: the forecasting performance of EWP QPs, computed as in Giacomini & Komunjer (2005), is significantly better than, or in few cases at least as good as, each individual QP forecast.

Focusing on EWP forecasts, Table II compares the forecasting performance of factor-augmented QPs and QPs *without* factors, reporting the ratio of the quantile scores of the  $\text{VaR}_\alpha$  of factor-augmented QPs over the quantile scores of the  $\text{VaR}_\alpha$  of the EWP QPs without factors, where the significance of the relevant DM statistics for  $\alpha \in \{0.05, 0.10\}$  is marked in bold. It turns out that the  $\text{VaR}_\alpha$  forecasts of factor-augmented QPs are strictly superior to those of the QPs without factors for every variable and

Table II. Ratios of  $\text{VaR}_\alpha$  scores of factor-augmented QPs to  $\text{VaR}_\alpha$  scores of QPs without factors and DM tests

Forecast horizon	$\alpha = 0.05$			$\alpha = 0.10$		
	3 months	6 months	12 months	3 months	6 months	12 months
IPG	<b>0.69</b>	<b>0.62</b>	<b>0.63</b>	<b>0.73</b>	<b>0.75</b>	<b>0.63</b>
EMG	<b>0.76</b>	<b>0.67</b>	<b>0.68</b>	<b>0.81</b>	<b>0.79</b>	<b>0.67</b>
BDI	<b>0.92</b>	<b>0.92</b>	<b>0.80</b>	<b>0.95</b>	<b>0.94</b>	<b>0.86</b>
CDI	1.00	<b>0.94</b>	<b>0.88</b>	0.91	<b>0.93</b>	<b>0.92</b>
DNFCI	<b>0.63</b>	<b>0.59</b>	<b>0.53</b>	<b>0.81</b>	<b>0.72</b>	<b>0.65</b>

Entries in bold indicate a DM test of equal predictive performance rejected at 5% probability.

forecasting horizon, with two exceptions just indicating not significantly different forecasting performance. Thus the use of factors in QPs appears to enhance forecasting performance even more strongly than in the case of AR models.

### 5.3. Comparing $\text{VaR}_\alpha$ Forecasts of EWPs of Factor-Augmented AR and QP Models

We now compare the forecasting power of the best EWP factor-augmented AR forecasts, renamed EWPAR, with the EWP of factor-augmented QPs, renamed EWPQP. Table III reports the ratio of the  $\text{VaR}_\alpha$  scores of the EWPQP over the  $\text{VaR}_\alpha$  scores of the EWPAR for  $\alpha \in \{0.05, 0.10\}$ , and each variable and forecasting horizon, with marks of significance associated with the relevant DM tests.

It is apparent that the  $\text{VaR}_\alpha$  forecasts of the EWPQP are significantly more accurate than those of the EWPFAR for all variables and horizons, the only exception being the  $\text{VaR}_\alpha$  forecast of BDI, where the forecasting performance of EWPQP and EWPFAR does not differ significantly.

To evaluate our tail risk forecasts as early warning signals, we also compare  $\text{VaR}_\alpha$  forecasts of EWPFAR and EWPQP using *coverage ratios* or  *$\text{VaR}_\alpha$  violations*, defined as the percentage of cases where the realized value of a variable is *lower* than the predicted  $\text{VaR}_\alpha$ . The finding of a coverage ratio *higher* than the target probability level would indicate that the relevant forecast *underestimates* tail risk, since it does not capture all adverse risk realizations at that given target probability level (a Type I error). Conversely, the finding of a  $\text{VaR}_\alpha$  forecast whose coverage ratio is lower than the target probability level would indicate that it *overestimates* tail risk, since it would issue a percentage of ‘signals’ that are ‘false alarms’ relative to the target probability level (a Type II error). A risk-averse policymaker would likely consider a forecast that potentially overestimates rather than underestimates tail risks as more reliable, since the consequences of missing future adverse tail realizations may entail significantly larger costs than those associated with false alarms.

Table III. Ratios of  $\text{VaR}_\alpha$  scores of EWP factor-augmented QPs to  $\text{VaR}_\alpha$  scores of EWP factor-augmented VAR models and DM tests

Forecast horizon	$\alpha = 0.05$			$\alpha = 0.10$		
	3 months	6 months	12 months	3 months	6 months	12 months
IPG	<b>0.75</b>	<b>0.66</b>	<b>0.61</b>	<b>0.82</b>	<b>0.78</b>	<b>0.69</b>
EMG	<b>0.94</b>	<b>0.83</b>	<b>0.70</b>	<b>0.96</b>	<b>0.92</b>	<b>0.78</b>
BDI	1.04	<b>0.94</b>	0.99	0.99	1.03	1.02
CDI	<b>0.70</b>	<b>0.76</b>	<b>0.79</b>	<b>0.91</b>	<b>0.92</b>	<b>0.86</b>
DNFCI	<b>0.65</b>	<b>0.62</b>	<b>0.68</b>	<b>0.67</b>	<b>0.65</b>	<b>0.69</b>

Entries in bold indicate a DM test of equal predictive performance rejected at 5% probability.

Table A8 in the online Appendix reports coverage ratios of  $\text{VaR}_\alpha$  forecasts of EWPFAR and EWPQP for  $\alpha=0.05$  and  $\alpha=0.10$ , further broken down into statistics for the whole sample and the sample starting in 2007, which includes all observations related to the 2007–2009 financial crisis.

With the exception of  $\text{VaR}_\alpha$  forecasts for IPG at the 3-month horizon, all EWPFAR coverage ratios significantly exceed the relevant target coverage, often by a worrisomely large magnitude for all variable and forecasting horizons. In addition, EWPQP forecasts become significantly worse during the sample period that includes the 2007–2009 financial crisis, indicating that their accuracy becomes poorer when accuracy is needed most.

A totally different picture emerges from the forecasting results of EWPQP. First, the improvement in the precision of EWPQP forecasts relative to the EWPQP is fairly dramatic. Second, the coverage ratios of the EWPQP forecasts are *not higher* than the target coverage for about three-quarters of predictions with a 5% target coverage (Table A8, panel A), and about half predictions with a 10% target coverage (Table A8, panel B). Importantly, this is the case for both the full sample and the subsample including the 2007–2009 financial crisis. Third, in all cases where coverage ratios are higher than the target, they are such by a much smaller magnitude than those associated with EWPQP forecasts. All in

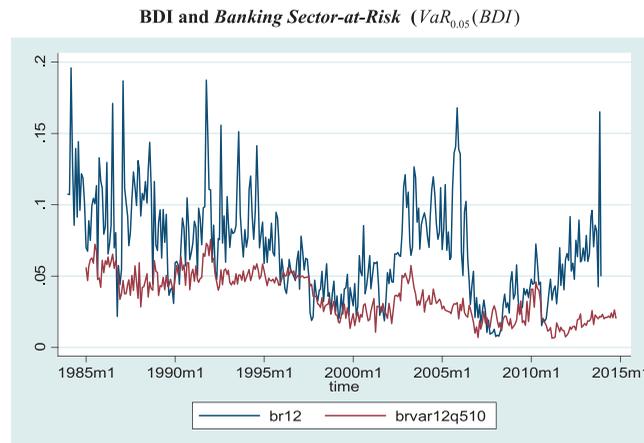


Figure 1. BDI and Banking sector at-risk ( $\text{VaR}_{0.05}(\text{BDI})$ ). This figure is available in color online at [wileyonlinelibrary.com/journal/jae](http://wileyonlinelibrary.com/journal/jae)

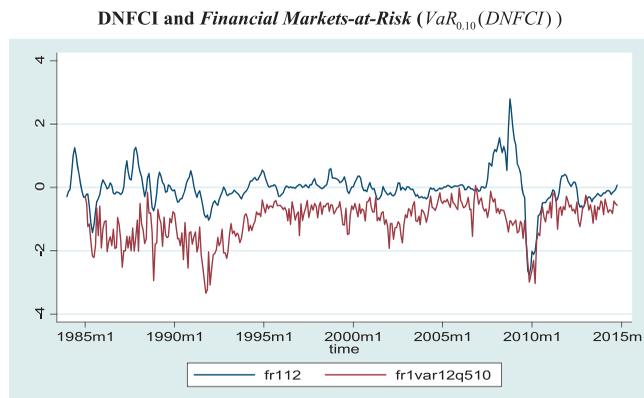


Figure 2. DNFCI and financial markets at-risk ( $\text{VaR}_{0.10}(\text{DNFCI})$ ). This figure is available in color online at [wileyonlinelibrary.com/journal/jae](http://wileyonlinelibrary.com/journal/jae)

all, the EWPQP forecasts are generally reliable, and deliver significantly more reliable early warning signals than those associated with the EWPAR.

The usefulness of the EWPQP  $\text{VaR}_\alpha$  forecasts as early warning signals is also visually illustrated in Figures 1 and 2, which depict actual and  $\text{VaR}_{0.05}$  12-month forecasts for banking sector at-risk and financial markets at-risk.

A decline in  $\text{VaR}_{0.05}$  forecasts *predicts* 12-month-ahead declines in each of these indicators fairly accurately, and the anticipation of these declines prior to the 2007–2009 financial crisis is particularly striking. Indeed, this is also true for industrial production at-risk, employment at-risk and corporate sector at-risk, as shown in Figure A1 in the online Appendix. We conclude that factor-augmented EWP QPs deliver fairly reliable forecasts and useful early warning signals for our measures of tail real and financial risks. Their superiority over forecasts obtained with AR models is due to their ability to capture and anticipate asymmetric changes in the distribution of a variable that may shift its probability mass to the left tail.

## 6. CONCLUSIONS

In this paper we developed a novel early warning system for tail real and financial risks. Using measures of tail risks consistent with a risk management approach, we: (i) assessed the predictive role of PCA factors in improving tail forecasts obtained with workhorse forecasting models such as AR models and quantile projections; (ii) gauged the additional predictive power of equally weighted pools of forecasts for both sets of models; and (iii) compared the accuracy of tail forecasts of these models in terms of their ability to issue reliable early warning signals.

Our results regarding significant improvement in the accuracy of tail forecasts arising from the use of PCA factors as predictors, and the dominance of equally weighted pools of forecasts over single model forecasts, complement the results of a large portion of the existing literature that has not specifically focused on tail forecasts. Our positive result concerning the ability of factor-augmented quantile projections to deliver reliable early warning signals for tail real and financial risk is encouraging and motivates several potentially useful extensions of our EWS. Important extensions include tailoring our modeling framework to different countries or sets of countries in a region, and designing procedures useful in identifying the economic drivers of shifts in the probability distribution of tail risks that would allow us to uncover the information content of changes in factors as related to observables. Some of these extensions are already part of our research agenda.

## ACKNOWLEDGEMENTS

We gratefully acknowledge helpful comments from the Editor, the referees, Gianni Amisano, Fabio Canova, Stijn Claessens, Robert McNown, and from participants at the September 2014 conference on ‘Macroeconomic Stability, Banking Supervision and Financial Regulation’ at the European University Institute, and at the February 2015 CESifo Area Conference on Macro, Money and International Finance. The views expressed in this paper are those of the authors and do not necessarily represent the views of the International Monetary Fund.

## REFERENCES

- Acemoglu, D, Ozdaglar A, Tahbaz-Salehi A. 2015. Microeconomic origins of macroeconomic tail risks. NBER Working Paper No. 20865.
- Acharya V, Pedersen L, Philippon T, Richardson M. 2012. Measuring systemic risk. CEPR Discussion Papers No. 8824.
- Ahn S, Horenstein A. 2013. Eigenvalue ratio test for the number of factors. *Econometrica* **81**(3): 1203–1227.

- Amisano G, Giacomini R. 2007. Comparing density forecasts via weighted likelihood tests. *Journal of Business and Economic Statistics* **25**(2): 177–190.
- Andersen T, Bollerslev T, Diebold F. 2010. Parametric and non-parametric volatility measurement. In *Handbook of Financial Econometrics*, Ait-Sahalia Y, Hansen LP (eds.). Elsevier: Amsterdam ch. 2; 67–138.
- Aramonte S, Rosen S, Schinler J. 2013. Assessing and combining financial conditions indexes. FEDS Working Paper 2013-39, Federal Reserve Board.
- Atkinson A, Eisfeldt A, Weill P-O. 2013. Measuring the financial soundness of U.S. firms. NBER Working Paper No. 19204.
- Bai J, Ng S. 2002. Determining the number of factors in approximate factor models. *Econometrica* **70**(1): 191–221.
- Bisias D, Flood M, Lo A, Valavanis S. 2012. A survey of systemic risk analytics. Working Paper No. 0001, Office of Financial Research.
- Boivin J, Ng S. 2006. Are more data always better for factor analysis? *Journal of Econometrics* **132**: 169–194.
- Brave S, Butters A. 2011. Monitoring financial stability: a financial conditions index approach. *Economic Perspectives* **35**(1): 22.
- Brownlees C, Engle R. 2011. Volatility, correlation and tails for systemic risk measurement. Working paper, Department of Finance, NYU.
- Corradi V, Swanson N. 2006. Predictive density evaluation. In *Handbook of Economic Forecasting*, Elliott G, Granger CWJ, Timmermann A (eds.). North-Holland: Amsterdam; 197–284.
- De Nicolò G, Lucchetta M. 2012. Systemic real and financial risks: measurement, forecasting and stress testing. IMF Working Paper 12/58.
- De Nicolò G, Lucchetta M. 2013. Systemic risks and the macroeconomy. NBER working paper No. 16998. Published. In *Quantifying Systemic Risk*, Haubrich J, Lo A (eds.). National Bureau of Economic Research: Cambridge, MA.
- Diebold FX, Mariano R. 1995. Comparing predictive accuracy. *Journal of Business and Economic Statistics* **13**: 253–263.
- Diks C, Panchenko V, van Dijk D. 2011. Likelihood-based scoring rules for comparing density forecasts in tails. *Journal of Econometrics* **163**: 215–230.
- Geweke J, Amisano G. 2011. Optimal prediction pools. *Journal of Econometrics* **164**(1): 130–141.
- Geweke J, Amisano G. 2012. Predictions with misspecified models. *American Economic Review: Papers and Proceedings* **102**(3): 482–486.
- Ghysels E, Rubia A, Valkanov R. 2009. Multi-period forecasts of volatility: direct, iterated and mixed-data approaches, EFA 2009 Bergen Meetings Paper.
- Giacomini R, Komunjer I. 2005. Evaluation and combination of conditional quantile forecasts. *Journal of Business Economics and Statistics* **23**(4): 416–431.
- Gneiting T, Ranjan R. 2011. Comparing density forecasts using threshold- and quantile-weighted scoring rules. *Journal of Business and Economic Statistics* **29**(3): 411–422.
- Hatzius J, Hooper P, Mishkin F, Schoenholtz KL, Watson MW. 2010. Financial conditions indexes: a fresh look after the financial crisis, U.S. Monetary Policy Forum. NBER paper.
- Komunjer I. 2013. Quantile prediction. In *Handbook of Economic Forecasting*, Vol. 2, Elliott G, Timmermann A (eds.). North-Holland: Amsterdam; 961–994.
- Leland H. 1994. Corporate debt value, bond covenants, and optimal capital structure. *Journal of Finance* **49**(4): 1213–1252.
- Marcellino M, Stock J, Watson M. 2006. A comparison of direct and iterated multistep AR methods for forecasting macroeconomic series. *Journal of Econometrics* **135**: 499–526.
- McCracken M, Ng S. 2014. FRED-MD: a monthly database for macroeconomic research. Available: <http://www.columbia.edu/~sn2294/papers/freddata.pdf>
- Pesaran H, Pick A, Timmerman A. 2011. Variable selection, estimation and inference for multi-period forecasting problems. *Journal of Econometrics* **164**: 173–187.
- Stock J, Watson M. 2002. Macroeconomic forecasting using diffusion indexes. *Journal of Business Economics and Statistics* **20**(2): 147–162.
- Stock J, Watson M. 2005. Implications of dynamic factor models for VAR analysis. NBER Working Paper No. 11467.
- Stock J, Watson M. 2006. “Forecasting with Many Predictors” Chapter 10. In *Handbook of Economic Forecasting*, Graham E, Clive WJ, Granger and Allan Timmermann (eds.). North Holland, Amsterdam; 516–554.
- Stock J, Watson M. 2011. Dynamic factor models. In *Oxford Handbook of Economic Forecasting*, Clements MP, Hendry DF (eds.). Oxford University Press: Oxford; 55–86.
- Stock J, Watson M. 2012. Generalized shrinkage methods for forecasting using many predictors. *Journal of Business Economics and Statistics* **30**(4): 481–493.