Revisiting Targeted Factors

Jack Fosten
University of East Anglia, UK*

May 3, 2016

Abstract

This paper proposes new methods for 'targeting' factors estimated from a big dataset. We suggest that forecasts of economic variables can be improved by tuning factor estimates so that both: (i) they are more relevant for a specific target variable, and (ii) so that variables with considerable idiosyncratic noise are down-weighted prior to factor estimation. Existing targeted factor methodologies are limited to estimating the factors with only one of these two objectives in mind. We therefore combine these ideas by providing new weighted Principal Components Analysis (PCA) procedures and a Targeted Generalized PCA (TGPCA) procedure. These methods offer a flexible combination of both types of targeting which is new to the literature. We illustrate this empirically by forecasting a range of U.S. macroeconomic variables, finding that our combined approach yields important improvements over competing methods, consistently surviving elimination in the Model Confidence Set procedure.

JEL Classification: C13, C22, C38, C53

Keywords: Forecasting, factor estimation, targeted predictors, LASSO, data reduction

1 Introduction

This paper revisits the idea of 'targeting' the factors estimated from a big dataset, when the purpose is to use the factors for economic forecasting. The principle of targeted factors is to down-weight or remove selected variables prior to factor estimation in order to improve the forecasts based on those factor estimates for a particular forecast variable of interest. This literature has evolved along two separate paths. On the one hand, Boivin and Ng (2006) suggested to down-weight variables which have noisy idiosyncratic variation as these can worsen the precision of factor estimates. On the other hand, Bai and Ng (2008) suggested to use LASSO-type methods to pre-select a subset of variables, targeted to a specific forecast variable, from which to estimate the factors. These are both in contrast with the seminal work of Stock and Watson (2002a,b) who suggest to use all available variables in the dataset, and weight these variables equally in the process of factor estimation. In this paper, we explore the idea that both types of targeting might be used together. We therefore propose methods which allow us to target the factor estimation procedure with both the forecast variable and the factor model properties in mind.

The first main contribution of this paper proposes a method to directly combine the existing methods of Bai and Ng (2008) and Boivin and Ng (2006) for targeting factors. Our approach uses elements of both of these methods to produce estimation weights for weighted Principal Components Analysis (PCA). The weight assigned to each variable depends both on its ability to predict a given forecast variable, and its properties with regards to idiosyncratic noise within the factor model structure. This method is implemented by first removing the weak predictor variables based on a LASSO-based selection procedure, as in Bai and Ng (2008), and then performing weighted PCA on the surviving variables. The implication of this method is that if there are two variables with similar predictive power for the forecast variable, but one is noisy and the other is not, then both variables will be retained for factor estimation, but the former will be down-weighted. This is not possible using either the methodologies of Bai and Ng (2008) or Boivin and Ng (2006) alone.

The second proposal we make is to extend the Bai and Ng (2008) method to use weighted PCA, rather than standard PCA, in order to reflect the rela-

tive strength of predictive power of different variables on the target variable. The existing approach of Bai and Ng (2008) uses the Elastic Net LASSO-based method of Zou and Hastie (2005) simply as a selection device and then estimates the factors using standard PCA by placing equal weight on the surviving variables which have non-zero Elastic Net coefficients. Similarly, the extensions of Kim and Swanson (2014) and Bulligan et al. (2015) follow the approach of pre-selecting variables prior to using standard PCA. We suggest that, after the LASSO-based pre-selection phase, the coefficient values are retained and used as weights in performing weighted PCA, rather than discarding the magnitude of these coefficients.

We finally propose a method which uses the idea of targeting to allow the implementation of a Generalized Least Squares analogue to Principal Components Analysis. We call this Targeted Generalized Principal Components Analysis (TGPCA). The paper of Boivin and Ng (2006) first suggested a Generalized PCA procedure, but noted that this was not feasible because a typical estimator of the variance-covariance matrix of the idiosyncratic errors is of reduced rank and therefore not invertible. We overcome this limitation by suggesting a method which uses the LASSO-based pre-selection phase to reduce the dimension of the problem and select a subset of variables whose error variance-covariance matrix can be inverted. This method is therefore also a combination of the two types of targeting, and additionally lets us solve the problem found by Boivin and Ng (2006) regarding the Generalized PCA procedure.

We expect that these proposed methods will provide empirical forecasting improvements in a wide variety of situations. Previous empirical studies such as Schumacher (2010) and Eickmeier and Ng (2011) found that using the Bai and Ng (2008) method provided improvements over other forecasting methods. We envisage that using our combined method of targeting which also targets factors based on factor model performance may provide yet further improvements. On the other hand, other studies such as den Reijer (2012) and Castle et al. (2013) find less evidence in favour of the Bai and Ng (2008) targeting approach. It is possible that the results of these studies are adversely affected because the targeted predictor method retains variables which give noisy factor estimates. This point would be addressed by using our proposed methodologies.

To this end, we provide an empirical illustration of our proposed method-

ologies to forecasting a range of macroeconomic and financial variables for the United States, based on the Stock and Watson (2002a,b) dataset. We compare these new methods to the existing targeted factor methodologies. As a preview of the results, we find that our combined targeted methodologies prove to perform better than all other methods in terms of the Mean Squared Forecast Error ($MSFE$) from a pseudo out-of-sample forecast experiment. We confirm this feature with evidence from the Model Confidence Set (MCS) procedure of Hansen et al. (2011).

The rest of the paper is organised as follows. Section 2 introduces the general framework for factor estimation which allows us to describe the spectrum of different targeted factor methodologies. Section 3 outlines our new proposed methods for targeted factors. Section 4 describes the data, the different competing models we use, and the pseudo out-of-sample forecasting experiment. Section 5 provides the results. Finally, Section 6 concludes the paper.

2   Targeted Factors Set-Up

In forecasting a target variable y_{t+h} at a forecast horizon $h > 0$, the literature of targeted factors is underpinned by the "diffusion index", or factor-augmented forecasting model of Stock and Watson (2002a,b). This method assumes that a high-dimensional $N \times 1$ vector of candidate predictors X_t has a common factor structure:

$$X_t = \Lambda F_t + u_t \tag{1}$$

where F_t is an $r \times 1$ vector of unobserved factors, Λ is an $N \times r$ matrix of factor loadings and u_t is an $N \times 1$ vector of idiosyncratic error terms. The diffusion index model uses the factors as predictors in the forecasting model instead of X_t as this performs substantial data reduction when $r << N$. The model can be written:

$$y_{t+h} = \beta' F_t + \varepsilon_{t+h} \tag{2}$$

Since the factors, F_t, are unknown, they must be estimated from the data in order to make forecasting using Equation (2) feasible. Stock and Watson (2002a,b) show that using standard Principal Components Analysis (PCA)

gives consistent factor estimates up to a rotation of the true factors. Standard PCA estimates the $T \times r$ matrix of factors, F, as the r eigenvectors corresponding to the r largest eigenvalues of the $T \times T$ covariance matrix XX', under the identifying normalization that $F'F/T = I$.

The idea of targeted factors is that we may wish to give more or less weight to certain variables in X_t when estimating the factors, in order to 'target' a specific scenario. At its most general, the estimation of targeted factors is a form of Generalized Principal Components Analysis (GPCA), solving the optimization problem:

$$\min_{\Lambda, F_1, \dots, F_T} \frac{1}{NT} \sum_{t=1}^{T} (X_t - \Lambda F_t)' W (X_t - \Lambda F_t) \tag{3}$$

subject to the identifying normalization $F'F/T = I$, and where W is an $N \times N$ weighting matrix whose form will be discussed throughout this paper.[1] When $W = I$, this optimization coincides with standard PCA.

In many of the targeting approaches we will discuss, the weighting matrix W has the diagonal form:

$$W = \text{diag}(w)$$

where w is an $N \times 1$ vector of weights to be chosen by the researcher, possibly only containing the values 1 and 0. In this case the estimation procedure reduces to Weighted Principal Components Analysis (WPCA), and the GPCA objective function in Equation (3) can be rewritten as:

$$\min_{\Lambda, F_1, \dots, F_T} \frac{1}{NT} \sum_{i=1}^{N} w_i \sum_{t=1}^{T} \left(X_{it} - \lambda_i' F_t \right)^2 \tag{4}$$

where the $r \times 1$ matrix λ_i corresponds to the ith row of Λ. WPCA can be implemented easily by performing standard PCA using each of the series X_{it}, weighted by $w_i^{1/2}$. We now formally describe how the targeted factor methods of Bai and Ng (2008) and Boivin and Ng (2006) fit into this general set-up.

Bai and Ng (2008) suggest to use a pre-selection phase which generates a binary 1/0 weight vector for WPCA, based on the non-zero coefficients from penalized regressions of y_{t+h} onto X_t. They base their approach on

---

[1] When W is data-dependent, it should be indexed by the panel dimensions as W_{NT}, though we drop these indices so as to simplify the notation.

the LASSO technique of Tibshirani (1996) because of its ability to shrink coefficients in high-dimensional regressions to zero. In practice, they use the the related Elastic Net method of Zou and Hastie (2005) as this is better able to deal with highly correlated series than the LASSO. The weights assigned to each variable i can be described as:

$$w_i^{EN} = \mathbf{1}\left\{\widehat{\theta}_i\left(EN, \tau_1, \tau_2\right) \neq 0\right\} \tag{5}$$

where $\mathbf{1}\{.\}$ is the indicator function, and $\widehat{\theta}_i\left(EN, \tau_1, \tau_2\right)$ are the Elastic Net coefficients based on the following penalized least squares optimization problem:

$$\widehat{\theta}_i\left(EN, \tau_1, \tau_2\right) = \arg\min_{\theta}\left(\frac{1}{T}\sum_{t=1}^{T}\left(y_{t+h} - \theta'X_t\right)^2 + \tau_1\|\theta\|_1 + \tau_2\|\theta\|^2\right) \tag{6}$$

The choice of τ_1 and τ_2, the penalties on the L_1 and L_2 norms, cannot be chosen optimally when the objective is to use the regression output only as a pre-selection device for Principal Components. Instead, Bai and Ng (2008) successfully choose a rule-based approach by fixing the parameter τ_2 and then choosing τ_1 in such a way which allows 30 variables to enter the targeted dataset for factor estimation. They choose 30 as a small but appropriate number for factor estimation based on previous Monte Carlo simulation evidence. Selection of the 'top 30' is made simple by using the least angle regression (LAR) algorithm of Efron et al. (2004) which gives a full ordering of the X_t variables for a given y_{t+h}.

The second approach to targeting, proposed by Boivin and Ng (2006), is to down-weight or eliminate variables with 'noisy' properties for factor estimation. They suggest that an ideal solution would be to use the analogue of Generalized Least Squares and setting the GPCA weighting matrix to be:

$$W^{GLS} = \Omega^{-1} \tag{7}$$

where Ω is the variance-covariance matrix of the vector of idiosyncratic errors u_t. However, they note that there is no feasible analogue to this problem, as the $N \times N$ estimator $\widehat{\Omega}$ from an r-factor model is of rank $N - r$, and is not invertible. Therefore it is not feasible to use the GPCA procedure of Equation (3) with the weight matrix of Equation (7).

Boivin and Ng (2006) suggest several ways to overcome this. The first approach, which they call "Rule SWa", suggests to take only the principal diagonal of the matrix $\widehat{\Omega}$ and use the inverse of these elements to form a diagonal weight matrix with entries:

$$w_i^{SWa} = \widehat{\Omega}_{ii}^{-1} \tag{8}$$

Their second approach, "Rule SWb", gives a weight to variable i equal to the inverse of the average correlation of that idiosyncratic error with all other errors:

$$w_i^{SWb} = \left(\frac{1}{N} \sum_{j=1}^{N} |\widehat{\Omega}_{ij}| \right)^{-1} \tag{9}$$

Unlike Rule SWa, Rule SWb uses all of the estimated idiosyncratic variances and covariances, but it only weights the variances in the estimation procedure. They also consider another set of methods, "Rule 1" and "Rule 2", specifying a binary 1/0 selection vector which drops series whose errors are most correlated with some other series.

Both Bai and Ng (2008) and Boivin and Ng (2006) find there to be forecasting gains to targeted factors over non-targeted factors when applied to a wide range of U.S. economic series. However, subsequent empirical studies have found more mixed conclusions. On the one hand, Schumacher (2010) and Eickmeier and Ng (2011) use factor models on big international datasets in forecasting German and New Zealand GDP growth respectively, and report success of targeting relative to using the whole dataset. On the other hand, den Reijer (2012) finds no gains to pre-selection in forecasting Dutch GDP and inflation.

Since Bai and Ng (2008) only target with the forecast variable in mind, and Boivin and Ng (2006) only target with factor model properties in mind, we suggest that a combined approach to targeting may help to improve forecasts in overcoming the limitations of using either these targeting methods individually. In the next section we propose forecasting methodologies which allow for targeting to occur both with the forecast variable and the factor model properties in mind.

3 Methodology

3.1 Combined Targeted Principal Components Analysis

The first contribution of this paper is to provide procedures which combine the benefits of both Bai and Ng (2008) and Boivin and Ng (2006) by targeting the factors both with respect to the factor model *and* the forecast model. We additionally provide a method which allows researchers to flexibly choose how much to target factor estimation based on the factor model and the forecast model. This is not possible in the methodologies of Bai and Ng (2008) or Boivin and Ng (2006) which do one form of targeting but not both. We also relax the procedure of Bai and Ng (2008) so that the magnitude of the LASSO-type coefficients are used to give varying weights to each variable.

We first propose a method which combines the existing weighting schemes from the targeting methods of Boivin and Ng (2006) and Bai and Ng (2008) by forming weights for WPCA which are a product of the weights of both methods. From the definition of the weights w_i^{EN}, w_i^{SWa} and w_i^{SWb} in Equations (5), (8) and (9), we suggest combined weights w_i^1 and w_i^2 which combine w_i^{EN} respectively with w_i^{SWa} and w_i^{SWb}:

$$
\begin{aligned}
w_i^1 &= w_i^{EN} \times w_i^{SWa} \\
&= \mathbf{1}\left\{ \widehat{\theta}_i \left(EN, \tau_1, \tau_2 \right) \neq 0 \right\} \times \widehat{\Omega}_{ii}^{-1}
\end{aligned}
\tag{10}
$$

and

$$
\begin{aligned}
w_i^2 &= w_i^{EN} \times w_i^{SWb} \\
&= \mathbf{1}\left\{ \widehat{\theta}_i \left(EN, \tau_1, \tau_2 \right) \neq 0 \right\} \times \left(\frac{1}{N} \sum_{j=1}^{N} |\widehat{\Omega}_{ij}| \right)^{-1}
\end{aligned}
\tag{11}
$$

The weights w_i^1 and w_i^2 have the dual effect of removing variables which are weak predictors for y_{t+h} while also down-weighting those variables whose idiosyncratic errors are noisy.

However, as suggested earlier, it may be useful to retain information regarding the strength of predictive power of each variable for y_{t+h}. In other words, rather than using the indicator function as in Equation (5) and giving equal weights to the targeted variables in factor estimation, we may

use the actual (absolute) values of $\widehat{\theta}\left(EN, \tau_1, \tau_2\right)$:

$$w_i^\theta = |\widehat{\theta}_i\left(EN, \tau_1, \tau_2\right)|$$

To combine this with the SWa and SWb methods of Boivin and Ng (2006), we suggest to use a Cobb-Douglas style function to calculate the weights, with a parameter α which controls the degree to which the researcher targets based on predictive ability or targets for the factor model:

$$
\begin{aligned}
w_i^3 &= \left(w_i^\theta\right)^\alpha \left(w_i^{SWa}\right)^{1-\alpha} \\
&= \left(|\widehat{\theta}_i\left(EN, \tau_1, \tau_2\right)|\right)^\alpha \left(\widehat{\Omega}_{ii}^{-1}\right)^{1-\alpha}
\end{aligned}
\tag{12}
$$

and finally:

$$
\begin{aligned}
w_i^4 &= \left(w_i^\theta\right)^\alpha \left(w_i^{SWb}\right)^{1-\alpha} \\
&= \left(|\widehat{\theta}_i\left(EN, \tau_1, \tau_2\right)|\right)^\alpha \left(\left(\frac{1}{N}\sum_{j=1}^{N}|\widehat{\Omega}_{ij}|\right)^{-1}\right)^{1-\alpha}
\end{aligned}
\tag{13}
$$

where $\alpha \in [0,1]$ reflects the importance placed on targeting the factors to the forecast model as in Bai and Ng (2008), and therefore $1-\alpha$ reflects the importance placed on targeting the factors for factor model properties. To the best of our knowledge, this is the first paper to allow researchers this flexibility. Note that this method still eliminates some of the variables prior to factor estimation as the Elastic Net method sets some of the weights exactly to zero. If we, instead, wished to retain all N variables in this framework, we could instead use estimates from Ridge estimation which is a special case of the Elastic Net where $\tau_1 = 0$ in the penalized least squares objective function in Equation (6). This would give all variables non-zero weight, with the weights being a combination of the two types of targeting.

3.2 Targeted Generalized Principal Components Analysis

Our final proposed methodology is an estimation procedure which we call Targeted Generalized Principal Components Analysis (TGPCA). In this method we attempt to address the problem of non-invertability of $\widehat{\Omega}$. This allows us use Generalized PCA, unlike in Boivin and Ng (2006).

To describe this method, we first of all introduce some notation. For the Elastic Net pre-selection stage with tuning parameters τ_1 and τ_2,[2] let $\mathcal{M}(EN, \tau_1, \tau_2)$ be the set of variables corresponding to non-zero coefficients in the estimator $\widehat{\theta}(EN, \tau_1, \tau_2)$:

$$\mathcal{M}(EN, \tau_1, \tau_2) = \left\{ i : \widehat{\theta}_i(EN, \tau_1, \tau_2) \neq 0 \right\}$$

let $M(EN, \tau_1, \tau_2)$ be the number of non-zero coefficients in the estimator $\widehat{\theta}(EN, \tau_1, \tau_2)$.[3]

The Targeted Generalized Principal Components Analysis approach we suggest forms an $M \times M$ matrix, $\widehat{\Omega}(\mathcal{M})$, constructed by deleting the rows and columns for which $j \notin \mathcal{M}$ from the non-invertible matrix $\widehat{\Omega}$. The dependence of M and $\mathcal{M}$ on τ_1 and τ_2 is suppressed for notational convenience. The estimate $\widehat{\Omega}$ can be obtained using the standard PCA estimates $\widehat{u}_{it}$ as in Boivin and Ng (2006).

With the matrix $\widehat{\Omega}(\mathcal{M})$, the estimation procedure for TGPCA is the following optimization:

$$\min_{\Lambda, F_1, \ldots, F_T} \frac{1}{MT} \sum_{t=1}^{T} (X_t(\mathcal{M}) - \Lambda(\mathcal{M}) F_t)' \left[\widehat{\Omega}(\mathcal{M}) \right]^{-1} (X_t(\mathcal{M}) - \Lambda(\mathcal{M}) F_t) \tag{14}$$

subject to $F'F/T = I$, where the $M \times 1$ vector $X_t(\mathcal{M})$ and the $M \times r$ matrix $\Lambda(\mathcal{M})$ are similarly equal to X_t and Λ with rows $j \notin \mathcal{M}$ removed.

Clearly this methodology combines the best aspects of both types of targeted factor methodologies. The reliance of the objective function on $\mathcal{M}$ means that only the most relevant variables for the target variable y_{t+h} are retained. Furthermore, the weighting matrix $\left[\widehat{\Omega}(\mathcal{M}) \right]^{-1}$ gives lower weight to the variables with high idiosyncratic correlation. This results in the estimated factors being different for each forecast variable, but in a way which takes the properties of the factor model into account.

The most important implication, however, is that we can choose the tuning parameter τ_1 (or equivalently M) in such as way that the matrix

---

[2]We could, of course, do likewise for any other method used in the pre-selection phase, such as LASSO, or bagging and boosting as in Kim and Swanson (2014).

[3]Note that in Bai and Ng (2008) they choose $M = 30$ directly and select the tuning parameter by inverting the equation $30 = \#\left(i : \widehat{\theta}_i(EN, \tau_1, \tau_2) \neq 0 \right)$

$\widehat{\Omega}(\mathcal{M})$ is invertible, by setting $M << N - r$. One difficulty is that, even if $M << N - r$, it is still possible that the matrix $\widehat{\Omega}(\mathcal{M})$ has reduced rank and is not invertible. However, in practice this does not happen often, and this problem can be overcome by a simple algorithm which removes the row and column which gives the smallest minimum eigenvalue of the matrix.

Using this methodology, it is possible to weight both the variances and the covariances in the objective function, which is an improvement upon previous methodologies.

4   Data and Forecasting Methodology

We will perform a pseudo out-of-sample forecasting exercise to assess the relative forecasting performance of the methods proposed in Section 3, applied to a range of U.S. macroeconomic and financial variables. We compare the performance of our methods to the existing targeted factor methodologies of Bai and Ng (2008) and Boivin and Ng (2006), described in Section 2, and the standard PCA procedure of Stock and Watson (2002a,b). Each of these methods can be written in terms of the feasible factor-augmented regression analogue to Equation (2) with additional autoregressive components:

$$y_{t+h} = \beta' \widehat{F}_t + \alpha(L)\, y_t + \varepsilon_{t+h} \tag{15}$$

where $\alpha(L)$ is the lag operator. We will consider as a benchmark the autoregressive (AR) model which has $\beta = 0$. For the remainder of the models we will use factor-augmented regressions where the factors $\widehat{F}_t$ are estimated by the different methods mentioned above.

As the α parameter is new to this paper, used in constructing the WPCA weights w_i^3 and w_i^4 above, we will consider a grid of values of $\alpha = \{1/4, 1/2, 3/4\}$. As α increases from $1/4$ to $3/4$, we place more weight on targeting the factor estimates based on their predictive ability for the target variable and less for the properties of the factor model. Using different values of α will allow us to explore the sensitivity of our methods to this parameter.[4]

In total there are 14 different forecasting models, which are summarized in Table 1.

---

[4]We thank an anonymous referee for encouraging us to explore this further.

[Insert Table 1 here]

It is important to note that, of the factor-based models PCA through to Method 5 in Table 1, each of these will produce different factor estimates. Furthermore, due to the type of targeting, LA(PC) and Methods 1 through 5 will produce a different ('targeted') set of factors for each forecast variable. In contrast, the factors are the same for all forecast variables under PCA, SWa and SWb.

We will forecast a range of U.S. macroeconomic and financial variables taken from the Stock and Watson (2002a,b) dataset. This dataset was extended by Kim and Swanson (2014)[5] and contains monthly observations on 144 variables, for which we use the observations from 1964:M1 to 2009:M7. The forecast variables we are interested in are: the consumer price index (CPI), the producer price index (PPI), total employees on non-farm payrolls, the index of total industrial production (IP), the S&P 500 index and the 10-year treasury bills rate.

For the pseudo out-of-sample forecasting exercise, we split the sample into $T = R + P - 1$, where R is the estimation sample size and we make P pseudo out-of-sample forecasts. After taking lags of the dependent variable for the direct forecasting scheme we have $T = 545$ observations and we let $R = 246$ so that $P = 300$ forecasts are made for 25 years over the period 1984:M6 to 2009:M5. We use the rolling scheme as in Kim and Swanson (2014), so that the estimation window length is held fixed at R in each pseudo out-of-sample horizon. This means that at the first horizon we use data from $1 : R$, make a forecast of $R + h$, and in the second horizon we use data from $2 : R + 1$, make a forecast of period $R + h + 1$ and so on. Since this sample spans the year 1984, which is seen by many as a structural break point coinciding with the start of the "Great Moderation", we will also run results where we only estimate using data post-1984. This is motivated by studies of Breitung and Eickmeier (2011) and Stock and Watson (2009) who find evidence of factor loading instability around these dates.

For all variables we will use the cumulative h period growth for the dependent variable using the logarithmic transformation $y_{t+h} = 100 \left(\log\left(Y_{t+h}\right)\right.$ $\left. -\log\left(Y_t\right)\right)$, with the exception of the 10-year Treasury Bill where we specify $y_{t+h} = (Y_{t+h} - Y_t)$. We will focus on the one-year ahead forecast horizon

---

[5]We thank these authors for making their data available to us.

with $h = 12$. Regarding model specification, we will set the number of autoregressive lags at $p = 6$ in line with other studies, and set the number of factors equal to that chosen by the BIC_3 criterion of Bai and Ng (2002). We keep these parameters fixed rather than re-estimating them at each horizon as this facilitates the use of a Diebold-Mariano type testing procedure.

For the Elastic Net parameters used in the LA(PC) method and all of our competing Methods 1-5 in Table 1, as in Bai and Ng (2008) we set τ_1 at the level which allows $M = 30$ variables to have non-zero coefficients, and we fix the L_2 parameter at $\tau_2 = 0.5$. The results are not sensitive to changing τ_2, as mentioned in Bai and Ng (2008). This is because, while τ_2 may play some role in determining the ordering of variables *within* the top 30, it plays virtually no role in determining the full set of top 30 variables, and therefore has little or no impact factor estimation and therefore on the forecasts.[6]

The metric we use to compare forecasts is the mean squared forecast error ($MSFE$) loss function. For each model i, the pseudo out-of-sample forecast experiment gives rise to a string of P pseudo out-of-sample forecasts $\widehat{\varepsilon}_{t+h}(i) = y_{t+h} - \widehat{y}_{t+h}(i)$. The $MSFE$ for this model is estimated as the average of the squared forecast errors:

$$MSFE(i) = \frac{1}{P} \sum_{t=R}^{T} \widehat{\varepsilon}_{t+h}(i)^2$$

To facilitate comparison across models, we will report the $MSFE$ measure only for the autoregressive model, and for all other models we report the relative $MSFE$:

$$RMSFE(i) = \frac{MSFE(i)}{MSFE(AR)}$$

for $i = 2, ..., 14$. A value of $RMSFE(i)$ less than 1 indicates that model i has lower $MSFE$ than the AR model.

It is important to assess the statistical significance of these differences in $MSFE$. In order to do this we will use tests similar to those of Diebold and Mariano (1995) and West (1996). However, since we have multiple models under consideration, we control for the multiple testing problem by using the Model Confidence Set (MCS) approach of Hansen et al. (2011). The MCS procedure aims to 'estimate' the best set of models $\mathcal{M}^*$ from the total

---

[6]Some results for different values of τ_2 are available from the author on request.

set of alternative models $\mathcal{M}^0$, which in our case contains 14 members. The procedure starts with all 14 models and eliminates the worst models, according to rejection of the null hypothesis of equal predictive ability using the Diebold-Mariano test, until it arrives at a set $\widehat{\mathcal{M}}^*$. The main contribution of Hansen et al. (2011) is that they provide conditions under which it can be shown that $\lim_{n\to\infty} \mathrm{P}\left(\mathcal{M}^* \subset \widehat{\mathcal{M}}^*\right) \leq 1 - \alpha$, where α is the significance level of each of the tests.

It is possible that our various models are nested to some degree, as they all use factor estimates which should converge to the same true factors. However, we feel that use of this procedure is still justified as it was used for similar models in the empirical application of Hansen et al. (2011). The performance of the MCS procedure based on test statistics involving estimated factors remains an open research question which we leave for future work. To implement the test, we use the R package MCS, written by Bernardi and Catania (2014).

5 Results

Table 2 presents the results for the pseudo out-of-sample forecasting experiment described in the previous section. These results are based on the rolling estimation procedure using the full dataset from 1964 to 2009. From these results a few key findings emerge. The first main finding is that one of our proposed Methods 1-4 yields the lowest $MSFE$ for all of the variables considered. While the 'best' method is not the same for all of the variables, it can be seen that Methods 3 and 4 are the only ones of all methods considered which beat the AR model for every forecast variable. While the LA(PC) method of Bai and Ng (2008) also performs relatively well, our proposed Method 4 beats LA(PC) in all but one case. This means that there appears to be improvement in our combined targeting approach over the Bai and Ng (2008) approach, which only targets the factors for their predictive properties.

[**Insert Table 2 here**]

On the other hand, the standard PCA factor estimation method, methods SWa and SWb of Boivin and Ng (2006), and our proposed Method 5 do

14

not provide an improvement over the AR benchmark in any of the 6 cases. These results imply that in terms of forecasting, it appears to be more important to use factors which change with each dependent variable. This is in contrast to the methods PCA, SWa and SWb which give the same factor estimates regardless of the forecast variable, and do not perform as well.

The $MSFE$ improvements over the naïve AR benchmark model are at their largest for forecasting CPI inflation, producer prices and employment. In the case of employment this gain is as large as 26%, and for CPI and PPI this difference is 17% and 16% respectively.

Table 2 also provides the results for the Model Confidence Set at both the 90% and 75% levels, which are the levels used by Hansen et al. (2011). These results confirm the strong performance of our newly proposed methods with regards to statistical significance.[7] For each dependent variable, one of Methods 1 to 4 is included in the MCS. On the other hand, the methods which are most frequently eliminated from the MCS are the naïve AR model, the standard PCA model and the SWa and SWb methods of Boivin and Ng (2006). The fact that SWa and SWb are regularly eliminated appears to indicate that targeting factor estimation only for factor model properties does not yield significant forecast improvements. For the S&P 500, the MCS fails to eliminate even a single model, meaning that no method is very informative at predicting long-term stock market returns. This result could have been expected *a priori*.

Turning to the interpretation of the parameter α, we see from Table 2 that there are some differences in the $RMSFE$ statistics for a given method across different levels of α. However, when we look at the results of the MCS procedure, in the cases where one of Method 3 or 4 remains in the MCS, it is generally the case that all three, or in some cases two, of the combinations of α are included in the MCS. This indicates that there is little consequence of selecting a particular level of α for these methods and for the set of forecast variables we consider.

We also present the results from re-running the analysis only using post-1984 data. This involves using $T + h = 293$ observations and we set $R = 132$ and $P = 162$ so as to have a similar fraction P/R as in the full-sample case.

---

[7]In an earlier version of the paper, the MCS results were slightly different. We found that adding in certain models to the initial model set could sometimes alter the conclusions of the MCS procedure for the rest of the models. We leave a full analysis of the MCS, particularly involving estimated factors, to further study.

The results for this sample split are displayed in Table 3.

[**Insert Table 3 here**]

In these post-1984 results, many of the key features remain unchanged relative to the full-sample results. The best model for each forecast variable in terms of $MSFE$ is one of the newly proposed Methods. Each of Methods 2, 3, 4 and LA(PC) improve over the AR model in 5 of the 6 forecast variables. Once again, the standard PCA method, along with SWa and SWb of Boivin and Ng (2006) are among the worst-performing models. In the case of Treasury Bills, however, we now see that the Model Confidence Set procedure fails to eliminate a single model, whereas for the S&P 500 variable, there is a very sparse MCS including only Method 4. These results indicate that for these financial variables, the conclusion is rather sensitive to the choice of sample and so we might treat these results with some caution.

For CPI, PPI and Employment and IP, the only methods which survive elimination are either LA(PC) or Methods 1 through 4, but not the AR, PCA or SWa and SWb. This, again, indicates that there is merit in targeting factor estimates to a particular forecast variable, and that our methods which additionally target for factor model properties perform strongly. It also indicates that the economic information contained in the targeted factors is useful in predicting these variables over and above a naïve autoregressive specification.

Overall, the conclusions we draw from these results is that the best performing methods in most cases tends to be our proposed methods which advocate targeting factors both for the forecast variable and for factor model properties. We proposed 4 methods which were successful at forecasting (Method 5 did not seem to be successful), in the sense that these had the lowest $MSFE$ in almost every single case. Since no single method of the 4 was always the 'winner', one might consider averaging the forecasts from these methods. On the other hand, the results for financial variables such as stock prices and treasury bill yields were somewhat sensitive to changing the sample.

6   Conclusion

In this paper we have proposed new methods of targeting factor estimates from big datasets for use in economic forecasting. We suggest that factor-based forecasts may be improved if we adjust factor estimation to up-weight the variables which are strong predictors for a certain target forecast variable, and down-weight variables which are noisy and may worsen the precision of factor estimates. This is in contrast to existing methods like Bai and Ng (2008) and Boivin and Ng (2006) which are only capable of adjusting factor estimates for one of these two purposes. In Section 3 we presented new weighted Principal Components Analysis procedures where the weights reflected both of these two targeting ideas. We also proposed a Targeted Generalized PCA procedure which allowed us to overcome the problem of feasible Generalized PCA in non-targeted cases in which the idiosyncratic error variance-covariance matrix is not invertible.

We applied our new forecasting methodologies to a wide range of U.S. macroeconomic and financial variables using a pseudo out-of-sample forecasting experiment. We find strong evidence that our proposed methods work better than competing targeted factor methods, and non-targeted methods. Particularly in forecasting variables like CPI inflation, we find that our methods out-perform other candidate methods, as evidenced by their survival in the Model Confidence Set procedure of Hansen et al. (2011). Future work would apply these methods to a wider range of variables and countries, to determine whether or not they may also be useful in situations other than forecasting the U.S. economy.

References

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 (2008). Forecasting economic time series using targeted predictors. *Journal of Econometrics 146*(2), 304–317.

Bernardi, M. and L. Catania (2014). The model confidence set package for R. *arXiv preprint arXiv:1410.8504*.

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

Breitung, J. and S. Eickmeier (2011). Testing for structural breaks in dynamic factor models. *Journal of Econometrics 163*(1), 71–84.

Bulligan, G., M. Marcellino, and F. Venditti (2015). Forecasting economic activity with targeted predictors. *International Journal of Forecasting 31*(1), 188–206.

Castle, J. L., M. P. Clements, and D. F. Hendry (2013). Forecasting by factors, by variables, by both or neither? *Journal of Econometrics 177*(2), 305–319.

den Reijer, A. J. (2012). Forecasting dutch gdp and inflation using alternative factor model specifications based on large and small datasets. *Empirical Economics*, 1–19.

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

Efron, B., T. Hastie, I. Johnstone, and R. Tibshirani (2004). Least angle regression. *The Annals of Statistics 32*(2), 407–451.

Eickmeier, S. and T. Ng (2011). Forecasting national activity using lots of international predictors: An application to new zealand. *International Journal of Forecasting 27*(2), 496–511.

Hansen, P. R., A. Lunde, and J. M. Nason (2011). The model confidence set. *Econometrica 79*(2), 453–497.

Kim, H. H. and N. R. Swanson (2014). Forecasting financial and macroeconomic variables using data reduction methods: New empirical evidence. *Journal of Econometrics 178*(1), 352–367.

Schumacher, C. (2010). Factor forecasting using international targeted predictors: The case of german gdp. *Economics Letters 107*(2), 95–98.

Stock, J. and M. Watson (2009). Forecasting in dynamic factor models subject to structural instablility. In J. Castle and N. Shephard (Eds.), *The Methodology and Practice of Econometrics: a Festschrift in Honour of David F. Hendry.* Oxford: Oxford University Press.

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). Macroeconomic forecasting using diffusion indexes. *Journal of Business and Economic Statistics 20*(2), 147–162.

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

West, K. D. (1996). Asymptotic inference about predictive ability. *Econometrica 64*(5), 1067–1084.

Zou, H. and T. Hastie (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society. Series B (Statistical Methodology) 67*(2), 301–320.

7 Tables

Table 1: Description of Forecasting Methods

	Method	Description
1	AR	Autoregressive Model
2	PCA	Standard PCA - Stock and Watson (2002a,b)
3	LA(PC)	Targeted PCA - Bai and Ng (2008)
4	SWa	Weighted PCA - Boivin and Ng (2006) SWa
5	SWb	Weighted PCA - Boivin and Ng (2006) SWb
6	Method 1	Weighted PCA - Weights w_i^1 in Equation (10)
7	Method 2	Weighted PCA - Weights w_i^2 in Equation (11)
8	Method 3 ($\alpha = 0.25$)	Weighted PCA - Weights w_i^3 in Equation (12)
9	Method 3 ($\alpha = 0.5$)	Weighted PCA - Weights w_i^3 in Equation (12)
10	Method 3 ($\alpha = 0.75$)	Weighted PCA - Weights w_i^3 in Equation (12)
11	Method 4 ($\alpha = 0.25$)	Weighted PCA - Weights w_i^4 in Equation (13)
12	Method 4 ($\alpha = 0.5$)	Weighted PCA - Weights w_i^4 in Equation (13)
13	Method 4 ($\alpha = 0.75$)	Weighted PCA - Weights w_i^4 in Equation (13)
14	Method 5	Targeted Generalized PCA

Table 2: Pseudo Out-of-Sample Forecasting Results - Full Sample

	CPI	PPI	Employment	IP	S&P 500	10 Year T-Bill
			MSFE			
AR	1.85	7.35	1.98	12.04	334.71**	1.56
			Relative *MSFE*			
PCA	1.10	1.08	1.01	1.16	1.11**	1.02
LA(PC)	0.84	0.92	0.77**	0.95**	1.02**	0.98
Swa	1.05	1.06	1.09	1.23	1.11**	1.03
SWb	1.08	1.05	1.04	1.18	1.11**	1.03
Method 1	0.96	1.02	0.94	1.06	1.04**	0.90**
Method 2	0.83**	0.94	0.78**	0.97	1.01**	0.93**
Method 3 ($\alpha = 0.25$)	0.96	0.97	0.81**	0.99	0.97**	0.88**
Method 3 ($\alpha = 0.5$)	0.88	0.89**	0.74**	0.95**	0.91**	0.91**
Method 3 ($\alpha = 0.75$)	0.85	0.86**	0.81**	0.92**	0.84**	0.91**
Method 4 ($\alpha = 0.25$)	0.92	0.90**	0.71**	0.94**	0.93**	0.93**
Method 4 ($\alpha = 0.5$)	0.91	0.88**	0.74**	0.91**	0.86**	0.91**
Method 4 ($\alpha = 0.75$)	0.89	0.84**	0.81**	0.90**	0.84**	0.91**
Method 5	1.18	1.13	1.04	1.11	1.11**	1.07

Notes: For the AR model, the *MSFE* is reported. This *MSFE* is used to calculate the Relative *MSFE* reported for the remaining models, as described in the text. Description of each of the 14 forecasting methods are provided in Table 1. The forecasts in the sets $\widehat{\mathcal{M}}^*_{90\%}$ and $\widehat{\mathcal{M}}^*_{75\%}$ are denoted * and ** respectively.

Table 3: Pseudo Out-of-Sample Forecasting Results - Post-1984

	CPI	PPI	Employment	IP	S&P 500	10 Year T-Bill
				MSFE		
AR	1.35	8.02	2.13	16.84	458.31	0.66**
				Relative *MSFE*		
PCA	1.23	1.19	0.75	0.99	1.10	1.19**
LA(PC)	0.93**	0.79	0.54**	0.55	0.98	1.40**
Swa	1.29	1.25	0.76	0.97	1.13	1.22**
SWb	1.25	1.22	0.74	0.95	1.11	1.20**
Method 1	1.04	0.98	0.54**	0.63	1.01	1.32**
Method 2	0.96*	0.85	0.55**	0.58	0.97	1.28**
Method 3 ($\alpha = 0.25$)	0.92**	0.88	0.50**	0.56	0.87	1.31**
Method 3 ($\alpha = 0.5$)	0.87**	0.79	0.51**	0.47	0.79	1.34**
Method 3 ($\alpha = 0.75$)	0.85**	0.73**	0.54**	0.42	0.75**	1.21**
Method 4 ($\alpha = 0.25$)	0.88**	0.74**	0.53**	0.50	0.87	1.46**
Method 4 ($\alpha = 0.5$)	0.86**	0.73**	0.55**	0.44	0.75	1.21**
Method 4 ($\alpha = 0.75$)	0.81**	0.70**	0.55**	0.41**	0.77	1.19**
Method 5	1.21	1.11	0.65	0.85	1.09	1.22**

Notes: Results are run using data post-1984. For the AR model, the *MSFE* is reported. This *MSFE* is used to calculate the Relative *MSFE* reported for the remaining models, as described in the text. Description of each of the 14 forecasting methods are provided in Table 1. The forecasts in the sets $\widehat{\mathcal{M}}^*_{90\%}$ and $\widehat{\mathcal{M}}^*_{75\%}$ are denoted * and ** respectively.