Predictive Regressions Under Asymmetric Loss: Factor Augmentation and Model Selection*

Matei Demetrescu[†]
University of Kiel

Sinem Hacıoğlu Hoke[‡]
Bank of England

This version: February 5, 2018

Abstract

The paper discusses the specifics of forecasting with factor-augmented predictive regressions under general loss functions. In line with the literature, we employ principal component analysis to extract factors from the set of predictors. We additionally extract information on the volatility of the series to be predicted, since volatility is forecast-relevant under non-quadratic loss functions. To ensure asymptotic unbiasedness of forecasts under the relevant loss, we estimate the predictive regression by minimizing the in-sample average loss. Finally, to select the most promising predictors for the series to be forecast, we employ an information criterion tailored to the relevant loss. Using a large monthly data set for the US economy, we assess the proposed adjustments in a pseudo out-of-sample forecasting exercise for various variables. Expectedly, the use of estimation under the relevant loss is found to be effective. But using an additional volatility proxy as predictor and conducting model selection tailored to the relevant loss function enhances forecasts significantly.

Keywords: Predictive regressions; Many predictors; Cost-of-error function; Latent variables; Volatility

JEL classification: C53 (Forecasting and Prediction Methods), C55 (Modeling with Large Data Sets)

*The views expressed here are those of authors and do not necessarily reflect those of Bank of England or the Monetary Policy Committee, the Financial Policy Committee or the Prudential Regulation Committee.

[†]**Corresponding Author:** Institute for Statistics and Econometrics, Christian-Albrechts-University of Kiel, Olshausenstr. 40-60, D-24118 Kiel, Germany. E-mail address: mdeme@stat-econ.uni-kiel.de.

[‡]Bank of England, Threadneedle Street EC2R 8AH, London, England. E-mail address: sinem.hacioglu@bankofengland.co.uk. Previous affiliation: University of Bonn.

1 Motivation

In forecasting macroeconomic series, the past decade has witnessed the increased availability and use of comprehensive data sets consisting of a large number of predictor time series. When forecasting macroeconomic aggregates like inflation or GDP growth especially in central banks, the appeal of such auxiliary data-rich sets is understandable: the additional informational content of the series helps improving forecasts compared to a benchmark (vector) autoregression of the variable to be predicted. At the same time, dealing with an increased number of predictor series poses problems, since the number of time observations is typically comparable with the number of series in such sets. This leads to imprecise coefficient estimates in an augmented predictive autoregression, and consequently to a trade-off between availability and usability of information. The literature has therefore focused on complexity reduction and information extraction. Factor-based forecasting models, for which it is assumed that unobserved common components of the auxiliary series are good predictors for the variable of interest, are particularly popular in this respect.

Since the predictors are not observed directly for factor-based forecasts, the forecasting procedure boils down to estimating a feasible predictive regression using lags of the dependent variable and *extracted* factors as right-hand side variables. Several contributions have shown that a relatively small number of estimated factors successfully summarize the contemporaneous information in the data set of predictors. Stock and Watson (2002c) show that Principal Component Analysis [PCA] of the predictors produces consistent estimates of the space spanned by the common factors. Their factor model forecasts outperforms other benchmark models to forecast personal income and output growth; see also the earlier work in Stock and Watson (1998). Focussing on estimation and inference in approximate factor models, Bai (2003) derives asymptotic distributions and uniform convergence results while Bai and Ng (2002) provide information criteria for estimating the number of factors; see also Alessi et al. (2010).

The popularity of factor models in forecasting is reflected by the large number of contributions in the applied literature. Ludvigson and Ng (2010, 2009) use factors from a large number of macroeconomic series to predict excess bond returns and to show that the predictability of future excess returns is related to macroeconomic activity. These are just the tip of the iceberg; see Marcellino et al. (2003), Artis et al. (2005), den Reijer (2005), Forni et al. (2005), Banerjee et al. (2008), Engel et al. (2015) or Godbout and Lombardi (2012) to name but a few more contributions to the literature on factor-based forecasting. While there are alternative approaches such as soft/hard thresholding or forecast combinations, they appear to be less popular than factor-based models. One reason to prefer factor-based forecasting procedures may be their interpretability; see e.g. the discussion in Ludvigson and Ng (2010, 2009). For instance, Ludvigson and Ng (2009) regress each macroeconomic variable in their data set on the PCA-extracted factors. The R^2s of these regressions are informative of the relations between the factors and the variables. They are thus able to identify e.g. stock market, inflation or real factors. More recently, Hacıoğlu Hoke and Tuzcuoğlu (2016) work on factor augmented VAR models with a threshold structure of the loadings (which are dynamic in their setup). The periods where the loadings are induced to zero or where the factors load more heavily on the variables are also informative on the relations between factors and variables. The point is that predictors with economic meaning prevent the interpretation of forecasting procedures as "crystal-ball" or "black-box" econometrics and are more likely to produce forecasts understandable by wider audiences.

The focus of the work cited above is on forecasts which are optimal in the mean squared-error [MSE] sense, i.e. on procedures minimizing the expected squared forecast error. The literature, however, documents a significant number of cases where more general – and in particular asymmetric – cost-of-error functions are employed. For instance, Elliott et al. (2005) propose formal methods of inference on the degree of asymmetry of the loss function and testing the rationality of forecasts; see also Patton and Timmermann (2007b).

Macroeconomic forecasting plays an important role at central banks' operation to conduct policy actions. The use of symmetric loss functions in forecasting is unrealistic, or at least it has to be tested, as central banks might have a particular type of aversion against positive and negative deviations from their targets. Therefore, forecasting under relevant loss is an on-going debate. Moreover, numerous papers investigate the asymmetry of central banks' loss functions alongside that of international organizations'. For instance, IMF and OECD forecasts of the deficit of G7 countries are found by Artis and Marcellino (2001) to be systematically biased towards over or under-prediction when compared with MSE-optimal forecasts. Building on the work of Elliott et al., Christodoulakis and Mamatzakis (2008, 2009) find asymmetric preferences of EU institutional forecasts. Clements et al. (2007) discuss the loss function of the Federal Reserve and Capistrán (2008) even finds that, for inflation, the forecasting preferences of the Fed are time-varying. While Pierdzioch et al. (2012) analyzes the loss function of Bank of Canada, Wang and Lee (2014) examines the forecasts of Greenbook and the Survey of Professional Forecasters. More recently, Tsuchiya (2016) examines the asymmetry of the loss functions of the Japanese government, the IMF and private forecasters for Japanese growth and inflation forecasts. This is, not unexpectedly, even more so for individual forecasters; see e.g. Elliott et al. (2008), Boero et al. (2008), Aretz et al. (2011), Clatworthy et al. (2012) or Fritsche et al. (2015).

We therefore study factor-augmented forecasting under asymmetric loss. For a given predictive model, there is little debate as to how to obtain point forecasts under a given loss function: it has been known since Weiss and Andersen (1984) and Weiss (1996) that the forecast model should be estimated under the relevant loss.[1] Estimation of the feasible predictive regression under the relevant loss would therefore improve forecasts. This prompts the question, first, whether such

[1] An alternative, more demanding, procedure is to model the entire predictive distribution and derive the point forecasts based on it; see e.g. McCullough (2000) for an ingenious bootstrap-based version.

estimation may indeed be conducted with estimated factors in a manner analogous to the MSE-optimal case. Less obvious however, is the second question of whether the forecast model should be the same under any asymmetric loss function. To put it bluntly, are the PCA-extracted factors still forecast-relevant under an asymmetric loss function? Considering the theory of forecasting under asymmetric loss functions, see Granger (1969), Granger (1999), Weiss (1996), Christoffersen and Diebold (1996), McCullough (2000), Elliott and Timmermann (2004), Elliott et al. (2005), Patton and Timmermann (2007a) or Patton and Timmermann (2007b), the least what may be expected for the extracted factors (or even for the lags of the dependent variable in the augmented predictive autoregression) is that their relative importance as a predictor changes. So, rather than relying on the summarizing power of, say, the first principal component, one may have to select the predictors (lagged dependent variables or factors) that are most informative under the relevant loss.[2] Third, perhaps even more importantly, one should ask whether the usual factor extraction does actually capture all information relevant under the given loss function. PCA essentially delivers linear combinations of the "many predictors" data set. In a linear predictive model under squared-error loss, this may be a convenient dimensionality reduction procedure. The optimal forecast function under an asymmetric loss function may, however, depend on the auxiliary series in a non-linear fashion, even if the optimal forecast function is linear in the MSE-optimal case. Thus, the informational content of the data set may not be fully exploited under an asymmetric loss function.

Our contributions are as follows. We show in Section 2 that, regularity conditions provided, one may indeed use PCA-extracted factors as predictors even when estimating forecast regressions using the relevant loss function. To make sure that relevant information is not wasted, we make use in Section 3 of the insight that the optimal point forecast under a general loss depends on the conditional variance of the variable to be predicted (Christoffersen and Diebold, 1996, and Patton and Timmermann, 2007b). Thus, in some cases, adding information on the *volatility* of the series to be predicted in the forecasting model improves forecasts under asymmetric loss. While the volatility of interest is not observed directly, it is plausibly related to the variability of the auxiliary series. The relation is not a forced one, since the volatility of the overall economic environment should be reflected – at least to some extent – by the volatility of all series involved. This common component can in turn be extracted from the auxiliary data set. Concretely, we extract additional factors from the log-squared residuals of the factor model to increase the quality of the forecasts under the relevant loss. This delivers a larger number of predictors, of which not all need be equally relevant. To find the ones with the highest predictive power, we resort to a suitable information criterion.[3]

[2]In fact, focusing on extracting the factors with the highest associated eigenvalue might not be a good idea in the MSE-case either, since a factor even if explaining most of the variance of the raw predictor series, need not capture the information relevant for forecasting.

[3]The issue of model selection is not restricted to our setup: e.g. Schumacher (2007) compares the forecast accuracy of variety of factor models to MSE-predict German GDP, and finds that results may change when different information criteria to select factors are used.

We then illustrate the proposed procedure in Section 4 by means of a forecasting exercise with US personal income, industrial production, unemployment rate and retail sales. We use the regularly updated version of the data set which has become widely known as the "Stock and Watson" data set (Stock and Watson, 2005). This monthly data set is currently referred to as FRED-MD (Federal Reserve Economic Data, A Monthly Database for Economic Research). Its features are detailed in McCracken and Ng (2015).[4] Here, we are interested in one-year-ahead forecasts. We compare the average forecast losses of all four variables in different forecast procedures we look into. We find, expectedly, that average losses of forecasts produced under the relevant loss function give smaller losses compared to the losses produced by forecasts obtained via OLS estimation of the predictive regression. At the same time, we also show that adding information from the volatility of the series and having parsimonious models by assessing the relevance of the extracted factors improve the average losses in some cases.

The final section concludes; some technical details and additional results have been gathered in the Appendix.

2 The basic forecasting problem

Let y_t be the series for which an h-step ahead forecast is required. Given the available information set $\mathcal{F}_t$, the optimal forecast is given by

$$y_{t+h}^{opt} = \arg\min_{y_{t+h}^*} \mathrm{E}\left(\mathcal{L}\left(y_{t+h} - y_{t+h}^*\right) | \mathcal{F}_t\right), \tag{1}$$

where $\mathcal{L}(\cdot)$ is the relevant loss function quantifying the cost incurred by discrepancies between a given forecast y_{t+h}^* of the variable y at some time $t+h$ and the actual realization y_{t+h}. According to Granger (1999), loss functions should be uniquely minimized at the origin, and be quasi-convex. We shall work with the popular class of loss functions introduced by Elliott et al. (2005); a forecast y_{t+h}^* is thus evaluated by means of

$$\mathcal{L}\left(y_{t+h} - y_{t+h}^*\right) = \left(\alpha + (1-2\alpha)\,\mathbb{I}\left(y_{t+h} - y_{t+h}^* < 0\right)\right) \left|y_{t+h} - y_{t+h}^*\right|^p. \tag{2}$$

This class of loss functions is quite flexible: it includes as special cases the widely used symmetric (for $\alpha = 0.5$) and asymmetric (for $0 < \alpha < 0.5$ or $0.5 < \alpha < 1$); linear and quadratic loss functions (for $p = 1$ and $p = 2$). Moreover, it only requires mild moment conditions on y_t, in contrast e.g. to the well-known linex loss.

We take the information set $\mathcal{F}_t$ to contain the variable of interest at all available times and additional predictors, $\mathcal{F}_t = \{y_t, y_{t-1}, \ldots, y_1, f_{t,1}, \ldots, f_{t,r}\}$, and start with the usual linear forecasting

[4] The use of this particular data set has been quite popular in the literature; see e.g. Belviso and Milani (2006), Boivin and Ng (2006), D'Agostino and Giannone (2006), Ludvigson and Ng (2010, 2009) and Bai and Ng (2011).

model

$$y_{t+h} = c + \sum_{j=1}^{q} a_j y_{t-j+1} + \sum_{k=1}^{r} b_k f_{t,k} + v_{t+h}, \quad t = 1, 2, \ldots, T\,, \tag{3}$$

where the forecast error v_{t+h} is unpredictable under the loss function $\mathcal{L}$. This does not imply, however, that v_{t+h} could not be forecast under another loss function. The lack of predictability of v_{t+h} under $\mathcal{L}$ implies that the so-called generalised forecast error $\mathcal{L}'(v_{t+h})$ is uncorrelated with the predictors y_{t-j+1} and $f_{t,k}$; see Granger (1999) and Patton and Timmermann (2007a). The optimal forecast is thus given by

$$y_{t+h}^{opt} = c + \sum_{j=1}^{q} a_j y_{t-j+1} + \sum_{k=1}^{r} b_k f_{t,k}. \tag{4}$$

In practice, one resorts to a two-stage procedure, given that observations on N auxiliary variables $x_{t,i}$ are available, from which $f_{t,k}$ may be estimated in a first stage. Maintaining the typical assumption of linear measurement equations for the factors, we have that

$$x_{t,i} = \sum_{k=1}^{r} \lambda_{i,k} f_{t,k} + u_{t,i}. \tag{5}$$

With additional conditions on $\lambda_{i,k}$ and $u_{t,i}$ (in particular orthogonality of the common and idiosyncratic components $f_{t,k}$ and $u_{t,i}$), extraction of the unknown factors can be conducted, leading to $\hat{f}_{t,k}$. To estimate the factors, we resort to PCA. This ultimately takes us to the feasible predictive regression

$$y_{t+h} = c + \sum_{j=1}^{q} a_j y_{t-j+1} + \sum_{k=1}^{r} b_k \hat{f}_{t,k} + v_{t+h}, \tag{6}$$

to be estimated under the relevant loss in a second stage, i.e.

$$\tilde{c}, \tilde{a}_j, \tilde{b}_k = \underset{c^*, a_j^*, b_k^*}{\arg\min} \frac{1}{T} \sum_{t=q}^{T-h} \mathcal{L}\left(y_{t+h} - c^* - \sum_{j=1}^{q} a_j^* y_{t-j+1} - \sum_{k=1}^{r} b_k^* \hat{f}_{t,k} \right), \tag{7}$$

from which the forecast is obtained as

$$\tilde{y}_{t+h}^{opt} = \tilde{c} + \sum_{j=1}^{q} \tilde{a}_j y_{t-j+1} + \sum_{k=1}^{r} \tilde{b}_k \hat{f}_{t,k}. \tag{8}$$

Its quality hinges on the precision of the factor approximation, as we have some regressors observed with measurement error; recall that factors cannot be consistently estimated in a fixed-N setup.

Note that the loss function does not play any role in estimating the factors, but only in the subsequent forecasting step. The main reason to proceed this way is to maintain the interpretability of the factors as economic driving forces (not depending on individual loss preferences), but we also wish to stay in line with the literature on factor-based forecasting. While Tran et al. (2016) discuss estimation of factors under asymmetric linear and asymmetric quadratic losses, these losses refer to the idiosyncratic components and not to the actual forecast errors, so this way of extracting factors is not directly relevant for forecast under asymmetric loss. We leave the integration of the two approaches to further work.

The justification to use the feasible forecast from (8) is provided by the following proposition establishing its consistency for the unfeasible optimal forecast from (4) as $T, N \to \infty$.

Proposition 1 *Let the lag polynomial* $1 - \sum_{j=1}^{q} a_j L^j$ *be causally invertible. Then, let* v_{t+h}, $f_{t,k}$ *and* $i_{t,i}$ *be piecewise locally stationary and weakly dependent as defined in Appendix B. Let furthermore the generalised forecast errors* $\mathcal{L}'(v_{t+h})$ *satisfy* $\mathrm{E}(\mathcal{L}'(v_{t+h})| y_t, y_{t-1}, \ldots, f_{t,k}) = 0$ *and have a distribution with no atom at 0. Finally, let* v_{t+h} *and* $f_{t,k}$ *have finite moments of order* $\min\{2p, 8\}$, *with* p *from* (2) *integer and positive. It then holds for the estimated optimal forecast from* (8) *that, for each* t,

$$\tilde{y}_{t+h}^{opt} \xrightarrow{p} y_{t+h}^{opt}$$

as $N, T \to \infty$ *such that* $T/N \to 0$.

Proof: *See Appendix C.*

Remark 2 *The technical assumptions given in Appendix B essentially impose weak serial dependence and weak cross-correlation as required for large-*N *large-*T *factor estimation and for convergence of sample averages. At the same time, they allow all predictor and forecast error series to exhibit time-varying variance and means, with both smooth and abrupt changes in time; see Zhou (2013). This is considerably less restrictive than the often made assumption of weak stationarity (see e.g. Stock and Watson, 2002c), but comes at the expense of a more detailed specification of the model compared to restrictions on (auto-)covariances; all in all, it is a small price to pay for being able to use non-MSE loss functions in a nonstationary environment. The critical requirement is that the generalised forecast error is a martingale difference sequence, which is a standard condition in the literature on forecasting under asymmetric loss (Patton and Timmermann, 2007a): in a nutshell, the forecast errors must be unforecastable under the relevant loss.*

Remark 3 *In factor models, the factors are only identified up to a rotation. But it follows from the proof that rotations do not affect the result: essentially,* $\sum_{k=1}^{r} \tilde{b}_k \hat{f}_{t,k}$ *consistently estimates* $\sum_{k=1}^{r} b_k f_{t,k}$ *which is the quantity required for forecasting* y_{t+h}. *E.g. Bai and Ng (2006) consider this explicitly; to keep notational effort at a minimum, we assume identification directly.*

3 Extracting additional relevant information and model selection

The two-step procedure for forecasting under asymmetric loss discussed in the previous section is the natural extension of the original method of Stock and Watson (2002c), for which the second step – i.e. estimation of the predictive regression – has been modified to account for the use of a specific loss function. But we should ask at this point whether the first stage – i.e. extracting the information carried by the auxiliary variables $x_{t,i}$ – is to be left unmodified. In other words, is the factor model (5) exhausting the possibilities of finding predictors for y_{t+h} under the relevant loss?

It should be pointed out that the linear model (5) is only sufficient under conditions which are not plausible for macroeconomic data sets. Namely, Patton and Timmermann (2007b) show that, for loss functions of the type given in (2), the optimal forecast has the form

$$y_{t+h}^{opt} = \mathrm{E}\left(y_{t+h} \middle| y_t, y_{t-1}, \ldots, x_{t,i}\right) + C\sqrt{\mathrm{Var}\left(y_{t+h} \middle| y_t, y_{t-1}, \ldots, x_{t,i}\right)} \qquad (9)$$

for some constant C depending on the loss function and the shape of the conditional distribution.[5] The first summand on the r.h.s. of (9) is nothing else that the conditional mean which the original factor-based model does indeed capture. The coefficient C, and thus the second summand, is zero e.g. when $\alpha = 0.5$ and $p = 2$, or when $\alpha = 0.5$ and the conditional distribution of y_{t+h} is symmetric, but not in general. When estimated under the relevant loss, the intercept c of the predictive regression (3) only captures the *average* of the so-called bias term $C\sqrt{\mathrm{Var}\left(y_{t+h} \middle| y_t, y_{t-1}, \ldots, x_{t,i}\right)}$ and misses the fact that the conditional standard deviation of y_{t+h}, *if time-varying*, is actually a predictor for y_{t+h} under $\mathcal{L}$.

And indeed, the volatility of macroeconomic variables is not constant in general. The Great Moderation is the perhaps best known case of time-varying volatility. The term coins the downward trend in the variance of inflation and economic growth since the 1980s (e.g., Stock and Watson, 2002b); Clark (2009) finds that the recent financial crisis has reversed the trend, thus strengthening the evidence of time-varying volatility. Along the same lines, Sensier and van Dijk (2004) find that four out of five of over two hundred U.S. macroeconomic time series exhibit unconditional volatility changes during the period 1959-1999.

What is more, it is expected that such volatility variations are common to the variables in the data set used for forecasting: the series stem, after all, from the same economic environment.[6] Thus, we may resort to *the same* data set $\{x_{t,i}\}$ in order forecast the conditional standard deviation of y_{t+h}.

[5]Their result actually holds for any homogenous loss function.

[6]This was e.g. exploited in the context of stock market volatilities by Barigozzi and Hallin (2016).

To exploit the above insight, we assume a stochastic volatility model of the form

$$v_{t+h} = e_t \, e^{\frac{1}{2}\left(g_t + \sum_{l=1}^{s} \xi_l h_{t,l}\right)}.$$

We follow Nelson (1991) in using the exponential "link" function, since it allows us to avoid positivity restrictions on the components g_t and $h_{t,l}$ and assume – in line with the very idea of factor-based forecasting – that $h_{t,l}$ could be forecast using information from the auxiliary series $x_{t,i}$; g_t is an unforecastable component. When the conditional variance of the idiosyncratic components in the factor model depend in a similar manner on $h_{t,l}$, we write

$$u_{t,i} = e_{t,i} \, e^{\frac{1}{2}\left(g_{t,i} + \sum_{l=1}^{s} h_{t,l} \xi_{l,i}\right)},$$

where $g_{t,i}$ are individual volatility components specific for $x_{t,i}$. As usually, e_t and $e_{t,i}$ are standardised variables, mutually independent and independent of $h_{t,l}$, g_t and $g_{t,i}$, and we take them and their logs to satisfy regularity conditions of the type outlined in Appendix B. Then,

$$\log u_{t,i}^2 = \log e_{t,i}^2 + g_{t,i} + \sum_{l=1}^{s} \xi_{l,i} h_{t,l},$$

which is nothing else than a factor model for the log squares of $u_{t,i}$ with $h_{t,l}$ the common components and $\log e_{t,i}^2 + g_{t,i}$ the idiosyncratic ones.

Since the variables $u_{t,i}$ are not observed directly, we resort to the idiosyncratic components extracted in the first-stage PCA. Thus we are now able to extract $h_{t,l}$ from $\log \hat{u}_{t,i}^2$ using a second-stage PCA, leading to $\hat{h}_{t,l}$. Note that the factors $f_{t,k}$ themselves may be (conditionally) heteroskedastic; we assume that they do not bear additional predictive power for the conditional variance of y_{t+h}, but one may of course consider their log squares when extracting $h_{t,l}$.

This is related to decomposition of the yield spreads in Ludvigson and Ng (2010, 2009). In both papers, additional information carried by the yield risk premium (or term premium) is acknowledged, due to the inability of the yield curve to explain business cycle variations in bond risk premia. The yield risk premium can be seen as an idiosyncratic error which should be constant under the expectation hypothesis. Ludvigson and Ng estimate this term via the average multi-step estimates of bond returns. They show that the predictive factors are not sufficient to display the counter-cyclical form of bond risk premia since the predictive power of these factors does not imply explaining the yield curve. In this respect, the additional information used, namely the yield risk premium, parallels the volatility factor we use in this paper.

Equation (9) shows that a nonlinear forecast may be better suited in an asymmetric loss context. Clearly, extracting factors from $\log u_{t,i}^2$ is not the only way to consider nonlinearities; for instance Bai and Ng (2008b) employ quadratic PCA. But Equation (9) motivates us to look directly for variables driving the volatility.

Ideally, we would include a term of the form $Ce^{\frac{1}{2}\sum_{l=1}^{s}\xi_l \hat{h}_{t,l}}$ in the predictive regression with additional parameters ξ_l (with g_t not being predictable, $e^{1/2 g_t}$ is absorbed in the error component e_t multiplicatively). But a non-linear regression equation is perhaps too cumbersome to deal with numerically, even if we must anyway resort to numerical optimization under non-MSE loss.[7] We therefore linearize the exponential, $e^x \approx 1 + x$, and trade some misspecification in exchange for increased clarity of the final procedure.

The component g_t is in principle not predictable, at least not using $x_{t,i}$, and we treat it as such by absorbing it in the forecast error. We thus obtain as estimated predictor for y_{t+h}

$$\tilde{y}_{t+h}^{opt} = \tilde{c} + \sum_{j=1}^{q} \tilde{a}_j y_{t-j+1} + \sum_{k=1}^{r} \tilde{b}_k \hat{f}_{t,k} + \sum_{l=1}^{s} \tilde{\xi}_l \hat{h}_{t,l}, \tag{10}$$

where the parameter estimates are obtained like before by minimising the observed forecast loss.

Due to the linearisation, the estimators $\tilde{\xi}_l$ in (10) do not converge to the population values. The following proposition guarantees that the fitted predictor is the best *linear* predictor under the given loss.

Proposition 4 *Define the (unfeasible) linear predictor*

$$\pi(y_t, f_t, h_t) = c^* + \sum_{j=1}^{q} a_j^* y_{t-j+1} + \sum_{k=1}^{r} b_k^* f_{t,k} + \sum_{l=1}^{s} \xi_l^* h_{t,l}$$

and assume that $\sup_t \left|\hat{h}_{t,l} - h_{t,l}\right| = o_p(1)$. *Under the above assumptions, it holds for* $\tilde{y}_{t+h}^{opt}$ *from* (10) *that*

$$\tilde{y}_{t+h}^{opt} - \underset{c^*, a_j^*, b_k^*, \xi_l^*}{\arg\min} \frac{1}{T} \sum_{t=p+1}^{T-h} \mathrm{E}\left(\mathcal{L}\left(y_{t+h} - \pi(y_t, f_t, h_t)\right)\right) \xrightarrow{p} 0$$

for each t.

Proof: *Analogous to the proof of Proposition 1 and omitted.*

Remark 5 *In the case of the squared-error loss, the bias-variance decomposition of the MSE indicates that the fitted linear model minimizes the expected squared difference between the linear fit and the nonlinear regression curve (where the expectation is taken with respect to the marginal distribution of the predictors). While such a clean decomposition is not available in the case of asymmetric power expected losses, the interpretation of the proposition remains the same.*

Remark 6 *The quality of the linear approximation depends on the signal-to-noise ratio in the series* $\sum_{l=1}^{s} \xi_l \hat{h}_{t,l}$. *One could improve it by taking a quadratic approximation for the exponential,*

[7] See Demetrescu (2006) for a tailored optimization method.

$e^x \approx 1 + x + x^2/2$. *When not imposing the coefficient restrictions resulting from the quadratic approximation of the exponential function to avoid further numerical complications, this results in a linear model with interactions,*

$$\tilde{y}_{t+h}^{opt} = \tilde{c} + \sum_{j=1}^{q} \tilde{a}_j y_{t-j+1} + \sum_{k=1}^{r} \tilde{b}_k \hat{f}_{t,k} + \sum_{l=1}^{s} \sum_{m=1}^{s} \tilde{\xi}_l \tilde{\xi}_m \hat{h}_{t,l} \hat{h}_{t,m}.$$

This is different from, though related to quadratic PCA as e.g. employed by Bai and Ng (2008b): Bai and Ng apply PCA directly onto $x_{t,i}$ and squares thereof, while we extract information on volatility from the squares of the idiosyncratic components $u_{t,i}$.

To sum up, the factor-based forecasting procedure is modified under asymmetric loss as follows.

1. Clean/prepare the auxiliary data set and the variable to be predicted.
2. Extract factors from auxiliary series (PCA).
3. Extract factors (demean, standardise, PCA) from log-squared extracted idiosyncratic components.
4. Augment the predictive autoregression with the factors extracted in steps 2 and 3.
5. Estimate under the relevant loss.
6. Suitably select the predictors to enter the predictive model.

Compared to the usual factor-based forecasting approach, steps 3 and 5 are new and specific to forecasting under a general loss function. Step 6 should of course be conducted even under squared-error loss, but requires here a careful consideration of the used selection tool. Concretely, to conduct predictor selection in (10), we resort to an information criterion, but tailored to the relevant forecasting problem. Let k_1 denote the number of directly observed regressors and k_2 the number of extracted factors. For a model of complexity $k = k_1 + k_2$, we thus compute

$$AIC_{\mathcal{L}}(k) = \frac{2}{p} \ln \left(\sum \mathcal{L} \left(\hat{v}_{t+h}(k) \right) \right) + \frac{2k_1}{T} + \frac{2k_2}{T} \left(1 + \frac{T}{N} \right)$$

with $\hat{v}_{t+h}(k)$ in-sample fitted errors from the respective model. As usually, choose then the model minimizing the criterion. See Appendix A for a justification of this particular choice.

We work with an information criterion because of the widespread use of information criteria in general, but partly also for computational convenience; we also examined the numerically more involved least absolute shrinkage and selection operator (LASSO) (Tibshirani, 1994) as an alternative, alongside with refinements due to Belloni and Chernozhukov (2013). We present in the following section the empirical results obtained using only the tailored information criterion $\text{AIC}_{\mathcal{L}}$ for model selection.[8]

4 Forecasting under asymmetric loss

The goal of the empirical exercise is to forecast several macroeconomic variables, such as Personal Income (PI), Industrial Production (IP), Unemployment Rate (UN) and Retail Sales (SL), under asymmetric loss. Expanding the target variables beyond the standard variables, such as industrial production as in Stock and Watson (2002a), reflects our desire to forecast variables that are not widely considered in the literature. We evaluate the out-of-sample forecasts of these four variables that use the factors recursively extracted by PCA from the auxiliary data. We pursue the forecast analysis by taking them as observable. In this regard, the set up of the forecasting exercise shows resemblance to that of Ludvigson and Ng (2009).

Throughout the empirical exercise, the loss function in Equation (2) is quadratic with $p = 2$. We allow for different degrees of asymmetry by considering five αs, $\alpha \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$. Note that for $p = 2$ and $\alpha = 0.5$, the symmetric quadratic loss is recovered for this particular loss function.

4.1 Setup

The data set used for the forecasting exercise is often referred as the Stock and Watson data set, Stock and Watson (2005). It is being updated every month by the Federal Reserve Economic Data (FRED) and therefore referred as FRED-MD by McCracken and Ng (2015). It consists of 128 monthly US macroeconomic aggregates and spans the time period of March 1959–October 2017. For missing observations we follow McCracken and Ng (2015), and use an iterative expectation-maximization algorithm for imputation, see also Stock and Watson (2002c).

The consistency of the estimated forecast function relies, among others, on the assumption that observable series are stationary. The series are therefore transformed to stationarity by taking differences, by taking logarithms – and in some cases by doing both; see McCracken and Ng (2015) for details. Finally, all transformed variables are standardized to have zero sample mean and unit sample variance for factor extraction.

We perform a recursive pseudo out-of-sample forecasting procedure. In each step, one-year-ahead forecasts are constructed and the forecast horizon is $h = 12$. Concretely, we start with data from 1959:03 through 1981:03; we run the forecasting regression with dependent variables

[8] The corresponding LASSO and post-fit LASSO results are available upon request; in short, the LASSO does not outperform the information criterion.

from 1960:03 to 1981:03 and factors as predictors from 1959:03 to 1980:03. The outcome is used to forecast PI, IP, UN and SL for 1982:03. We then expand the data set by one period to obtain the forecasts for 1983:03. The procedure is iterated until we obtain the last forecast for 2017:10. (In the last step, the independent variables from 1959:10 through 2015:10 and dependent variables from 1960:10 to 2016:10 are used to run the forecasting regressions to forecast 2017:10.) All steps also include the first lag of the dependent variable. Appendix D presents additional results for $h = 3$ and $h = 6$.

4.2 Extracted factors

One of the common issues associated with factor-based forecasting approaches is the number of factors to be extracted from the auxiliary data set. To set this in stone, we start by performing the information criteria developed by Bai and Ng (2002), and used by Ludvigson and Ng (2010, 2009) and Bai and Ng (2011).[9] The criteria find eight factors in the data set. Factors are identified up to a rotation, so a comprehensive interpretation of extracted factors is not straightforward. Ludvigson and Ng (2009) and McCracken and Ng (2015) report marginal R^2s of the regressions of each of the series against each of the eight factors. They relate these factors with broad classes of economic activity. Note, however, that the forecasting procedure does not hinge on this classification. These eight factors might not all improve the forecast accuracy. Similarly, factors beyond the first eight might appear to be forecast-relevant.

For a closer look on the number of factors, we employ the tailored $\text{AIC}_{\mathcal{L}}$ for a preliminary check of the number of factors in the data for the *full time span*. This preliminary exercise starts with selecting among all possible combinations of the first 8 PCA-extracted factors which are chosen by the Bai and Ng (2002) information criteria. In the second step, 9 factors are extracted from the auxiliary data and selection is conducted among all combinations of these 9, and so on. We stop at selection among the first 12 PCA-extracted factors. The factors in each step, along with the first lag of the respective dependent variable, are used in the predictive regressions to forecast all four variables of interest after being subject to the model selection. Figure 1 reports the predictors chosen by minimizing $\text{AIC}_{\mathcal{L}}$ among all predictors, by aggregating the results for all αs.

Figure 1 shows that not all factors selected as predictors among the first eight common factors. For example, factors 1, 3 and 6 seems to be irrelevant for forecasting PI. Factor 8 does not contribute to the forecast of SL. On the other hand, factors 9 to 12 collectively appear to be forecast relevant for most of the variables.[10] When we contemplate all 12 PCA-extracted factors,

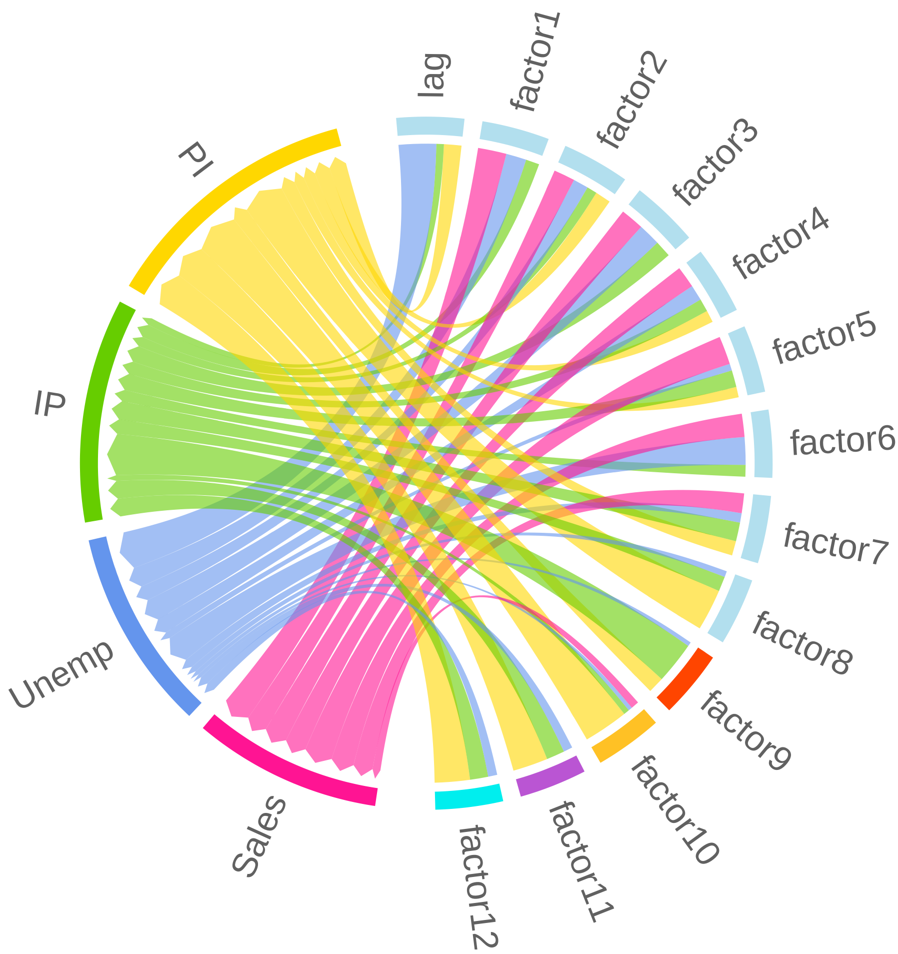
[9] Bai and Ng (2002) information criteria do *not* consider generalised loss functions. We apply these criteria to give a preliminary idea about the number of the factors.

[10] The objective function of the $\text{AIC}_{\mathcal{L}}$ targets the dependent variable whereas the PCA analysis aims to maximize the variance explained by factors. Due to the difference in the objective functions, the factors selected by the information criteria do not always appear to be forecast relevant.

we observe that some of the commonly used first 8 factors do not always appear to be forecast relevant while some additional ones do.

The table leading to Figure 1 is given in Appendix D.3. Table D.6 presents evidence that increasing the number of factors to select from one by one, the factors that are already selected occasionally change but new ones appear to be selected. Moreover, the selected factors tend to change depending on the asymmetry of the loss function and the set of selected factors frequently does not include all the first eight but further PCA-extracted factors.

Figure 1: Factors selected for all predictor series by $\text{AIC}_{\mathcal{L}}$; full data span including the first lag of the dependent variable and number of factors increasing gradually from 8 to 12



Notes: Asymmetric quadratic loss; see the text for details. The analysis includes the first 8 to 12 factors and additionally the first lag of the dependent variable. In each step, the lag and the first 8 factors are included in the analysis. Additional factors 9, 10, 11 and 12 are added consecutively without replacing the previous factors. The results aggregated over different αs for a given variable. The analysis is conducted for the whole time span. The forecast relevant factors for each variable are as follows. PI: lag, factor 2, 4, 5, 7 to 12. IP: lag and all 12 factors. $Unemp$: lag and all 12. $Sales$: factors 1 to 7 and 10.

Evidence from this preliminary exercise suggests to also consider for forecasting factors beyond the ones selected by the Bai and Ng (2002) information criteria. Thus, we use 12 factors (the largest PCA-extracted ones) as benchmark rather than the first 8 factors found by the information criteria. To keep the complexity tractable and computation time reasonable, we do not consider other classical factors beyond these. This example also emphasizes the need for model selection amongst the considered factors and lagged dependent variable.

The extracted volatility factor(s) give(s) information which is not (linearly) contained in the original series. According to the mentioned information criteria,[11] the PCA of the log-squared residuals from the first-step factor analysis generally leads to only one additional factor to be taken into account. In the main exercise in Section 4.3, we consider it as an additional predictor. When selecting the concrete predictive model for a given span of observations, the volatility factor is subject to model selection with the tailored AIC alongside the other factors. Following Remark 6, we also consider the squared volatility factor.

4.3 Forecast Results

In this section, we discuss the forecast results for different asymmetry parameters, α, and forecast procedures. For each α, we first estimate the respective predictive regression by ordinary least squares (OLS) relying on the regressors in the benchmark models in recursive manner. We construct one-year-ahead forecasts in each recursive step and evaluate the occurring loss via the forecast errors under the relevant loss function. This approach is henceforth named as OLS-Asymmetric Loss ($OLS - AL$). The second route to take here is estimating the regression coefficients numerically directly to use them to construct the forecasts and evaluate the forecast errors. This approach is named as Asymmetric Loss (AL) henceforth.

Concentrating on the evaluation of the forecasts obtained using OLS vs. those obtained via estimation under the relevant loss, each recursive step provides the forecast losses of $OLS - AL$ and AL for associated variables. We simply compare the *average* forecast over all recursive steps. One expects the average forecast loss of AL to be smaller than the average loss which occurs under $OLS - AL$; see the early work of Weiss and Andersen (1984).

We consider six forecast procedures in total. Procedure I uses only 12 factors for the forecasting exercise. Procedure II includes the factor extracted from the log-squared idiosyncratic components $\widehat{u}_{t,i}$. Thus, there are in total 13 factors in procedure II. Procedure III adds the squared volatility factor after which we have 14 factors. We do not conduct model selection for procedures I, II and III. The forecast procedure IV is the counter-party of procedure I with model selection by the tailored $\text{AIC}_{\mathcal{L}}$ for the all possible combinations of 12 factors. Similarly, procedures V and VI are model selection versions of procedures II and III, respectively. Note that

[11] Following Bai and Ng (2002), we rely on PC_{p_2} and PC_{p_3} as the other criteria tend to – unrealistically – over-parameterize the model in our case.

the model selection is performed in each recursive step. Moreover, the first lag of the dependent variable is added to the set of predictors in all procedures (and is subject to model selection in IV, V and VI). Further lags did not improve forecasting ability for any of the loss functions so we do not present those results here.

For each of the six forecast procedures we consider, the goal is to predict PI, IP, UN and SL under two alternatives of forecast evaluation. Table 1 summarizes the pseudo out-of-sample average forecast losses for each of the six forecast procedures. For each variable of interest, first $OLS - AL$ losses are presented and followed by the AL losses. Evaluating the forecasts by the asymmetric loss function of choice leads to lower average forecast losses with only a few exceptions.[12] AL provides forecast loss improvements in the range of 8% to 48% against $OLS - AL$.

We shape our analysis to forecast four macroeconomic variables with 'forecast-relevant predictors'. As shown in Table 1, selecting among all the factors included in the system results with smaller forecast losses. Comparisons of procedures I and IV, procedures II and V, and procedures III and VI point out that variable selection generally leads to smaller forecast losses under AL. The degree of the improvement in the forecast losses depends on the asymmetry parameter and the variable of interest. For instance, forecast procedures IV, V and VI lead smaller forecast errors than procedures I, II and III, except for IP when $\alpha = 0.9$ and UN when $\alpha = 0.1$. Given the small number of exceptions, the analysis provides reliable evidence for variable selection by the tailored information criterion being useful.

As we add volatility factors, namely when we compare procedures II and V, and procedures III and VI, we observe improvement in the forecast accuracy, e.g. all variables when $\alpha = 0.3$ under AL. In some cases, however, these additional factors combined with model selection does not change the forecast losses, e.g. for IP when $\alpha = 0.1, 0.3, 0.5$. This simply happens as the model selection eliminates the additional volatility factors. Therefore, procedure IV seems to provide the smallest forecast losses for most of the cases.

Our analysis is not designed to select an optimal asymmetry parameter, since α is imposed by the beneficiary of the forecast i.e. the corresponding loss preferences. Yet, our results can still deliver some useful insight on the matter. For forecasting personal income, $\alpha = 0.1$ appears to be the optimal value which leads to the smallest forecast losses for all cases. For the other three variables, $\alpha = 0.9$ results with the smaller forecast errors for industrial production and retail sales for all cases. One, of course, conduct a more detailed search to identify the optimal asymmetry parameter by grid search in our set up or a lá Elliott et al. (2005).

We additionally compare the $OLS - AL$ and AL forecasts with the help of the Diebold-Mariano

[12] The exceptions are some cases of IP and Sales when $\alpha = 0.7$. As reported in Table 2, the differences in the $OLS - AL$ and AL forecast losses of these exceptions are not statistically significant according to the Diebold-Mariano test. The test results are presented in Table 2.

Table 1: Forecast Losses (×100) Evaluated for OLS-Asymmetric Loss and Asymmetric Loss

α	Forecast Procedure	PI_{OLS-AL}	PI_{AL}	IP_{OLS-AL}	IP_{AL}	UN_{OLS-AL}	UN_{AL}	SL_{OLS-AL}	SL_{AL}
0.1	I	15.34	10.37	28.37	15.61	20.54	11.35	81.56	48.46
	II	14.85	10.50	27.37	15.77	20.42	11.42	84.32	48.27
	III	14.92	10.46	27.46	15.93	20.42	11.30	83.97	48.69
	IV	14.84	9.61	28.93	14.80	19.30	11.63	79.00	44.06
	V	14.75	9.73	28.93	14.80	19.30	11.63	79.00	44.09
	VI	14.79	9.72	28.93	14.80	19.30	11.63	79.01	44.08
0.3	I	15.53	14.16	25.02	20.98	20.13	18.13	74.84	65.27
	II	15.33	14.29	24.55	21.16	20.10	18.20	76.29	65.27
	III	15.37	14.28	24.65	21.28	20.11	18.17	76.18	65.61
	IV	15.32	13.82	24.88	19.91	19.41	17.44	71.97	62.04
	V	15.41	14.00	24.88	19.91	19.41	17.44	71.94	62.12
	VI	15.43	14.00	24.88	19.91	19.41	17.44	71.91	62.08
0.5	I	15.71	15.71	21.68	21.68	19.72	19.72	68.12	68.12
	II	15.81	15.81	21.73	21.73	19.79	19.79	68.27	68.27
	III	15.81	15.81	21.85	21.85	19.81	19.81	68.39	68.39
	IV	15.35	15.35	20.79	20.79	18.83	18.83	65.07	65.07
	V	15.55	15.55	20.79	20.79	18.83	18.83	65.57	65.57
	VI	15.53	15.53	20.79	20.79	18.83	18.83	65.52	65.52
0.7	I	15.90	15.68	18.33	18.76	19.32	17.69	61.40	61.38
	II	16.29	15.76	18.92	18.65	19.48	17.77	60.25	61.65
	III	16.25	15.77	19.04	18.77	19.51	17.83	60.60	61.33
	IV	15.41	15.34	16.85	18.28	18.28	16.86	59.19	59.29
	V	15.69	15.44	16.82	18.12	18.28	16.86	58.19	60.00
	VI	15.69	15.48	16.82	18.12	18.28	16.86	58.19	60.00
0.9	I	16.09	13.26	14.98	11.02	18.91	10.94	54.68	41.30
	II	16.77	13.42	16.10	10.75	19.16	11.08	52.22	41.49
	III	16.69	13.45	16.24	10.79	19.21	11.17	52.81	40.82
	IV	15.85	12.64	15.11	10.82	17.73	10.29	52.56	39.96
	V	16.28	12.73	17.13	10.86	17.73	10.29	49.98	40.39
	VI	16.22	12.72	17.25	10.90	17.73	10.29	49.85	40.25

Notes: The losses are evaluated using asymmetric quadratic loss functions within a recursive pseudo out of sample setup. See the text for details. All the forecast losses are multiplied by 100. The data set for factor extraction includes the dependent variables. Forecast procedures are as follows. I: 12 factors; II: 12 factors + volatility factor; III: 12 factors + volatility factor + squared volatility factor; IV: model selection with the information criteria on I; V: model selection with the information criteria on II; VI: model selection with the information criteria on III. The predictive regression includes the first lag of the dependent variables and it is subject to model selection in IV, V and VI. The forecast losses are the averages of forecast losses occurring in each recursive step.

[DM] test for predictive accuracy (Diebold and Mariano, 1995).[13] Since we compute the differences between AL and $OLS - AL$, we may expect test statistics to be smaller than -1.645 at the 5% significance level when AL is superior. The results are presented in Table 2. For all variables, the DM test statistics indicate the statistical superiority of AL for all αs except for PI, IP and SL when $\alpha = 0.7$. For $\alpha = 0.7$, the DM test does not favor one forecast procedure over another. These results are consistent with the observations from Table 1.

[13] The null hypothesis is that the expected forecast loss is equal for both procedures of interest, $\tilde{y}^{(1)}_{t+h}$ and $\tilde{y}^{(2)}_{t+h}$. The losses implied by these forecasts are $\mathcal{L}(\tilde{v}^{(1)}_{t+h})$ and $\mathcal{L}(\tilde{v}^{(2)}_{t+h})$. Under the null hypothesis, $H_0 : \mathrm{E}\left(\mathcal{L}(\tilde{v}^{(1)}_{t+h})\right) = \mathrm{E}\left(\mathcal{L}(\tilde{v}^{(2)}_{t+h})\right)$ or $H_0 : E(d_t) = 0$ where $d_t = \mathcal{L}(\tilde{v}^{(1)}_{t+h}) - \mathcal{L}(\tilde{v}^{(2)}_{t+h})$ is the loss differential, the DM test statistic is $S = \bar{d}/(\widehat{LRV}(\bar{d})/\bar{T})^{0.5} \sim N(0,1)$ where $\bar{T}$ is the number of forecast errors available for comparison and $\widehat{LRV}$ is an estimate of the asymptotic (long-run) variance of $\sqrt{\bar{T}}\bar{d}$.

Table 2: Test statistics of equal predictive accuracy of OLS and AL based forecasts for model selection with $AIC_{\mathcal{L}}$

α	Forecast Procedure	PI	IP	UN	SL
0.1	I	−3.16	−2.96	−3.42	−4.03
	II	−2.72	−2.67	−3.29	−4.33
	III	−2.81	−2.67	−3.32	−4.26
	IV	−3.13	−3.04	−3.48	−4.32
	V	−2.97	−3.04	−3.48	−4.32
	VI	−3.00	−3.04	−3.48	−4.32
0.3	I	−2.66	−2.43	−1.99	−3.20
	II	−1.92	−2.02	−1.88	−3.77
	III	−2.04	−2.02	−1.90	−3.66
	IV	−2.95	−2.99	−2.42	−3.55
	V	−2.65	−2.99	−2.42	−3.59
	VI	−2.69	−2.99	−2.42	−3.59
0.7	I	−0.67	0.40	−1.89	−0.01
	II	−1.40	−0.24	−2.00	0.76
	III	−1.29	−0.25	−1.98	0.40
	IV	−0.25	1.35	−1.83	0.06
	V	−0.70	1.22	−1.83	1.06
	VI	−0.61	1.22	−1.83	1.06
0.9	I	−3.32	−1.98	−4.00	−3.38
	II	−3.63	−2.72	−4.16	−2.91
	III	−3.55	−2.77	−4.16	−3.24
	IV	−3.93	−2.35	−3.81	−2.56
	V	−3.89	−3.01	−3.81	−2.05
	VI	−3.81	−3.01	−3.81	−2.05

Notes: The null hypothesis for the test is $H_0 : E[d_t] = 0$ where $d_t = \mathcal{L}(\hat{v}_{t+h}) - \mathcal{L}(\tilde{v}_{t+h})$ with $\hat{v}_{t+h}$ the forecast errors from OLS based forecasts and $\tilde{v}_{t+h}$ the asymmetric loss forecast errors. For the one-sided test with the alternative hypothesis $H_0 : E[d_t] > 0$, the test statistic should be smaller than -1.645 for 5% significance. Forecast procedures are as follows. I: 12 factors; II: 12 factors + volatility factor; III: 12 factors + volatility factor + squared volatility factor; IV: model selection with the information criteria on I; V: model selection with the information criteria on II; VI: model selection with the information criteria on III. The predictive regression includes the first lag of the dependent variables and it is subject to model selection in IV, V and VI. The table does not include the test statistics for $\alpha = 0.5$ because the results of $OLS - AL$ and AL are identical.

To explore on whether the results are generalisable to further predictands and specifications, we provide additional results in Appendix D.2. We compare our forecast model with commonly used benchmarks such as AR(4) and factor-augmented autoregressive predictor. In addition, we explore the performance of a quadratic principal component analysis in the spirit of Bai and Ng (2008c) as an additional way of dealing with the potential nonlinearity implied by our model of the volatility.

5 Concluding remarks

The forecasting literature often focusses on MSE-optimal forecasts. Yet there is evidence emphasising the relevance of more general loss functions in concrete situations. In this paper, we incorporate some aspects of forecasting under asymmetric loss functions in factor-based predictive regressions.

First, we show that one may estimate predictive regressions under the relevant loss by plugging in factors extracted from a data set by means of a first-step principal components analysis. The estimated optimal forecast from the feasible regression converges in probability to the theoretical optimal forecast.

Second, we address the relevance of the estimated factors by assessing whether they are forecast-relevant under *a given* loss function. To this end, we employ *tailored* information criteria and consider the factors with highest predictive powers for forecasting. Moreover, we argue that principal component analysis does not always extract all relevant information: we analyze *the variability* of the predictor series and include corresponding additional information in the forecasting model, namely a factor extracted from the log-squared idiosyncratic components estimated in the first-step PCA. Refinements such as targeting the predictors à la Bai and Ng (2008b) (see also Dias et al., 2010) are not considered here, but may of course be incorporated in the forecasting procedure.

We then illustrate the discussion by forecasting the Personal Income, Industrial Production, Unemployment Rate and Retail Sales. We resort to a recursive pseudo out-of-sample forecast evaluation procedure where the factors are extracted from a subset of the large data set in each step and used for forecasting one-year-ahead values of four variables under several asymmetric power loss functions. We compare six forecasting procedures for different parameter values when the $p = 2$ is fixed. Expectedly, fitting the forecasting model under the relevant loss function leads to smaller averaged losses compared to the case when we use MSE. Model selection taking the relevant loss into account leads to smaller forecast losses. Adding volatility information sometimes improves the forecasts.

Both our theoretical and empirical results underscore the importance of using forecast-relevant information by estimating factors from an auxiliary data set to exploit the additional information (i.e. the volatility factor in our case). Also relevant, if not even more so, is the issue of choosing the most relevant information for the particular loss function used to define optimality of the forecast.

Our results have important implications for practitioners especially in central banks. Our findings support the debate on evaluating the forecasts of macroeconomic variables under relevant loss function. Additionally, model selection should be considered as standard practise in factor-augmented forecasting exercises.

Acknowledgements

The authors would like to thank two anonymous referees, the former editor (Graham Elliott), as well as Andy Blake, Jörg Breitung, Gernot Müller, Marco del Negro, Gianni De Nicoló, Kerem Tuzcuoğlu, Mike West and the conference participants of the Central Bank Forecasting Conference 2017 for very helpful comments and suggestions. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. Petrow and F. Csake (Eds.), *Second International Symposium on Information Theory*, pp. 267–281. Akademiai Kiado.

Alessi, L., M. Barigozzi, and M. Capasso (2010). Improved penalization for determining the number of factors in approximate factor models. *Statistics & Probability Letters 80*(23), 1806–1813.

Andersen, P. and R. Gill (1982). Cox's regression model for counting processes: A large sample study. *Annals of Statistics 10*(4), 1100–1120.

Aretz, K., S. M. Bartram, and P. F. Pope (2011). Asymmetric loss functions and the rationality of expected stock returns. *International Journal of Forecasting 27*(2), 413–437.

Artis, M. and M. Marcellino (2001). Fiscal forecasting: The track record of the IMF, OECD and EC. *The Econometrics Journal 4*(1), 20–36.

Artis, M. J., A. Banerjee, and M. Marcellino (2005). Factor forecasts for the UK. *Journal of Forecasting 24*(4), 279–298.

Ayebo, A. and T. J. Kozubowski (2003). An asymmetric generalization of Gaussian and Laplace laws. *Journal of Probability and Statistical Science 1*(2), 187–210.

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics 12*(2), 171–178.

Bai, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica 71*(1), 135–171.

Bai, J. and S. Ng (2002). Determining the number of factors in approximate factor models. *Econometrica 70*(1), 191–221.

Bai, J. and S. Ng (2006). Confidence intervals for diffusion index forecasts and inference for factor-augmented regressions. *Econometrica 74*(4), 1133–1150.

Bai, J. and S. Ng (2008a). Extremum estimation when the predictors are estimated from large panels. *Annals of Economics & Finance 9*(2), 201–222.

Bai, J. and S. Ng (2008b). Forecasting economic time series using targeted predictors. *Journal of Econometrics 146*(2), 304–317.

Bai, J. and S. Ng (2008c). Forecasting economic time series using targeted predictors. *Journal of Econometrics 146*(2), 304 – 317. Honoring the research contributions of Charles R. Nelson.

Bai, J. and S. Ng (2011). Principal components estimation and identification of the factors. Working Paper 1112-02, Department of Economics, Columbia University.

Banerjee, A., M. Marcellino, and I. Masten (2008). Forecasting macroeconomic variables using diffusion indexes in short samples with structural change. Working Papers 334, IGIER (Innocenzo Gasparini Institute for Economic Research), Bocconi University.

Barigozzi, M. and M. Hallin (2016). Generalized dynamic factor models and volatilities: recovering the market volatility shocks. *The Econometrics Journal 19*(1), C33–C60.

Belloni, A. and V. Chernozhukov (2013). Least squares after model selection in high-dimensional sparse models. *Bernoulli 19*(2), 521–547.

Belviso, F. and F. Milani (2006). Structural factor-augmented VARs (SFAVARs) and the effects of monetary policy. *The B.E. Journal of Macroeconomics 6*(3), 1–46.

Boero, G., J. Smith, and K. Wallis (2008). Evaluating a three-dimensional panel of point forecasts: the Bank of England Survey of External Forecasters. *International Journal of Forecasting 24*(3), 354–367.

Boivin, J. and S. Ng (2006). Are more data always better for factor analysis? *Journal of Econometrics 132*(1), 169–194.

Capistrán, C. (2008). Bias in Federal Reserve inflation forecasts: Is the Federal Reserve irrational or just cautious? *Journal of Monetary Economics 55*(8), 1415–1427.

Christodoulakis, G. and E. Mamatzakis (2008). An assessment of the EU growth forecasts under asymmetric preferences. *Journal of Forecasting 27*(6), 483–492.

Christodoulakis, G. and E. Mamatzakis (2009). Assessing the prudence of economic forecasts in the EU. *Journal of Applied Econometrics 24*(4), 583–606.

Christoffersen, P. and F. Diebold (1996). Further results on forecasting and model selection under asymmetric loss. *Journal of Applied Econometrics 11*(5), 561–571.

Clark, T. E. (2009). Is the Great Moderation over? An empirical analysis. *Economic Review 94*(4), 5–42.

Clatworthy, M. A., D. A. Peel, and P. F. Pope (2012). Are analysts' loss functions asymmetric? *Journal of Forecasting 31*(8), 736–756.

Clements, M., F. Joutz, and H. Stekler (2007). An evaluation of the forecasts of the Federal Reserve: A pooled approach. *Journal of Applied Econometrics 22*(1), 121–136.

D'Agostino, A. and D. Giannone (2006). Comparing alternative predictors based on large-panel factor models. Working Paper Series 0680, European Central Bank.

Demetrescu, M. (2006). An extension of the Gauss–Newton algorithm for estimation under asymmetric loss. *Computational Statistics & Data Analysis 50*(2), 379–401.

den Reijer, A. H. J. (2005). Forecasting Dutch GDP using large scale factor models. DNB Working Papers 28, Netherlands Central Bank, Research Department.

Dias, F., M. Pinheiro, and A. Rua (2010). Forecasting using targeted diffusion indexes. *Journal of Forecasting 29*(3), 341–352.

Diebold, F. X. and R. S. Mariano (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics 13*(3), 253–263.

Elliott, G., I. Komunjer, and A. Timmermann (2005). Estimation and testing of forecast rationality under flexible loss. *Review of Economic Studies 72*(4), 1107–1125.

Elliott, G., I. Komunjer, and A. Timmermann (2008). Biases in macroeconomic forecasts: Irrationality or asymmetric loss? *Journal of the European Economic Association 6*(1), 122–157.

Elliott, G. and A. Timmermann (2004). Optimal forecast combinations under general loss functions and forecast error distributions. *Journal of Econometrics 122*(1), 47–79.

Engel, C., N. C. Mark, and K. D. West (2015). Factor model forecasts of exchange rates. *Econometric Reviews 34*(1-2), 32–55.

Fernandez, C., J. Osiewalski, and M. F. Steel (1995). Modeling and inference with v-spherical distributions. *Journal of the American Statistical Association 90*(432), 1331–1340.

Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2005). The generalized dynamic factor model: One-sided estimation and forecasting. *Journal of the American Statistical Association 100*(471), 830–840.

Fritsche, U., C. Pierdzioch, J.-C. Rülke, and G. Stadtmann (2015). Forecasting the Brazilian real and the Mexican peso: Asymmetric loss, forecast rationality, and forecaster herding. *International Journal of Forecasting 31*(1), 130–139.

Godbout, C. and M. J. Lombardi (2012). Short-term forecasting of the Japanese economy using factor models. Working Paper Series 1428, European Central Bank.

Granger, C. (1999). Outline of forecast theory using generalized cost functions. *Spanish Economic Review 1*(2), 161–173.

Granger, C. W. J. (1969). Prediction with a generalized cost of error function. *Operational Research Society 20*(2), 199–207.

Groen, J. J. J. and G. Kapetanios (2013). Model selection criteria for factor-augmented regressions. *Oxford Bulletin of Economics and Statistics 75*(1), 37–63.

Hacıoğlu Hoke, S. and K. Tuzcuoğlu (2016). Interpreting the latent dynamic factors by threshold FAVAR model. Staff Working Paper Series 622, Bank of England.

Komunjer, I. (2007). Asymmetric power distribution: Theory and applications to risk measurement. *Journal of Applied Econometrics 22*(5), 891–921.

Ludvigson, S. C. and S. Ng (2009). Macro factors in bond risk premia. *The Review of Financial Studies 22*(12), 5027–5067.

Ludvigson, S. C. and S. Ng (2010). *A Factor Analysis of Bond Risk Premia*. Number 313-372. Handbook of Empirical Economics and Finance.

Marcellino, M., J. H. Stock, and M. W. Watson (2003). Macroeconomic forecasting in the Euro area: Country specific versus area-wide information. *European Economic Review 47*(1), 1–18.

McCracken, M. W. and S. Ng (2015). FRED-MD: A Monthly Database for Macroeconomic Research. Working Papers 2015-12, Federal Reserve Bank of St. Louis.

McCullough, B. (2000). Optimal prediction with a general loss function. *Journal of Combinatorics, Information and System Sciences 25*(1–4), 207–221.

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica 59*(2), 347–370.

Patton, A. and A. Timmermann (2007a). Properties of optimal forecasts under asymmetric loss and nonlinearity. *Journal of Econometrics 140*(2), 884–918.

Patton, A. and A. Timmermann (2007b). Testing forecast optimality under unknown loss. *Journal of the American Statistical Association 102*(480), 1172–1184.

Pierdzioch, C., J. Rülke, and G. Stadtmann (2012). On the loss function of the Bank of Canada: A note. *Economics Letters 115*(2), 155–159.

Schumacher, C. (2007). Forecasting German GDP using alternative factor models based on large datasets. *Journal of Forecasting 26*(4), 271–302.

Sensier, M. and D. van Dijk (2004). Testing for volatility changes in U.S. macroeconomic time series. *The Review of Economics and Statistics 86*(3), 833–839.

Stock, J. H. and M. W. Watson (1998). Diffusion indexes. NBER Working Papers 6702, National Bureau of Economic Research.

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

Stock, J. H. and M. W. Watson (2002b). Has the business cycle changed and why? *NBER Macroeconomics Annual 17*(1), 159–218.

Stock, J. H. and M. W. Watson (2002c). Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics 20*(2), 147–62.

Stock, J. H. and M. W. Watson (2005). Implications of dynamic factor models for VAR analysis. NBER Working Papers 11467, National Bureau of Economic Research.

Stock, J. H. and M. W. Watson (2008, September). Phillips curve inflation forecasts. Working Paper 14322, National Bureau of Economic Research.

Tibshirani, R. (1994). Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society, Series B 58*(2), 267–288.

Tran, N. M., P. Burdejova, M. Osipenko, and W. K. Härdle (2016). Principal component analysis in an asymmetric norm. Working paper, SSRN.

Tsuchiya, Y. (2016). Assessing macroeconomic forecasts for japan under an asymmetric loss function. *International Journal of Forecasting 32*(2), 233 – 242.

Wang, Y. and T.-H. Lee (2014). Asymmetric loss in the Greenbook and the Survey of Professional Forecasters. *International Journal of Forecasting 30*(2), 235 – 245.

Weiss, A. A. (1996). Estimating time series models using the relevant cost function. *Journal of Applied Econometrics 11*(5), 539–560.

Weiss, A. A. and A. P. Andersen (1984). Estimating time series models using the relevant forecast evaluation criterion. *Journal of the Royal Statistical Society. Series A (General) 147*(3), 484–487.

Zhou, Z. (2013). Heteroscedasticity and autocorrelation robust structural change detection. *Journal of the American Statistical Association 108*(502), 726–740.

Appendix

A An information criterion

Following Akaike (1973), the definition of the information criterion in form of a penalized log-likelihood leads to

$$AIC\left(k\right) = -2\ln\left(\hat{L}\left(k\right)\right) + 2k$$

with $\hat{L}\left(k\right)$ denoting the maximum of the likelihood function for model complexity k.

Suppose now that the error term in the model of interest follows an asymmetric (exponential) power distribution as characterized by Ayebo and Kozubowski (2003) and Komunjer (2007)[14] with density function

$$f\left(v\right) = \frac{\delta^{\frac{1}{\lambda}}}{\sigma\Gamma\left(1+\frac{1}{\lambda}\right)} e^{-\delta\left(\frac{1}{\alpha_*^\lambda}I(v\leq 0)+\frac{1}{(1-\alpha_*)^\lambda}I(v>0)\right)\left|\frac{v}{\sigma}\right|^\lambda}$$

where $\delta = \frac{2\alpha_*^\lambda(1-\alpha_*)^\lambda}{\alpha_*^\lambda+(1-\alpha_*)^\lambda}$. Quasi-ML estimation of a regression model assuming $v_t \sim f$ is then easily shown to be equivalent to estimation under the loss function $\mathcal{L}$ with parameters $p = \lambda$ and $\alpha = \frac{(1-\alpha_*)^p}{(1-\alpha_*)^p+\alpha_*^p}$.

After concentrating out σ, some algebra leads to

$$AIC_{\mathcal{L}}\left(k\right) = \frac{2}{p}\ln\left(\sum\mathcal{L}\left(\hat{v}_{t+h}\right)\right) + \frac{2k}{T}$$

with $\hat{v}_t$ the residuals from estimation of the predictive regression under the relevant loss $\mathcal{L}$.

This reduces to the AIC when $\mathcal{L}$ is the squared-error loss function. Note that AIC$_{\mathcal{L}}$ differs from the IC proposed by (Weiss, 1996, Section 5) in two important respects. First, Weiss focusses on comparing forecasts from models based on different loss functions, while we are interested in selecting the best forecasting model for a given loss function; second, the expression he arrives at is not scale invariant, whereas, for the loss function in (2), AIC$_{\mathcal{L}}$ is.

To account for the fact that the predictive regression also uses extracted factors, which are noisy proxies of the true $X_{t,i}$, we follow Groen and Kapetanios (2013) and strengthen the penalty term by $\left(1+\frac{T}{N}\right)$. Denoting by k_1 the number of observed regressors and by k_2 the number of PCA-estimated ones (obviously, $k_1 + k_2 = k$), the final version of the tailored AIC is given by

$$AIC_{\mathcal{L}}\left(k\right) = \frac{2}{p}\ln\left(\sum\mathcal{L}\left(\hat{v}_{t+h}\right)\right) + \frac{2k_1}{T} + \frac{2k_2}{T}\left(1+\frac{T}{N}\right).$$

[14] They introduce asymmetry in the exponential power (also generalized power, or generalized error) distribution by using the method discussed in Fernandez et al. (1995). An alternative way of "skewing" the exponential power distribution is based on the approach of Azzalini (1985).

B Technical assumptions

Let v_{t+h}, $f_{t,k}$, $k = 1, \ldots, r$, and $u_{t,i}$, $i = 1, \ldots, N$, be piecewise locally stationary in the sense that $\exists 0 = \tau_0 < \tau_1 < \ldots < \tau_M < \tau_{M+1} = 1$ such that, for $\tau_m T < t \leq \tau_{m+1} T$,

$$u_{t,i} = G_{i,m}\left(\frac{t}{T}; \mathcal{F}_{t,i}\right) \quad \text{and} \quad (v_{t+h}, f_{t,1}, \ldots, f_{t,r})' = F_m\left(\frac{t}{T}; \mathcal{G}_t\right),$$

where $\mathcal{F}_{t,i} = \{\varepsilon_{i,t}, \varepsilon_{i,t-1} \ldots\}$ and $\mathcal{G}_t = \{\nu_t, \nu_{t-1} \ldots\}$ for $(\varepsilon_{1,t}, \ldots, \varepsilon_{N,t})' \in \mathbb{R}^N$ and $\nu_t \in \mathbb{R}^{r+1}$ mutually independent zero-mean iid sequences, and $G_{i,m} \mapsto \mathbb{R}$ and $F_m \mapsto \mathbb{R}^{r+1}$, $m = 1, \ldots, M$, are nonlinear filters satisfying the uniform (in i) Lipschitz condition

$$\begin{aligned} |G_{i,m}(r_1, \mathcal{F}_{0,i}) - G_{i,m}(r_2, \mathcal{F}_{0,i})| &\leq C|r_2 - r_1| \\ ||F_m(r_1, \mathcal{G}_0) - F_m(r_2, \mathcal{G}_0)|| &\leq C|r_2 - r_1| \quad \forall r_1, r_2 \in [\tau_m; \tau_{m+1}]. \end{aligned}$$

Furthermore, assume that v_{t+h}, $f_{t,k}$ and $u_{t,i}$ are uniformly (in i and t) L_q-bounded for some $q \geq \min\{2p, 8\}$. Also, let for $\kappa > 0$

$$\delta_u(\kappa, i, j) = \max_m \sup_{\tau_m T < t \leq \tau_{m+1} T} \left\| G_{i,m}\left(\frac{t}{T}; \mathcal{F}_{\kappa,i}\right) - G_{j,m}\left(\frac{t}{T}; \mathcal{F}^*_{\kappa,j}\right) \right\|_{\min\{2p,8\}}$$

with $\mathcal{F}^*_{t,j} = \left\{\varepsilon^*_{j,t}, \varepsilon^*_{j,t-1}, \ldots, \varepsilon^*_{j,1}, \mathcal{F}_{-1,j}\right\}$, where $\left(\varepsilon^*_{1,t}, \ldots, \varepsilon^*_{N,t}\right)'$ is an independent copy of the sequence $(\varepsilon_{1,t}, \ldots, \varepsilon_{N,t})'$, and

$$\delta_{v,f}(\kappa) = \max_m \sup_{\tau_m T < t \leq \tau_{m+1} T} \left\| F_m\left(\frac{t}{T}; \mathcal{G}_\kappa\right) - F_m\left(\frac{t}{T}; \mathcal{G}^*_\kappa\right) \right\|_{\min\{2p,8\}}$$

with analogous definition of $\mathcal{G}^*_\kappa$, and assume that

$$\delta_u(\kappa, i, j) \leq Ce^{-\kappa|i-j|} \quad \text{and} \quad \delta_{v,f}(\kappa) \leq Ce^{-\kappa}.$$

These weak serial dependence conditions allow e.g. for a law of large numbers as formalised in

Lemma 7 *Let z_t be piecewise locally stationary in the above sense and assume that $\mathrm{E}\left(g^2(z_t)\right)$ is uniformly bounded for some measurable function g. Then, $\frac{1}{T}\sum_{t=1}^T g(z_t) - \frac{1}{T}\sum_{t=1}^T \mathrm{E}(g(z_t)) \xrightarrow{p} 0$.*

Proof: *It is easily seen that $q_t = g(z_t) - \mathrm{E}(g(z_t))$ is zero-mean piecewise locally stationary in the above sense. The proof of Theorem 1 of Zhou (2013) establishes the stronger weak convergence of partial sums of the form $\frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]} q_t$ to a Gaussian process; this suffices for the result.*

Finally, let the loadings $\lambda_{i,k}$ satisfy Assumption B of Bai (2003). Together with the restrictions on cross-sectional dependence implied by the conditions on $\delta_u(\kappa, i, j)$, this will allow for consistent estimation of the factor space.

C Proof of Proposition 1

We first note that Assumptions A-E of Bai (2003) are fulfilled. Assumption A follows from Lemma 7 since $f_{t,k}^2$ have finite 4th order moments. The exponential decay of the serial and spatial dependence of the idiosyncratic errors translates to the covariances in Assumptions C and E, which are then easily seen to hold. Assumption D is fulfilled for factors independent of the idiosyncratic errors under our assumptions. This implies that $\sup_t \left| f_{t,k} - \hat{f}_{t,k} \right| \xrightarrow{p} 0$ for all k; see Bai's Proposition 2.

We now show that this vanishing estimation error does not affect the consistency of the parameter estimators in 7. Given the non-smoothness of the loss function for $p = 1$ and $p = 2, \alpha \neq 0.5$, we cannot apply e.g. the result of Bai and Ng (2008a) directly (their setup assumes smoothness) and must modify their arguments accordingly. The target function is given by

$$\begin{aligned}
&\mathcal{Q}\left(a_j^*, b_k^*, c^*, a_j, b_k, c\right) \\
&= \frac{1}{T} \sum_{t=p+1}^{T-h} \mathcal{L}\left(y_{t+h} - c^* - \sum_{j=1}^{q} a_j^* y_{t-j+1} - \sum_{k=1}^{r} b_k^* \hat{f}_{t,k} \right) \\
&= \frac{1}{T} \sum_{t=p+1}^{T-h} \mathcal{L}\left(v_{t+h} - (c^* - c) - \sum_{j=1}^{q} \left(a_j^* - a_j\right) y_{t-j+1} - \sum_{k=1}^{r} \left(b_k^* - b_k\right) f_{t,k} + \sum_{k=1}^{r} b_k^* \left(f_{t,k} - \hat{f}_{t,k}\right) \right).
\end{aligned}$$

In a first step, we show that

$$\begin{aligned}
&\mathcal{Q}\left(a_j^*, b_k^*, c^*, a_j, b_k, c\right) \\
&= \frac{1}{T} \sum_{t=p+1}^{T-h} \mathcal{L}\left(v_{t+h} - (c^* - c) - \sum_{j=1}^{q} \left(a_j^* - a_j\right) y_{t-j+1} - \sum_{k=1}^{r} \left(b_k^* - b_k\right) f_{t,k} \right) + o_p(1)
\end{aligned}$$

where the $o_p(1)$ term is uniform in t as follows.

Let

$$q_t = v_{t+h} - (c^* - c) - \sum_{j=1}^{q} \left(a_j^* - a_j\right) y_{t-j+1} - \sum_{k=1}^{r} \left(b_k^* - b_k\right) f_{t,k}$$

and $\Delta q_t = \sum_{k=1}^{r} b_k^* \left(f_{t,k} - \hat{f}_{t,k}\right)$.

For $p = 1$, $\mathcal{L}$ is Lipschitz such that

$$\left|\mathcal{L}\left(q_t + \Delta q_t\right) - \mathcal{L}\left(q_t\right)\right| \leq C \left|\Delta q_t\right|,$$

which can be re-written as

$$\mathcal{L}\left(q_t + \Delta q_t\right) = \mathcal{L}\left(q_t\right) + C\xi_t$$

where $|\xi_t| \leq |\Delta q_t|$.

For $p > 1$, use a Taylor expansion of order $p-1$ with the rest term in differential form, we have that

$$\mathcal{L}\left(q_t+\Delta q_t\right)=\mathcal{L}\left(q_t\right)+\mathcal{L}'\left(q_t\right)\Delta q_t+\ldots+\frac{1}{(p-1)!}\mathcal{L}^{(p-1)}\left(q_t+\xi_t\right)\left(\Delta q_t\right)^{p-1}$$

where again $|\xi_t| \leq |\Delta q_t|$ (for $p=2$, this is just the Mean Value Theorem). Summing up, we obtain

$$\begin{aligned}&\left|\mathcal{Q}\left(a_j^*,b_k^*,c^*,a_j,b_k,c\right)-\frac{1}{T}\sum_{t=p+1}^{T-h}\mathcal{L}\left(q_t+\Delta q_t\right)\right|\\&\quad\leq\sum_{j=1}^{p-2}\frac{1}{j!}\frac{1}{T}\sum_{t=p+1}^{T-h}\mathcal{L}^{(j)}\left(q_t\right)\left|\left(\Delta q_t\right)^j\right|+\frac{1}{(p-1)!}\frac{1}{T}\sum_{t=p+1}^{T-h}\mathcal{L}^{(p-1)}\left(q_t+\xi_t\right)\left|\left(\Delta q_t\right)^{p-1}\right|.\end{aligned}$$

Note that $\mathcal{L}^{(p-1)}$ is Lipschitz continuous, so we have that

$$\left|\mathcal{L}^{(p-1)}\left(q_t+\xi_t\right)-\mathcal{L}^{(p-1)}\left(q_t\right)\right|\leq C\left|\xi_t\right|\leq C\left|\Delta q_t\right|$$

and it follows that

$$\begin{aligned}&\left|\mathcal{Q}\left(a_j^*,b_k^*,c^*,a_j,b_k,c\right)-\frac{1}{T}\sum_{t=p+1}^{T}\mathcal{L}\left(q_t+\Delta q_t\right)\right|\\&\quad\leq C\sum_{j=1}^{p-2}\frac{1}{T}\sum_{t=p+1}^{T-h}\mathcal{L}^{(j)}\left(q_t\right)\left|\Delta q_t\right|^j+C\frac{1}{T}\sum_{t=p+1}^{T-h}\mathcal{L}^{(p-1)}\left(q_t\right)\left|\Delta q_t\right|^{p-1}+C\frac{1}{T}\sum_{t=p+1}^{T-h}\left|\Delta q_t\right|^p.\end{aligned}$$

We have from Proposition 2 in Bai (2003) that $\sup_t\left|f_{t,k}-\hat{f}_{t,k}\right|\stackrel{p}{\to}0$ for all k, so we immediately obtain that $\sup_t\left|\Delta q_t\right|^j\stackrel{p}{\to}0$, such that

$$\frac{1}{T}\sum_{t=p+1}^{T}\left|\Delta q_t\right|^p\stackrel{p}{\to}0.$$

Moreover, for all $1\leq j\leq p-1$,

$$\frac{1}{T}\sum_{t=p+1}^{T-h}\mathcal{L}^{(p-1)}\left(q_t\right)\left|\Delta q_t\right|^{p-1}\leq\sup_t\left|\Delta q_t\right|^{p-1}\frac{1}{T}\sum_{t=p+1}^{T-h}\mathcal{L}^{(j)}\left(q_t\right)\stackrel{p}{\to}0$$

since $\mathcal{L}^{(j)}\left(q_t\right)\leq C\left|q_t\right|^j$ for suitable C, and q_t has finite pth order moments (because v_{t+h}, y_t and $f_{t,k}$ do), such that, thanks to the Markov's inequality, $\frac{1}{T}\sum_{t=p+1}^{T-h}\left|q_t\right|^j$ is uniformly bounded in probability.

Then, we resort to a suitable law of large numbers to establish that

$$\frac{1}{T}\sum_{t=p+1}^{T}\mathcal{L}\left(q_t\right)-\frac{1}{T}\sum_{t=p+1}^{T-h}\mathrm{E}\left(\mathcal{L}\left(q_t\right)\right)\stackrel{p}{\to}0$$

pointwise in the parameter space. To this end note that y_t is a stable AR filtering of v_{t+h} and $\sum_{k=1}^{r} b_k f_{t,k}$, so $y_t, y_{t-1}, \ldots, f_{t,k}, v_{t+h}$ is easily checked to be a piecewise locally stationary process in the sense of Appendix B, and that the finiteness of $\mathrm{E}\left(\mathcal{L}^2\left(q_t\right)\right)$ is given since

$$\mathcal{L}^2\left(q_t\right) \leq C\left|q_t\right|^{2p},$$

where the expectation of the r.h.s. is finite whenever $\left\|q_t\right\|_{2p} = \sqrt[2p]{\mathrm{E}\left(\left|q_t\right|^{2p}\right)}$ is finite. But Minkowski's inequality indicates that $\left\|q_t\right\|_{2p}$ is finite whenever the L_{2p} norm of y_t and $f_{t,k}$ is finite, which is the case under the assumptions in Appendix B.

Hence, Lemma 7 indicates that

$$\mathcal{Q}\left(a_j^*, b_k^*, c^*, a_j, b_k, c\right) - \frac{1}{T}\sum_{t=p+1}^{T-h} \mathrm{E}\left(\mathcal{L}\left(v_{t+h} - \left(c^* - c\right) - \sum_{j=1}^{q}\left(a_j^* - a_j\right) y_{t-j+1} - \sum_{k=1}^{r}\left(b_k^* - b_k\right) f_{t,k}\right)\right) \xrightarrow{p} 0$$

pointwise. Since $\mathcal{L}$ is convex, Lemma II.1 of Andersen and Gill (1982) applies such that the above convergence is uniform on any compact set.

Finally, we only have to check that the above expectation is minimized for $a_j^* = a_j$, $b_k^* = b_k$ and $c^* = c$. This is a standard argument; given the continuity of $\mathcal{Q}$, consistency of the estimators $\tilde{a}_j$, $\tilde{b}_k$ and $\tilde{c}$ follows via the continuity of the argmin operator w.r.t. the sup norm. To this end, note that, since the generalized forecast error is a martingale difference sequence with no atom at the origin, it holds that

$$\arg\min_{v^*} \mathrm{E}\left(\mathcal{L}\left(v_{t+h} - v^* \middle| y_{t-j}, f_{t,k}\right)\right) = 0$$

uniquely, implying that, for any $v^* \neq 0$ and all t,

$$\begin{aligned} \mathrm{E}\left(\mathcal{L}\left(v_{t+h} - v^*\right)\right) &= \mathrm{E}\left(\mathrm{E}\left(\mathcal{L}\left(v_{t+h} - v^* \middle| y_{t-j}, f_t\right)\right)\right) \\ &> \mathrm{E}\left(\mathrm{E}\left(\mathcal{L}\left(v_{t+h} \middle| y_{t-j}, f_{t,k}\right)\right)\right) = \mathrm{E}\left(\mathcal{L}\left(v_{t+h}\right)\right) \end{aligned}$$

such that $\mathrm{E}\left(\mathcal{L}\left(v_{t+h} - \sum_{j=1}^{q}\left(a_j^* - a_j\right) y_{t-j+1} - \sum_{k=1}^{r}\left(b_k^* - b_k\right) F_{t,k}\right)\right)$ must be minimized at each t for $\sum_{j=1}^{q}\left(a_j^* - a_j\right) y_{t-j+1} - \sum_{k=1}^{r}\left(b_k^* - b_k\right) f_{t,k} = 0$ which, with y_t and $f_{t,k}$ linearly independent stochastic processes, is only the case when $a_j^* - a_j = b_k^* - b_k = 0$ for all $1 \leq j \leq q$ and $1 \leq k \leq r$. Since the minimizers a_j^* and b_k^* are the same at each t, they will also minimize

$$\frac{1}{T}\sum_{t=p+1}^{T-h} \mathrm{E}\left(\mathcal{L}\left(v_{t+h} - \left(c^* - c\right) - \sum_{j=1}^{q}\left(a_j^* - a_j\right) y_{t-j+1} - \sum_{k=1}^{r}\left(b_k^* - b_k\right) f_{t,k}\right)\right).$$

The consistency of the forecast function then follows.

D Additional empirical results

D.1 Different forecast horizons

This section presents additional results for the forecasting under asymmetric loss function. The exercise is set up exactly the same as in Section 4 except for the forecast horizon. The same variables are now forecast under the same asymmetric loss function for different forecast procedures for forecast horizons $h = 3$ and $h = 6$. The forecast results and associated DM tests are presented in Table D.1 to Table D.4.

In summary, evaluating the forecast under asymmetric loss leads to smaller forecast losses in the recursive out-of-sample forecasting exercise even under different forecast horizons. All the other results, i.e. the comparison of different cases and the DM statistics, remain qualitative the same.

Table D.1: Forecast Losses (×100) Evaluated for OLS-Asymmetric Loss and Asymmetric Loss for $h = 3$

α	Forecast Procedure	PI_{OLS-AL}	PI_{AL}	IP_{OLS-AL}	IP_{AL}	UN_{OLS-AL}	UN_{AL}	SL_{OLS-AL}	SL_{AL}
0.1	I	15.03	10.41	23.98	14.45	20.07	10.42	74.06	44.83
	II	15.38	10.53	23.57	14.23	19.65	10.50	76.72	44.45
	III	15.23	10.23	23.19	14.07	19.45	10.50	77.10	45.71
	IV	15.54	9.81	28.44	16.43	18.74	10.63	73.40	41.77
	V	15.54	9.81	28.44	16.43	18.74	10.63	73.40	41.77
	VI	15.43	9.65	28.44	16.43	18.74	10.63	73.40	41.77
0.3	I	15.65	14.51	22.76	20.24	19.11	16.62	70.27	63.17
	II	15.85	14.61	22.64	20.36	18.89	16.64	71.74	63.17
	III	15.72	14.46	22.35	20.11	18.81	16.68	72.10	64.06
	IV	15.65	14.22	23.71	20.11	18.88	17.01	68.39	60.10
	V	15.65	14.22	23.71	20.11	18.88	17.01	68.78	60.76
	VI	15.65	14.23	23.71	20.11	18.88	17.01	68.77	60.72
0.5	I	16.26	16.26	21.54	21.54	18.16	18.16	66.48	66.48
	II	16.32	16.32	21.71	21.71	18.13	18.13	66.77	66.77
	III	16.20	16.20	21.52	21.52	18.18	18.18	67.09	67.09
	IV	15.77	15.77	20.80	20.80	18.44	18.44	64.68	64.68
	V	15.77	15.77	20.80	20.80	18.44	18.44	64.75	64.75
	VI	15.78	15.78	20.80	20.80	18.44	18.44	65.34	65.34
0.7	I	16.88	16.36	20.32	19.61	17.20	16.36	62.70	59.56
	II	16.79	16.36	20.79	19.70	17.37	16.28	61.80	60.09
	III	16.69	16.28	20.68	19.54	17.54	16.32	62.09	59.85
	IV	15.90	15.55	17.20	18.72	18.00	16.58	60.64	59.09
	V	15.90	15.55	17.20	18.72	18.00	16.58	60.51	59.15
	VI	15.90	15.55	17.20	18.72	18.00	16.58	60.78	59.31
0.9	I	17.49	13.86	19.10	13.62	16.25	10.10	58.91	39.34
	II	17.26	13.79	19.86	13.57	16.62	9.99	56.83	39.60
	III	17.17	13.86	19.85	13.51	16.90	10.01	57.09	38.85
	IV	15.90	12.77	13.37	11.74	17.73	10.28	54.79	39.00
	V	15.91	12.78	13.48	11.65	17.73	10.28	54.10	38.93
	VI	15.91	12.78	13.55	11.68	17.73	10.28	54.10	38.77

Notes: The losses are evaluated using asymmetric quadratic loss functions within a recursive pseudo out of sample setup. See the text for details. All the forecast losses are multiplied by 100. The data set for factor extraction includes the dependent variables. Forecast procedures are as follows. I: 12 factors; II: 12 factors + volatility factor; III: 12 factors + volatility factor + squared volatility factor; IV: model selection with the information criteria on I; V: model selection with the information criteria on II; VI: model selection with the information criteria on III. The predictive regression includes the first lag of the dependent variables and it is subject to model selection in IV, V and VI. The forecast losses are the averages of forecast losses occurring in each recursive step.

Table D.2: Test statistics of equal predictive accuracy of OLS and AL based forecasts for $h = 3$ for model selection with $AIC_{\mathcal{L}}$

α	Cases	PI	IP	UN	SL
0.1	I	−3.99	−4.02	−7.31	−4.84
	II	−4.20	−4.07	−7.03	−5.36
	III	−4.12	−4.09	−6.92	−5.27
	IV	−5.10	−3.83	−7.29	−4.82
	V	−5.10	−3.83	−7.29	−4.82
	VI	−5.02	−3.83	−7.29	−4.82
0.3	I	−2.56	−2.42	−5.07	−3.56
	II	−2.80	−2.28	−4.67	−4.34
	III	−2.80	−2.27	−4.42	−4.03
	IV	−4.07	−2.91	−4.31	−3.88
	V	−4.07	−2.91	−4.31	−4.08
	VI	−4.06	−2.91	−4.31	−4.10
0.7	I	−1.18	−0.72	−1.84	−2.01
	II	−0.96	−1.11	−2.42	−1.11
	III	−0.92	−1.16	−2.69	−1.50
	IV	−1.12	1.76	−3.26	−1.03
	V	−1.12	1.76	−3.26	−0.92
	VI	−1.10	1.76	−3.26	−0.99
0.9	I	−3.02	−2.78	−5.05	−4.48
	II	−2.73	−3.21	−5.62	−4.04
	III	−2.64	−3.21	−5.79	−4.34
	IV	−3.06	−1.01	−6.45	−3.74
	V	−3.06	−1.13	−6.45	−3.62
	VI	−3.06	−1.18	−6.45	−3.66

Notes: The null hypothesis for the test is $H_0 : E[d_t] = 0$ where $d_t = \mathcal{L}(\hat{v}_{t+h}) - \mathcal{L}(\tilde{v}_{t+h})$ with $\hat{v}_{t+h}$ the forecast errors from OLS based forecasts and $\tilde{v}_{t+h}$ the asymmetric loss forecast errors. For the one sided test with the alternative hypothesis $H_0 : E[d_t] > 0$, the test statistic should be smaller than -1.645 for 5% significance. Forecast procedures are as follows. I: 12 factors; II: 12 factors + volatility factor; III: 12 factors + volatility factor + squared volatility factor; IV: model selection with the information criteria on I; V: model selection with the information criteria on II; VI: model selection with the information criteria on III. The predictive regression includes the first lag of the dependent variables and it is subject to model selection in IV, V and VI. The table does not include the test statistics for $\alpha = 0.5$ because the results of $OLS - AL$ and AL are identical.

Table D.3: Forecast Losses (×100) Evaluated for OLS-Asymmetric Loss and Asymmetric Loss for $h = 6$

α	Cases	PI_{OLS-AL}	PI_{AL}	IP_{OLS-AL}	IP_{AL}	UN_{OLS-AL}	UN_{AL}	SL_{OLS-AL}	SL_{AL}
0.1	I	14.82	10.62	26.53	15.54	19.36	10.44	78.55	48.58
	II	14.85	10.77	25.03	15.58	18.96	10.49	79.71	48.20
	III	15.06	10.78	25.01	15.51	19.18	10.51	79.01	48.69
	IV	15.06	10.17	31.63	15.39	18.78	10.83	76.64	44.36
	V	15.06	10.17	31.63	15.39	18.78	10.80	76.85	43.38
	VI	15.06	10.17	31.63	15.39	18.78	10.84	76.73	43.43
0.3	I	15.53	14.70	25.07	21.98	18.73	16.68	72.96	65.34
	II	15.57	14.78	24.21	21.94	18.52	16.66	73.39	64.75
	III	15.73	14.84	24.34	22.10	18.68	16.76	73.09	65.12
	IV	15.64	14.38	27.38	21.79	18.84	16.93	71.00	61.43
	V	15.64	14.40	27.38	21.79	18.84	16.93	70.95	61.49
	VI	15.64	14.40	27.38	21.79	18.84	16.93	70.68	61.20
0.5	I	16.24	16.24	23.62	23.62	18.09	18.09	67.37	67.37
	II	16.29	16.29	23.40	23.40	18.09	18.09	67.06	67.06
	III	16.39	16.39	23.67	23.67	18.18	18.18	67.18	67.18
	IV	15.98	15.98	23.01	23.01	18.29	18.29	64.88	64.88
	V	15.96	15.96	23.01	23.01	18.29	18.29	64.90	64.90
	VI	16.01	16.01	23.01	23.01	18.29	18.29	64.44	64.44
0.7	I	16.95	16.15	22.16	21.70	17.45	16.12	61.79	59.67
	II	17.01	16.18	22.58	21.26	17.65	16.15	60.74	59.62
	III	17.05	16.35	23.00	21.58	17.68	16.22	61.27	59.44
	IV	16.10	15.75	18.79	20.35	17.74	16.37	58.51	58.20
	V	16.15	15.79	18.80	20.22	17.74	16.37	58.51	58.20
	VI	16.16	15.84	18.80	20.22	17.74	16.37	58.30	57.67
0.9	I	17.66	13.67	20.70	14.31	16.81	9.89	56.20	38.88
	II	17.73	13.66	21.76	13.61	17.22	9.94	54.42	38.96
	III	17.72	13.96	22.33	14.17	17.18	9.98	55.35	38.85
	IV	15.80	12.64	13.97	11.60	17.19	10.05	52.43	38.60
	V	15.93	12.70	16.06	11.27	17.19	10.05	52.41	38.64
	VI	15.87	12.65	16.06	11.27	17.19	10.05	52.86	38.33

Notes: The losses are evaluated using asymmetric quadratic loss functions within a recursive pseudo out of sample setup. See the text for details. All the forecast losses are multiplied by 100. The data set for factor extraction includes the dependent variables. Forecast procedures are as follows. I: 12 factors; II: 12 factors + volatility factor; III: 12 factors + volatility factor + squared volatility factor; IV: model selection with the information criteria on I; V: model selection with the information criteria on II; VI: model selection with the information criteria on III. The predictive regression includes the first lag of the dependent variables and it is subject to model selection in IV, V and VI. The forecast losses are the averages of forecast losses occurring in each recursive step.

Table D.4: Test statistics of equal predictive accuracy of OLS and AL based forecasts for $h = 6$ for model selection with $AIC_{\mathcal{L}}$

α	Cases	PI	IP	UN	SL
0.1	I	−2.99	−3.43	−5.12	−5.04
	II	−2.87	−2.99	−4.97	−5.35
	III	−2.93	−2.96	−5.10	−5.07
	IV	−3.50	−3.87	−5.37	−4.87
	V	−3.50	−3.87	−5.40	−4.86
	VI	−3.50	−3.87	−5.36	−4.84
0.3	I	−1.86	−2.39	−3.38	−3.81
	II	−1.71	−1.77	−3.14	−4.47
	III	−1.88	−1.72	−3.24	−3.94
	IV	−2.89	−3.63	−3.63	−4.50
	V	−2.83	−3.63	−3.63	−4.45
	VI	−2.83	−3.63	−3.63	−4.43
0.7	I	−2.00	−0.38	−2.87	−1.48
	II	−2.02	−1.08	−3.25	−0.85
	III	−1.69	−1.16	−3.20	−1.31
	IV	−1.08	1.52	−2.89	−0.20
	V	−1.11	1.39	−2.89	−0.20
	VI	−0.96	1.39	−2.89	−0.39
0.9	I	−3.14	−2.36	−6.10	−4.28
	II	−3.11	−2.94	−6.46	−3.96
	III	−2.79	−3.02	−6.48	−4.22
	IV	−3.08	−1.63	−5.83	−2.90
	V	−3.11	−2.57	−5.83	−2.89
	VI	−3.08	−2.57	−5.83	−3.03

Notes: The null hypothesis for the test is $H_0 : E[d_t] = 0$ where $d_t = \mathcal{L}(\hat{v}_{t+h}) - \mathcal{L}(\tilde{v}_{t+h})$ with $\hat{v}_{t+h}$ the forecast errors from OLS based forecasts and $\tilde{v}_{t+h}$ the asymmetric loss forecast errors. For the one sided test with the alternative hypothesis $H_0 : E[d_t] > 0$, the test statistic should be smaller than -1.645 for 5% significance. Forecast procedures are as follows. I: 12 factors; II: 12 factors + volatility factor; III: 12 factors + volatility factor + squared volatility factor; IV: model selection with the information criteria on I; V: model selection with the information criteria on II; VI: model selection with the information criteria on III. The predictive regression includes the first lag of the dependent variables and it is subject to model selection in IV, V and VI. The table does not include the test statistics for $\alpha = 0.5$ because the results of $OLS - AL$ and AL are identical.

D.2 Comparing forecasting models

The forecasting model we explore in Section 4 is a way of dealing with the forecasting under asymmetric loss function. To check the robustness of our finding, we present the forecast errors of different forecasting models under the generalised loss function in Equation (2). Additionally we further explore a different strategy to account for the potential nonlinearity leading to the inclusion of the volatility factor.

We consider four models. Model 1 is an autoregressive regression with 4 lags, AR(4), as näive predictor, see Stock and Watson (2002a), Stock and Watson (2002c), Stock and Watson (2008). Model 2 is an AR(1) model which also includes a factor extracted in a given recursive step. Under Model 3, we have two sub-cases. In Model 3.1, we have 12 factors extracted from the data set composed by the original series and their squared values, alongside the AR(1) component (see Bai and Ng (2008c)).[15] Model 3.2 performs model selection out of 12 factors extracted in Model 3.1 and the first lag by making use of our tailored selection criteria. The rest follows the set up in Section 4. The results are presented in Table D.5.

Our interest lies in finding out which model performs better, i.e. which one leads to smaller forecast losses. Our analysis in Section 4 concludes that forecasting under asymmetric loss combined with model selection leads to the smallest forecast errors in most of the cases. Here the same conclusion holds. Model 3.2 provides the smallest forecast errors in majority of the cases. In comparison with Model 1, AR(4), which is considered as the forecasting benchmark, Model 3.2 provides smaller forecast errors in 28 cases out of 40, for both $OLS - AL$ and AL combined for all variables and all αs.

[15] The number of factors here is an arbitrary choice to stay consistent with the analysis in Section 4. The same analysis can be as well done with different number of factors.

Table D.5: Forecast Losses (×100) of Different Models Evaluated for OLS-Asymmetric Loss and Asymmetric Loss

α	PI_{OLS-AL}	PI_{AL}	IP_{OLS-AL}	IP_{AL}	UN_{OLS-AL}	UN_{AL}	SL_{OLS-AL}	SL_{AL}
				Model 1: AR(4)				
0.1	15.18	10.02	30.98	16.60	17.91	10.90	81.78	41.76
0.3	15.35	13.89	26.50	21.79	18.72	17.46	73.23	60.14
0.5	15.51	15.51	22.03	22.03	19.54	19.54	64.68	64.68
0.7	15.68	15.46	17.55	18.70	20.35	18.26	56.13	59.79
0.9	15.85	12.94	13.08	10.57	21.16	11.92	47.57	40.82
				Model 2: AR(1) and 1 factor				
0.1	15.34	9.62	29.40	15.29	19.92	11.19	78.18	42.05
0.3	15.37	13.73	25.52	21.00	19.37	17.41	71.23	60.25
0.5	15.39	15.39	21.64	21.64	18.83	18.83	64.29	64.29
0.7	15.41	15.32	17.76	18.65	18.28	16.86	57.34	58.86
0.9	15.43	12.66	13.88	10.90	17.73	10.29	50.40	39.79
				Model 3.1: 12 factors extracted from the sample and squared sample				
0.1	15.76	11.20	29.12	19.02	20.48	11.75	82.08	47.92
0.3	15.91	14.71	26.71	23.88	20.36	18.72	75.35	65.83
0.5	16.07	16.07	24.30	24.30	20.24	20.24	68.63	68.63
0.7	16.23	16.01	21.89	21.32	20.12	17.98	61.90	61.65
0.9	16.38	14.14	19.48	13.25	20.00	11.51	55.17	41.67
				Model 3.2: Model selection for Model 3.1				
0.1	14.78	9.95	29.93	14.78	19.76	11.65	77.30	43.70
0.3	15.19	13.56	24.16	19.23	19.85	17.99	71.84	62.10
0.5	15.06	15.06	19.98	19.98	18.97	18.97	65.33	65.33
0.7	15.15	14.99	16.43	17.11	18.20	16.89	59.74	59.80
0.9	15.57	12.43	16.39	10.13	17.44	10.19	52.26	39.99

Notes: *Model 1*: AR(4), *Model 2* : AR(1) and 1 factor, *Model 3.1* : 12 factors extracted from the data composed by the original series and their squares and *Model 3.2* : factor selection among data in Model 3.1. Estimation is done in recursive manner and average losses are calculated for all variables given α for the loss function Equation (2) under fixed $p = 2$.

D.3 Extracted factors, detailed table for Figure 1

Table D.6 shows the factors selected for each variable and α in the exercise outlined in Section 4.2. As shown in the columns of Table D.6, not all factors in each step are selected as predictors. For example, for forecasting PI, when $\alpha = 0.1$, only second, fourth, seventh and eighth factors are selected among the first eight PCA-extracted factors. For the same variable, when $\alpha = 0.5$, only the second, fourth and eighth are identified as forecast relevant. For all α, it turns out that the first 8 factors given by the information criteria are not all forecast-relevant. Increasing the number of factors to select from one by one, the factors that are already selected occasionally change but new ones appear to be selected. In the last step of our exercise, we contemplate all 12 PCA-extracted factors and note that some of the additional ones appear to be forecast relevant, while some of the commonly used first 8 factors do not. The set of selected factors change within different αs.

Table D.6: Factors selected for all predictor series by $\text{AIC}_{\mathcal{L}}$; full data span
including the first lag of the dependent variable and number of factors increasing gradually from 8 to 12

	$\alpha = 0.1$					$\alpha = 0.3$					$\alpha = 0.5$					$\alpha = 0.7$					$\alpha = 0.9$				
	8	**9**	**10**	**11**	**12**	**8**	**9**	**10**	**11**	**12**	**8**	**9**	**10**	**11**	**12**	**8**	**9**	**10**	**11**	**12**	**8**	**9**	**10**	**11**	**12**
Personal Income	y_{t-1}	y_{t-1}	y_{t-1}	y_{t-1}	y_{t-1}	y_{t-1}	y_{t-1}	y_{t-1}	y_{t-1}	y_{t-1}	2	2	2	y_{t-1}	y_{t-1}	2	2	2	2	2	y_{t-1}	y_{t-1}	y_{t-1}	y_{t-1}	y_{t-1}
	2	2	2	2	2	2	2	2	2	2	4	4	4	2	2	5	5	5	5	5	5	5	5	5	5
	4	4	4	4	4	4	4	4	4	4	8	9	10	4	4	8	9	10	10	10	8	8	8	8	8
	7	7	7	7	7	7	7	7	7	7				10	10				11	12		9	10	10	10
	8	9	10	10	10	8	9	10	10	10				11	12									11	12
				11	12				11	12															
Industrial Production	y_{t-1}	y_{t-1}	y_{t-1}	y_{t-1}	y_{t-1}	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	y_{t-1}	y_{t-1}	y_{t-1}	y_{t-1}	y_{t-1}
	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	4	4	4	4	4
	2	2	2	2	2	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	5	5	5	5	5
	3	3	3	3	3	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	7	6	6	6	6
	4	4	4	4	4	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5		7	7	7	7
	5	5	5	5	5	7	7	7	7	7	6	6	6	6	6	6	6	6	6	6		8	9	9	9
	8	9	6	6	6	8	9	9	9	9	7	7	7	7	7	7	7	7	7	7					
			9	9	9			10	11	12	8	9	9	9	9	8	9	9	9	9					
			10	11	11								10	11	12			10	11	12					
					12																				
Unemployment Rate	y_{t-1}	y_{t-1}	y_{t-1}	y_{t-1}	y_{t-1}	y_{t-1}	y_{t-1}	y_{t-1}	y_{t-1}	y_{t-1}	y_{t-1}	y_{t-1}	y_{t-1}	y_{t-1}	y_{t-1}	y_{t-1}	y_{t-1}	y_{t-1}	y_{t-1}	y_{t-1}	y_{t-1}	y_{t-1}	y_{t-1}	y_{t-1}	y_{t-1}
	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	3	3	3	3	3	3	3	3	3	3
	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	4	4	4	4	4	4	4	4	4	4
	5	5	5	5	5	6	6	6	6	6	3	3	3	3	3	6	6	6	6	6	6	6	6	6	6
	6	6	6	6	6						4	4	4	4	4	8	9	10	11	12					
	7	7	7	7	7																				
Retail Sales	1	1	1	1	1	1	1	1	1	1	5	5	5	5	5	2	2	2	2	2	2	2	2	2	2
	3	3	3	3	3	4	4	4	4	4						5	5	5	5	5	5	5	5	5	5
	4	4	4	4	4											6	6	6	6	6	6	6	6	7	7
																					7	7	7	10	10

Notes: Asymmetric quadratic loss; see the text for details. The number of factors considered given a particular α are given in bold. The analysis includes the first 8 to 12 factors and additionally the first lag of the dependent variable. The analysis is conducted for the whole time span.