

TECHNICAL APPENDIX TO “FORECASTING REAL GDP GROWTH FOR AFRICA”

Philip Hans Franses Max Welz
franses@ese.eur.nl welz@ese.eur.nl

ECONOMETRIC INSTITUTE
ERASMUS SCHOOL OF ECONOMICS

March 12, 2021

Abstract

This is a technical appendix to the working paper “Forecasting Real GDP Growth for Africa” by the same authors. In this technical appendix, we thoroughly describe every step of the proposed methodology of the main paper.

1 SEFM Estimation

Let there be N countries for each of which we have T time series observations. Let $\mathcal{N} = \{1, 2, \dots, N\}$ be the set of all N countries, which are indexed by integers. For each country $i \in \mathcal{N}$, we have information about T periods of a time-series variable of interest, $y_{i,t}$, called the *outcome*, for $t = 1, \dots, T$. The overall goal is to estimate a Single Equation Forecasting Model (SEFM) for each country.

1.1 Step 1

For a fixed country $i \in \mathcal{N}$, run $N - 1$ OLS time series regressions of the type

$$y_{i,t} = \delta_i + \pi_i y_{j,t} + \omega_{i,t}^{(j)}, \quad t = 1, \dots, T, \quad (1)$$

for all $j \in \mathcal{N} \setminus \{i\}$. Compute with the estimated OLS residuals $\hat{\omega}_{i,t}^{(j)}$ the (estimated) first order autocorrelation $\hat{\rho}_i^{(j)} = \text{Cov}(\hat{\omega}_{i,t}^{(j)}, \hat{\omega}_{i,t-1}^{(j)}) / \text{Var}(\hat{\omega}_{i,t}^{(j)})$ for all $j \in \mathcal{N} \setminus \{i\}$, and use this quantity to calculate the Cointegration Regression Durbin-Watson (CRDW) test statistic by

$$\text{CRDW}_i^{(j)} = 2(1 - \hat{\rho}_i^{(j)}), \quad j \in \mathcal{N} \setminus \{i\}. \quad (2)$$

Then, fix some threshold $\tau \in \mathbb{R}$. If $\text{CRDW}_i^{(j)} > \tau$, we retain the country j associated with regressor $y_{j,t}$ in (1). Formally, denote the set that holds all retained countries for country i (including i itself) by

$$\mathcal{R}_i \equiv \{j \in \mathcal{N} \setminus \{i\} : \text{CRDW}_i^{(j)} > \tau\} \cup \{i\} \subseteq \mathcal{N}.$$

1.2 Step 2

For each country contained in \mathcal{R}_i , run the Johansen cointegration method. Let $\mathcal{J}_i \subseteq \mathcal{R}_i \setminus \{i\}$ contain all countries for which the Johansen cointegration method returned a cointegration relation (excluding i itself). Let the variable that emerges from the exact cointegration relation be denoted $c_{i,t}$, $t = 1, \dots, T$, which is a linear function of the outcome variables contained in \mathcal{J}_i and the outcome of i . Define furthermore $|\mathcal{J}_i|$ -dimensional vectors $\mathbf{j}_{i,t} \equiv \{y_{k,t}\}_{k \in \mathcal{J}_i}$, $t = 1, \dots, T$, that contain the outcomes of all countries for which a cointegration relation with country i was found.

1.3 Step 3

For a given country $i \in \mathcal{N}$, compute all pairwise differentiated correlations over the time dimension, that is,

$$\varrho_{i,j} \equiv \text{Corr}(\Delta y_{i,t}, \Delta y_{j,t-1}), \quad (3)$$

for all $j \in \mathcal{N} \setminus \{i\}$. Then, fix by $\tau_{pc} \geq 0$ and $\tau_{nc} \leq 0$ thresholds for the positive or negative correlations, respectively. Next, let the sets

$$\begin{aligned} \mathcal{C}_i^+ &= \{j \in \mathcal{N} \setminus \{i\} : \varrho_{i,j} > \tau_{pc}\} \quad \text{and} \\ \mathcal{C}_i^- &= \{j \in \mathcal{N} \setminus \{i\} : \varrho_{i,j} < \tau_{nc}\} \end{aligned}$$

contain all countries with a sufficiently strong positive and negative differentiated autocorrelation, respectively, as in (3).

1.4 Step 4

Fix an integer K_{pc} (K_{nc}), which denotes how many of the—in absolute value—most positively (negatively) correlated countries¹ shall be considered.

Moreover, denote by $\text{sort}_+(\cdot)$ a map that rearranges the country indices contained in \mathcal{C}_i^+ such that the country $j \in \mathcal{C}_i^+$ with the strongest (positive) correlation $\varrho_{i,j}$ becomes the first country in $\text{sort}_+(\mathcal{C}_i^+)$, the country with the second-strongest correlation in \mathcal{C}_i^+ becomes the second country in $\text{sort}_+(\mathcal{C}_i^+)$, and so on. The map $\text{sort}_+(\cdot)$ is defined similarly, so that the first element in $\text{sort}_-(\mathcal{C}_i^-)$ is the country that has the strongest (negative) correlation in \mathcal{C}_i^- , the second element in $\text{sort}_-(\mathcal{C}_i^-)$ is the country that has the second strongest correlation in \mathcal{C}_i^- , and so on.

¹Henceforth, the term “correlation” refers to the differentiated autocorrelation as defined in (3).

Abusing notation, the operation $\{j\}_{j \in \text{sort}_+(\mathcal{C}_i^+)}^{K_{pc}}$ means that the first K_{pc} elements of the set $\text{sort}_+(\mathcal{C}_i^+)$ form a vector, that is, the K_{pc} countries with the strongest (positive) correlations form a vector. The operation $\{j\}_{j \in \text{sort}_-(\mathcal{C}_i^-)}^{K_{nc}}$ is defined similarly.

Then, for a given country $i \in \mathcal{N}$, collect the K_{pc} countries that are strongest positively correlated with country i in K_{pc} -dimensional vectors $\mathbf{p}_{i,t}$. Analogously, the K_{nc} -dimensional vectors $\mathbf{n}_{i,t}$ contain the K_{nc} strongest negatively correlated countries with i . Formally,

$$\mathbf{p}_{it} \equiv \{y_{j,t}\}_{j \in \text{sort}_+(\mathcal{C}_i^+)}^{K_{pc}} \quad \text{and} \quad \mathbf{n}_{it} \equiv \{y_{j,t}\}_{j \in \text{sort}_-(\mathcal{C}_i^-)}^{K_{nc}}, \quad (4)$$

for $t = 1, \dots, T$.

2 Model Building and Estimation

Finally, we can obtain estimates for the SEFM where we only have to fix the values of five parameters: $\tau, \tau_{pc}, \tau_{nc}, K_{pc}$ and K_{nc} . We can do so by means of classical OLS and NLS. Before we proceed, let us recall what we have obtained in the previous steps for a given country $i \in \mathcal{N}$:

- the variable $c_{i,t}$, which is the cointegration variable for i ,
- the vector $\mathbf{j}_{i,t}$, which contains the outcomes of the countries for which a cointegration relationship with i was found,
- the vector $\mathbf{p}_{i,t}$, which contains the outcomes of the K_{pc} countries that are most strongly positively correlated with i (ordered by strength of correlation),
- the vector $\mathbf{n}_{i,t}$, which contains the outcomes of the K_{nc} countries that are most strongly negatively correlated with i (ordered by strength of correlation).

2.1 OLS Estimation

Using the obtained variables, run the following OLS regression for a given country $i \in \mathcal{N}$:

$$\Delta y_{it} = \mu_i + \gamma_i c_{i,t-1} + \lambda_i (\Delta y_{i,t-1}) + \boldsymbol{\theta}_{pc,i}^\top (\Delta \mathbf{p}_{i,t-1}) + \boldsymbol{\theta}_{nc,i}^\top (\Delta \mathbf{n}_{i,t-1}) + \varepsilon_{i,t}, \quad (5)$$

for $t = 3, \dots, T$. Note that $\boldsymbol{\theta}_{pc,i}$ is a K_{pc} -vector of coefficients and $\boldsymbol{\theta}_{nc,i}$ is a K_{nc} -vector of coefficients.

We then drop all variables whose OLS coefficients are not statistically significant at 95% confidence (with robust standard errors) and re-estimate the OLS model to arrive at our final OLS model.

2.2 NLS Estimation

For a given country $i \in \mathcal{N}$, fit the following model by means of NLS estimation:

$$\begin{aligned}
\Delta y_{i,t} = & \mu_i + \\
& + \gamma_i \left(y_{i,t-1} - \beta_i^\top \mathbf{j}_{i,t-1} \right) + \\
& + \lambda_i (\Delta y_{i,t-1}) + \\
& + \theta_{pc,i} \sum_{j=1}^{K^{pc}} \alpha_{pc,i}^{j-1} (\Delta p_{i,t-1,j}) + \\
& + \theta_{nc,i} \sum_{j=1}^{K^{nc}} \alpha_{nc,i}^{j-1} (\Delta n_{i,t-1,j}) + \\
& + \varepsilon_{i,t},
\end{aligned} \tag{6}$$

for $t = 3, \dots, T$. The scalars $p_{i,t,j}$ and $n_{i,t,j}$ are the j -th elements of the vectors $\mathbf{p}_{i,t}$ and $\mathbf{n}_{i,t}$, respectively.

The parameter β_i in (6) is a $|\mathcal{J}_i|$ -vector of coefficients. Further (scalar) parameters are $\mu_i, \gamma_i, \lambda_i, \theta_{pc,i}, \theta_{nc,i}, \alpha_{pc,i}$. Hence, the NLS needs to numerically estimate $7 + |\mathcal{J}_i|$ parameters. Note that the problem associated with (6) is non-convex,² so we cannot guarantee that the returned local optimum is also the global optimum, let alone that the chosen optimization algorithm even converges (especially given the number of parameters to estimate here).

We use the following initial values for the numerical optimization:

- $\mu_i = 0$,
- $\gamma_i = -0.3$,
- $\lambda_i = 0.5$,
- $\theta_{pc,i} = 0.6$,
- $\alpha_{pc,i} = 0.5$,
- $\theta_{nc,i} = -0.6$,
- $\alpha_{nc,i} = 0.5$,
- β_i is initialized at the Johansen coefficients from Step 2.

We perform variable selection for the NLS model (6) as follows. We test the significance of all parameters in model (6) except the intercept μ_i and γ_i . We use robust standard errors. If a parameter is not significant at 95% confidence, we set it equal to zero.

²The objective function to optimize the parameters by means of NLS is the usual sum of squared residuals, i.e. $\sum_{t=3}^T \varepsilon_{i,t}^2$ in (6).

2.3 Special Cases

There are two special cases that prevent NLS estimation of (6):

Case 1: No Johansen cointegration relationships are found in Step 2. Hence, the vector $\mathbf{j}_{i,t}$ cannot be constructed. We overcome this problem by setting $\mathbf{j}_{i,t} = \boldsymbol{\beta}_i = \mathbf{0}$ and $\gamma_i = 0$ in the NLS model (6). We perform variable selection as described in Section 2.2.

Case 2: The NLS model (6) does not converge. In this case, we estimate the OLS model (5) and perform variable selection as in Section 2.1.