Multi-step forecasting with large vector autoregressions[*]

Andreas Pick[†] Matthijs Carpay[‡]

December 2020

Abstract

This paper investigates the performance of different dimension reduction approaches for large vector autoregressions in multi-step ahead forecasts. We consider factor augmented VAR models using principal components and partial least squares, random subset regression, random projection, random compression, and estimation via LASSO and BVAR. We compare the accuracy of iterated and direct multi-step point and density forecasts. The comparison is based on macroeconomic and financial variables from the FRED-MD data base. Our findings suggest that random subspace methods and LASSO estimation deliver the most precise forecasts.

Keywords Multi-step forecasting, VAR, dimension reduction, density forecasting.

1 Introduction

This paper investigates the ability of different dimension reduction techniques for large vector autoregressions to deliver accurate multi-step point and density forecasts. The background to this study is the increasing number of methods that have been proposed for dimension reduction for vector autoregressive (VAR) models. We provide a comprehensive review of their forecast accuracy.

An early use of dimension reduction in a large VAR is the paper by Bernanke et al. (2005) who use factors estimated via principal components methods in a VAR model together with the variable of interest. They call

[*] The authors would like to thank participants at the RMSE workshop and the ISCEF conference, in particular our discussant, Gilles Dufrénot, for valuable comments. Additionally, we thank Tom Boot, Agnieszka Markiewicz, Didier Nibbering, and Richard Paap for helpful discussions. We thank SURFsara for access to the Lisa Computer Cluster. The opinions expressed are those of the authors and should not be attributed to DNB or Egeria.

[†] Corresponding author. Erasmus University Rotterdam and De Nederlandsche Bank. andreas.pick@cantab.net

[‡] Egeria

this the factor augmented VAR (FAVAR). Factor extraction by principal components implies that the factors reflect the information in the regressor set but not their importance for the dependent variable. Partial least squares (PLS), in contrast, extract factors that target the dependent variable, and the results of Groen and Kapetanios (2016) suggest that PLS may deliver better forecasts in the absence of a strong factor structure.

An alternative to factor models is compressing the variables of the large data set via random compression introduced by Donoho (2006), which was extended to a Bayesian setting by Guhaniyogi and Dunson (2015). Random draws of compression matrices are used to reduce the dimension of the data set. It can be shown that the information that is of importance for forecasts is retained with only limited loss. This approach has been used by Koop et al. (2019) in a Bayesian VAR. In place of the non-standard distribution of the random compression weights, Boot and Nibbering (2019) propose the use of standard normally distributed weights in the random matrix for dimension reduction and call this the random projection approach.

Boot and Nibbering (2019) also investigate complete and random subset regression proposed by Elliott et al. (2013). Complete subset regression constructs forecasts from large data sets by averaging the forecasts from all possible combinations of small dimensional models that can be selected from the large data set. The downside of complete subset regression is that the number of small dimensional models can be prohibitively large. Random subset regression draws a number of these small dimensional models at random and averages the forecasts. Boot and Nibbering (2019) show theoretically that the loss from using only a number of models decreases quickly in the number of randomly drawn models. Our work sheds light on the ability of these different random subspace methods to produce accurate forecasts.

The LASSO estimator of Tibshirani (1996) combines the estimation and dimension reduction step by using a penalized estimation where the L_1 norm of the parameter vector is constrained. This causes small parameters to be set to zero, which is equivalent to eliminating the corresponding variables and leads to automatic dimension reduction.

Bayesian estimation for VAR models has been extended to large VARs by Mol et al. (2008) and Bańbura et al. (2010). They show that using appropriate priors, Bayesian methods can deliver similar forecast performance compared to factor models. The choice of priors for Bayesian VAR (BVAR) models has been investigated by Giannone et al. (2015). Based on this work, Koop et al. (2019) evaluate the forecast performance of Bayesian VAR using random compression and time varying parameters on seven variables using iterated forecasts. Their results suggest that BVARs provide more accurate forecasts than factor models.

While the literature has considered the forecasting ability of subsets of these approaches at different horizons, the multi-step nature of the forecasts has typically not been investigated. This contrasts with the literature on univariate AR and small VAR models, which has dealt with this topic extensively. Multi-step forecasts can be constructed in an iterative or a direct manner and Marcellino et al. (2006) and Pesaran et al. (2011) evaluate the forecasting performance of the two approaches empirically. They find that, while on average the iterated approach tends to be preferred, no approach dominates for all variable categories. In this paper, we revisit this issue for the case of large VAR models. It is generally thought that direct forecasts are more precise than iterated forecasts when the model is misspecified. However, the results of Pesaran et al. (2011) suggest that the misspecification has to be very large for this to be the case. Given that dimension reduction potentially introduces an additional level of model misspecification, it is important to assess in how far this changes the trade-off between the two multi-step forecasting methods.

McCracken and McGillicuddy (2019) compare direct and iterated forecasts for conditional forecasts in VAR models, that is, the forecasts of, say, inflation conditional on an assumed future path of monetary policy. They find that conditional forecasts generally have similar properties to unconditional forecasts. However, McCracken and McGillicuddy (2019) also find that, when restricting the sample to the Great Moderation, the direct approach yields more precise forecasts.

A source of possible model misspecification is structural instability, which has been suggested as a source of forecast failure by, among others, Stock and Watson (1996), Pesaran et al. (2006), Koop and Potter (2007), Giacomini and Rossi (2009) and Inoue and Rossi (2011). We evaluate in how far allowing for structural breaks influences the predictive ability of dimension reduction techniques for multi-step forecasting. We use the robust optimal weighting scheme introduced by Pesaran et al. (2013) to account for structural breaks. Our findings suggest that accounting for structural breaks does not change the relative performance of the different methods.

Much of the literature on VAR forecasting has focused on point forecasts. Over the recent years, however, a large literature has developed that considers the properties of density forecasts. See Tay and Wallis (2000) and Corradi and Swanson (2006) for surveys. These developments were partially driven by the adoption of density forecasts as a communication tool for monetary policy, such as the Bank of England's fan charts. We will therefore consider the density forecasting accuracy of the dimension reduction methods in addition to their point forecast accuracy.

This paper is structured as follows. In the next section, we introduce the dimension reduction methods, robust optimal weights for structural breaks, and the construction of density forecasts. Section 3 discusses the empirical setup of our investigation, including a description of the data set, and Section 4 contains the results. The conclusion in Section 5 summarizes the findings.

2 Forecasting approaches

Consider the VAR model

$$\boldsymbol{\Phi}(L)\mathbf{z}_t = \boldsymbol{\mu} + \boldsymbol{\varepsilon}_t$$

where $\boldsymbol{z}_t$ is a $K \times 1$ vector of endogenous variables, $\boldsymbol{\mu}$ is a $K \times 1$ intercept vector, and $\boldsymbol{\varepsilon}_t$ is a $K \times 1$ vector of disturbances. A complication for estimation and forecasting is that the number of parameters in the polynomial lag matrices, $\boldsymbol{\Phi}(L)$, increases non-linearly in K to a point where the number of observations in standard data sets do not allow efficient estimation. As a result, a number of approaches have been developed to mitigate the estimation problem by reducing the dimension of the data set while retaining most of its information. For a review of the statistical literature on dimension reduction see Ma and Zhu (2013).

Assume that the purpose is to forecast one variable, y_t, using its own lags and the lags of the K_x dimensional vector $\boldsymbol{x}_t$, such that $\boldsymbol{z}_t = (y_t, \boldsymbol{x}_t')'$. The idea is then to forecast y_t using a lower dimensional vector $\tilde{\boldsymbol{x}}_t = \boldsymbol{R}\boldsymbol{x}_t$ that retains most of the information in $\boldsymbol{x}_t$, where $\boldsymbol{R}$ is an $M \times K_x$ matrix with $M < K_x$,. The different methods discussed in this paper differ in the matrix $\boldsymbol{R}$ that is used to reduce the dimensionality of $\boldsymbol{x}_t$. The first two methods, the factor augmented VAR and partial least squares, use dimension reduction matrices that are determined by the properties of the data. The three methods described thereafter use random dimension reduction matrices. Next, we discuss the LASSO, which uses a penalized estimation procedure for the parameter vector, which can however be reinterpreted as zero-restrictions on $\boldsymbol{R}$. Finally, we discuss Bayesian estimation where priors address the estimation problem of the large dimensional data set.

The factor models and the random subspace methods allow for dimension reduction at different stages. First, common factors can be extracted from the variables, $\boldsymbol{x}_t$, before they enter the VAR. We refer to this as dimension reduction in the variable space. Second, the extraction of the factors can be done from the regressor matrix of each equation in the VAR, that is, variables enter with their lags into the dimension reduction. We refer to this as dimension reduction in the variable-lag space. Both procedures have been used in the literature. For example, Bernanke et al. (2005) extract factors using the principal components of the variable space and Boot and Nibbering (2019) employ random subset regression and random projection in the variable-lag space. A priori it is not clear which method is superior. Extracting factors, for example, in the variable space will retain the temporal lag structure in the VAR. However, the dimension reduction methods may be more efficient in extracting the relevant temporal structure from the larger regressor matrix. In the empirical application below, we will use both approaches and compare their relative forecast accuracy.

2.1 Dimension reduction techniques

2.1.1 Factor augmented VAR

The factor augmented VAR (FAVAR) model of Bernanke et al. (2005) replaces the full vector $\boldsymbol{x}_t$ by a set of factors $\boldsymbol{f}_t$. The factors are extracted using principal components as suggested by Stock and Watson (2002). The included factors correspond to the largest eigenvalues of the covariance matrix of $\boldsymbol{x}_t$ and summarize the variation of the data in the direction of the main axes of the space spanned by the covariance matrix of $\boldsymbol{x}_t$.

The resulting factor augmented VAR model is

$$\boldsymbol{\Phi}^{(f)}(L)\mathbf{z}_t^{(f)} = \boldsymbol{\mu} + \nu_t \tag{1}$$

where $\mathbf{z}_t^{(f)} = (y_t, \boldsymbol{f}_t')'$ and $\boldsymbol{\Phi}^{(f)}(L)$ is the corresponding polynomial lag matrix. The dimension reduction matrix, $\boldsymbol{R}_{(f)}$, is a function of the eigenvectors associated with the largest eigenvalues of the covariance matrix of $\boldsymbol{x}_t$.

The number of factors to include in the FAVAR model is a crucial ingredient. This choice can be made based on economic considerations or statistical measures, such as information criteria or cross-validation. In our empirical application, we will use cross-validation to determine the number of factors.

2.1.2 Partial least squares

A potential weakness of the FAVAR approach is that the factors summarize the main variation of $\boldsymbol{x}_t$ without regard to the importance of the factors for y_t. If y_t is largely determined by a factor that is less important for the variables in $\boldsymbol{x}_t$ it will likely be omitted from the FAVAR. PLS, in contrast, selects the factors that are most highly correlated with the dependent variable.

Similar to the factors in the FAVAR approach above, the factors in the PLS approach are weighted averages of the variables in $\boldsymbol{x}_t$. However, while in the FAVAR approach the weights are the eigenvectors associated with the largest eigenvalues of the covariance matrix of $\boldsymbol{x}_t$, PLS uses the correlations with the dependent variable as the weights. The most common representation of PLS is the following algorithm by Helland (1990).

Initialize by setting $v_t = y_t - \frac{1}{T}\sum_{t=1}^{T} y_t$ and $q_{it} = x_{it} - \frac{1}{T}\sum_{t=1}^{T} x_t,\ i = 1, 2, \ldots, K_x$. Then iterate of the following steps M times, where M is the number of factors used in the forecast.

(i) Calculate $\boldsymbol{w}_j = (w_{j1}, w_{j2}, \ldots, w_{jK_x})$, $w_{ji} = \boldsymbol{v}'\boldsymbol{q}_i/(T-1)$, and construct factor j as $\boldsymbol{f}_j = \boldsymbol{X}\boldsymbol{w}_j$, where $\boldsymbol{v} = (v_1, v_2, \ldots, v_T)'$, $\boldsymbol{q}_i = (q_{i1}, q_{i2}, \ldots, q_{iT})'$, and $\boldsymbol{X} = (\boldsymbol{x}_1', \boldsymbol{x}_2', \ldots, \boldsymbol{x}_T')'$.

(ii) Calculate the residuals from regressing $\boldsymbol{v}$ and $\boldsymbol{q}_i$ on factor $\boldsymbol{f}_j$, $\tilde{\boldsymbol{v}} = \left(\boldsymbol{I} - \boldsymbol{f}_j(\boldsymbol{f}_j'\boldsymbol{f}_j)^{-1}\boldsymbol{f}_j'\right)\boldsymbol{v}$ and $\tilde{\boldsymbol{q}}_i = \left(\boldsymbol{I} - \boldsymbol{f}_j(\boldsymbol{f}_j'\boldsymbol{f}_j)^{-1}\boldsymbol{f}_j'\right)\boldsymbol{q}_i$. Set $\boldsymbol{v} = \tilde{\boldsymbol{v}}$ and $\boldsymbol{q}_i = \tilde{\boldsymbol{q}}_i$, for $i = 1, 2, \ldots K_x$.

By calculating the residuals in each step the orthogonality of the factors is ensured. The factors are then used in a VAR such as (1). Again, the number of factors is required, which we determine via cross-validation.

2.1.3 Random compression

Random compression uses a compression matrix, $\boldsymbol{R}_{(c)}^{(r)}$, to reduce the dimension of $\boldsymbol{x}_t$, where $r = 1, 2, \ldots, R$ are different draws of the compression matrix. Forecasts resulting using $\tilde{\boldsymbol{x}}_t^{(r)} = \boldsymbol{R}_{(c)}^{(r)}\boldsymbol{x}_t$ are averaged over the R draws with equal weights to yield the random compression forecast.

Each element $\rho_{ij}^{(r)}$ of $\boldsymbol{R}_{(c)}^{(r)}$ is drawn from the following distribution

$$\begin{aligned} \Pr\left(\rho_{ij}^{(r)} = \frac{1}{\sqrt{\varphi^{(r)}}}\right) &= \varphi^{(r)2} \\ \Pr\left(\rho_{ij}^{(r)} = 0\right) &= 2(1-\varphi^{(r)})\varphi^{(r)} \\ \Pr\left(\rho_{ij}^{(r)} = -\frac{1}{\sqrt{\varphi^{(r)}}}\right) &= (1-\varphi^{(r)})^2 \end{aligned}$$

where $i = 1, 2, \ldots, M_{(c)}^{(r)}$ and $j = 1, 2, \ldots, K_x$, $\varphi^{(r)}$ is drawn from a uniform distribution, $\mathrm{U}(0.02, 0.98)$ (Achlioptas 2003; Guhaniyogi and Dunson 2015). The columns of $\boldsymbol{R}_{(c)}^{r}$ are then orhonormalized using Gram-Schmidt orhonormalization to achieve unit lengths in the rows. Guhaniyogi and Dunson (2015) show in a Bayesian settings that predictions using the above distribution converge to the true predictive density. The row dimension, $M_{(c)}$ can be drawn from a uniform distribution. However, we obtained better results by using cross-validation to determine $M_{(c)}$, which also makes the three random methods comparable.

2.1.4 Random projection

A similar approach to random compression has been introduced by Boot and Nibbering (2019). They suggest to draw the elements, $\rho_{ij}^{(r)}$, of $\boldsymbol{R}_{(p)}^{(r)}$ from a standard normal distribution

$$\rho_{ij}^{(r)} \sim \mathrm{N}(0, 1)$$

where the dimension of $\boldsymbol{R}_{(p)}^{(r)}$, $M_{(p)}$ is now a choice parameter, which we select via cross-validation in the application.

Random compression and random projection both use random weights to achieve dimension reduction. However, while random compression excludes some variables in each draw, random projection uses weighted averages of all variables in each draw. Both methods require standardization of the regressors.

2.1.5 Random subset regression

Random subset regression randomly draws $M_{(s)}$ predictors from the data set. Each draw uses a $M_{(s)} \times K_x$ selection matrix, $\boldsymbol{R}_{(s)}^{(r)}$. Random subset regression is a randomized version of complete subset regression proposed by Elliott et al. (2013). Complete subset regression uses every possible combination of $M_{(s)}$ regressors from the set of K_x regressors to create forecasts and the forecasts are then averaged with equal weights. With large K_x this approach may be computationally infeasible. Boot and Nibbering (2019) show that drawing R subsets approximates the forecast from complete subset regression even for moderately large R.

2.1.6 LASSO

The LASSO of Tibshirani (1996) minimizes the squared residuals subject to a penalization of the sum of the absolute values of the coefficients

$$\min_{\boldsymbol{\theta}} \left(\frac{1}{T} \sum_{t=1}^{T} (y_t - \boldsymbol{\beta}_0 - \boldsymbol{\beta}_1' \boldsymbol{z}_{t-1} - \cdots - \boldsymbol{\beta}_p' \boldsymbol{z}_{t-p})^2 + \kappa \, |\boldsymbol{\theta}| \right)$$

where $\boldsymbol{\theta} = (\boldsymbol{\beta}_0', \boldsymbol{\beta}_1', \ldots, \boldsymbol{\beta}_p')'$ and $| \cdot |$ denotes the L_1 norm. The use of the L_1 penalization implies that small coefficients are set to zero, which is equivalent to zero restrictions in the dimension reduction matrix $\boldsymbol{R}$. The penalization constant κ needs to be set by the researcher and we choose it via cross-validation.

2.1.7 BVAR

The use of priors to facilitate the estimation of large VAR models has been discussed by Mol et al. (2008) and Bańbura et al. (2010). Their results suggest that the Minnesota prior of Doan et al. (1984) and Litterman (1986) produces accurate forecasts. The original specification of the Minnesota prior assumes that the variables are independent random walks. The first order autoregressive coefficient is therefore a priori one and all other coefficients a priori zero. An alternative specification sets the first order autoregressive coefficient to zero, which implies that the variable is a priori white noise.

Formally, the prior is defined as

$$\mathrm{E}[\boldsymbol{\Phi}(1)_{ij}] = \begin{cases} \delta_i, & j = i \\ 0, & \text{otherwise} \end{cases} \qquad \text{and} \qquad \mathrm{Var}[\boldsymbol{\Phi}(1)_{ij}] = \begin{cases} \frac{\lambda^2}{k^2}, & j = i \\ \vartheta \frac{\lambda^2 \sigma_i^2}{k^2 \sigma_j^2}, & \text{otherwise} \end{cases}$$

where $\boldsymbol{\Phi}(1)_{ij}$ is the ij-element of the first lag coefficient matrix. The BVAR requires a number of parameters to be set a priori: the choice between random walk and white noise, that is $\delta_i = 1$ or 0, the amount of shrinkage of the remaining coefficient is determined by λ, which ranges between 0 and ∞, and the importance of variable across equations given by $\vartheta \in (0, 1)$. We set $\vartheta = 1$ a priori as do, for example, Bańbura et al. (2010). We determine the prior parameters, δ_i and λ via cross-validation. The estimation of the BVAR with Minnesota prior can be achieved via OLS in combination with dummy observations, which makes it computationally efficient.

Our iterated forecasts from the BVAR are constructed by, first, estimating the parameter and, second, plugging them into the forecasting equation. In a Bayesian setting, this is a pseudo-iterated forecast that approximates the mean of the posterior predictive density, which would take the distribution of the parameters into account. Bańbura et al. (2010) find, however, that the forecasting performance of the pseudo-iterated forecast is essentially the same as evaluating the posterior predictive density. As this is also the approach to construct the remaining, frequentist forecast, it allows us to compare the forecasts on an equal basis.

Several other classes of priors exist for BVAR models, such as the Dirichlet-Laplace prior, the horseshoe prior, the normal-gamma prior, and the stochastic search variable selection prior. Cross et al. (2020) show that none of these priors beats the Minnesota prior when forecasting macroeconomic data. For this reason, we restrict our attention to the Minnesota prior.

2.2 Iterated and direct forecasts

When forecasting multiple periods ahead with a VAR, direct and iterated versions of the forecast can be obtained. The two options have been discussed by a large literature summarized by Marcellino et al. (2006) and Pesaran et al. (2011). Iterated forecasts will use a given VAR model for all horizons and iterate the model to obtain forecasts beyond the one-step ahead forecast. Direct forecasts, in contrast, use a different model for each horizon. As iterated forecasts use the one-step ahead model, they use the largest amount of data whereas direct forecasts typically require a larger pre-sample for larger forecast horizons. Given a correctly specified model, the iterated forecast is therefore more efficient in finite samples. However, it is thought that direct forecasts may be more robust to misspecification because iterated forecasts use powers of the autoregressive matrices, which could exacerbate potential biases. Empirically Marcellino et al. (2006) and

Pesaran et al. (2011) find that iterated forecast deliver more precise forecast for a majority of time series. Given that the dimension reduction techniques potentially introduce a source of misspecification as they only deliver approximations to the true DGP, it will be interesting to see whether this improves direct forecasts relative to their iterated counterparts.

2.3 Time varying parameter forecasts

Stock and Watson (1996) investigate the prevalence of structural breaks among macroeconomic and financial time series and find that a substantial number of series have one or more structural breaks. While the modeling of structural breaks is not the focus of this paper, we need to ensure that our results are not driven by structural breaks. We therefore construct forecasts that use robust optimal weighting of observations developed by Pesaran et al. (2013) to ensures that forecasts are robust against possible structural breaks.

The idea of robust optimal weighting is as follows. For a linear regression model with a break in the parameter vector, one can determine an optimal weighting scheme for the observations such that the mean square forecast error is minimized in expectations. However, the weights will depend on the time and size of the break, which in practice are unknown. One can estimate the time and size of the break using, for example, the test of Boot and Pick (2020). However, the parameter uncertainty that is introduced by using estimated break parameters in the weights leads to a deterioration of forecast accuracy.

An alternative to putting point estimates of the time and size of the break into the weights, is to integrate the weights with respect to a uniformly distributed break time. This leads to the following weights, which do not depend on any of these parameters,

$$w_t^* = \frac{-\log(1 - t/T)}{T-1}, \text{ for } t = 1, 2, \ldots, T-1 \tag{2}$$

$$w_T^* = \frac{\log(T)}{T-1} \tag{3}$$

and $w_t = \frac{w_t^*}{\sum_{s=1}^{T} w_s^*}$. An observation at time t is then multiplied by the weight w_t and the parameter estimates of the forecast equation are obtained as normally, for example, using least squares or maximum likelihood estimators.

The weights in (2) and (3) give the highest weight to the most recent observation and reduce the weight smoothly for observations further in the past. The intuition is that the further an observation is in the past, the more likely it is that a break in the parameters has occurred after this period. Therefore less weight should be placed on observations further in the past compared to more recent ones, which are less likely to be before a break point.

The weights bear resemblance to exponential smoothing weights in the tradition of Holt (1957). However, in contrast to exponential smoothing, the robust optimal weights do not require the choice of a nuisance parameter, the down-weighting coefficient. In our application, we also experimented with exponential smoothing. While the results are qualitatively similar, they are sensitive to the choice of down-weighting coefficient. We therefore restrict attention to the results using robust optimal weights.

2.4 Density forecasting

So far, the discussion implicitly focused on point forecasts. An alternative, however, is to consider density forecasts. Density forecasts require a decision on the implied distribution. A popular distribution for density forecasts is the two-piece normal distribution, which is, for example, the basis for the Bank of England fan charts (Elliott and Timmermann 2016). The two-piece normal density uses the formulation of the normal distribution but with different variances on either side of the mean.

The density is given as

$$p(y_{T+h}|\hat{y}_{T+h|T}, \sigma_{h,1}, \sigma_{h,1}) = \begin{cases} \frac{\exp\left[-(y_{T+h}-\hat{y}_{T+h|T})^2/2\sigma^2_{h,1}\right]}{\sqrt{2\pi}(\sigma_{h,1}+\sigma_{h,2})/2} & \text{for } y_{T+1} \leq \hat{y}_{T+h|T} \\ \frac{\exp\left[-(y_{T+h}-\hat{y}_{T+h|T})^2/2\sigma^2_{h,2}\right]}{\sqrt{2\pi}(\sigma_{h,1}+\sigma_{h,2})/2} & \text{for } y_{T+1} > \hat{y}_{T+h|T} \end{cases}$$

The two-piece normal requires the additional estimation of the two variances, $\sigma^2_{h,1}$ and $\sigma^2_{h,2}$, which are estimated as

$$\begin{aligned} \hat{\sigma}^2_{h,1} &= \omega \left[\sum_{y_{t+h}:y_{t+h}<\hat{y}_{t+h|t}} (y_t - \hat{y}_{t+h|t})^2 \right]^{2/3} \\ \hat{\sigma}^2_{h,2} &= \omega \left[\sum_{y_{t+h}:y_{t+h}\geq\hat{y}_{t+h|t}} (y_t - \hat{y}_{t+h|t})^2 \right]^{2/3} \end{aligned}$$

where

$$\omega = \left[\sum_{y_{t+h}:y_{t+h}<\hat{y}_{t+h|t}} (y_t - \hat{y}_{t+h|t})^2 \right]^{1/3} + \left[\sum_{y_{t+h}:y_{t+h}\geq\hat{y}_{t+h|t}} (y_t - \hat{y}_{t+h|t})^2 \right]^{1/3}$$

see John (1982). Note that the variances are estimated for each variable and forecasting method separately but for simplicity of notation we do not add further subscripts. We start the estimation of the variances with a sample of forecasts that starts two years before the forecast evaluation period and, for subsequent forecasts we add the additional variances in the expanding estimation sample.

The two-piece normal distribution nests the normal distribution. So if the two variances were, in fact, identical it would be more efficient to use the normal distribution. We can, ex post, evaluate the equality of the variances using a sign test. The equality of the variances is rejected for all the variables and all forecasting methods at the 5% level. A concern could be that the estimates of the variances are autocorrelated, and we repeat the sign test for skip sampled variances, sampling a variance every twelve months. The sign tests still reject the null of equal variances in 84% of forecasting methods and variables. Allowing for different variances and using the two-piece normal therefore appears the prudent choice.

An alternative to the two-piece normal would be to derive the distribution from the forecasting model, which requires making distributional assumption at that stage. Given the results from the sign tests, we feel the more conservative approach is justified.

Finally, for the evaluation of density forecasts we need to consider the choice of loss function. Here, we use log score of the density forecasts, $\log p(y_{T+h}|\hat{y}_{T+h|T}, \sigma^2_{h,1}, \sigma^2_{h,2})$, which is the most popular scoring rule in economic forecasting according to Elliott and Timmermann (2016). However, a range of alternative loss functions is available that weight forecast errors in different parts of the distribution differently; see the discussions by Gneiting and Raftery (2007) and by Elliott and Timmermann (2016).

3 Forecasting exercise

We apply the different forecasting methods to data from the FRED-MD data base of McCracken and Ng (2016). We use the vintage from September 2017 and transform the data as suggested by McCracken and Ng (2016). The data set contains 126 variables with monthly observations from January 1959 to August 2017. We focus on forecasting 14 variables: (1) industrial production, (2) unemployment rate, (3) non-farm employees, (4) federal funds rate, (5) ten-year treasury bond yield, (6) PPI, (7) CPI, (8) RPI, (9) housing starts, (10) real personal consumption expenditure, (11) real M2, (12) trade weighted dollar exchange rate, (13) S&P 500, (14) VXO. In order to forecast these variables, we use all of the 126 series as regressors. The period up to June 1986 spans the first estimation window and we use expanding estimation windows for subsequent forecasts. This yields 363 forecasts for each forecasting method and forecast horizon.

Each forecasting method requires a number of tuning parameters, which we determine by cross-validation: FAVAR and PLS require the number of factors, random subset regression, random projection, and random compression require the number of variables selected in each draw or generated in the projection or compression, LASSO requires the shrinkage coefficient κ, and the BVAR requires the parameters δ and λ. We use the cross-validation sample from July 1981 to June 1986, generate pseudo-out-of-sample forecasts for a range of settings of the tuning parameters and select the tuning parameter for each method that minimizes the MSFE for each h-step ahead forecast separately. In the baseline results, we use tuning parameters by pooling the MSFE of all series relative to that of the prevailing mean forecast. This has the advantage that a few odd forecasts cannot dominate the choice of tuning parameters for a given series. Only for the BVAR do we use series specific tuning parameters as the choice of random walk versus white noise cannot be averaged across series. We also check in how far the results differ when using separate tuning parameters for each series for the other methods. To anticipate the results: the differences are minor.

For all methods we include 13 lags and leave it to the dimension reduction methods to make the best use of this information. The first lag of the dependent variable is always included separately in the factor and random subspace methods. For the random subspace methods we use 1000 iterations, which was also used by Boot and Nibbering (2019). We conducted several trial forecasts with more iterations but changes of the forecasts are minimal after 1000 iterations.

We add two benchmark models to the forecasting exercise. The first is the prevailing mean model

$$\hat{y}^{\mathrm{PM}}_{T+h|T} = \frac{1}{T}\sum_{t=1}^{T} y_t$$

The second is a univariate autoregressive model of order 13, where the lag order is the same as the large VAR methods.

All point forecasts are evaluated using the MSFE

$$\mathrm{MSFE}_{ij} = \frac{1}{T - T_0 - h + 1}\sum_{t=T_0+h}^{T}\left(y_{t+h,j} - \hat{y}^{(i)}_{t+h,j}\right)^2$$

where i denotes the forecasting method: prevailing mean, AR, FAVAR, PLS, random subset, random projection, random compression, LASSO, and BVAR, T_0 is the last observation in the estimation sample, and j denotes the different series considered. We report the MSFE relative to that of the prevailing mean model, which results in the relative MSFE

$$\mathrm{RelMSFE}_{ij} = \mathrm{MSFE}_{ij}/\mathrm{MSFE}_{\mathrm{PM},j}$$

Tables 1 and 2 report the average of the relative MSFEs across series, $\frac{1}{N}\sum_{j=1}^{N}\mathrm{RelMSFE}_{ij}$.

Density forecasts are evaluated using the log score, $\log p(y_{T+h}|\hat{y}_{T+h}, \hat{\sigma}_{1h}, \sigma_{2h})$. Similar to the MSFE above, we report the log score for each method as a ratio of the average log score of the benchmark, prevailing mean model. Tables 5 and 6 again report the average of the relative log scores across series.

For both evaluation criteria, we calculate the test statistic of equal predictive accuracy of Diebold and Mariano (1995) and report the number of series where, for a given pair of methods, the null of equal predictive accuracy was rejected.

4 Results

4.1 Point forecasts

4.1.1 Average forecast accuracy

Table 1 reports the average relative MSFE of the different forecasting methods with the results for the equally weighted, iterated multi-step ahead forecasts in the top panel and results for equally weighted direct forecasts in the second panel. The most accurate one-step ahead forecast, given in the first line, is from the the random subset regression in the variable space with an average relative MSFE of 0.711. Random subset regression in the variable-lag space, random projection in the variable space, and LASSO produce similarly accurate forecasts. FAVAR and PLS produce noticeably less precise forecasts.

The most accurate multi-step ahead forecasts are random subset regression ($h = 3$) and random projection where the best $h = 6$ forecast is a direct forecast but the best $h = 12$ forecast an iterated forecast. LASSO, BVAR, and univariate AR yield forecasts that are close in precision. The forecasts of the FAVAR and PLS are again considerably less precise.

The comparison of iterated against direct forecasts is quite balanced. AR and LASSO are more precise when using the iterated approach and BVAR is similar for both approaches. For the factor models and random subspace models an interesting pattern emerges: methods that reduce dimension in the variable-lag space favor direct forecasts, methods that reduce dimension in the variable space favor the iterated approach.

Compared to the univariate AR, only random projection—iterated when reducing dimension of the variable space and direct when reducing dimension of the variable-lag space—produces more precise forecasts over all horizons. Random subset selection and LASSO produce more precise forecasts for three out of four forecast horizons.

When considering factor models, FAVAR and PLS are near identical when extracting factors in the variable space and the iterated approach is more accurate. The most precise forecasts among the factor approaches are, however, produced by the direct FAVAR with factors from the variable-lag space, and this is the only set of factor forecasts that are comparable to the random subset forecast in precision.

The third and fourth panel show the results when using robust optimal weights such that forecasts are robust against structural breaks. Compared

Table 1: Average relative MSFE

	AR	FAVAR		PLS		Rand. SubSet		Rand. Proj.		Rand. Compr.		LASSO	BVAR
h		var	var/lag	var	var/lag	var	var/lag	var	var/lag	var	var/lag		
Equal weights forecasts													
Iterated forecasts relative to prevailing mean benchmark													
1	0.737	0.905	0.792	0.905	0.951	0.711	0.729	0.724	0.759	0.761	0.772	0.726	0.771
3	0.872	1.011	0.914	1.012	0.901	0.877	0.846	0.865	0.874	0.904	0.885	0.854	0.854
6	0.902	0.993	0.970	0.994	0.971	0.906	0.901	0.898	0.921	0.946	0.924	0.907	0.907
12	0.946	0.999	1.046	0.999	1.057	0.950	0.957	0.942	0.952	0.973	0.952	0.943	0.975
Direct forecasts relative to prevailing mean benchmark													
3	0.877	1.050	0.871	1.050	0.942	0.896	0.878	0.870	0.853	0.915	0.860	0.887	0.859
6	0.909	1.056	0.910	1.056	0.969	0.911	0.901	0.905	0.890	0.967	0.893	0.946	0.903
12	0.968	1.120	0.973	1.120	1.014	0.965	0.952	0.962	0.944	1.022	0.944	0.984	0.964
Robust optimal weights forecasts													
Iterated forecasts relative to equal weights, prevailing mean benchmark													
1	0.759	0.880	0.776	0.881	0.850	0.715	0.726	0.743	0.755	0.773	0.768	0.738	0.770
3	0.899	1.016	0.911	1.016	0.899	0.889	0.834	0.893	0.853	0.932	0.860	0.853	0.853
6	0.948	1.041	0.980	1.041	0.980	0.949	0.907	0.946	0.906	1.014	0.908	0.900	0.906
12	0.994	1.039	1.067	1.039	1.072	0.991	0.985	0.984	0.948	1.043	0.949	0.946	0.975
Direct forecasts relative to equal weights, prevailing mean benchmark													
3	0.908	1.056	0.859	1.056	0.894	0.914	0.878	0.903	0.847	0.956	0.853	0.884	0.857
6	0.970	1.108	0.904	1.108	0.935	0.974	0.909	0.970	0.888	1.061	0.890	0.941	0.903
12	1.024	1.160	0.973	1.160	0.998	1.021	0.952	1.016	0.951	1.092	0.951	0.985	0.964
Individual specific tuning parameters, equal weights forecasts													
Iterated forecasts relative to prevailing mean benchmark													
1	0.737	0.893	0.798	0.905	0.970	0.710	0.726	0.727	0.754	0.765	0.765	0.726	0.771
3	0.872	1.005	0.910	1.012	0.922	0.877	0.846	0.870	0.874	0.907	0.884	0.854	0.854
6	0.902	0.983	0.988	0.994	0.987	0.917	0.900	0.898	0.921	0.943	0.924	0.907	0.907
12	0.946	0.993	1.058	0.999	1.070	0.943	0.956	0.941	0.952	0.970	0.952	0.943	0.975

Note: The table reports the averages of the MSFE relative to that of the prevailing mean benchmark in the top panel for the iterated forecasts and in the second panel for the direct forecasts. The third and fourth panel report the MSFE of the iterated and direct forecasts using robust optimal weights. These panels use common tuning parameters, except BVAR. The fifth panel reports the MSFE relative to the prevailing mean benchmark for the iterated forecasts for individual specific tuning parameters. The forecasting methods in the columns are: univariate AR, FAVAR, PLS, random subset, random projection, and random compression, LASSO, and BVAR. 'var' denotes forecast with dimension reduction in the variable space and 'var/lag' those with dimension reduction in the variable-lag space. The numbers are the average results over 14 series. The forecasts are $h = 1, 3, 6, 12$ month ahead.

to the forecasts in the first two panels, the forecast are more precise in a number of, but far from all, cases. This suggests that, for the series considered here, structural breaks are not a major driver of forecast performance. Importantly for our analysis, the relative results of the different forecasting methods remains unchanged. In the following, we will therefore focus on the results from the equally weighted forecasts.

The fifth panel gives the results when using individual specific tuning parameters. For conciseness, we only report the iterated forecast results; those for direct forecasts follow the same pattern. The average relative MSFEs are very similar to those obtained using pooled cross-validation to obtain the tuning parameters; about as many forecasts are marginally more precise as are marginally less precise. Again, the relative ranking of the different methods is not affected. The results we obtain are therefore neither sensitive to structural breaks nor the way we obtain tuning parameters.

McCracken and McGillicuddy (2019) observe that the relative performance of iterated and direct forecasts depends on the period under consideration. Table 2 reports the average of the relative MSFE for three sub-periods: the great moderation until July 2007, the financial crisis from August 2007 until July 2009, and the recent recovery from August 2009 onwards. In the last subsample, we excluded forecasts for the federal funds rate as it was near constant for this period and the resulting good forecasts of the prevailing mean model led to odd behavior of some MSFE ratios.

The first panel in Table 2 shows the results for the iterated forecasts in the great moderation period. The relative ranking of the dimension reduction methods is the same as that of the entire forecast sample, which is not surprising given that most of the forecast sample is in this subsample. Again, only random subset regression and random projection consistently beat the benchmark and the univariate AR. The second panel contains the average relative MSFEs of the direct forecasts. The pattern is similar to the one for the entire forecast period: iterated forecasts are more precise when combined with dimension reduction over the variable space and direct forecasts are more precise when combined with dimension reduction over the variable-lag space. LASSO favors iterated forecasts, whereas the BVAR is equally precise with the pseudo-iterated and direct approach.

The third and fourth panel contain the results for the financial crisis. Here, the value of additional variables for forecasting appears to be reduced: for $h = 1$ and $h = 3$ the (direct) univariate AR produces the most precise forecasts. For $h = 6$ the direct BVAR and for $h = 12$ iterated forecasts from the LASSO are more precise. In this sub-period the pattern between direct and iterated forecasts changes and the direct forecast is the more precise version of nearly all forecasting methods; only the LASSO retains a clear preference for iterated forecasts.

The results for the period since the financial crisis are displayed in the

Table 2: Average relative MSFE over subsamples

	AR	FAVAR		PLS		Rand. SubSet		Rand. Proj.		Rand. Compr.		LASSO	BVAR
h		var	var/lag	var	var/lag	var	var/lag	var	var/lag	var	var/lag		
Great moderation: July 1986 to July 2007													
Iterated forecasts relative to prevailing mean benchmark													
1	0.760	0.969	0.828	0.969	1.029	0.714	0.750	0.743	0.786	0.763	0.799	0.753	0.795
3	0.897	1.068	0.938	1.068	0.936	0.878	0.860	0.888	0.884	0.905	0.893	0.873	0.869
6	0.912	1.033	0.971	1.033	0.986	0.907	0.908	0.907	0.931	0.950	0.933	0.925	0.930
12	0.964	1.028	1.042	1.028	1.034	0.958	0.960	0.959	0.973	0.993	0.973	0.956	0.987
Direct forecasts relative to prevailing mean benchmark													
3	0.904	1.118	0.885	1.119	0.976	0.888	0.878	0.894	0.868	0.917	0.874	0.901	0.876
6	0.923	1.115	0.932	1.116	1.005	0.914	0.904	0.916	0.903	0.974	0.906	0.962	0.936
12	0.980	1.173	1.005	1.173	1.046	0.972	0.949	0.972	0.961	1.036	0.961	0.996	0.987
Financial crisis: August 2007 to July 2009													
Iterated forecasts relative to prevailing mean benchmark													
1	0.665	0.681	0.730	0.682	0.670	0.743	0.713	0.668	0.711	0.741	0.729	0.669	0.702
3	0.811	0.848	0.888	0.848	0.822	0.965	0.876	0.823	0.865	0.948	0.881	0.827	0.827
6	0.917	0.936	1.064	0.936	1.031	0.987	1.015	0.926	0.932	1.021	0.936	0.909	0.933
12	0.961	1.001	1.197	1.002	1.224	1.080	1.360	0.961	0.963	1.045	0.963	0.953	1.051
Direct forecasts relative to prevailing mean benchmark													
3	0.801	0.853	0.833	0.853	0.827	1.030	0.922	0.815	0.820	0.961	0.826	0.856	0.822
6	0.907	0.945	0.902	0.944	0.918	0.962	0.947	0.915	0.896	1.034	0.897	0.942	0.863
12	1.003	1.067	0.968	1.067	1.030	1.038	1.051	0.994	0.968	1.079	0.968	0.987	0.980
Post-crisis period: August 2009 to August 2017													
Iterated forecasts relative to prevailing mean benchmark													
1	0.813	0.981	0.840	0.981	1.029	0.808	0.790	0.809	0.821	0.871	0.831	0.800	0.875
3	0.911	1.087	1.007	1.088	1.061	0.896	0.878	0.900	0.902	0.925	0.911	0.899	0.940
6	0.898	1.043	1.005	1.044	1.047	0.887	0.899	0.888	0.911	0.910	0.917	0.911	0.935
12	0.942	1.139	1.082	1.140	1.185	0.938	0.956	0.939	0.926	0.957	0.927	0.955	1.017
Direct forecasts relative to prevailing mean benchmark													
3	0.931	1.113	0.945	1.113	1.091	0.928	0.913	0.920	0.916	0.947	0.923	0.915	0.944
6	0.908	1.052	0.929	1.052	1.015	0.910	0.910	0.901	0.898	0.940	0.903	0.933	0.943
12	0.988	1.115	0.972	1.116	1.069	0.966	1.020	0.990	0.935	1.005	0.935	0.994	1.101

Note: Forecasts use common tuning parameters and equal weights. For further details see the footnote of Table 1.

bottom two panels of Table 2. The pattern from the pre-crisis period reemerges. Random subset regression and random projection deliver the most accurate results followed by LASSO. Iterated forecasts are now even more clearly favored. Thus, similar to McCracken and McGillicuddy (2019), we observe a shift in relative forecasting performance between iterated and direct forecasts. Iterated forecasts are generally preferred but direct forecasts outperforming during the financial crisis. This also coincides with the intuition that direct forecasts are more robust to misspecification, which will be larger in periods of extreme events such as the financial crisis.

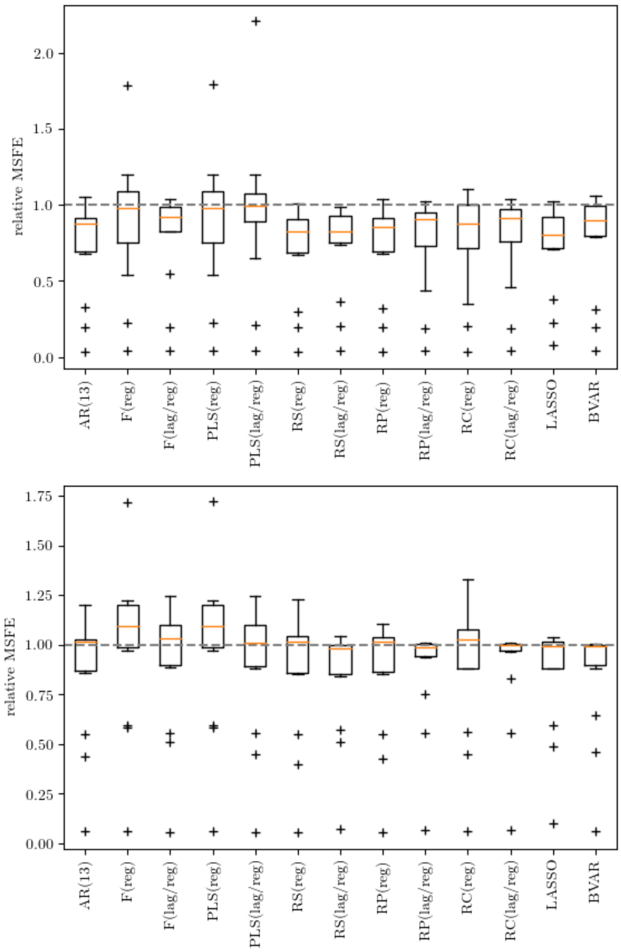
4.1.2 Forecast accuracy for individual series

The average performance of the forecasting techniques could potentially mask heterogeneity of performance across series. We generate box plots of the relative MSFE of each series and forecasting method for $h = 1$ and $h = 3$ in Figure 1. Tables for the individual series are in the online appendix. The plots show whether forecasting methods are more or less precise than the benchmark for a substantial proportion of the series. For $h = 1$, only random subset regression results in forecasts that are more precise than the benchmark for all series. Random projection and LASSO result in forecasts that are more precise for nearly all series or with only minimal loss. Random compression, BVAR, and the univariate model are similar with mildly larger loss. FAVAR and PLS, in contrast produce forecasts that are less accurate than the benchmark for a larger number of series.

The results for $h = 3$ in the bottom plot of Figure 1 show that a similar pattern emerges for multi-step forecasts. Random subset regression, random projection, LASSO and BVAR produce forecasts that do not lead to substantial losses in accuracy for any of the series, where dimension reduction in the variable-lag space reduces upwards outliers. The factor models, in contrast, produce forecasts that are worse than the benchmark for a majority of series. The same pattern continues for larger horizons.

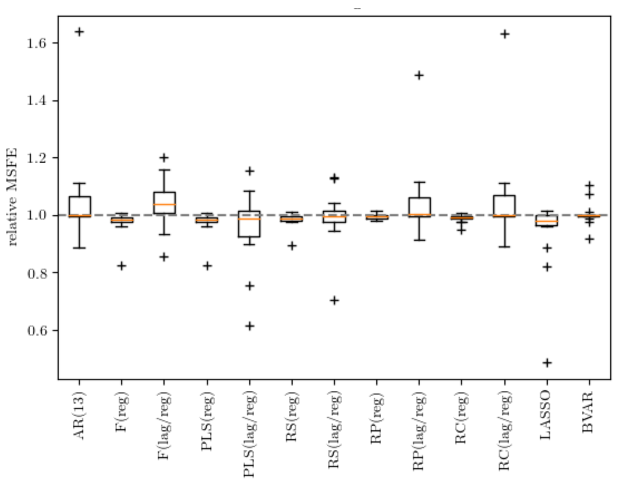
In order to compare iterated and direct forecasts, Figure 2 shows box plots of the ratio of MSFEs of iterated forecasts over that of direct forecasts. A ratio greater than one implies that the direct forecast is more precise and a ratio below one that the iterated forecast is more precise. While all methods have more precise iterated forecasts for some series and direct for others, some interesting patterns emerge. The univariate model and dimension reduction methods that work on the variable-lag space tend to deliver large gains for many series from direct models. Dimension reduction methods that work on the variable space are more likely to deliver improvements over the benchmark with iterated forecasts but these improvements are relatively small. This is also true for the LASSO but the improvements of the iterated forecasts are large for a few series. Again, these patterns continue for larger horizons.

Figure 1: Relative MSFE of iterated forecasts for $h = 1$ and 3



Note: The box plots show the relative MSFEs per forecasting method, where the elements in the box plot are the MSFE of each method relative to the benchmark, prevailing mean per series. The top plot is for $h = 1$ and the bottom plot for $h = 3$. The order of methods corresponds to that in the tables.

Figure 2: Relative MSFE of iterated against direct forecasts for $h = 3$



Note: The figures displays box plots for each forecasting method, where the elements in the box plot are the ratio of MSFE of iterated over direct forecasts per series.

4.1.3 Tests for equal predictive accuracy

Given the distribution of MSFE across series, are some of the forecasting methods significantly more accurate? Table 3 and 4 report the number of series for which the test statistic of Diebold and Mariano (1995) rejects the null of equal forecast accuracy. Rejection means that the methods in columns have a significant smaller square loss than the methods in rows. Table 3 contains the results for $h = 1$. For most combinations of methods only very few series have significant differences in forecast accuracy. Notable exceptions, however, are the factor models, which are significantly worse than the univariate AR, the random subspace methods, LASSO, and BVAR for a sizable number of series. Additionally, random projection in the variable-lag space delivers results that are more accurate than random compression and BVAR for five and six series.

The results for $h = 3$ are in Table 4. For each forecasting methods we now have the choice between iterated and direct forecast. In order to keep the table concise, we use the previous results that the dimension reduction methods in the variable space are more precise with the iterated approach and those in the variable-lag space with the direct approach and report the results that correspond to the method in the table. The results mirror those of the one-step ahead forecast: the factor based methods are significantly worse than the other methods. An interesting exception, however, is the variable-lag space FAVAR model, which is significantly worse only for few series but which is significantly better than the other factor based methods for six or eight series. This does not depend on whether the other factor

Table 3: Diebold-Mariano test statistics for square loss, $h = 1$

	AR	FAVAR		PLS		RS		RP		RC		L	B
		v	v/l	v	v/l	v	v/l	v	v/l	v	v/l		
AR	-	0	1	0	1	2	2	5	2	0	2	2	2
FAVAR(v)	8	-	3	2	2	4	2	8	4	4	4	5	4
FAVAR(v/l)	7	3	-	3	5	6	6	8	7	4	7	6	5
PLS(v)	8	0	3	-	2	4	2	8	4	4	4	5	4
PLS(v/l)	7	3	1	3	-	6	6	7	5	5	5	5	5
RS(v)	2	0	1	0	2	-	1	2	1	1	1	0	2
RS(v/l)	1	1	1	1	1	2	-	1	2	1	2	2	3
RP(v)	0	0	1	0	1	2	2	-	1	0	1	1	1
RP(v/l)	1	0	2	0	1	4	2	3	-	1	2	4	1
RC(v)	5	0	2	0	2	2	3	7	5	-	5	3	4
RC(v/l)	2	0	2	0	1	4	2	2	6	1	-	4	1
LASSO	1	0	1	0	1	2	1	1	2	0	2	-	2
BVAR	3	1	1	1	2	3	4	3	6	1	6	3	-

Note: The table reports the number of series for which the test statistic of Diebold and Mariano (1995) rejects the null of equal forecast accuracy. Rejection means that the methods in columns have a significant smaller square loss than the methods in rows. The methods are: univariate AR, FAVAR, PLS, random subset regression (RS), random projection (RP), random compression (RC), LASSO (L), and BVAR (B). 'v' denotes dimension reduction in the variable space and 'v/l' dimension reduction in the variable-lag space.

methods use iterated or direct forecast. This confirms the view that if factor based methods are preferred, factors should be extracted from the variable-lag space. For larger forecast horizons, the pattern remains the same even if fewer series show significant differences.

4.2 Density forecasts

4.2.1 Average forecast accuracy

Table 5 shows the average ratios of the log score of the different forecasting methods to that of the prevailing mean, where, in contrast to the previous results for the MSFE, a higher number indicated better forecast performance. The top panel represents the iterated forecasts and the second panel the direct forecasts. The results show that iterated random subset regression and BVAR yield the highest log scores. Iterated forecast generally are more accurate even for many forecast with dimension reduction in the variable-lag space. The choice of dimension reduction of variable space versus variable-lag space is less clear cut: for most methods dimension reduction over the

Table 4: Diebold-Mariano test statistics for square loss, $h = 3$

	AR	FAVAR		PLS		RS		RP		RC		L	B
		v	v/l	v	v/l	v	v/l	v	v/l	v	v/l		
AR	-	0	1	0	0	2	0	2	2	0	2	1	1
FAVAR(v)	12	-	6	0	4	13	2	13	6	9	6	10	8
FAVAR(v/l)	1	0	-	0	0	2	0	1	3	0	1	1	0
PLS(v)	12	1	6	-	4	13	2	13	6	9	6	10	8
PLS(v/l)	2	0	8	0	-	4	1	6	9	2	9	5	8
RS(v)	0	0	0	0	0	-	0	0	1	0	0	1	2
RS(v/l)	0	0	0	0	0	0	-	0	0	0	0	0	0
RP(v)	0	0	1	0	0	1	0	-	1	0	1	2	1
RP(v/l)	1	0	0	0	0	2	0	1	-	0	0	0	0
RC(v)	0	0	1	0	0	3	0	3	2	-	2	3	3
RC(v/l)	1	0	0	0	0	2	0	1	3	0	-	1	1
LASSO	1	1	2	1	1	2	1	2	2	1	2	-	2
BVAR	0	0	0	0	0	1	0	0	0	0	1	0	-

Note: The methods that reduce dimension in the variable space are iterated forecasts and those that reduce dimension in the variable-lag space are direct forecasts. For further details see footnote of Table 3.

variable-lag space is more accurate but the opposite is true for random projection. However, a constant result throughout is that the best factor model is the direct FAVAR with factors extracted from the variable-lag space.

The averages in the top two panels of Table 5 include two series, which have relatively extreme log scores for all methods: ‘housing starts’ and ‘VXO’. Unlike the MSFE, where these series had small relative MSFEs, which would not overly influence the average, some log scores are now very large and can influence the averages unduly. In order to investigate the robustness of our results, the third and fourth panel of Table 5 report the averages that exclude those series. For $h = 1$, LASSO is now the most accurate, followed by random subset regression, which reverses the finding when including ‘housing starts’ and ‘VXO’. The multi-step forecasts for $h = 6$ and 12 are most precise using random subset regression and the BVAR. For $h = 12$, no method beats the benchmark. Across forecast horizons, iterated forecasts are more precise for the majority of forecasting methods. The result for the factor models remains untouched: Direct FAVAR with factors extracted from the variable-lag space is the most accurate factor model and not too far off the best methods.

Table 6 reports the average relative log score over the three sub-samples, where we again exclude ‘housing starts’ and ‘VXO’. The first panel shows the results for the great moderation period. The most precise forecasts for all horizon are from the iterated random subset regression in the variable-lag

Table 5: Average relative log score

	AR	FAVAR		PLS		Rand. SubSet		Rand. Proj.		Rand. Compr.		LASSO	BVAR
h		var	var/lag	var	var/lag	var	var/lag	var	var/lag	var	var/lag		
All series													
Iterated forecasts relative to prevailing mean benchmark													
1	1.143	1.099	1.130	1.099	1.088	1.160	1.156	1.150	1.138	1.138	1.139	1.143	1.135
3	1.095	1.061	1.094	1.061	1.100	1.104	1.113	1.101	1.094	1.097	1.103	1.100	1.118
6	1.074	1.055	1.067	1.055	1.069	1.084	1.080	1.079	1.069	1.071	1.078	1.065	1.089
12	1.037	1.024	1.022	1.024	1.026	1.044	1.037	1.041	1.036	1.040	1.042	1.041	1.021
Direct forecasts relative to prevailing mean benchmark													
3	1.092	1.045	1.104	1.045	1.083	1.097	1.100	1.097	1.089	1.098	1.106	1.067	1.113
6	1.069	1.028	1.079	1.028	1.062	1.077	1.083	1.074	1.060	1.076	1.082	1.032	1.088
12	1.031	0.979	1.033	0.979	1.018	1.037	1.031	1.034	1.019	1.035	1.037	1.005	1.038
All series excluding 'housing starts' and 'VXO'													
Iterated forecasts relative to prevailing mean benchmark													
1	1.024	0.977	1.018	0.977	0.971	1.039	1.043	1.032	1.019	1.027	1.028	1.047	1.021
3	0.990	0.953	0.988	0.953	0.995	0.993	1.014	0.994	0.988	0.997	1.004	1.007	1.011
6	0.993	0.973	0.979	0.973	0.982	0.994	1.004	0.996	0.984	0.994	1.001	1.001	1.004
12	0.995	0.981	0.969	0.981	0.974	0.994	0.995	0.997	0.992	0.998	1.000	0.998	1.000
Direct forecasts relative to prevailing mean benchmark													
3	0.988	0.938	1.004	0.938	0.982	0.986	0.996	0.992	0.985	0.996	1.004	0.998	1.007
6	0.990	0.950	1.001	0.950	0.984	0.992	1.005	0.993	0.978	0.995	1.002	0.995	1.004
12	0.992	0.939	0.994	0.939	0.978	0.993	0.994	0.994	0.977	0.994	0.997	1.000	0.994

Note: The table reports the average of the log score per method relative to that of the prevailing mean benchmark in the top panel for the iterated forecasts and in the second panel for the direct forecasts. Further details are given in the footnote of Table 1.

Table 6: Average relative log score over subsamples

	AR	FAVAR		PLS		Rand. SubSet		Rand. Proj.		Rand. Compr.		LASSO	BVAR
h		var	var/lag	var	var/lag	var	var/lag	var	var/lag	var	var/lag		
Great moderation: July 1986 to July 2007													
Iterated forecasts relative to prevailing mean benchmark													
1	1.019	0.970	1.014	0.970	0.961	1.040	1.044	1.027	1.018	1.022	1.025	1.043	1.017
3	0.983	0.942	0.983	0.942	0.990	0.993	1.016	0.988	0.988	0.995	1.003	1.003	1.011
6	0.989	0.966	0.976	0.966	0.980	0.992	1.008	0.992	0.983	0.992	1.001	0.999	1.006
12	0.994	0.978	0.967	0.978	0.977	0.996	1.003	0.997	0.993	0.997	1.000	0.997	1.003
Direct forecasts relative to prevailing mean benchmark													
3	0.981	0.925	1.005	0.925	0.979	0.988	0.998	0.985	0.986	0.992	1.002	0.994	1.006
6	0.985	0.941	1.001	0.941	0.984	0.990	1.006	0.989	0.976	0.992	1.000	0.992	1.002
12	0.992	0.930	0.995	0.930	0.977	0.994	0.994	0.994	0.977	0.994	0.997	0.999	0.996
Financial crisis: August 2007 to July 2009													
Iterated robust optimal weights forecasts													
1	1.146	1.131	1.100	1.131	1.132	1.120	1.141	1.155	1.121	1.124	1.105	1.195	1.152
3	1.080	1.092	1.048	1.092	1.123	1.010	1.060	1.084	1.036	1.031	1.013	1.086	1.050
6	1.025	1.046	0.995	1.046	1.017	1.013	0.996	1.031	1.001	1.007	1.003	1.039	1.008
12	0.995	1.000	0.960	1.000	0.924	0.985	0.943	1.000	0.992	1.001	1.003	1.005	0.979
Direct forecasts relative to prevailing mean benchmark													
3	1.081	1.080	1.050	1.080	1.084	0.972	1.039	1.083	1.019	1.064	1.054	1.074	1.056
6	1.024	1.029	1.022	1.029	1.029	1.013	1.028	1.030	0.994	1.028	1.022	1.028	1.044
12	0.987	0.986	0.999	0.986	0.998	0.984	0.987	0.991	0.989	0.995	0.995	1.009	0.988
Post-crisis period: August 2009 to August 2017													
Iterated forecasts relative to prevailing mean benchmark													
1	1.022	1.000	1.013	1.000	1.003	1.029	1.027	1.024	1.020	1.018	1.015	1.030	1.014
3	1.003	0.984	0.992	0.984	0.992	1.004	1.007	1.004	1.000	1.001	1.002	1.006	1.008
6	1.002	0.988	0.984	0.988	0.981	1.002	1.003	1.003	0.998	0.999	1.000	1.002	1.003
12	0.999	0.993	0.981	0.993	0.972	0.997	0.998	1.000	0.995	1.000	1.000	0.998	0.997
Direct forecasts relative to prevailing mean benchmark													
3	1.001	0.980	1.002	0.980	0.994	1.001	1.003	1.002	0.998	1.003	1.004	1.002	1.008
6	0.999	0.981	1.001	0.981	0.993	0.999	1.004	1.000	0.996	1.003	1.004	0.999	1.003
12	0.995	0.978	0.994	0.978	0.985	0.996	0.999	0.997	0.992	0.999	0.999	0.999	0.997

Note: For details see the footnote of Table 5.

space. The iterated forecasts are more precise the vast majority of cases: 32 out of 39 combinations of method and forecast horizon. The second panel gives the results for the financial crisis. Here the picture is more mixed. The best forecasts for $h = 1$ is from the LASSO. For $h = 3$ and 6, PLS with factors from the variable-lag space is most accurate followed by FAVAR and PLS with factors from the variable space and iterated LASSO. At $h = 12$ the direct LASSO is most precise. During the financial crisis, direct forecasts are more competitive: the direct forecasts beat the iterated in 16 cases and are more precise than the benchmark for the majority methods and horizons. In the last sub-period, direct density forecasts remain more competitive compared to the point forecast: in 13 out of 39 cases they beat the corresponding iterated forecast and for $h = 3$ and 6 they provide the most accurate forecast with the BVAR and random subset in the variable-lag space.

4.2.2 Forecast accuracy for individual series

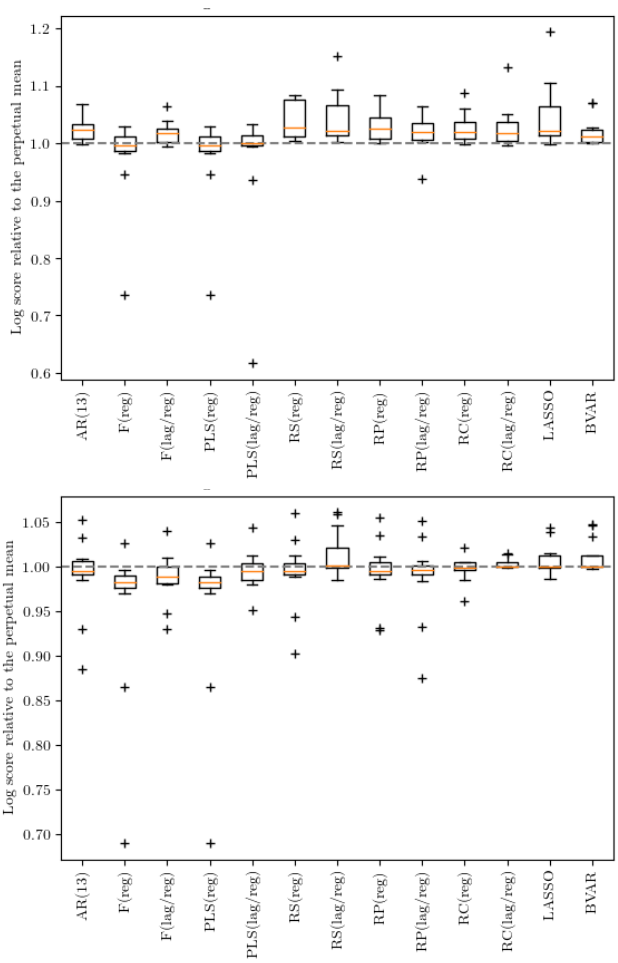
Figure 3 shows box plots for the relative log scores excluding the series 'housing starts' and 'VXO'. The top plot is for $h = 1$ and the bottom plot for $h = 3$. The box plot for $h = 1$ shows that the log scores of the random subspace methods, LASSO, BVAR, and AR improve upon the benchmark for (nearly) all series. The performance of the factor models is more mixed with improvements for some variables but deterioration for others. The results for $h = 3$ indicate that beating the benchmark is more difficult at larger horizons. The factors models produce iterated forecasts that are more likely to be inferior to the benchmark, whereas the other methods beat the benchmark for some series and are beaten for others. The results for larger h follow the same pattern.

4.2.3 Tests for equal predictive accuracy

Table 7 shows the number of series for which the test statistic of equal forecast accuracy of Diebold and Mariano (1995) rejects the null for (negative) log score loss for $h = 1$. Rejection means that the methods in columns have a significant smaller loss (i.e. larger log score) than the methods in rows. As this is a variable by variable comparison, we included all series in Diebold-Mariano test statistics. The pattern is very similar to the one for the point forecast: Factor models are significantly worse for a large number of series. The random subset and random projection statistically beat other methods for a large number of series.

The results of the Diebold-Mariano statistics for $h = 3$ are in Table 8. Again, the patterns are similar to those of the point forecasts. The factor models are significantly less accurate than other methods with the exception of the FAVAR model in the variable-lag space. For $h = 3$, the random

Figure 3: Relative log score of iterated against prevailing mean for $h = 1$ and 3



Note: The box plots show the relative log score per forecasting method, where the elements in the box plot are the log score of each method relative to the benchmark, prevailing mean per series. The top plot is for $h = 1$ and the bottom plot for $h = 3$. The box plots exclude the results for the series ‘housing starts’ and ‘VXO’.

Table 7: Diebold-Mariano test statistics for log scores, $h = 1$

	AR	FAVAR		PLS		RS		RP		RC		L	B
		v	v/l	v	v/l	v	v/l	v	v/l	v	v/l		
AR	-	0	4	0	0	10	7	11	2	4	5	4	2
FAVAR(v)	13	-	7	0	6	13	10	13	13	9	8	9	7
FAVAR(v/l)	4	1	-	1	2	7	10	5	4	7	6	6	4
PLS(v)	13	6	7	-	6	13	10	13	13	9	8	9	7
PLS(v/l)	11	4	10	4	-	12	13	11	9	11	11	9	6
RS(v)	0	0	2	0	0	-	5	0	0	1	3	2	0
RS(v/l)	2	0	2	0	0	2	-	2	0	1	2	4	1
RP(v)	0	0	2	0	0	7	6	-	1	2	3	4	1
RP(v/l)	4	0	3	0	0	7	6	5	-	4	4	3	1
RC(v)	4	0	1	0	0	6	8	6	6	-	5	6	4
RC(v/l)	4	0	0	0	0	5	7	4	4	5	-	6	5
LASSO	4	1	4	1	2	4	6	4	4	5	5	-	3
BVAR	7	3	7	3	2	10	9	9	6	8	8	9	-

Note: The table reports the number of series for which the test statistic of Diebold and Mariano (1995) rejects the null of equal forecast accuracy for negative log score loss. Rejection means that the methods in columns have a significant larger log score than the methods in rows. For further details see footnote of Table 3.

Table 8: Diebold-Mariano test statistics for log scores, $h = 3$

	AR	FAVAR		PLS		RS		RP		RC		L	B
		v	v/l	v	v/l	v	v/l	v	v/l	v	v/l		
AR	-	0	5	0	2	6	0	8	1	5	5	5	6
FAVAR(v)	12	-	12	1	11	13	5	13	11	10	13	11	11
FAVAR(v/l)	1	0	-	0	0	2	0	2	0	0	4	2	3
PLS(v)	12	7	12	-	11	13	5	13	11	10	12	11	11
PLS(v/l)	4	0	10	0	-	6	2	6	2	7	11	9	9
RS(v)	1	0	4	0	0	-	0	2	0	4	5	3	5
RS(v/l)	0	0	0	0	0	0	-	0	0	0	0	1	1
RP(v)	1	0	4	0	0	4	0	-	0	4	4	4	5
RP(v/l)	1	0	5	0	2	7	0	7	-	5	6	5	6
RC(v)	2	0	4	0	1	4	1	3	2	-	10	5	7
RC(v/l)	1	0	1	0	0	2	0	2	1	0	-	1	5
LASSO	2	1	3	1	2	3	1	3	2	3	4	-	4
BVAR	1	0	1	0	0	3	0	2	0	0	1	1	-

Note: The methods that reduce dimension in the variable space are iterated forecasts and those that reduce dimension in the variable-lag space are direct forecasts. For further details see footnote of Table 7.

subspace methods are less dominant compared to $h = 1$. The BVAR, which was beaten for many series at $h = 1$, is significantly more accurate for more series at $h = 3$. At larger horizons, the factor models remain significantly worse for most series and random subset models significantly more accurate.

5 Conclusion

This paper compares dimension reduction methods for large vector autoregressions for multi-step ahead point and density forecasts. A few patterns emerge from the analysis. The first is that random subspace methods, in particular random subset regression and random projection, generally deliver the most accurate point and density forecasts. This is true irrespective of whether the randomization is over the variable space or the variable/lag space. The LASSO is the next most accurate method.

The decision between iterated and direct forecasts appears to be connected to the decision of dimension reduction in the variable or variable/lag space. The combinations of iterated forecasts and dimension reduction in the variable space consistently beat their direct counter parts. When the dimension reduction is in the variable/lag space, the iterated forecast looses its advantage over the direct forecast. This is most notably so for the factor model. The reason for this could be that the dimension reduction introduces a model misspecification when done in the variable/lag space, which effects the direct forecast to a lesser extent. In a Monte Carlo experiment, reported in the online appendix, we show that this pattern can be replicated and does not depend on the choice of DGP.

Finally, when factor models are preferred, for example due to the interpretability of estimated factors, the direct FAVAR model with principal components extracted from the full variable-lag regressor matrix delivers the most accurate point and density forecasts.

References

Achlioptas, D. (2003). Database-friendly random projections: Johnson-Lindenstrauss with binary coins. *Journal of Computer & System Sciences 66*(4), 671–687.

Bańbura, M., D. Giannone, and L. Reichlin (2010). Large Bayesian vector auto regressions. *Journal of Applied Econometrics 25*(1), 71–92.

Bernanke, B. S., J. Boivin, and P. Eliasz (2005). Measuring the effect of monetary policy: A factor-augmented vector autoregressive (FAVAR) approach. *Quarterly Journal of Economics 120*(1), 387–422.

Boot, T. and D. Nibbering (2019). Forecasting using random subspace methods. *Journal of Econometrics 209*(2), 391–406.

Boot, T. and A. Pick (2020). Does modeling a structural break improve forecast accuracy? *Journal of Econometrics 215*(1), 35–59.

Corradi, V. and N. R. Swanson (2006). Predictive density evaluation. In G. Elliott, C. Granger, and A. Timmermann (Eds.), *Handbook of Economic Forecasting, Volume 1*, Chapter 5, pp. 197–284. Amsterdam: North-Holland.

Cross, J. L., C. Hou, and A. Poon (2020). Macroeconomic forecasting with large Bayesian VARs: Global-local priors and the illusion of sparcity. *International Journal of Forecasting 36*(3), 899–915.

Diebold, F. X. and R. S. Mariano (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics 13*(3), 134–144.

Doan, T., R. B. Litterman, and C. A. Sims (1984). Forecasting and conditional projection using realistic prior distributions. *Econometric Reviews 3*(1), 1–30.

Donoho, D. (2006). Compressed sensing. *Transaction of Information Theory 52*(4), 1289–1306.

Elliott, G., A. Gargano, and A. Timmermann (2013). Complete subset regressions. *Journal of Econometrics 177*(2), 357–373.

Elliott, G. and A. Timmermann (2016). *Economic Forecasting*. Princeton: Princeton University Press.

Giacomini, R. and B. Rossi (2009). Detecting and predicting forecast breakdowns. *Review of Economic Studies 76*(2), 669–705.

Giannone, D., M. Lenza, and G. E. Primiceri (2015). Prior selection for vector autoregressions. *Review of Economics & Statistics 97*(2), 436–451.

Gneiting, T. and A. E. Raftery (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association 102*(477), 359—378.

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

Guhaniyogi, R. and D. B. Dunson (2015). Bayesian compressed regression. *Journal of the American Statistical Association 110*(512), 1500–1514.

Helland, I. S. (1990). Partial least squares regression and statistical models. *Scandinavian Journal of Statistics 17*(2), 97–114.

Holt, C. (1957). Forecasting trends and seasonals by exponential weighted averages. *ONR Memorandum 52/1957. Carnegy Mellon University*.

Inoue, A. and B. Rossi (2011). Identifying the sources of instabilities in macroeconomic fluctuations. *Review of Economics & Statistics 164*(4), 158–172.

John, S. (1982). The three-parameter two-piece normal family of distributions and its fitting. *Communications in Statistics - Theory and Methods 11*(8), 879—885.

Koop, G., D. Korobilis, and D. Pettenuzzo (2019). Bayesian compressed vector autoregressions. *Journal of Econometrics 210*(1), 135–154.

Koop, G. and S. M. Potter (2007). Estimation and forecasting in models with multiple breaks. *Review of Economic Studies 74*(3), 763–789.

Litterman, R. B. (1986). Forecasting with Bayesian vector autoregressions: Five years of experience. *Journal of Business & Economic Statistics 4*(1), 25–38.

Ma, Y. and L. Zhu (2013). A review on dimension reduction. *International Statistical Review 81*(1), 134–150.

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

McCracken, M. W. and J. McGillicuddy (2019). An empirical investigation of direct and iterated multistep conditional forecasts. *Journal of Applied Econometrics 2*(34), 181–204.

McCracken, M. W. and S. Ng (2016). FRED-MD: A monthly database for macroeconomic research. *Journal of Business & Economic Statistics 34*(4), 574–589.

Mol, C. D., D. Giannone, and L. Reichlin (2008). Forecasting using a large number of predictors - is Bayesian regression a valid alternative to principal components? *Journal of Econometrics 146*(2), 318–328.

Pesaran, M. H., D. Pettenuzzo, and A. Timmermann (2006). Forecasting time series subject to multiple structural breaks. *Review of Economic Studies 73*(4), 1057–1084.

Pesaran, M. H., A. Pick, and M. Pranovich (2013). Optimal forecasts in the presence of structural breaks. *Journal of Econometrics 177*(2), 134–152.

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

Stock, J. H. and M. W. Watson (1996). Evidence on structural instability in macroeconomic time series relations. *Journal of Business & Economic Statistics 14*(1), 11–30.

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

Tay, A. S. and K. F. Wallis (2000). Density forecasting: A survey. *Journal of Forecasting 19*(4), 124–143.

Tibshirani, R. (1996). Regression shrinkage and selection via the LASSO. *Journal of Royal Statistical Society, Series B 58*(1), 267–288.