ERASMUS UNIVERSITY ROTTERDAM

Erasmus School of Economics

Forecasting Nigerian GDP growth trough factor models and machine learning techniques

Bachelor Thesis

March 2022

Author: Daan de Roode(482793)

Supervisor: P.H.B.F. Franses

Second Assessor: C. Carvicchia

Abstract

This paper investigates whether a hybrid forecasting model is useful in forecasting the gross domestic product (GDP) growth of Nigeria. A data set is used that consists of the yearly GDP levels of 51 other African countries, for the period 1960-2016. For the forecasts, the lagged GDP growth of Nigeria is used, combined with the factors produced by the hybrid model. The forecasts are evaluated by comparing them with a standard autoregressive model of order 1 (AR(1)), by means of their mean squared forecasting errors (MSFE). In addition, the model specifications are evaluated through simulated data, using numerous data generating processes. The results showed that the hybrid forecasting models outperform the standard AR(1) model, leading to the conclusion that machine learning techniques and factor models are useful in forecasting the GDP growth of Nigeria.

Contents

1 Introduction

Big-data is becoming a more and more important tool in life nowadays. Take for example the self-driving Tesla, or personalized advertisements on social media. They are all conducted through big-data analyses. And with this big-data available, there is also a large variety of different research methods to analyze this big-data, especially in the field of machine learning and dimension reduction.

Because big-data often contains a lot of variables, it can be costly to evaluate all of these variables in the models. Hence, Kim and Swanson (2018) reduce the dimensions of the models by constructing diffusion indices. They do so by implementing numerous factor analysis methods, such as principal component analysis (PCA). They then combine these factor models with various machine learning and shrinkage methods to construct 'hybrid' forecasts of 11 different macroeconomic variables in the United States. These hybrid forecasts are compared to a few benchmark models using the mean squared forecast error (MSFE). From the results of table 4 of Kim and Swanson (2018), it follows that the method Boosting often results in the MSFE-best forecast. Boosting was first mentioned by Schapire (1990), who used it to classify problems in *probably approximately correct* learning models. Later on, Bai and Ng (2009) used boosting for selecting the predictors is factor augmented autoregressions. Since boosting seems to perform well according to Kim and Swanson (2018), this paper uses boosting for predictor selection.

This paper uses similar methods as mentioned in Kim and Swanson (2018) to forecast economical growth of Nigeria. A popular measure of wealth of a country is the Gross Domestic Product (GDP). Africa belongs to the poorest regions of the world. Hope Sr (2009) found that in 2005 51% of Sub-Saharan Africa lived of less than 1.25 US dollar per day, and 73% of less than 2.00 US dollar. However, the GDP of most African countries has improved over the last 50 years, and even though Africa is still one of the poorest regions of the world, it is developing. This paper tries to find out if there is a connection in the growth of GDP between the different countries. A data set is available for 52 different African countries, containing yearly GDP values from 1960 till 2016. With this data set, a forecast is made of the GDP growth of Nigeria, based on the GDP growth of the other African countries and on the lagged GDP growth of Nigeria. Based on the different models, this paper tries to answer the question: *Are factor models combined with the shrinkage method boosting useful for forecasting the Nigerian GDP growth, compared to standard autoregressive models?*

Suppose the factor models and machine learning techniques do make better forecasts of the GDP growth rates than the standard autoregressive models, this could be very interesting for

Nigeria. Nigeria is the highest populated country of Africa, and according to Varrella (2020) around 40% of the people live in poverty. When the models are able to accurately forecast the GDP growth rates, monetary policies can be adjusted to these forecasts and possibly even try to lower this level of poverty.

In the remainder of this paper, first some literature about the subject will be discussed in section 2. Then the data that is used in this paper will be evaluated in section 3. Next the different methods and models that are used to construct the forecasts will be elaborated on in section 4. In addition, the results will be reviewed in section 5, and finally a conclusion will be drawn in section 6.

2 Literature Review

Over the years a lot of research has been conducted on efficient ways of dimension reduction when a high dimensional set of predictor variables is available.

Stock and Watson (2002a) invented the so-called diffusion index model. In this model, a data set with more predictors than time observations is used. These predictors are then explained by relatively few diffusion indexes, which are constructed through a dynamic factor model. Stock and Watson (2002a) used principal component analysis as the the dynamic factor model. They forecast the dependent variable through a linear regression, consisting of the principal components, and a set of pre-determined explanatory variables. These pre-determined explanatory variables are variables that are often used by other researchers to forecast this particular dependent variable. Stock and Watson (2002a) found that the forecasts of the diffusion index model are asymptotically efficient and consistent. Furthermore, they also conducted empirical evidence, by forecasting the industrial production index using 149 explanatory macroeconomic variables. The results showed that the models based on principal components and autoregressive components outperform the benchmark autoregressive models. Stock and Watson (2002b) presented further evidence on the superiority of diffusion index models above autoregressive models.

Despite the improvement of forecasting accuracy that the diffusion models have shown over the years, Bai and Ng (2009) found two major flaws in the model. The first shortcoming is that the existing information criteria on selecting the number of factors assume that the components are ordered. It is true that the components are ordered. However, the components are ordered by part of the variance they explain of the set of explanatory variables. This does not have to mean that the same ordering applies to the importance to the dependent variable. In addition, Bai and Ng (2009) argued that the specification of the forecasting equation is not right. If the p^*th lag of the autoregression is chosen, all the previous lags $p = 1, ..., p^*$ are chosen as well. The same rule applies for choosing the right amount of principal components, since the factors are ordered by importance to explaining the explanatory variables.

To select the right number of lags, while selecting the appropriate lags to include as well, Bai and Ng (2009) started to use boosting to pre-select the variables that are important for the dependent variables, and perform PCA on these variables afterwards. Boosting was first introduced by Schapire (1990). Schapire (1990) used boosting to increase the accuracy of predictions, by combining the results of many so-called 'weak learners'. Ordinary least squares (OLS) is an example of a 'weak learner'. In other words, a prediction is formed out of a linear combination of many individually estimated 'weak learners'. Schapire (1990) only used boosting for classification problems. Later on Ridgeway et al. (1999) introduced boosting to regression type problems. They do so by casting the regression problem into a classification problem. However, they did not modify the boosting algorithm in such a way that it can be used for time series. Bai and Ng (2009) started with using boosting in a time series regression with factor analysis. Boosting is used to select the most valuable variables from a large set of predictor variables, while factor analysis compresses the large set of variables into a smaller set of variables. There are two methods of combining factor analysis with boosting: First applying boosting, and then factor analysis or first applying factor analysis, and then applying boosting on the new predictors. Bai and Ng (2009) argued that the best method of combining strongly depends on the data set. Hence, researchers have to think carefully about what method should be applied to their data set.

Kim and Swanson (2018) used numerous methods of factor analysis and machine learning techniques to forecast 11 macroeconomic variables in the United States. Besides the well-known PCA, Kim and Swanson (2018) also used Sparse principal component analysis (SPCA) and Independent principal component analysis (ICA). They combined these factor analysis methods with different machine learning techniques, such as bagging, boosting, and ridge regression. Furthermore they combined these different factor models and machine learning techniques into four different specifications. In specification 1 they first construct factors from the different factor models, and then perform machine learning techniques on these factors. In specification 2 they first make a subset of the variables by performing the machine learning techniques on the data set, and then perform the factor models on this subset. In specifications 3 and 4 only machine learning techniques are used. The forecasts were made using the diffusion index methods that are introduced by Stock and Watson (2002a). The MSFE's of the different forecasts were compared afterwards. The results showed that PCA computes better forecasts for longer time horizons, while SPCA and ICA compute better forecasts for smaller time horizons. In addition, models combining the factor models and machine learning techniques almost always outperformed the models that solely use machine learning techniques.

3 Data

The data set that is used in this paper contains yearly GDP values of 52 different African countries from 1960 till 2016. The values are stated in indices, meaning that every country has value 100.00 in 1960. The values in the years following represent the GDP level relative to the level in 1960. Hence, the growth levels are calculated by $\text{growth} = 100 * \log(\frac{y_t}{y_{t-1}})$ where y_t equals the GDP index at time t.

Figure 1 shows the graph of the growth rates for Nigeria over time. The figure shows a clear negative peak around 1967. O'Brien (1973) states that this decline in GDP growth is caused by the Nigerian civil war, which started in 1967. In 1967, the Republic of Biafra declared its independence from Nigeria. The consequence of this declaration of independence is that the Biafran national accounts were excluded from Nigeria's national accounts, causing the decline in GDP growth. On the contrary, a large positive peak is observed around 1970. Nigeria was winning the civil war, leading to the re-inclusion of Biafra's national accounts to Nigeria's national accounts (O'Brien (1973)). After 1970, the figure shows a lot of fluctuations in the GDP growth of Nigeria. From 1995 onwards the GDP growth levels stayed positive, meaning that the GDP level kept on growing since then.

Table 1: Descriptive stats growth rates Nigeria

	observations	**mean**	**max**	**min**	**st.dev**
1961 - 1972	12	5.005	22.314	-17.079	10.862
1973 - 1994	22	1.044	12.045	-14.041	7.239
1995 - 2015	21	5.893	29.043	0.300	5.820

Table 1 contains some descriptive statistics about the growth rates of the Nigerian GDP and gives some more explanation to figure 1. Table 1 shows that even around the time of the Nigerian civil war, the GDP of Nigeria still increased with an average level of 5.05% per year. In the period from 1973 to 1994, a lot of fluctuations showed, leading to a smaller increase in GDP of averagely 1.04%, and from 1995 onwards the GDP kept on increasing with an average level of 5.89%.

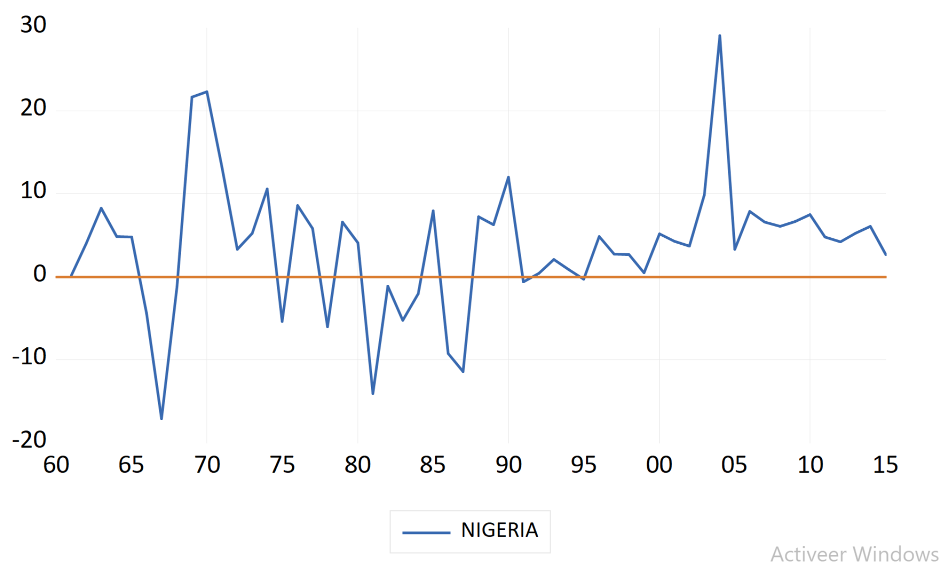


Figure 1. Graph of growth rates of Nigerian GDP over time, where the x-axis indicates the year, and the y-axis the GDP growth

4 Methodology

4.1 Forecasting Method

This paper uses the following model to forecast the GDP growth of Nigeria, similar as in Stock and Watson (2002a), Stock and Watson (2002b), and Kim and Swanson (2018):

$$Y_{t+h} = W_t \beta_W + F_t \beta_F + \epsilon_{t+h} \tag{1}$$

Here Y_t is equal to the GDP growth of Nigeria at time t, h is equal to the forecast horizon. W_t is equal to a $1 * s$ vector of 'extra' explanatory variables, which in this paper is equal to Y_t. F_t are the factors that are constructed through PCA (section 4.2.1) or SPCA (section 4.2.2) and Boosting (section 4.3). ϵ_{t+h} is an error term. In addition, β_W and β_F are the coefficients of the 'extra' explanatory variables and the constructed factors respectively.

To make a forecast of Y_{t+h}, first $\hat{\beta}_W$ and $\hat{\beta}_F$ have to be estimated. To estimate $\hat{\beta}_F$, first the factors are constructed based on Y_t by performing boosting and PCA/SCPA. Subsequently Y_t is regressed on F_{t-h} to estimate $\hat{\beta}_F$, which can then be used to estimate Y_{t+h}. The estimation of $\hat{\beta}_W$ is a lot more straightforward, since W_t is equal to Y_t. Hence, to estimate $\hat{\beta}_W$, Y_t is simply regressed on W_{t-h}.

As mentioned before, Kim and Swanson (2018) discuss four different specifications for the forecasts. This papers uses the second specification, which means that first Boosting is applied on the explanatory variables to make a subset of important variables. Then, PCA and SPCA are performed on this subset of important variables to construct the final factors that are used in (1).

4.2 Factor models

In this section the factor models that are used to construct the common factors are discussed. First PCA is discussed, as it has been done in the papers written by Kim and Swanson (2014), Stock and Watson (2002a) and Bai and Ng (2008). Besides PCA, this paper also uses SPCA as mentioned in the papers written by Zou et al. (2006) and Kim and Swanson (2018).

4.2.1 Principal Component Analysis

The main idea of PCA is to reduce the number of predictors. Since this paper first performs boosting, and then PCA, the set of predictors exists of a subset of the original set of explanatory variables; the GDP levels of the other 51 African countries. PCA reduces the number of predictors by rewriting the set of predictors in a new, smaller set of common factors. These common factors are chosen in such a way that they contain as much information from the predictors as possible.

Let $X_{(TxN)}$ be the matrix of N potential predictors. PCA first finds a linear combination of the columns of X that explains as much variance of X as possible. This is the first principal component. Subsequently, PCA finds a second linear combination of the columns of X that explain as much variance as possible, but is uncorrelated with the first linear combination, this is the second principal component. This procedure is repeated until no more linear combinations can be formed. The general formula looks as follows.

$$X_t = \Lambda' F_t + \epsilon_t \tag{2}$$

Here, $X_t = (X_{t,1}, ..., X_{t,N})$ for t = 1...T. $F_t = (f_{t,1}, ..., f_{t,r})$ is the vector of factors, where r<N, and Λ are the associated factor loadings. ϵ_t is the error term. In this formula, $\Lambda' F_t$ is defined as the common factor of X_t.

4.2.2 Sparse Principal Component Analysis

The other factor model that is included in this paper is SPCA. With PCA it is sometimes difficult to interpret the results, because every factor is a linear combination of the original variables. The main idea of SPCA is to first adress PCA like a (ridge) regression, and then to use elastic net to obtain sparse principal components. Zou et al. (2006).

Since PCA is already performed, this paper uses the two-stage exploratory analysis mentioned in section 3.2 of Zou et al. (2006). That is, first PCA is considered as a regression

$$
\begin{aligned}
(\hat{\Delta}, \hat{\Lambda}) = & \min_{\Delta,\Lambda} \sum_{t=1}^{T} ||X_t - \Delta\Lambda' X_t||^2 + \eta \sum_{j=1}^{r} ||\lambda_j||^2 \\
& \text{subject to } \Delta'\Delta = I_r
\end{aligned} \tag{3}
$$

Where $\Delta\Lambda'$ equals a $r * r$ matrix of factor loadings. Because of the orthonormality constraint of Δ, if $\Lambda = \Delta$, $\hat{\Lambda}$ becomes the same r factor loadings as with the regular PCA. When adding a lasso penalty constraint to this equation, sparse factor loadings are obtained.

$$
\begin{aligned}
(\hat{\Delta}, \hat{\Lambda}) = & \min_{\Delta,\Lambda} \sum_{t=1}^{T} ||X_t - \Delta\Lambda' X_t||^2 + \eta \sum_{j=1}^{r} ||\lambda_j||^2 + \sum_{j=1}^{r} \eta_{1,j} ||\lambda_j||_1^2 \\
& \text{subject to } \Delta'\Delta = I_r
\end{aligned} \tag{4}
$$

Where $||\lambda_j||$ is equal to $\sum_{i=0}^{N} |\lambda_{ij}|$. Also η is the same for all r components, while different values of $\eta_{1,j}$ are allowed. $\hat{\lambda}_j' X$ is equal to the j^{th} principal component, where $\hat{\lambda}_j$ is the j^{th} column of $\hat{\Lambda}$.

4.2.3 Selecting the number of factors.

The aim of using factor models is to describe X_{tj} with a limited number of factors. Hence, it is introduce a selection criterion on this number of factors. One does not want to use too many factors, while the factors that are used do need to contain enough predictive performance. Many different methods for selecting the number of factors when using PCA have been conducted through the years, as is shown in Kim and Swanson (2018). This paper uses the method as described in Bai and Ng (2002). That is, the following formula is followed.

$$
IC(k) = \ln(V(k, \hat{F}^k)) + k * \left(\frac{N+T}{N*T}\right) \ln\left(\frac{N*T}{N+T}\right) \tag{5}
$$

Where k is the number of factors to select, $\hat{F}^k$ means that there are k factors allowed and V(.) minimizes the Euclidean distance between true variables and their factor representations. This formula is repeated for different values of k. The value of k that minimizes IC(k) is the number of factors that is used.

Since there has not been any research on how to select the number of factors when using SPCA

(Kim and Swanson (2018)), this paper uses the method from Bai and Ng (2002) for SPCA as well. That is, the number of factors that is determined for PCA is also used as the number of factors for SPCA.

4.3 Boosting

Besides the factor models described above, the boosting algorithm as described by Bai and Ng (2009) is also used. Boosting is a shrinkage method that produces many outputs for different models, called weak learners, and combines these to one output. These weak learners are user-determined, and are used repeatedly on modified data, usually from previous iterations (Kim and Swanson (2013)). The output comes from minimizing a loss function that is averaged over the training data, and is eventually used for the final boosting procedure, which is a linear combination of the different weak learners.

First, the set of potential predictors z_t has to be chosen. This z_t consist of all the variables that can be used to fit Y. In this paper, z_t consists of X_t and lagged values of X_t. That is, $z_t = (Z_t, Z_{t-1}, ..., Z_{t-pmax})$, where $Z_T = (X_{t,1}, ..., X_{t,N})$. Z_t is a 1∗N vector, meaning that z_t consist of N∗pmax = R vectors. pmax denotes the amount of lags of X that are used as predictors, this paper uses different values for pmax. The boosting algorithm is then performed for M iterations. In each iteration, the predictor that has the smallest sum of squared residuals amongst all predictors is chosen. The original set of predictors is then updated for the next iteration with steplength v. The algorithm of this procedure looks as follows.

Algorithm 1 Component-wise L2 Boosting

Initialize $\hat{\phi}_{t,0} = \bar{y}$ for every t
for m = 1,...,M **do**
  for t=1,...,T **do**
    Let $u_t = y_t$ - $\hat{\phi}_{t,m-1}$ be the current residual
  end for
  for i = 1,...,n **do**
    regress residual vector u on the ith regressor of Z (z_i) to obtain $\hat{b}_i$. With this $\hat{b}_i$, compute $\hat{e}_i$ = u - $z_i\hat{b}_i$ and SSR = $\hat{e}_i'$ $\hat{e}_i$
  end for
  Choose the i_m^* that minimizes the SSR, so $SSR_{i_m^*} = min_{i^* \in [1,...N]} SSR_i$
  Let $\hat{\eta}_m = z_{i_m^*} \hat{b_{i_m^*}}$
  for t = 1,...,T **do**
    update $\hat{\phi}_{t,m} = \hat{\phi}_{t,m-1} + \nu * \hat{\eta}_{t,m}$, with $0 < \nu < 1$ the steplength.
  end for
end for

Take P as the matrix containing all predictors that were chosen in the Boosting algorithm. Since it is possible for predictors to be chosen multiple times in the Boosting algorithm, P can contain duplicates. These duplicates are eventually removed from the matrix P.

4.3.1 Determining the number of predictors after the Boosting algorithm

To prevent the algorithm from over-fitting the model, a stopping rule has to be implied. That is, the number of iterations M that is chosen has to minimize the following Information Criterion.

$$IC(m) = \log(\hat{\sigma}^2) + \frac{A_T * dof_m}{T} \tag{6}$$

Here, $\hat{\sigma}^2 = \sum_{t=1}^{T}(y_t - \hat{\phi}_{t,m})^2$. The degrees of freedom are chosen as $dof_m = trace(B_m)$, where B_m looks as follows.

$$B_m = B_{m-1} + \nu * P^{(m)}(I_T - B_{m-1}) = I_T - \prod_{j=0}^{m}(I_T - P^{(j)}) \tag{7}$$

Where $P^{(m)} = z_{i_m^*}(z'_{i_m^*} z_{i_m^*})^{-1} z'_{i_m^*}$, and $B_0 = \frac{\iota_T \iota_T'}{T}$. ι_T is a 1xT vector of ones.
In (5), A_t is chosen in such a way that it equals the Bayesian Information Criterion (BIC). That is, $A_t = \log(T)$.

4.4 Forecast evaluation

4.4.1 Benchmark model

To evaluate the forecasts that were made using the discussed methods, the forecasts of a benchmark model are made. This model is autoregressive model (AR(p)) as discussed in Franses et al. (1998), which looks as follows

$$Y_t = \alpha + \beta_1 Y_{t-1} + ... + Y_{t-p} + \epsilon_t \tag{8}$$

This is called an autoregressive model of order p, since there are p lags included. The number of lags to include, are determined by the SIC as mentioned in Schwarz (1978).

To compare the forecasts made by the diffusion index models with the forecasts of the benchmark model, the Mean Squared Forecast Error (MSFE) is used, which is defined by

$$MSFE_h = \sum_{t=S-h+2}^{T-h+1} \left(Y_{t+h} - \hat{Y}_{t+h}\right)^2 \tag{9}$$

Where S equals the estimation sample, h the forecast horizon and $\hat{Y}_{t+h}$ the forecast of Y_{t+h}. The MSFE's of the different forecasts are then compared with one another.

To check whether the difference in MSFE is significant, the Diebold-Mariano test is used (Diebold and Mariano (2002)). This test looks as follows (Van Dijk (2019)).

$$\mathsf{DM} = \frac{\bar{d}}{\sqrt{V(\widehat{d_{t+1}})/P}} \tag{10}$$

Here, P is the forecast window, $\bar{d}$ is the sample mean of the loss differential $d_{t+1} = (y_{t+1} - \widehat{y^{DI}_{t+1}})^2 - (y_{t+1} - \widehat{y^{AR}_{t+1}})^2$ over the forecast window, and $V(\hat{d}_{t+1})$ is an estimation of the variance of d_{t+1}, which looks as follows.

$$V(\widehat{d_{t+1}}) = \frac{1}{P-1} \sum_{t=T}^{T+P-1} (d_{t+1} - \bar{d})^2 \tag{11}$$

The Diebold-Mariano test follows a standard normal distribution asymptotically. This paper uses the two-sided Diebold-Mariano test to test for significant differences. That is, the null hypothesis is as follows

H0: The forecast errors of the diffusion index model do not differ significantly from the forecast errors of the benchmark model.

Since the Diebold-Mariano test follows a N(0,1) distribution asymptotically, and since this paper uses the two-sided version of the test, the null hypothesis will not be rejected at a 95% confidence for test values that lie between -1.96 and 1.96.

4.4.2 Simulation

To evaluate whether the factor estimation methods and boosting depend on the number of variables and observations, multiple data generating processes (dgp) are used to simulate different data sets. To do so, K factors F are constructed through an AR(1) process, as shown in the equation below.

$$F_{i,t} = \phi_i F_{i,t-1} + \epsilon_{i,t} \qquad i = 1, ..., K \qquad t = 1, ..., T \tag{12}$$

Where $F_{i,0} = 0$, $\epsilon_{i,t} \sim$ N(0,1) and different values of ϕ_i, T and K will be used for the different data generating processes.

Next, the explanatory variables X and the dependent variable Y can be generated by the factors as follows.

$$X_{j,t} = \lambda_{j,1} F_{1,t} + ... + \lambda_{j,K} F_{K,t} + \eta_{X,j,t} \qquad j = 1, ..., N \qquad t = 1, ..., T \tag{13}$$

Where $\lambda_{j,i} \sim$ N(0,1) and $\eta_{X,j,t} \sim$ N(0,1). For Y_t, all factor loadings are set equal to 1, which looks as follows:

$$Y_t = F_{1,t} + ... + F_{K,t} + \eta_{Y,t} \qquad t = 1, ..., T \tag{14}$$

Now that the simulated data is generated, this data can be used to examine the effectiveness of the model, and how the model reacts to different data generating processes. The data generating processes will be used to test different sections of the model. First, the selection criterion is tested that uses only factor models. That is, data will be generated for different values of N, T and K. PCA will then be applied to these simulated datasets. The goal of these simulations is to test the effectiveness of the selection criterion for different input variables. The effectiveness will be evaluated based on hitrates. Each data generating process will be simulated 100 times. The number of times that the model selects the right number of factors will be counted as successes. The hitrate is then the number of successes divided by the number of simulations. The data generating processes look as follows. Next, the selection criterion of the Boosting

dgp1:	N $\in$(1,100)	T = 50	K = 3
dgp2:	N = 50	T $\in$ (1,100)	K = 3
dgp3:	N = 50	T = 50	K $\in$ (1,8)

algorithm is evaluated. To do so, data will again be simulated through different data generating processes. First of all the effect of the steplength v, which indicates the shrinkage coefficient in each iteration, will be evaluated. Again, different values of N and T will be evaluated as well. Compared to this paper, Kim and Swanson (2018) used a data set with a lot more variables. Hence it is important to test for different values of N and T again, because there is a possibility that the model does not perform well for data sets with dimensions similar to this paper. In that case, the factors in this paper could be unreliable. The data generating processes for the hybrid model look as follows.

dgp4:	N = 50	T = 50	K = 2	v $\in$ (0,1)
dgp5:	N $\in$(50, 200)	T = 50	K = 2	v = 0.5
dgp6:	N = 50	T $\in$ (50, 500)	K = 2	v = 0.5
dgp7:	N = 50	T = 50	K $\in$ (1,5)	v = 0.5

At last, the forecasting ability of the model has to be evaluated. This is done by comparing the MSFE of the benchmark model with the MSFE of the diffusion index model based on simulated data. In addition, the forecast errors of the models will be evaluated, to test whether they correspond with the residuals that were used to make X. This last data generating process looks as follows.

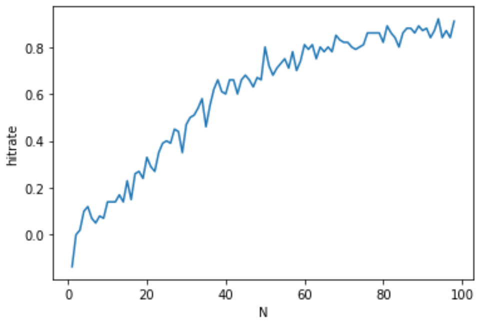
dgp8:	N = 50	T = 55	K = 2	v = 0.5

5 Results

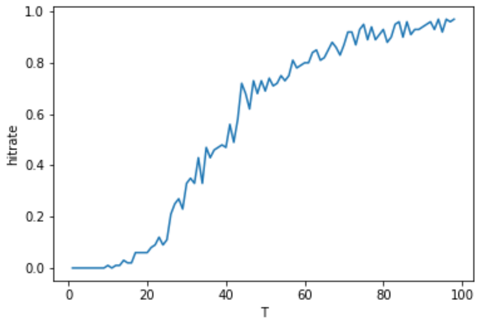
In this section the results of the forecasts will be discussed. First, the outcomes of the different data generating processes will be evaluated, and afterwards the forecast ability of the hybrid model will be compared to the forecast ability of the benchmark model.

5.1 Simulation

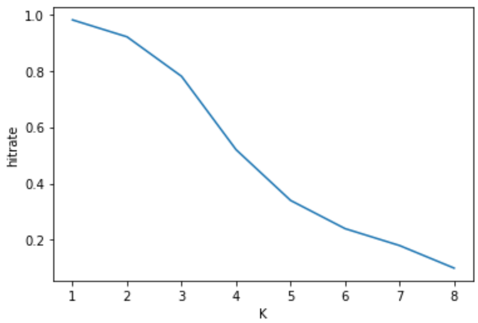
First, the data generating processes that were used to evaluate the selection criterion that is used only on factor models are discussed. The results of these first three data generating processes are shown in figure 2a, 2b and 2c below.



(a) *dgp1*: Hitrates for different values of N



(b) *dgp2* Hitrates for different values of T



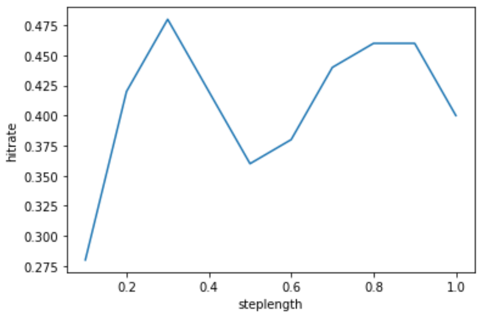
(c) *dgp3* Hitrates for different values of K

Figure 2. Hitrates for the first three dgp's, where the success rate is displayed on the y-axis, and the varying variables on the x-axis

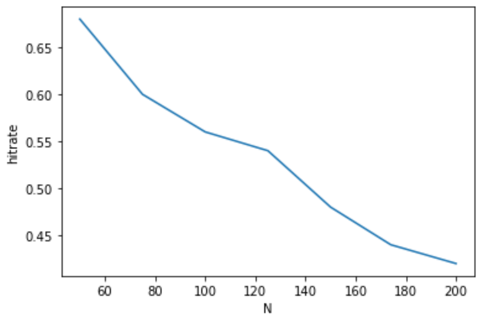
Figure 2a shows the hitrates for varying number of variables. The figure shows that the success rate of the selection criterion increases from around 0 to around 0.9 almost linearly. Around

N=50, the success rate of the criterion is approximately 0.8. For small values of N, the selection criterion often chooses less than three factors. Figure 2b shows the hitrates for varying number of observations. Again, the hitrates increase from around 0 to 0.95 when the number of observations increases. In this case, the hitrate around T=50 is approximately 0.75. At last, figure 2c shows the hitrates for varying number of pre-selected factors. The figure shows that the selection criterion performs quite well for a small amount of pre-selected factors. However, when the number of pre-selected factors increases, the accuracy of the selection criterion decreases. Based on these results it is assumed that the selection criterion used for the factor models is correctly specified.

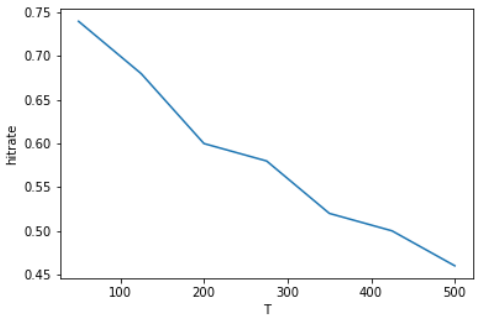
Next, the data generating processes that were used to check the selection criterion of the boosting algorithm are evaluated. The results of these four data generating processes are shown in figure 3.



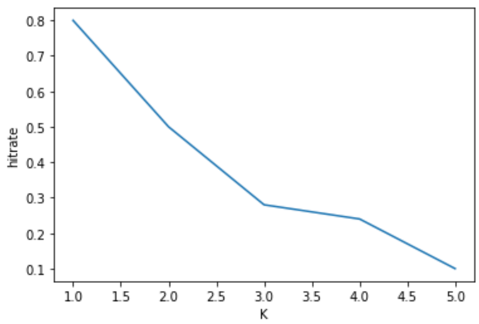
(a) *dgp4*: Hitrates for different steplengths v



(b) *dgp5* Hitrates for different values of N



(c) *dgp6* Hitrates for different values of T



(d) *dgp7* Hitrates for different values of K

Figure 3. Hitrates for the four dgp's testing the boosting criterion, where the success rate is displayed on the y-axis, and the varying variables on the x-axis

Figure 3a presents the hitrates of the boosting selection criterion for different steplengths v. The figure shows that the hitrates fluctuate around 0.43 for steplengths greater than 0.2. For smaller steplengths, the hitrate is lower. Figure 3b and 3c show the hitrates for varying number of variables N and observations T respectively. The figures show that the hitrate decreases when

N and T increase. This could be caused by the fact that the selection procedure of the boosting model uses other information than PCA. That is, boosting produces a new, smaller selection of predictors, where part of the important information for the PCA model could be excluded. That way, the decreasing hitrate does not have to indicate that the model is specified wrongly. Besides, the hitrates are still sufficiently high for values of N and T close to the dimensions of the African data set. The hitrates for varying number of factors K that are used to simulate the data are illustrated in figure 3d. Again, the hitrate decreases as the number of factors increases. This could again be caused by the fact that the boosting algorithm excludes important variables for the PCA model. When all the variables are generated out of one factor, this effect should be vanished. Based on the results of the first seven data generating processes, it is assumed that the model is correctly specified.

Now that it is assumed that the model is correctly specified, the forecast ability of the model is simulated. This is done by simulating data using dgp 8. With this data, five 1-step ahead forecasts will be made using the hybrid models and the benchmark AR(1) model. These forecasts are evaluated by means of their mean MSFE's. That is, the mean of the all the simulated MSFE's is evaluated. The results are presented in table 2. Table 2 shows both the hybrid models should

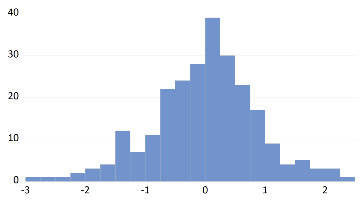
Table 2: MSFE's of the forecasts for the simulated data

	PCA	SPCA	AR(1)
MSFE	5.189	5.014	11.650

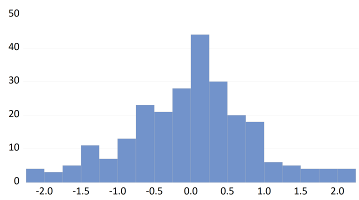
produce more accurate forecasts than the benchmark AR(1) model. The results show that the hybrid model using PCA produces more accurate forecasts in 80% of the simulations. For the hybrid model using SPCA it was even 84%. Besides the MSFE's of the forecast, the forecast errors are also evaluated. The results are illustrated in figure 4. From the figure it looks as if the forecast residuals of all three of the models are approximately normally distributed. When performing the Jarque-Bera test, the null hypothesis of normality is not rejected for all three the models.

5.2 Results African Data

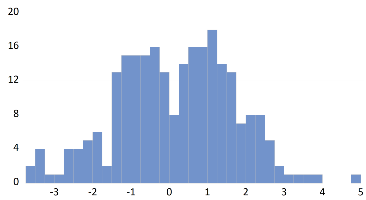
After evaluating the results of the different simulation studies, the hybrid model is used to make forecasts of the GDP growth of Nigeria. Five forecasts are made for the period 2011 - 2015 using a recursive window. As discussed in 4.3, this paper uses different values for *pmax* to obtain different sets of predictor variables for the boosting algorithm. The optimal value of *pmax* yields the lowest value of MSFE. The MSFE results for different values of pmax are shown



(a) forecast residuals hybrid model with PCA



(b) forecast redisuals hybrid model with SPCA



(c) forecasting residuals AR(1) model

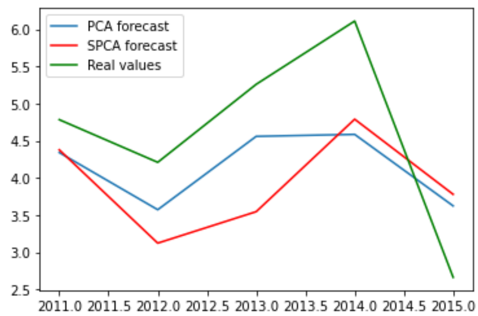
Figure 4. Forecast residuals of the three models

in table 3.

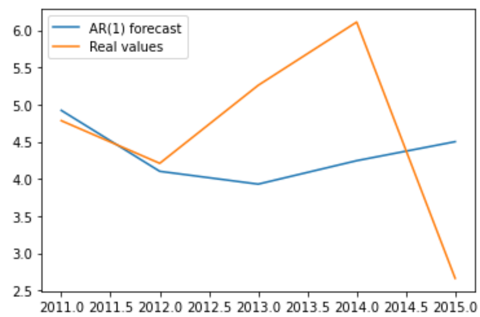
Table 3: MSFE values for different values of pmax

pmax:	1	2	3	4	5	6
PCA	17.183	**4.335**	7.849	30.950	26.768	21.020
SPCA	15.136	**7.263**	11.458	32.245	21.206	19.347

Table 3 shows that for both the PCA and the SPCA factor models $pmax$=2 yield the best MSFE. Hence, $pmax$=2 will be used for the forecasts of the GDP growth of Nigeria. Subsequently, the forecasts are made using the hybrid models using the different factor models, as well as forecasts using the benchmark model. Following the Schwarz Information Criterion (SIC) mentioned by Schwarz (1978) to determine the optimal number of lags in an autoregressive model, the benchmark model resulted in an autoregressive model using one lag. The forecasts that were made and the forecast errors are illustrated in figure 5 and table 4 respectively.



(a) Forecasts made using boosting combined with PCA and SPCA along with the real GDP growth



(b) Forecasts of the AR(1) model along with the real values of the GDP growth

Figure 5. Forecasts made by the different hybrid models compared with the forecasts made by the benchmark model

Figure 5 shows that the AR(1) model produces better forecasts for 2011 and 2012 than the

Table 4: Forecast errors per forecast for the different models

	2011	2012	2013	2014	2015
PCA	0.442	0.637	0.700	1.524	-0.960
SPCA	0.405	1.087	1.712	1.320	-1.116
AR(1)	-0.138	0.107	1.329	1.864	-1.837

hybrid model. This is substantiated by table 4. After 2012 however, the hybrid model using PCA outperforms the AR(1) model, and from 2014 onwards the AR(1) model is outperformed by both hybrid models. These forecast errors lead to the MSFE ratio's that are shown in table 5. Both ratio's are less than one, meaning that the hybrid models outperform the benchmark AR(1) model. Based on the p-values of the Diebold Mariano test, the null hypothesis of no significant difference between the forecast is rejected for the hybrid model that uses PCA. That is, the hybrid model based on PCA significantly outperforms the AR(1) model in forecasting based on the forecast errors. For the hybrid model that uses SPCA, this nullhypothesis has to be rejected, meaning that there is no significant evidence that the model outperforms the benchmark AR(1) model.

Table 5: MSFE ratio's for the two hybrid models compared to the AR(1) model, with the p-value of the Diebold Mariano test in brackets

	PCA	SPCA
MSFE ratio	**0.501** (0.047)	0.840 (0.661)

Now that it's clear that the hybrid model based on boosting and (sparse) principal component analysis outperforms the benchmark AR(1) model in forecasting, it is useful to check which predictors have the highest impact on the forecasts. That is, the factors are evaluated to check which countries the highest loadings in the factors. Principal components are hard to interpret, because every factor is a linear combinations of all original variables. Hence, the sparse principal components are investigated. For every forecast, only one factor was selected after boosting and PCA. Hence, only one sparse principal component is evaluated for each forecast. Figure 6 illustrates the factor loadings for the forecast of 2011. Figure 6 illustrates that Botswana with a lag of 2 years, Burkina Faso with a lag of 2 years and Ethiopia with no lags have the greatest impact on the forecast. The factor loadings of the different countries for the other forecasts are shown in figure 7 to 10 in the Appendix. When all the forecasts are taken into account cumulatively, different lags of the countries Zambia, Mozambique, Ghana and Ethiopia. Geographically, Ghana lies close to Nigeria adjacent to the west coast of Africa. This could be a reason for the impact on the forecasts of the Nigerian GDP growth. For the other countries

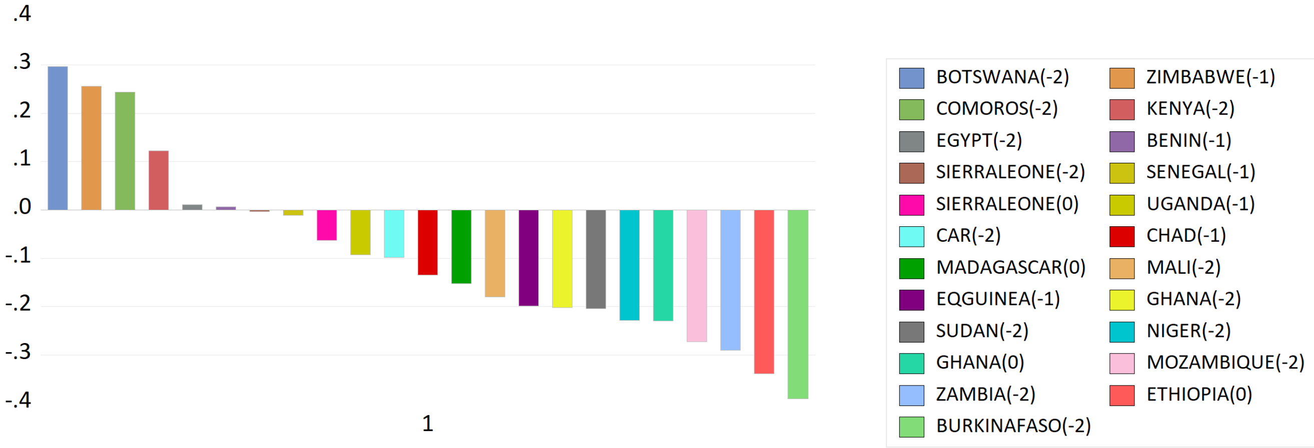


Figure 6. Factor loadings of the different countries on the factor for the forecast of 2011.
Note. The value between between the brackets behind the country names indicate the lags that are used. That is, 0 means there is no lag used, -1 indicates that there is a lag of 1 year, and -2 indicates that there is a lag of 2 years.

there is no clear indication why they have such a great impact on the forecasts of the Nigerian GDP growth. Another interesting result is that Cameroon has zero impact on all of the GDP growth forecasts, even though Cameroon is Nigeria's neighbour, and is also adjacent to the west coast of Africa.

6 Conclusion

This paper follows methods as discussed in (Kim and Swanson (2018)) to produce five one-step ahead forecasts of the GDP growth of Nigeria. This is done by introducing a hybrid model. That is, first boosting is applied to the data set to produce a new set of predictors. Subsequently principal component analysis and sparse principal component analysis are performed on the new set of predictors, to produce the final set of factors. These factors are then used to produce the five forecasts of the GDP growth of Nigeria. The data set that is used to perform the boosting model on consists of yearly GDP growth of 51 other African countries for the period 1961 to 2015. Besides the factors produced in the hybrid model, the lagged GDP growth of Nigeria is also added to the set of explanatory variables that is used to produce the forecasts. The forecasts are evaluated by means of their mean squared forecasting errors. To evaluate these mean squared forecasting errors, a benchmark autoregressive model of order one is introduced to produce benchmark forecasts. These results are used to answer the research question:
Are factor models combined with the shrinkage method boosting useful for forecasting the Nigerian GDP growth, compared to a standard autoregressive model?

Beside comparing the results of the forecasts to a benchmark model, the specification of the model is also evaluated. This is done by simulating different data sets through an AR(1) model, with varying input values. The results are used to examine whether the input values have impact on the model and in which way.

First, three different data generating processes were used to evaluated the PCA model. The results were promising, since the hitrates where high overall, and the hitrates for a data set of approximately the dimensions of the African data set are around 0.80. Next, data generating processes were used to evaluate the PCA and boosting model as a whole. The results showed that the hitrates decrease as the number of observations and variables increase. This could be due to the fact that the boosting model is hard to interpret when selecting its predictors. The model selects predictors in a different way than PCA, and the predictors that were chosen by the boosting model are then used for the PCA model. This way, information that is important for the PCA model could already have been lost in the boosting algorithm, which means that the lower hitrates do not necessarily need to indicate that the model is specified in the wrong way. Based on the simulation results as a whole, the model specification is assumed to be correct. At last, a simulation is used to evaluate forecast accuracy of the hybrid model. Based on the simulation results both of the hybrid models should outperform the benchmark model in forecasting. The forecast errors of the three different models are evaluated as well, and seem to be normally distributed for all of the three models. This is an important characteristic, since the data is simulated with normally distributed error terms as well.

After the evaluation of the model specification, the models are used to produce forecasts of the Nigerian GDP growth. Based on the results of the forecasts, this paper concludes that the hybrid model using factor models and boosting outperforms the benchmark AR(1) model in forecasting. Both the specifications of the hybrid model produce lower MSFE's than the benchmark model. However, only the hybrid model using PCA as factor model produces significantly better forecasts than the benchmark model, based on the Diebold Mariano test. The countries with the greatest impact on all of the forecasts cumulatively are Zambia, Mozambique, Ghana and Ethiopia.

To conclude, this paper used hybrid models, based on boosting, PCA and SPCA, to produce forecasts of the Nigerian GDP growth. The results showed that the hybrid model outperforms the benchmark AR(1) model in forecasting based on the MSFE, meaning that factor model combined with boosting are helpful in forecasting the Nigerian GDP growth. Especially Zambia, Mozambique, Ghana and Ethiopia have a great impact in forecasting the GDP growth go Nigeria.

For further research it could be interesting to examine the data more closely. This paper assumed that the loadings of the factors are approximately constant over time. Since there have been large negative and positive peaks in the data of Nigeria, it could be an improvement to test for structural breaks. In addition, this paper found which countries have great impact the forecasts of the Nigerian GDP growth. It could be interesting in the future to find out what economic indicators cause these impacts. At last, this paper used the shrinkage method boosting to filter the initial set of predictors. Kim and Swanson (2018) discuss many other shrinkage methods, such as bagging, least angle regression and Bayesian model averaging. These methods could lead to other MSFE results for the African data set, so this would be an area for future research as well.

References

[1] Jushan Bai and Serena Ng. “Boosting diffusion indices”. In: *Journal of Applied Econometrics* 24.4 (2009), pp. 607–629.

[2] Jushan Bai and Serena Ng. “Determining the number of factors in approximate factor models”. In: *Econometrica* 70.1 (2002), pp. 191–221.

[3] Jushan Bai and Serena Ng. “Forecasting economic time series using targeted predictors”. In: *Journal of Econometrics* 146.2 (2008), pp. 304–317.

[4] Francis X Diebold and Robert S Mariano. “Comparing predictive accuracy”. In: *Journal of Business & economic statistics* 20.1 (2002), pp. 134–144.

[5] Dick van Dijk. *Time Series Analysis: Chapter 3 - Basic concepts.* Erasmus School of economics. 2019.

[6] Philip Hans Franses et al. *Time series models for business and economic forecasting.* Cambridge University Press, 1998.

[7] Kempe Ronald Hope Sr. “Climate change and poverty in Africa”. In: *International Journal of Sustainable Development & World Ecology* 16.6 (2009), pp. 451–461.

[8] Hyun Hak Kim and Norman R Swanson. “Forecasting financial and macroeconomic variables using data reduction methods: New empirical evidence”. In: *Journal of Econometrics* 178 (2014), pp. 352–367.

[9] Hyun Hak Kim and Norman R Swanson. “Mining big data using parsimonious factor and shrinkage methods”. In: *Available at SSRN 2294110* (2013).

[10] Hyun Hak Kim and Norman R Swanson. “Mining big data using parsimonious factor, machine learning, variable selection and shrinkage methods”. In: *International Journal of Forecasting* 34.2 (2018), pp. 339–354.

[11] F. S. O’Brien. “The Economic Impact of the Nigerian Civil War: A Comment”. In: *The Journal of Modern African Studies* 11.3 (1973), pp. 460–463. ISSN: 0022278X, 14697777. URL: http://www.jstor.org/stable/159613.

[12] Greg Ridgeway, David Madigan, and Thomas S Richardson. “Boosting methodology for regression problems”. In: *Seventh International Workshop on Artificial Intelligence and Statistics*. PMLR. 1999.

[13] Robert E Schapire. “The strength of weak learnability”. In: *Machine learning* 5.2 (1990), pp. 197–227.

[14] Gideon Schwarz. "Estimating the dimension of a model". In: *The Annals of Statistics* (1978), pp. 461–464.

[15] James H Stock and Mark W Watson. "Forecasting using principal components from a large number of predictors". In: *Journal of the American statistical association* 97.460 (2002), pp. 1167–1179.

[16] James H Stock and Mark W Watson. "Macroeconomic forecasting using diffusion indexes". In: *Journal of Business & Economic Statistics* 20.2 (2002), pp. 147–162.

[17] Simona Varrella. *Nigeria: poverty rate, by state 2019*. Sept. 2020. URL: `https://www.statista.com/statistics/1121438/poverty-headcount-rate-in-nigeria-by-state/`.

[18] Hui Zou, Trevor Hastie, and Robert Tibshirani. "Sparse principal component analysis". In: *Journal of computational and graphical statistics* 15.2 (2006), pp. 265–286.

Appendix

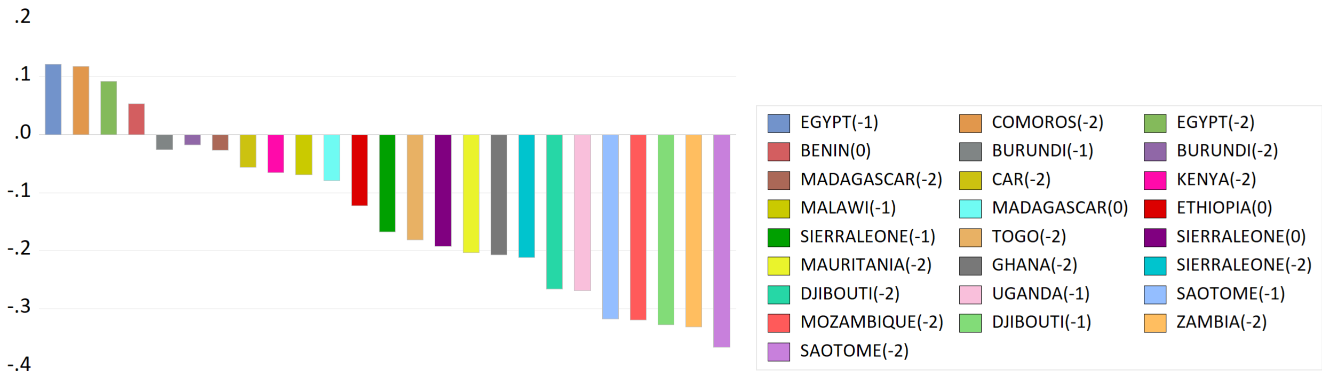


Figure 7. Factor loadings of the different countries on the factor for the forecast of 2012.

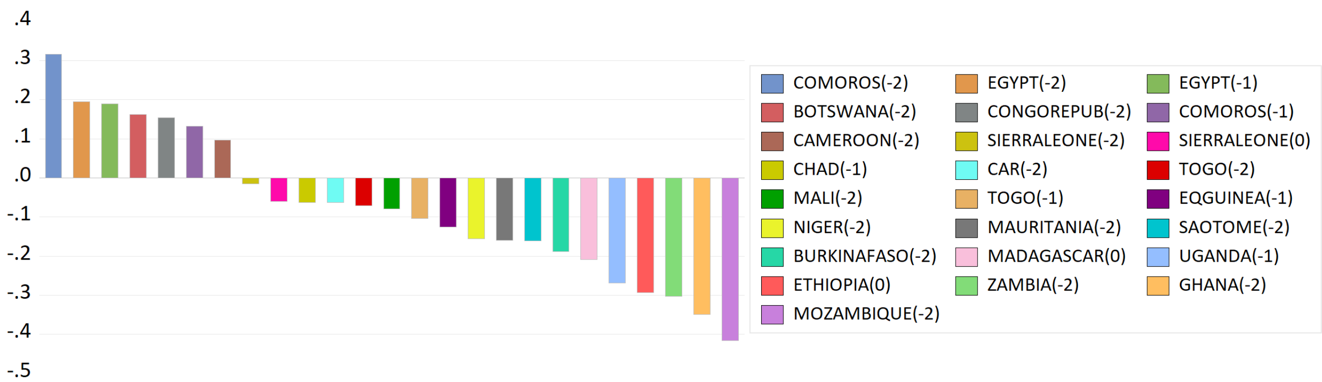


Figure 8. Factor loadings of the different countries on the factor for the forecast of 2013.

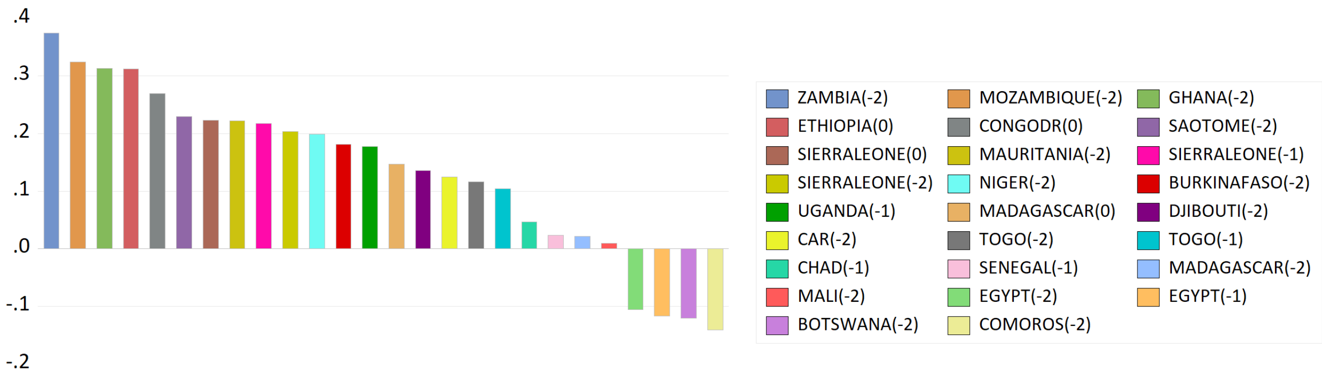


Figure 9. Factor loadings of the different countries on the factor for the forecast of 2014.

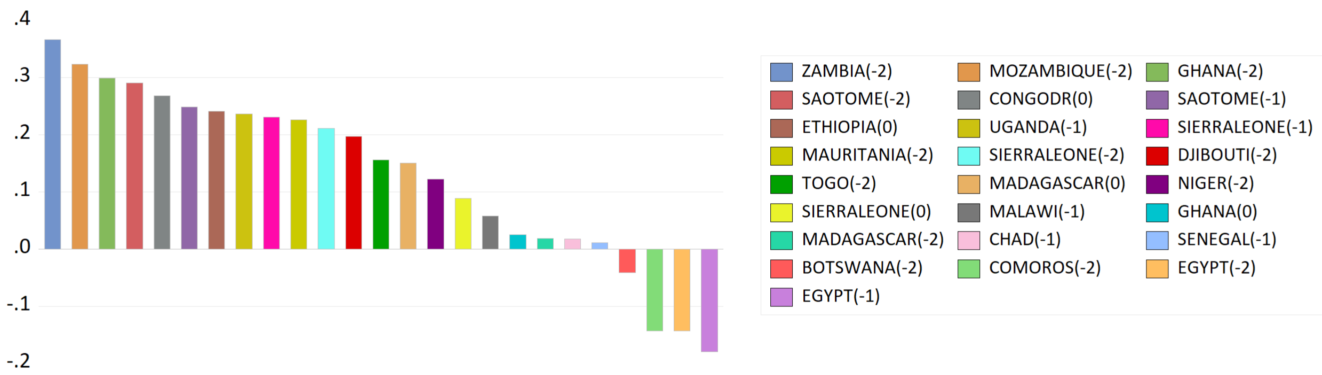


Figure 10. Factor loadings of the different countries on the factor for the forecast of 2015.

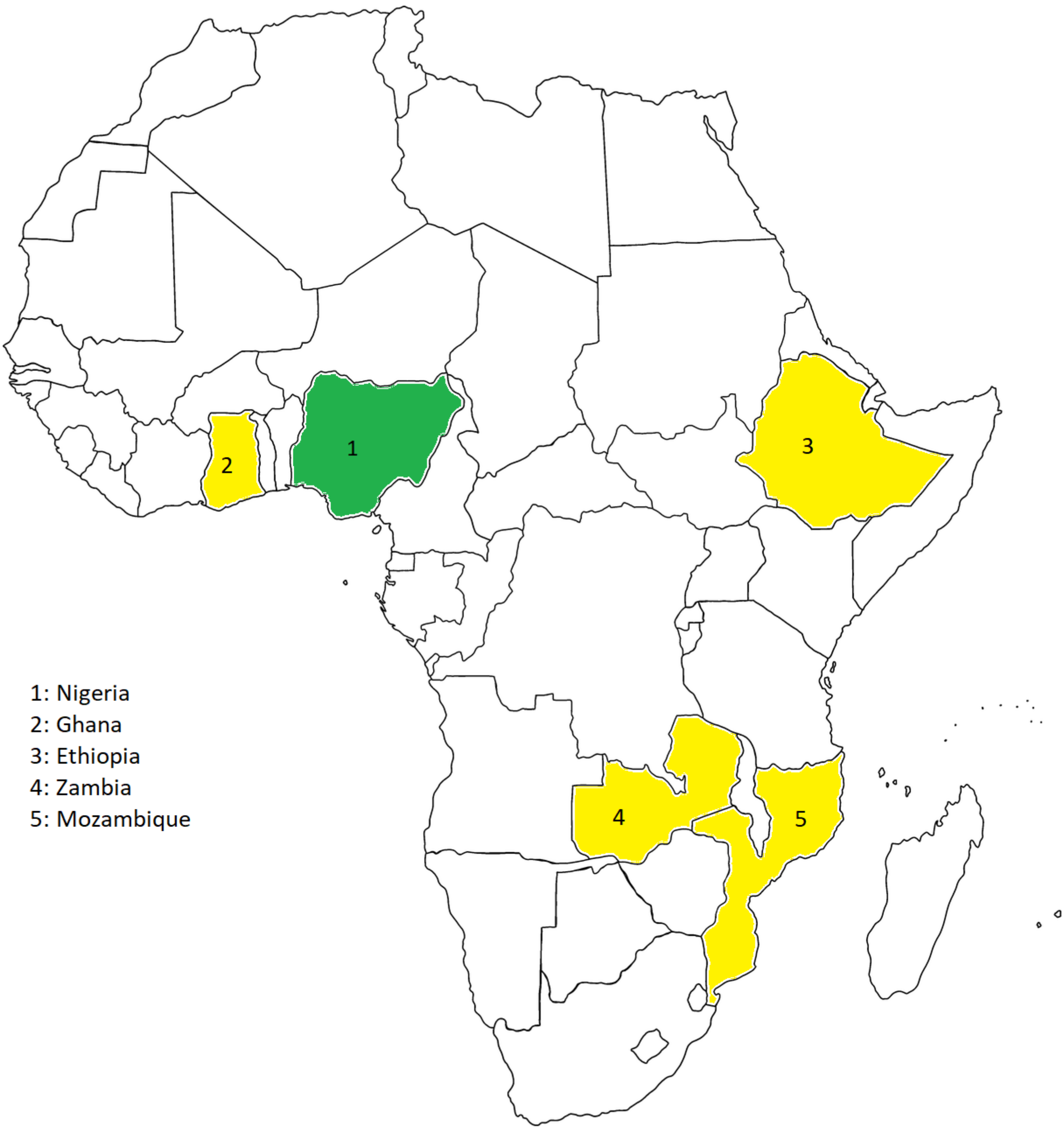


Figure 11. Map of africa with the countries have the greatest impacts cumulatively on all the forecasts highlighted.