

EUR Research Information Portal

Forecasting Real GDP Growth for Africa

Published in:
Econometrics

Publication status and date:
Published: 05/01/2022

DOI (link to publisher):
[10.3390/econometrics10010003](https://doi.org/10.3390/econometrics10010003)

Document License/Available under:
CC BY

Citation for the published version (APA):
Franses, P. H., & Welz, M. (2022). Forecasting Real GDP Growth for Africa. *Econometrics*, 10(1), Article 3.
<https://doi.org/10.3390/econometrics10010003>

[Link to publication on the EUR Research Information Portal](#)

Terms and Conditions of Use

Except as permitted by the applicable copyright law, you may not reproduce or make this material available to any third party without the prior written permission from the copyright holder(s). Copyright law allows the following uses of this material without prior permission:

- you may download, save and print a copy of this material for your personal use only;
- you may share the EUR portal link to this material.


In case the material is published with an open access license (e.g. a Creative Commons (CC) license), other uses may be allowed. Please check the terms and conditions of the specific license.

Take-down policy

If you believe that this material infringes your copyright and/or any other intellectual property rights, you may request its removal by contacting us at the following email address: openaccess.library@eur.nl. Please provide us with all the relevant information, including the reasons why you believe any of your rights have been infringed. In case of a legitimate complaint, we will make the material inaccessible and/or remove it from the website.

Article

Forecasting Real GDP Growth for Africa

Philip Hans Franses *  and Max Welz

Econometric Institute, Erasmus School of Economics, Burgemeester Oudlaan 50,
NL-3062 PA Rotterdam, The Netherlands; welz@ese.eur.nl

* Correspondence: franses@ese.eur.nl; Tel.: +31-10-4081273

Abstract: We propose a simple and reproducible methodology to create a single equation forecasting model (SEFM) for low-frequency macroeconomic variables. Our methodology is illustrated by forecasting annual real GDP growth rates for 52 African countries, where the data are obtained from the World Bank and start in 1960. The models include lagged growth rates of other countries, as well as a cointegration relationship to capture potential common stochastic trends. With a few selection steps, our methodology quickly arrives at a reasonably small forecasting model per country. Compared with benchmark models, the single equation forecasting models seem to perform quite well.

Keywords: forecasting; cointegration; annual real GDP growth; Africa

1. Introduction

This paper deals with forecasting low-frequency macroeconomic variables, when data are available for a reasonably large number of countries or states. As many macroeconomic variables have a stochastic trend, the forecasting methodology also addresses potentially common stochastic trends. In this paper the particular focus is on forecasting annual real GDP (Gross Domestic Product) growth rates in Africa.

The data that we analyze in this paper concern annual real GDP figures, for 52 countries in Africa, where these are all indexed at 100 in 1960, and where the sample ends in 2016. These data will be transformed to growth rates, and the levels will be included in cointegration relations. Figure 1 gives a first impression of the levels of the data at hand¹. We will create models for 1960–2010, and we will use 2011 up to and including 2016 to evaluate the accuracy of recursively created one-year-ahead forecasts for the growth rates.

The question that we address is whether it helps to include information from other countries to predict real GDP growth for a focus country. Such information can concern the past growth rates of each of these countries and a cointegration relationship. This last feature puts a challenge on the methodology as it shall not be feasible to consider a 52-variable vector autoregression due to the large number of parameters to be estimated, even already for a first-order vector autoregression.

One way to move forward could be to allocate countries into clusters, prior to any cointegration analysis, but it is unclear how such clusters should be created. Even neighboring countries in Africa can have very different economies, governments, infrastructures, resources, cultures, and colonial history, to mention just a few. At the same time, countries that are located at very different locations on the continent can still have similar properties and may, for example, similarly respond to local or worldwide economic shocks. In sum, it seems hazardous to put zeroes in heavily parameterized models from the onset. So, we consider all countries on the African continent, as it may be that groups of countries (in different constellations) may be expected to respond similarly to worldwide economic shocks, to changes in production, to weather conditions, and to political conditions.



Citation: Franses, Philip Hans, and Max Welz. 2022. Forecasting Real GDP Growth for Africa. *Econometrics* 10: 3. <https://doi.org/10.3390/econometrics10010003>

Academic Editors: Martin Wagner and Robert Kunst

Received: 23 August 2021

Accepted: 22 December 2021

Published: 5 January 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

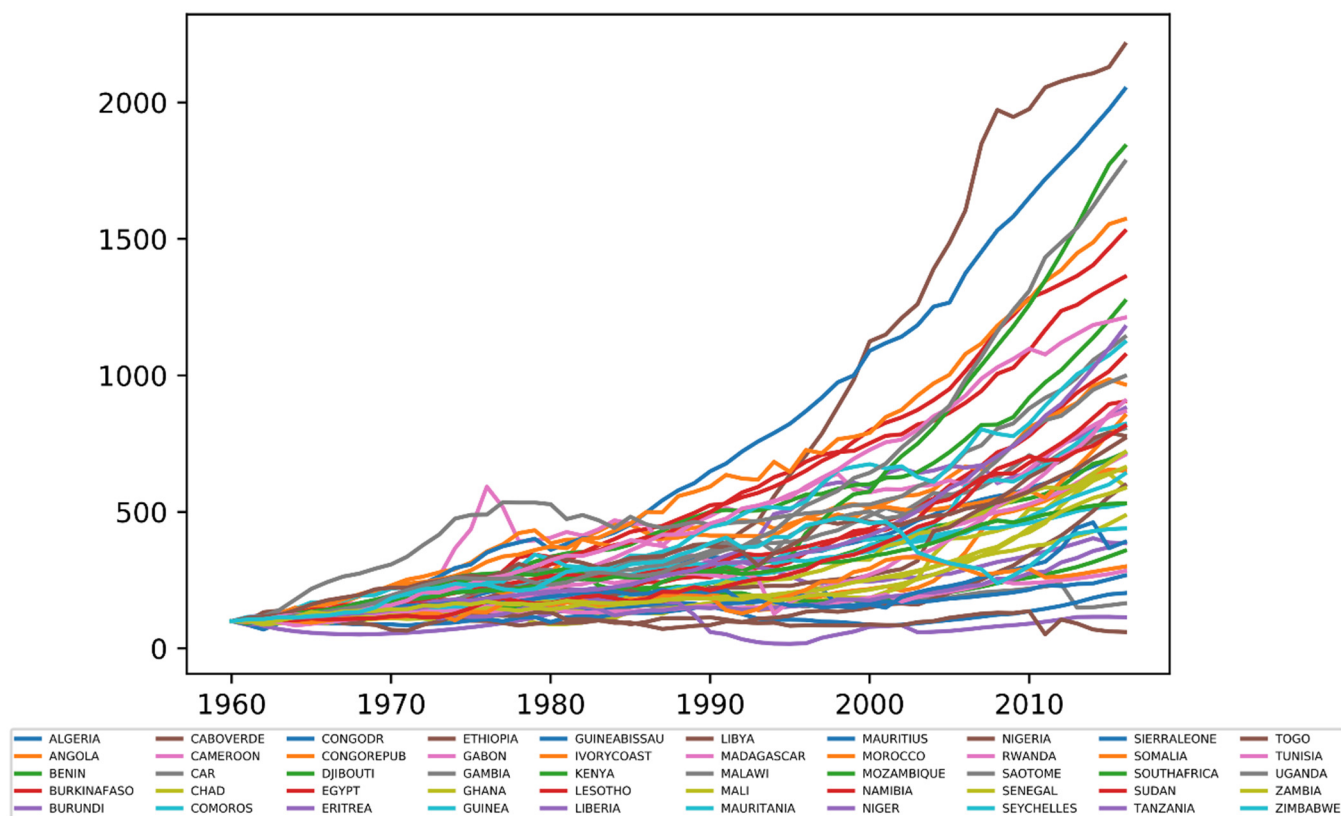


Figure 1. Levels of real GDP in 50 countries in Africa 1960–2016 (where Equatorial Guinea and Botswana are excluded due to very high growth rates), indexed at 1960 = 100. The data are obtained from [Franses and Vasilev \(2019\)](#). These authors used imputation techniques to create a complete data set.

In this paper we propose a simple and reproducible econometric methodology that allows the creation of forecasting models for real GDP growth for each of the 52 countries, while taking aboard potential cointegration relations between countries, and while including lagged growth rates of other countries. The methodology is based on a series of partial analyses², and this immediately suggests the weaker part of the proposed methodology. Yet, at the same time, as decisions in our analyses are all driven by verifiable criteria, and can be set by the researcher to see if different results could emerge when different choices are made, we believe that our methodology allows for sensible flexibility, while it remains easy to use and it relies only on standard econometric methods. Section 2 below will provide all the details of the methodology. The result for each country is what can be called a single equation forecasting model (SEFM), which can include a cointegration relationship with other countries, and which includes one-year-lagged growth rates of a selected set of other countries. A byproduct of our models, in contrast with many machine learning methods, is that we can construct one-step-ahead prediction intervals.

The paper continues as follows. Section 2 deals with the way to arrive at a single equation forecasting model. We show that just a few criteria need to be set, and we only use formal statistical tests to reduce the number of parameters in the final stage. Before that stage, all steps can be automated. Section 3 starts with a detailed illustration for one country, in this case the real GDP growth rates of Botswana, and after that, all other countries on the African continent will be considered. We first analyze forecast accuracy before we present estimation results. For illustration, we use a moving average model of order 1 for the growth rates as a benchmark model, as this extends the usually considered random walk by allowing past forecast errors to have a correcting impact on forecasts³. As an example, we find that, roughly speaking, our single equation forecasting models substantially improve

on the MA(1) model in close to about half of the cases. Alternative forecasting benchmarks give qualitatively similar results. Next, we present tables with results on cointegration and short-run effects, and as such, we learn that there could be some clusters of countries with similar behavior, while there are also countries with evident idiosyncratic economic growth (such as Nigeria and South Africa). Section 4 deals with simulations to see if our empirical findings are significant in general⁴. Our paper then concludes with a discussion of the limitations and some avenues for further research.

2. Methodology

This section deals with the methodology that we will follow to create forecasts. The basic premise is that we have data for T years and I countries (or states in other applications), where both T and I are about equally large like 50 or 100 but not much more. A similar setting is considered in Kim and Swanson (2018)⁵, who use modern methods such as parsimonious factor models, machine learning, variable selection, and shrinkage methods. At the same time, they do not consider common stochastic trends, and that is what we will also do below.

We make a few assumptions. In the single equation forecasting models, we allow for one-period lags of the growth rates of all other countries, and we use a simple and partial preselection method to see which variables to include in the models. At the same time, we allow in each equation a single cointegration variable. There seems to be no reason to include more than one cointegration variable as we, each time, are looking at a single variable to be predicted. Moreover, a system of 52 equations for growth rates can have 51 cointegration relations at the maximum, and hence allowing for more than one per country can cause inference problems.

The initial data concern:

$$y_{i,t} = \log(GDP_{i,t}),$$

where “log” is the natural logarithm, for $i = 1, 2, \dots, I$ countries (where we shall consider 52 countries in our study) and $t = 1, 2, \dots, T$ years (where we have data for the years 1960 up to and including 2016). We are interested in forecasting the growth rates, that is,

$$\Delta y_{i,t} = \log(GDP_{i,t}) - \log(GDP_{i,t-1})$$

Our model will include the lags of growth rates of the focus country and of other countries, as well as a cointegration relation amongst the levels $y_{i,t}$.

A single equation forecasting model (SEFM) for country i will look like:

$$\Delta y_{i,t} = \mu_i + \sum_{j=1}^I \alpha_{i,j} \Delta y_{j,t-1} + \gamma_i (y_{i,t-1} - \sum_{j=1, j \neq i}^I \beta_{i,j} y_{j,t-1}) + \varepsilon_{i,t}, \quad (1)$$

where $\varepsilon_{i,t}$ is a standard white noise process with mean 0 and variance σ_i^2 . We assume that the error terms are uncorrelated across equations⁶.

The term:

$$\sum_{j=1}^I \alpha_{i,j} \Delta y_{j,t-1} \quad (2)$$

reflects the contribution of the one-period lagged growth rates of all countries (also including the lag of the focal country's growth rate). The term:

$$\gamma_i (y_{i,t-1} - \sum_{j=1, j \neq i}^I \beta_{i,j} y_{j,t-1}) \quad (3)$$

reflects the error correction variable (cointegration relation) with its adjustment parameter γ_i . Note that this variable can be different for each country. When it is assumed that the error correction variable is stationary indeed, its statistical significance can be checked by a standard t -test on the parameter γ_i ; see Boswijk (1994).

Per equation, there are potentially $1 + I + 1 + (I - 1) = 2I + 1$ parameters. Given the sample size, straightforward OLS (Ordinary Least Squares) or NLS (Nonlinear Least Squares) per country, while including all regressors, is not possible. Therefore, we introduce a few selection steps that should reduce the number of variables.

Step 1:

Run $I - 1$ OLS regressions of the type

$$y_{i,t} = \delta_i + \pi_i y_{j,t} + \omega_{j,t} \quad (4)$$

for $j = 1, \dots, I$ and $j \neq i$. Compute for the estimated residuals $\hat{\omega}_{j,t}$ the CRDW (Cointegration Regression Durbin Watson) test statistic as:

$$CRDW_j = 2(1 - \hat{\rho}_j), \quad (5)$$

where $\hat{\rho}_j$ is the estimated first-order autocorrelation of $\hat{\omega}_{j,t}$. When $CRDW_j > \tau$, we keep the regressor $y_{j,t}$ for the next round. Thus, the first selection parameter in our methodology is τ . We can set it at 0.4 or 0.5 at our own choice. Formal critical values appear in [Sargan and Bhargava \(1983\)](#), and depending on the size of the regression model, these range from 0.4 to 0.7.

Step 2:

For all retained variables from step 1, also including $y_{i,t}$, we run the [Johansen \(1991\)](#) cointegration estimation method for the levels of the log GDP data, and retain the first estimated cointegration relationship for the single equation forecasting model. There is no need to check its statistical significance at this stage. If the estimated potential cointegration variable is not stationary, the final t -test on γ_i in (1) will not reject the associated null hypothesis.

Step 3:

Compute all pairwise correlations between $\Delta y_{i,t}$ and $\Delta y_{j,t-1}$ for $j = 1, \dots, I$ and $j \neq i$.

Rank the countries with the positive correlations from high to low, and with the negative correlations from (absolute) high to low.

Step 4:

To further reduce the number of parameters, impose two sets of restrictions, one for the positive correlated variables in step 3 (hence the growth rates) and one for the negatively correlated variables (again for the growth rates), as follows:

$$\alpha_{i,j} = \alpha_i + \alpha_i \theta + \alpha_i \theta^2 + \dots + \alpha_i \theta^k, \quad (6)$$

which mimics a geometric decay pattern. We can have a θ_p for the positive correlations and a θ_n for the negative correlations, and there is an associated $\alpha_{i,p}$ and an $\alpha_{i,n}$. The threshold k can be k_{pc} and k_{nc} for positive and negative correlations, respectively. These are two more selection parameters.

In sum, our methodology involves three selection parameters, that is, τ , k_{pc} , and k_{nc} , and these need to be set by the researcher. Finally, the parameters in the single equation forecasting model (1) can be estimated using NLS (Nonlinear Least Squares). Using t -tests on γ_i and $\alpha_{i,p}$ and $\alpha_{i,n}$, one can further reduce the number of parameters⁷. In case γ_i is not significant, and hence no significant cointegration is found, the parameters in the restricted version of (1) will be estimated using OLS. If no variables are retained from step 1, we set $\gamma_i = 0$ in (1). In a few cases where NLS does not converge, the cointegration relation found in step 2 will be fixed, and the parameters will be estimated using OLS. A technical appendix which illustrates the methodology can be found in the online appendix.

3. Forecasting GDP Growth in Africa

In this section we first illustrate our methodology in detail for Botswana, and after that we consider all countries at the same time. The estimation sample runs from 1960 to 2010 for all countries, and the forecast evaluation sample is 2011 to 2016.

3.1. Botswana

We choose to use in our analysis that $\tau = 0.4$ for the CRDW. For Botswana, this means that the countries that are to be retained are Algeria and Central African Republic (CAR)⁸. The Johansen estimation method gives as the first cointegrating variable (for 1960–2010):

$$\log(\text{Botswana}) - 1.253 \log(\text{Algeria}) - 2.177 \log(\text{CAR}).$$

A graph of this variable is presented in Figure 2.

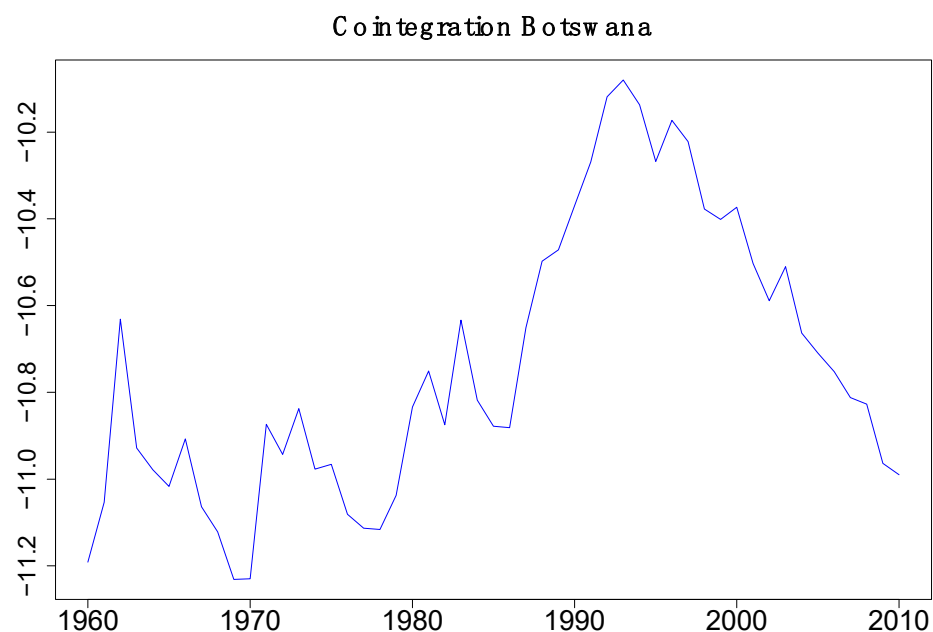


Figure 2. Cointegration relationship between $\log(\text{GDP})$ of Botswana, Algeria, and CAR (1960–2010).

Furthermore, when we adopt that $k_{pc} = k_{nc} = 3$ (which we will do throughout this paper⁹), the top three positive correlations between $\Delta y_{i,t}$ (Botswana) and $\Delta y_{j,t-1}$ are Zimbabwe (0.493), Kenya (0.382), and Mauritius (0.337), and the top three negative correlations concern Benin (−0.267), Burkina Faso (−0.238), and Ethiopia (−0.212).

Application of NLS to the initial version of the single equation forecasting model for Botswana gives the following estimation results (with estimated standard errors in parentheses):

$$\begin{aligned} \mu_{\text{Botswana}} &= -0.509 (0.292), \\ \gamma_{\text{Botswana}} &= -0.067 (0.020), \\ \beta_{\text{Botswana, Algeria}} &= 2.107 (0.467), \\ \beta_{\text{Botswana, CAR}} &= 0.737 (1.362), \\ \alpha_{\text{Botswana, Botswana}} &= 0.263 (0.101), \\ \alpha_{\text{Botswana, Zimbabwe, } p} &= 0.712 (0.193), \\ \theta_p &= 0.362 (0.067), \\ \alpha_{\text{Botswana, Benin, } n} &= 0.251 (0.279), \\ \theta_n &= -0.388 (0.148). \end{aligned}$$

Setting the insignificant parameter $\alpha_{Botswana, Benin, n}$ equal to zero results in

$$\begin{aligned}\mu_{Botswana} &= -0.460 (0.283), \\ \gamma_{Botswana} &= -0.069 (0.019), \\ \beta_{Botswana, Algeria} &= 2.161 (0.448), \\ \beta_{Botswana, CAR} &= 0.479 (1.268), \\ \alpha_{Botswana, Botswana} &= 0.278 (0.098), \\ \alpha_{Botswana, Zimbabwe, p} &= 0.708 (0.195), \\ \theta_p &= 0.358 (0.067).\end{aligned}$$

The significance of the error correction variable is based on the significance of $\gamma_{Botswana}$, and hence, the model is not reduced any further. The adjusted R^2 of this model with 7 parameters for 49 effective observations (estimation sample ends in 2010) is 0.698.

When this model is recursively estimated for samples 1960–2010 to 1960–2015, we can create six one-year-ahead forecasts for the years 2011 to 2016. The ME (mean error) of these forecasts is -4.53 and the MAE (mean absolute error) is 6.1.

When we estimate an MA(1) model for GDP growth in Botswana for 1960–2010, that is:

$$\Delta y_{Botswana, t} = \mu_{Botswana} + u_{Botswana, t} + \theta u_{Botswana, t-1} \quad (7)$$

where $u_{Botswana, t}$ is the error term and $\mu_{Botswana}$ the intercept, the θ is estimated as 0.540 with standard error 0.104, and an adjusted R^2 of 0.276. Again, six recursively created one-year-ahead forecasts can be created. The mean error of these forecasts for this case of Botswana is -2.93 while the mean absolute error is 3.92. Hence, in terms of both the mean error and the mean absolute error, the MA(1) model gives more accurate forecasts than the single equation forecasting model. Formal testing is not reliable here, due to the small number of forecasts.

3.2. 52 Countries in Africa

An overview of how the single equation forecasting model performs is provided in Table 1, where we present the results after applying our methodology outlined in Section 2 to the growth rates of 52 African countries, where we choose $\tau = 0.4$, and $k_{pc} = k_{nc} = 3$. Table 1 presents for 52 countries how the parameters in each model were estimated (NLS or OLS), and the ME, MAE and RMSE (Root Mean Squared Errors of the forecasts) of the final single equation forecasting models and those of the MA(1). A boldface number marks the cases where the ME of the single equation forecasting model is closer to 0 (in an absolute sense) and where the MAE and RMSE are closer to 0 for that model¹⁰. We see that in 29 of the 52 cases the ME of the single equation forecasting model is closer to 0 than that of the benchmark. For the MAE and RMSE this holds for 16 of the 52 cases. So, even though the MA(1) model is often found hard to beat, we see that it can be beaten in a substantial number of cases, also with larger numerical differences between the accuracy measures.

We also compare our forecasts with those of an AR(1) model, a no-change forecast, and forecasts from a Principal Component Regression including only the lagged growth rates, where we include the first 10 estimated principal components. In terms of RMSE, our models give more accurate forecasts than an AR(1) in 13 of the 52 cases, while the no-change forecast is beaten for all 52 cases. In 26 of the 52 cases our forecasts improve on those of a Principal Component Regression.

Table 1. Forecast performance comparison of single equation forecasting models (SEFM) with MA(1) models. A boldface number means that for the SEFM the MA (or MAE or RMSE) is closer to 0 than for the MA(1) model. In the column “Estimation”, “NLS” means that classical NLS estimation was applied. “OLS” means that OLS estimation without the nonlinear terms was applied, which is applicable in cases where the NLS estimation did not converge. “NLS w/o” means that NLS estimation was applied, but without the variables pertaining to cointegration relationships, which is applicable if no cointegration relationships were found. For each model we created six one-step-ahead forecasts, using the periods (1960–2010) to (1960–2015) to forecast the years 2011 to 2016, respectively.

	Estimation	SEFM			MA(1)		
		ME	MAE	RMSE	ME	MAE	RMSE
ALGERIA	NLS	−3.93	3.93	4.79	−0.48	0.51	0.63
ANGOLA	OLS	0.78	2.9	3.94	−0.01	1.48	2.01
BENIN	NLS	−0.71	2.44	3.09	0.98	1.66	2.00
BOTSWANA	NLS	−4.53	6.1	6.69	−2.93	3.92	4.80
BURKINAFASO	NLS	2.52	2.66	3.08	1.25	1.38	1.64
BURUNDI	NLS	0.59	2.14	3.29	−0.42	2.74	3.28
CABOVERDE	OLS	−4.28	4.28	4.53	−3.09	3.09	3.40
CAMEROON	NLS	0.57	1.17	1.37	1.16	1.16	1.28
CAR	NLS w/o	−7.30	13.59	2.01	−8.87	12.96	2.07
CHAD	NLS	−3.70	4.95	5.82	−1.17	4.69	5.58
COMOROS	NLS	−0.27	0.55	0.67	−0.65	0.78	1.00
CONGODR	NLS w/o	2.74	2.74	3.26	3.95	3.97	4.44
CONGOREPUB	NLS	1.08	3.33	3.48	−1.31	2.37	2.81
DJIBOUTI	NLS	1.06	1.26	1.68	2.33	2.33	2.36
EGYPT	NLS	−1.00	1.29	1.65	−1.42	1.42	1.84
EQGUINEA	NLS w/o	−8.39	8.39	9.29	−10.35	10.35	12.26
ERITREA	OLS	0.64	1.8	2.85	1.37	2.13	3.87
ETHIOPIA	NLS w/o	3.94	4.28	4.81	3.93	3.93	4.11
GABON	NLS	1.28	3.38	4.02	0.39	0.99	1.25
GAMBIA	NLS	1.03	2.74	3.17	−1.85	2.29	3.55
GHANA	OLS	0.23	2.81	3.28	2.38	2.56	3.92
GUINEA	NLS	1.55	1.55	1.84	1.48	1.48	1.89
GUINEABISSAU	NLS	0.48	4.05	4.54	2.48	3.28	4.00
IVORYCOAST	NLS	7.36	9.41	10.16	2.54	5.08	5.59
KENYA	NLS	0.73	1.28	1.33	0.84	0.86	0.95
LESOTHO	NLS	−3.78	6.55	8.97	−0.98	1.71	1.89
LIBERIA	NLS w/o	−0.99	4.36	4.65	2.28	3.70	4.28
LIBYA	NLS w/o	−8.32	43.53	57.26	−10.85	40.83	56.38
MADAGASCAR	NLS	1.02	3.69	4.09	1.11	1.23	1.39
MALAWI	NLS	2.48	5.12	8.41	−0.36	1.40	1.48
MALI	NLS	1.63	6.32	7.53	0.39	2.43	2.76
MAURITANIA	OLS	−0.88	2.38	2.72	0.46	1.89	1.95
MAURITIUS	OLS	−1.16	1.16	1.51	−1.77	1.77	1.78
MOROCCO	NLS	−0.41	1.46	1.63	−1.79	1.79	2.17
MOZAMBIQUE	NLS	−4.17	4.31	6.07	0.75	1.32	1.36
NAMIBIA	NLS	0.08	1.41	2.15	0.41	1.50	1.77
NIGER	NLS	2.67	2.85	3.69	3.26	3.34	4.45
NIGERIA	NLS	−2.57	2.57	3.30	−0.47	1.72	2.24
RWANDA	NLS	6.38	8.3	9.52	3.52	3.52	3.80
SAOTOME	NLS	4.31	4.69	6.12	−0.18	1.10	1.34
SENEGAL	OLS	1.90	2.40	2.86	2.25	2.36	2.88
SEYCHELLES	NLS	−2.24	2.54	3.44	0.78	1.44	1.74
SIERRALEONE	NLS	0.63	9.80	13.79	1.87	10.46	13.59
SOMALIA	NLS	−0.94	3.98	6.08	−1.32	3.17	5.61
SOUTHAFRICA	NLS	−1.44	1.69	2.11	−1.05	1.05	1.21
SUDAN	OLS	−3.93	3.93	5.26	−0.95	1.95	2.65
TANZANIA	NLS	1.02	1.49	2.19	1.71	1.86	2.07
TOGO	OLS	−0.48	0.98	1.40	1.20	1.20	1.32
TUNISIA	NLS	−2.48	2.55	3.80	−3.17	3.17	3.67
UGANDA	OLS	0.36	1.87	2.46	−0.04	1.34	2.05
ZAMBIA	NLS	−0.21	1.62	2.00	1.61	1.73	2.17
ZIMBABWE	NLS w/o	1.14	3.79	4.63	2.53	3.63	5.46

Figure 3 presents a histogram of the adjusted R^2 values of the models with estimation sample 1960 to 2015, and we find that the mean is 0.353, that the maximum is 0.710 (Rwanda), and the minimum is 0.039 (Somalia). Figure 4 presents a histogram of the mean errors (each time, the mean of the six one-year-ahead forecast errors), and there we find that the average mean error is -0.344 , where for some countries the mean error is very large (-8.39 for Equatorial Guinea, and 7.36 for Ivory Coast). Figure 5 illustrates a histogram of the mean absolute errors. Note that this histogram was created without Libya, as for this country the one-step-ahead forecasts are very poor (MAE of 43.63). The average of the mean absolute errors (excluding Libya) amounts to 3.62, which is relatively larger than what is typically found for OECD countries (usually within the range of 1 to 2). For the RMSE the results are qualitatively similar.

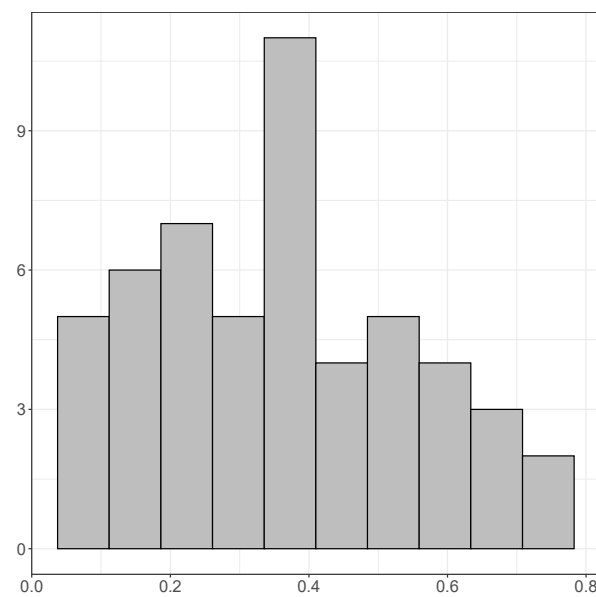


Figure 3. The adjusted R^2 values of the single equation forecasting models with estimation sample 1960 to 2015.

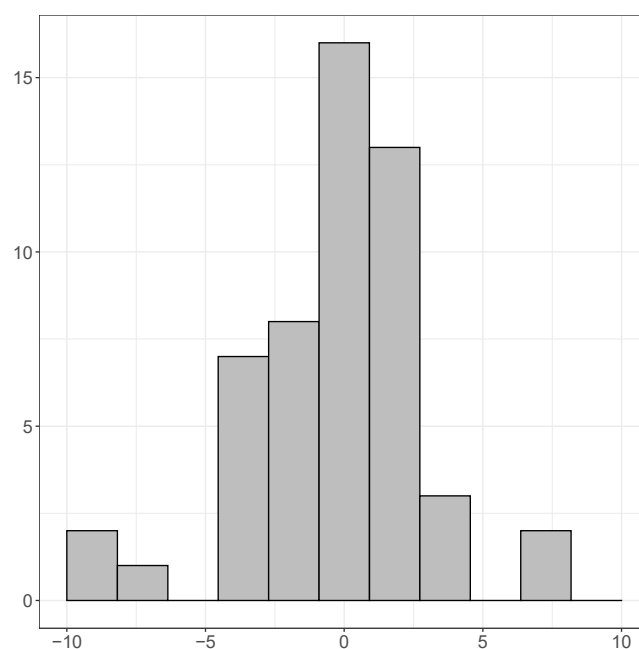


Figure 4. The mean forecast errors of the single equation forecasting models.

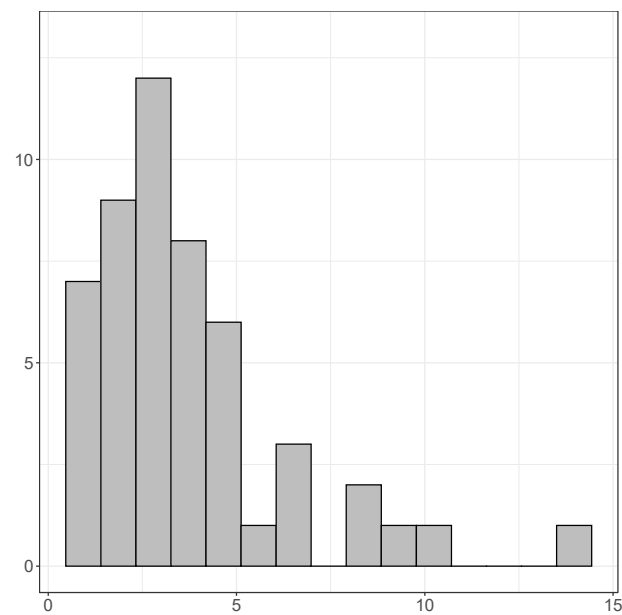


Figure 5. The mean absolute forecast errors of the single equation forecasting models (Libya not included).

A last conclusion from Table 1 is that for the single equation forecasting models it occurred four times that all six forecasts were too high, whereas there are three cases for which all forecasts were too low. These numbers become seven and six for the MA(1) model, and hence in that respect also the single equation forecasting models seem to do quite well.

Table 2 presents the obtained cointegration relations. The left-hand column is the focal country, and to the right we report all other countries which appear in the cointegration relation:

$$(y_{i,t-1} - \sum_{j=1, j \neq i}^I \beta_{i,j} y_{j,t-1}). \quad (8)$$

Table 2. Cointegration relationships of each “Home” country. An NA denotes the situation where the home country does not have cointegration relationships with other countries. The following abbreviations are used: CAR = Central African Republic, CONGODR = Democratic Republic of the Congo, CONGOREPUB = Republic of the Congo, EQGUINEA = Equatorial Guinea.

Home	Cointegration Relations with
ALGERIA	BOTSWANA, BURUNDI, CAMEROON, CAR, COMOROS, CONGOREPUB, GABON, GAMBIA, IVORYCOAST, KENYA, LESOTHO, MALAWI, MAURITIUS, MOROCCO, SEYCHELLES, SOUTHAFRICA, TOGO, TUNISIA
ANGOLA	NIGER, ZAMBIA
BENIN	BURKINAFASO, CABOVERDE, CAR, GUINEA, GUINEABISSAU, MALI, SENEGAL, SUDAN
BOTSWANA	ALGERIA, CAR
BURKINAFASO	BENIN, CAR, MALI, SENEGAL, SUDAN
BURUNDI	ALGERIA, CAR, CONGOREPUB
CABOVERDE	BENIN, CAR, MALI, SENEGAL
CAMEROON	ALGERIA, CAR, COMOROS, CONGOREPUB, RWANDA
CAR	NA
CHAD	GHANA, MADAGASCAR, NIGER
COMOROS	ALGERIA, CAR, CONGOREPUB
CONGODR	NA
CONGOREPUB	ALGERIA, CAMEROON, CAR, COMOROS
DJIBOUTI	CAR, RWANDA, SIERRALEONE

Table 2. Cont.

Home	Cointegration Relations with
EGYPT	CAR, GAMBIA, GUINEABISSAU, LESOTHO
EQGUINEA	NA
ERITREA	GUINEABISSAU
ETHIOPIA	NA
GABON	ALGERIA, CAR, IVORYCOAST
GAMBIA	ALGERIA, CAR, EGYPT, GUINEABISSAU, KENYA, LESOTHO, MALAWI, MALI, MAURITIUS, MOROCCO, SEYCHELLES, TUNISIA
GHANA	CHAD, MADAGASCAR, NIGER
GUINEA	BENIN, CAR, GUINEABISSAU, MALI, SENEGAL
GUINEABISSAU	BENIN, BOTSWANA, BURKINAFASO, CABOVERDE, CAR, EGYPT, ERITREA, GAMBIA, GUINEA, LESOTHO, MALAWI, MALI, MOROCCO, SEYCHELLES, SUDAN, TANZANIA, TUNISIA
IVORYCOAST	ALGERIA, CAR, GABON
KENYA	ALGERIA, CAR, GAMBIA, LESOTHO, MALAWI, MOROCCO, SEYCHELLES, TOGO
LESOTHO	ALGERIA, CAR, EGYPT, GAMBIA, GUINEABISSAU, KENYA, MALAWI, MALI, MAURITIUS, MOROCCO, SEYCHELLES, SOMALIA, TUNISIA
LIBERIA	NA
LIBYA	NA
MADAGASCAR	BURKINAFASO, CAR, CHAD, GHANA, MALI, MAURITANIA, MOZAMBIQUE, NIGERIA, RWANDA, SENEGAL, SUDAN, TANZANIA, UGANDA, ZAMBIA
MALAWI	ALGERIA, CAR, GAMBIA, GUINEABISSAU, KENYA, LESOTHO, MALI, MAURITANIA, MAURITIUS, MOROCCO, NAMIBIA, SEYCHELLES, SOUTHAFRICA, TOGO, TUNISIA
MALI	BENIN, BURKINAFASO, CABOVERDE, CAR, GAMBIA, GUINEA, GUINEABISSAU, LESOTHO, MALAWI, MAURITANIA, SENEGAL, SUDAN, TANZANIA, UGANDA
MAURITANIA	ALGERIA, BENIN, BURKINAFASO, CAR, GUINEA, MADAGASCAR, MALAWI, MALI, NAMIBIA, RWANDA, SENEGAL, SOMALIA, SOUTHAFRICA, SUDAN, TANZANIA, TOGO, UGANDA
MAURITIUS	ALGERIA, CAR, GAMBIA, LESOTHO, MALAWI, MOROCCO, SEYCHELLES, TOGO, TUNISIA
MOROCCO	ALGERIA, CAR, GAMBIA, GUINEABISSAU, KENYA, LESOTHO, MALAWI, MAURITIUS, SEYCHELLES, TOGO, TUNISIA
MOZAMBIQUE	CAR, MADAGASCAR
NAMIBIA	ALGERIA, CAR, MALAWI, MAURITANIA, SOUTHAFRICA, TOGO
NIGER	ANGOLA, CHAD, EQGUINEA, GHANA, MADAGASCAR, ZAMBIA
NIGERIA	CAR, MADAGASCAR, RWANDA, ZAMBIA
RWANDA	ALGERIA, CAMEROON, CAR, MADAGASCAR, MALI, MAURITANIA, MOROCCO, NAMIBIA, NIGERIA, SIERRALEONE, SOUTHAFRICA, TANZANIA, TOGO
SAOTOME	CAR
SENEGAL	BENIN, BURKINAFASO, CABOVERDE, CAR, GUINEA, MADAGASCAR, MALI, MAURITANIA, SUDAN, TANZANIA
SEYCHELLES	ALGERIA, CAR, GAMBIA, GUINEABISSAU, KENYA, LESOTHO, MALAWI, MAURITIUS, MOROCCO, TOGO, TUNISIA
SIERRALEONE	DJIBOUTI, RWANDA
SOMALIA	CABOVERDE, CAR, GAMBIA, LESOTHO, MALI, SEYCHELLES, TUNISIA
SOUTHAFRICA	ALGERIA, CAR, MALAWI, MALI, MAURITANIA, NAMIBIA, RWANDA, TOGO
SUDAN	BENIN, BURKINAFASO, CAR, MADAGASCAR, MALI, SENEGAL

Table 2. Cont.

Home	Cointegration Relations with
TANZANIA	CAR, GUINEABISSAU, MALI, MAURITANIA, SENEGAL, SUDAN
TOGO	ALGERIA, BURUNDI, CAR, COMOROS, GAMBIA, KENYA, LESOTHO, MALAWI, MALI, MAURITANIA, MAURITIUS, MOROCCO, NAMIBIA, RWANDA, SEYCHELLES, SOMALIA, SOUTHAFRICA, TUNISIA
TUNISIA	ALGERIA, CAR, GAMBIA, GUINEABISSAU, LESOTHO, MALAWI, MAURITIUS, MOROCCO, SEYCHELLES, SOMALIA
UGANDA	CAR, MADAGASCAR, MALI, MAURITANIA
ZAMBIA	ANGOLA, MADAGASCAR, NIGER, NIGERIA
ZIMBABWE	NA

Figure 6 presents a histogram of how often a country appears in such long-term relations. The maximum is obtained for the Central African Republic, which appears 38 times in a cointegrating relation. The minimum value of 0 is obtained for Democratic Republic of the Congo, Ethiopia, Liberia, Libya, Sao Tome and Principe, and Zimbabwe. These six countries apparently do not share a stochastic trend with any of the countries on the African continent.

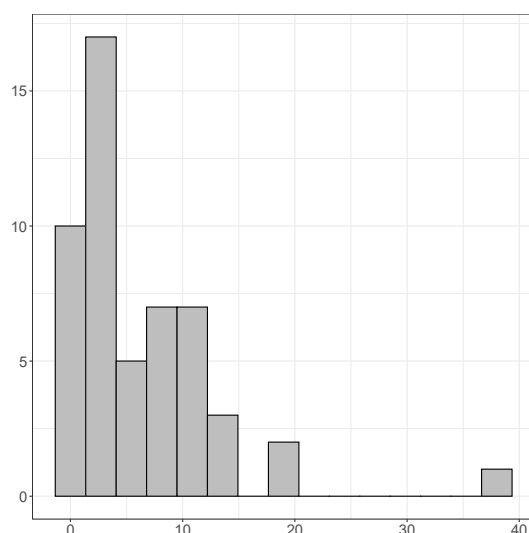


Figure 6. How often does a country appear in a cointegration relation? (51 is the maximum).

Furthermore, there are six countries in Table 2 which do not have a cointegration relation with other countries that can be meaningfully exploited for out-of-sample forecasting: these are CAR, Democratic Republic of the Congo, Equatorial Guinea, Ethiopia, Liberia, and Zimbabwe. For various reasons one might believe that these countries, at least many of them, behave in an idiosyncratic manner. For example, Zimbabwe entered standalone recessions, like CAR, while Equatorial Guinea witnessed huge growth. The same holds for Table 3, which reports the short-run effects of lagged growth rates, for which there are also three countries which do not appear as explanatory variables.

Finally, Figure 7 presents how often a one-year-lagged country's growth rate appears as a useful predictor for other countries. That is, how often was the country among the three strongest positive or negative autocorrelations with other countries, as computed in step 3 of Section 2. With 15 such appearances, Botswana is the frontrunner. There is no country that never appears as a predictor, but there are three countries that only appear once, and these are Libya, Namibia, and Nigeria. The Central African Republic appears only twice. Since the Central African Republic was most often in a cointegration relationship, this indicates that this country's predictive power is contained in the cointegration relationships rather than in the autocorrelations.

Yet, from an alternative perspective, the results in Figures 6 and 7 show that the countries in Africa do have various ties, which can be meaningfully exploited in forecasting models. These ties can be explained by closeness of the countries, similar economies, similar populations, similar links with non-African countries, and perhaps many other reasons.

Table 3. Strongest three autocorrelations of each “Home” country, as computed in step 3 of Section 2. These are the short-run relations that indicate which one-year-lagged growth rates have predictive value for the growth rates of the home country. The following abbreviations are used: CAR = Central African Republic, CONGODR = Democratic Republic of the Congo, CONGOREPUB = Republic of the Congo, EQGUINEA = Equatorial Guinea.

	Strongest Positive Correlations	Strongest Negative Correlations
ALGERIA	KENYA, CONGODR, MOROCCO	MAURITANIA, BENIN, SENEGAL
ANGOLA	ETHIOPIA, SIERRALEONE, CONGODR	COMOROS, ERITREA, GAMBIA
BENIN	TUNISIA, SEYCHELLES, CABOVERDE	MAURITIUS, CAR, BURKINAFASO
BOTSWANA	ZIMBABWE, KENYA, MAURITIUS	BENIN, BURKINAFASO, ETHIOPIA
BURKINAFASO	BENIN, TANZANIA, ANGOLA	MOROCCO, BOTSWANA, SAOTOME
BURUNDI	SIERRALEONE, EGYPT, DJIBOUTI	CABOVERDE, UGANDA, LESOTHO
CABOVERDE	EQGUINEA, ALGERIA, EGYPT	DJIBOUTI, CONGODR, BURUNDI
CAMEROON	LESOTHO, EGYPT, LIBERIA	SEYCHELLES, SENEGAL, GUINEABISSAU
CAR	CABOVERDE, MALI, SOMALIA	LIBYA, NIGER, SIERRALEONE
CHAD	SIERRALEONE, SUDAN, TANZANIA	SEYCHELLES, IVORYCOAST, BOTSWANA
COMOROS	CAMEROON, LIBERIA, GAMBIA	ETHIOPIA, UGANDA, ZAMBIA
CONGODR	DJIBOUTI, SIERRALEONE, ZAMBIA	CABOVERDE, EQGUINEA, TUNISIA
CONGOREPUB	CAMEROON, BOTSWANA, DJIBOUTI	GUINEA, SEYCHELLES, MOROCCO
DJIBOUTI	CONGODR, SIERRALEONE, LIBERIA	EGYPT, SUDAN, CABOVERDE
EGYPT	SUDAN, SENEGAL, CAMEROON	GHANA, UGANDA, BENIN
EQGUINEA	CABOVERDE, BURKINAFASO, GHANA	DJIBOUTI, BURUNDI, CONGODR
ERITREA	COMOROS, CONGOREPUB, BOTSWANA	SAOTOME, ANGOLA, SOUTHAFRICA
ETHIOPIA	NIGER, ANGOLA, GHANA	BOTSWANA, GABON, TUNISIA
GABON	CONGOREPUB, DJIBOUTI, BOTSWANA	NIGER, EGYPT, BENIN
GAMBIA	SAOTOME, ALGERIA, MAURITIUS	GHANA, ETHIOPIA, GUINEABISSAU
GHANA	TANZANIA, GUINEA, ETHIOPIA	CONGOREPUB, SOUTHAFRICA, GABON
GUINEA	MOZAMBIQUE, ETHIOPIA, ZIMBABWE	SIERRALEONE, CONGOREPUB, CAMEROON
GUINEABISSAU	CONGODR, CHAD, KENYA	LIBERIA, SEYCHELLES, EQGUINEA
IVORYCOAST	TOGO, SAOTOME, MOROCCO	ERITREA, CONGOREPUB, ETHIOPIA
KENYA	BURUNDI, MAURITANIA, NIGER	EQGUINEA, CABOVERDE, BENIN
LESOTHO	SOMALIA, EGYPT, BOTSWANA	GHANA, DJIBOUTI, MADAGASCAR
LIBERIA	ANGOLA, MOZAMBIQUE, CAMEROON	ERITREA, BURUNDI, BOTSWANA
LIBYA	CAR, UGANDA, MAURITIUS	GAMBIA, EGYPT, CHAD
MADAGASCAR	SEYCHELLES, LIBERIA, CABOVERDE	EGYPT, CONGOREPUB, TUNISIA
MALAWI	ANGOLA, TOGO, UGANDA	BURKINAFASO, RWANDA, COMOROS

Table 3. Cont.

	Strongest Positive Correlations	Strongest Negative Correlations
MALI	TANZANIA, COMOROS, SEYCHELLES	MAURITANIA, ZIMBABWE, BOTSWANA
MAURITANIA	ALGERIA, SAOTOME, ANGOLA	ERITREA, SUDAN, RWANDA
MAURITIUS	BOTSWANA, UGANDA, IVORYCOAST	NIGER, BENIN, ETHIOPIA
MOROCCO	SOUTHAFRICA, MALAWI, IVORYCOAST	ERITREA, NAMIBIA, NIGERIA
MOZAMBIQUE	GHANA, MADAGASCAR, ETHIOPIA	ERITREA, EGYPT, CONGOREPUB
NAMIBIA	SAOTOME, TANZANIA, MOZAMBIQUE	ERITREA, CABOVERDE, EGYPT
NIGER	GHANA, TANZANIA, LESOTHO	TUNISIA, BOTSWANA, ALGERIA
NIGERIA	UGANDA, SIERRALEONE, MOZAMBIQUE	ERITREA, ZIMBABWE, BENIN
RWANDA	ANGOLA, TOGO, CAMEROON	GUINEA, SEYCHELLES, COMOROS
SAOTOME	SOUTHAFRICA, IVORYCOAST, GABON	ERITREA, BURKINAFASO, MALI
SENEGAL	TANZANIA, GHANA, CHAD	BOTSWANA, EGYPT, TUNISIA
SEYCHELLES	TOGO, MAURITIUS, CHAD	BENIN, SENEGAL, CONGOREPUB
SIERRALEONE	DJIBOUTI, ZAMBIA, RWANDA	MALAWI, ERITREA, SUDAN
SOMALIA	SUDAN, GABON, LIBERIA	TANZANIA, ANGOLA, BURUNDI
SOUTHAFRICA	SAOTOME, MAURITANIA, CONGODR	ERITREA, GUINEABISSAU, BURKINAFASO
SUDAN	BURKINAFASO, GABON, TANZANIA	KENYA, LESOTHO, BOTSWANA
TANZANIA	CONGODR, DJIBOUTI, CHAD	BOTSWANA, CONGOREPUB, ZIMBABWE
TOGO	SAOTOME, ETHIOPIA, SOUTHAFRICA	CHAD, CONGOREPUB, ERITREA
TUNISIA	BOTSWANA, MALAWI, LESOTHO	BENIN, ZAMBIA, ALGERIA
UGANDA	MOZAMBIQUE, ETHIOPIA, CABOVERDE	CAMEROON, ERITREA, COMOROS
ZAMBIA	ANGOLA, TANZANIA, CONGODR	ERITREA, MAURITIUS, ZIMBABWE
ZIMBABWE	CONGOREPUB, COMOROS, MADAGASCAR	SUDAN, CHAD, BURKINAFASO

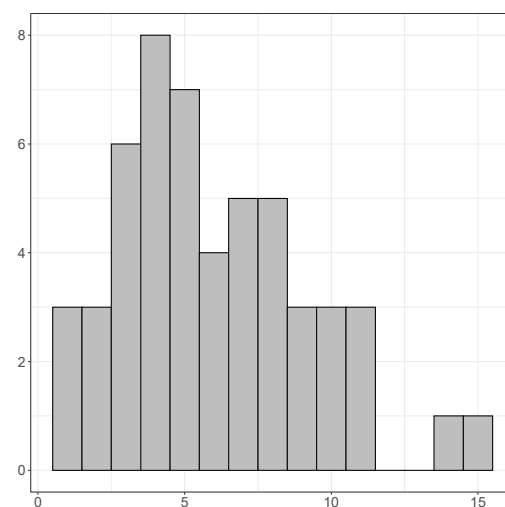


Figure 7. How often does a country's growth rate appear as a useful predictor? That is, how often was a country among the three strongest positive or negative autocorrelations with other countries?

4. A Simulation Study

To examine if our empirical results have any meaningful significance, we run some simulation experiments. We create 50 different time series of length 100 as:

$$y_{i,t} = \delta_i + y_{i,t-1} + \varepsilon_{i,t},$$

with $i = 1, 2, \dots, 50$, $t = 1, 2, \dots, T$, with $\varepsilon_{i,t} \sim N(0, 0.5)$, $y_{i,0} = 0$, and where δ_i is a draw from a continuous uniform distribution with the range 1 to 2. Hence, the DGP (Data Generating Process) amounts to independent random walks. We split the sample into T_1 and T_2 with $T_1 = T_2 = 50$.

For each series i we run 49 OLS regressions $y_{i,t} = \delta_i + \pi_i y_{j,t} + \omega_{j,t}$ for $j = 1, \dots, 50$ and $j \neq i$. We compute the residuals $\hat{\omega}_{j,t}$ and create the CRDW (Cointegration Regression Durbin Watson) test statistic as $CRDW_j = 2(1 - \hat{\rho}_j)$, where $\hat{\rho}_j$ is the estimated first-order autocorrelation of $\hat{\omega}_{j,t}$. When $CRDW_j > \tau$, the regressor $y_{j,t}$ is kept for the next round. We choose $\tau = 0.4$ and $\tau = 0.7$. Next, for all retained variables, also including $y_{i,t}$, we run the Johansen (1991) cointegration estimation method and retain the first estimated cointegration relationship for the single equation forecasting model. We compute all pairwise correlations between $\Delta y_{i,t}$ and $\Delta y_{j,t-1}$ for $j = 1, \dots, I$ and $j \neq i$. We rank the series with the positive correlations from high to low, and with the negative correlations from (absolute) high to low. To further reduce the number of parameters, we impose two sets of restrictions as in step 4 before. We choose k_{pc} and k_{nc} equal to 3. Finally, we estimate:

$$\Delta y_{i,t} = \mu_i + \sum_{j=1}^I \alpha_{i,j} \Delta y_{j,t-1} + \gamma_i (y_{i,t-1} - \sum_{j=1, j \neq i}^I \beta_{i,j} y_{j,t-1}) + \varepsilon_{i,t}$$

using NLS. We also estimate the MA(1) model $\Delta y_{i,t} = \omega_{i,t} + \theta_i \omega_{i,t-1}$. We create for these models, recursively, $T_2 = 50$ one-step-ahead forecasts for each $\Delta y_{i,t}$, and compute the RMSE (Root Mean Squared Error) of the forecasts. This is repeated 25 times.

The key results are presented in Table 4. From this table we can see that in case the SEFM methodology results in spuriously relevant models, these models give about equally accurate forecasts as the MA(1) model. Moreover, the average RMSPE is about equal. This provides further substantiation to our results in Table 1, where the differences between the two models can be large and relevant.

Table 4. Results from simulation experiments. The number of replications is 25.

	$\tau = 0.4$		$\tau = 0.7$	
	SEFM	MA(1)	SEFM	MA(1)
Average RMSE (over 25 replications)	0.888	0.851	0.807	0.795
Average number of cases (out of $I = 50$) where SEFM outperforms the MA(1)	21.4		21.6	

5. Conclusions

We have proposed a simple and reproducible methodology to create single equation forecasting models for low-frequency macroeconomic variables. Our methodology was illustrated for forecasting annual real GDP growth rates for 52 African countries. The models (potentially) included lagged growth rates of all countries, as well as a single cointegration relationship to capture potential common stochastic trends. With a few selection steps, our methodology quickly arrived at a reasonably small forecasting model. Compared with simple benchmark models, the single equation forecasting models seem to perform quite well.

A limitation of our methodology is of course that we must make decisions each time based on a partial analysis, that is, one variable with another. The partial analysis does not involve any statistical testing, so there is no multiple testing problem. A partial analysis could lead to incorrect inclusion of variables or an incorrect neglect of variables. The final model is analyzed using the familiar t -tests. One way or the other, one would have to decide which countries are associated more with each other than other countries, and that does not seem to be a simple task. Perhaps machine learning methods or shrinkage methods can become useful here; see Koo et al. (2020). More experience with our methodology should tell how useful it is.

Another limitation is that we assumed a diagonal covariance matrix for the errors. For our illustration this seemed reasonable, but for other applications one may wish to include cross-equation error dependence. A modification of our methodology to include such dependencies will be considered as interesting further work.

Finally, we believe that an exercise in searching for those values of τ , k_{pc} , and k_{nc} which minimize a certain criterion can be interesting. One could create a hold-out sample, for example, 2000–2010, fit the models for data up to 1999, and select the τ , k_{pc} , and k_{nc} which minimize, for example, the mean absolute error. It is however not guaranteed that the resulting “optimal” τ , k_{pc} , and k_{nc} also provide the most accurate forecasts for the true out-of-sample data. Of course, other choices of τ , k_{pc} , and k_{nc} can lead to other forecast outcomes, but this was not the main intention of our paper. We wanted to introduce a simple and reproducible modeling strategy, which contrasts with many currently used machine learning methods which need tuning of hyperparameters and involve all kinds of other choices, which, in our experience, make empirical results difficult to replicate. In our future research we will work more on this issue and will think of optimization strategies for finding appropriate values of τ , k_{pc} , and k_{nc} .

Author Contributions: Conceptualization, P.H.F. and M.W.; methodology, P.H.F.; software, M.W.; validation, P.H.F. and M.W.; formal analysis, P.H.F. and M.W.; investigation, P.H.F. and M.W.; resources, P.H.F.; data curation, P.H.F.; writing—original draft preparation, P.H.F. and M.W.; writing—review and editing, P.H.F.; visualization, P.H.F. and M.W.; supervision, P.H.F.; project administration, P.H.F. and M.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: We thank Peter Boswijk for helpful conversations, Eva van Rooijen for her help with the simulations, and two anonymous reviewers for their helpful suggestions.

Conflicts of Interest: The authors declare no conflict of interest.

Notes

- ¹ We are aware of the literature on the potential problems with macroeconomic data for Africa; see for example Jerven (2013). However, the data that we use are publicly available through the World Bank and for most countries start in 1960. For some series we have imputed missing values using regression models, including a range of available explanatory variables.
- ² In a sense, our approach bears similarities with a recent pairwise-based approach to model a macroeconomic variable that is the weighted average of many components, see Carlomagno and Espasa (2021).
- ³ The IMA(1,1) model for the levels of the data is recommended as a benchmark in Franses (2020). It is of course not certain that this univariate model fits the data best, and hence also alternative univariate benchmark models are considered. Given the data at hand, it is difficult to propose a multivariate benchmark model, although we do consider a Principal Components Regression for the growth rates.
- ⁴ In an online appendix, all results and the complete code that was used to write this paper are publicly available on <https://github.com/mwelz/sefm-africa>.

- 5 Their methods build on those proposed in Bai and Ng (2002) and Stock and Watson (2002), using modern tools like Lasso and Elastic Net to reduce the number of parameters.
- 6 In further research, one could think of allowing non-zero correlations where perhaps cluster techniques can be used on the estimated residuals to put zeroes in the large covariance matrix. For the particular illustration in the present paper, we estimated 52×51 divided by 2 is 1326 cross-equation correlations and these are almost all in between -0.2 and 0.2 . Given the large diversity across the African countries, this comes as no surprise. Our model bears similarities with the autoregressive distributed lag model in Pesaran et al. (2001).
- 7 Note that the θ_p and θ_n parameters are not identified under the null hypothesis that $\alpha_i = 0$.
- 8 If we would have used $\tau = 0.6$, there would not have been a cointegration relation, and if we would have chosen $\tau = 0.2$, more variables would have been included in the cointegration relation.
- 9 As mentioned, various other decisions can be made, and these may have an impact on forecast quality. It is not our intention in this paper to look for configurations that yield the highest forecast accuracy, although there may be search routines that can do that.
- 10 For 17 countries we have imputed data. The average MAE for these countries is 5.391, while the average MAE of all countries is 4.388. Hence, the differences are not substantial. A regression of MAE on a constant and a dummy for missingness of data give an estimate of 1.534 with a standard error of 1.784. Hence, we conclude that there are no significant differences in MAE.

References

- Bai, Jushan, and Serena Ng. 2002. Determining the number of factors in approximate factor models. *Econometrica* 70: 191–221. [CrossRef]
- Boswijk, H. Peter. 1994. Testing for an unstable root in conditional and structural error correction models. *Journal of Econometrics* 63: 37–60. [CrossRef]
- Carlomagno, Guillerma, and Antonio Espasa. 2021. Discovering specific common trends in a large set of disaggregates: Statistical procedures, their properties and an empirical application. *Oxford Bulletin of Economics & Statistics*. in print. Available online: <https://onlinelibrary.wiley.com/doi/epdf/10.1111/obes.12412> (accessed on 20 August 2021).
- Franses, Philip Hans. 2020. IMA(1,1) as a new benchmark for forecast evaluation. *Applied Economics Letters* 27: 1419–23. [CrossRef]
- Franses, Philip Hans, and Simeon Vasilev. 2019. *Real GDP Growth in Africa, 1963–2016*. Econometric Institute Report EI2019-23. Rotterdam: Erasmus School of Economics.
- Jerven, Morten. 2013. *Poor Numbers: How We are Misled by African Development Statistics and What to Do about It*. Ithaca: Cornell University Press.
- Johansen, Soren. 1991. Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica* 59: 1551–80. [CrossRef]
- Kim, Hyun Hak, and Norman R. Swanson. 2018. Mining big data using parsimonious factor, machine learning, variable selection and shrinkage methods. *International Journal of Forecasting* 34: 339–54. [CrossRef]
- Koo, Bonsoo, Heather Anderson, Myung Hwan Seo, and Wenying Yao. 2020. High-dimensional predictive regression in the presence of cointegration. *Journal of Econometrics* 219: 456–77. [CrossRef]
- Pesaran, M. Hashem, Yongcheol Shin, and Richard J. Smith. 2001. Bounds testing approaches to the analysis of level relationships. *Journal of Applied Econometrics* 16: 289–326. [CrossRef]
- Sargan, J. Dennis, and Alok Bhargava. 1983. Testing residuals from least squares regression for being generated by the Gaussian random walk. *Econometrica* 51: 153–74. [CrossRef]
- Stock, James H., and Mark W. Watson. 2002. Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association* 97: 1167–79. [CrossRef]