A G D I   Working Paper

WP/22/061

What Really Drives Economic Growth in Sub-Saharan Africa? Evidence from The Lasso Regularization and Inferential Techniques

Forthcoming: Journal of the Knowledge Economy

Isaac K. Ofori
(Corresponding Author)
Department of Economics, University of Insubria,
Via Monte Generoso 71, Varese, Italy.
E-mail: ikofori@outlook.com

Camara K. Obeng
Department of Economic Studies, School of Economics,
University of Cape Coast
Cape Coast, Ghana.
E-mails: cobeng@ucc.edu.gh

Simplice A. Asongu
(Corresponding author)
African Governance and Development Institute,
P.O. Box 8413, Yaoundé
E-mails: asongusimplice@yahoo.com / asongus@afridev.org

2022 African Governance and Development Institute WP/22/061

Research Department

What Really Drives Economic Growth in Sub-Saharan Africa? Evidence from The Lasso Regularization and Inferential Techniques

Isaac K. Ofori, Camara K. Obeng & Simplice A. Asongu

January 2022

Abstract

The question of what really drives economic growth in sub-Saharan Africa (SSA) has been debated for many decades now. However, there is still a lack of clarity on variables crucial for driving growth as prior contributions have been executed at the backdrop of preferential selection of covariates in the midst several of potential drivers of economic growth. The main challenge with such contribution is that even tenuous variables may be deemed influential under some model specifications and assumptions. To address this and inform policy appropriately, we train algorithms for four machine learning regularization techniques— *the Standard lasso, the Adaptive lasso, the Minimum Schwarz Bayesian information criterion lasso*, and *the Elasticnet* to study patterns in a dataset containing 113 covariates and identify the key variables affecting growth in SSA. We find that only 7 covariates are key for driving growth in SSA. Estimates of these variables are provided by running the lasso inferential techniques of *double-selection linear regression, partialing-out lasso linear regression,* and *partialing-out lasso instrumental variable regression*. Policy recommendations are also provided in line with the AfCFTA and the green growth agenda of the region.

Keywords: Economic growth; Elasticnet; Lasso; Machine learning; Partialing-out IV regression; sub-Saharan Africa

JEL Codes: C52, C53, C55, O11, O4, O55

1. Introduction

The debate on sources of growth continues to generate attention in the political and academic landscapes due to its relevance for policy formulations on welfare, international competition, and economic management. From the saving-oriented (Harrod, 1939; Domar, 1947) and technical progress neoclassical theories of economic growth (Solow,1956), to the imperfect market-augmented endogenous growth theories of Romer (1990), Aghion and Howitt (1992), Grossman and Helpman (1991), economists are still exploring which variables matter for growth. The need to identify the key drivers of economic growth has even become crucial than ever following the emergence of the coronavirus pandemic (IMF, 2020; World Bank, 2020). For developing economies such as those in sub-Saharan Africa (SSA), knowledge on the key drivers of economic growth is a great step in formulating and implementing policies to foster, sustained and share growth. Additionally, identifying the key drivers of economic growth would be a giant breakthrough on the parts of policymakers and developing partners in mapping out growth strategies in line with the '*green growth*[1] *agenda*' of the region.

A plethora of prior contributions on economic growth in SSA identify covariates such as trade openness, foreign direct investment, capital flows and innovation (Sakyi *et al.,* 2015; Agbloyor *et al.,* 2014), financial development (Opoku *et al.,* 2019; Peprah *et al*., 2019), macroeconomic management (Alagidede & Ibrahim, 2017), institutional quality (Berhane, 2018; Chakamera, 2018), human capital (Anyanwu, 2014; Gyimah-Brempong *et al.,* 2006), and ICT (Adeleye & Eboagu, 2019; Asongu & Odhiambo, 2019) as drivers of economic growth in SSA. A conspicuous lacuna in the extant scholarship, however, is that all these variables deemed crucial for economic growth are selected based on the researcher's discretion even in large datasets. The concern with preferential selection of covariates is that even *tenuous* drivers of growth may be deemed highly influential under certain assumptions, model specification or estimation technique. Another concern is that the preferential selection of covariates in the midst of several potential determinants of the outcome variable partly contributes to the inconclusive results in big-data regression problems. Addressing this challenge and thus informing policy appropriately can be done through the use of machine learning[2] (artificial intelligence) algorithms for regularization, prediction, and inference (see, Tibshirani, 1996; Zou & Hastie, 2005; Zou, 2006). Indeed, machine learning techniques have been applied in various fields including health (see, Mateen *et al.,* 2020; Doupe *et al.,* 2019;

[1] Green growth refers to achieving sustainable growth trajectories that is environmentally friendly (OECD 2017)
[2] Machine learning has gained attention in recent times due to its ability to detect relevant patterns in big data for prediction and analysis.

Beam & Kohane, 2018), transportation (Tizghadam *et al.*,2019; Bhavsar *et al.*,2017), games and psychology (Sandeep *et al.*, 2020; De Almeida-Rocha & Duarte, 2019; Luxton, 2016), and finance (Bredt, 2019; Bazarbash, 2019; Akbari *et al*., 2020).

Despite the growing application of machine learning algorithms for regularization and variable selection, prediction and inference in several fields, rigorous empirical works exploring its applicability and power in selecting variables crucial for economic growth in SSA are hard to find. This fundamentally forms the contribution of this paper. The first objective, therefore, is to train several machine learning algorithms to identify the main drivers of economic growth in SSA. The second objective is to provide reliable estimates and confidence intervals of the key drivers of economic growth taking into consideration possible endogeneity, multicollinearity, and modelling complexities. To the best of our knowledge, this study is the first of its kind in SSA to apply machine learning techniques in selecting key drivers of economic growth. Particularly, following renewed efforts to achieve, sustain, and share growth gains in line with implementation of the African Continental Free Trade Area (AfCFTA), and institution of the African Agenda 2063, our results could prove crucial to the course by aiding the planning, modelling and the targeting of growth.

Our choice of the study area is informed by a number of factors. First, as Kaufmann, Kraay and Mastruzzi (2010) note, SSA countries are fundamentally common in terms of institutions. Despite lags in several facets of governance such as rule of law, regulatory quality, and corruption-control, the quality of these indicators as gleaned from the World Governance Indicators is rising steadily across the region. However, macroeconomic challenges relating to inflation, exchange rate fluctuations, macroeconomic bailouts, and geopolitical fragilities are common among countries in the region. Second, SSA countries are remarkably similar in terms of structural or real sector setting (OECD/ILO, 2019; UNCTAD, 2021; World Bank, 2021a). For instance, most of the region's active workforce is employed in the agricultural sector and are more susceptible to political, financial and trade shocks. Also, worth mentioning is the common goal of SSA countries in using economic integration[3] as a vehicle to spur industrialisation, growth, poverty alleviation and equitable income distribution. Another peculiarity is the low industrial output but fast rising service sector, providing policymakers with opportunities to leapfrog classical development processes (IMF 2021). Third, as noted by the African Development Bank (2018), countries in the region are markedly common in infrastructural development. Particularly, SSA countries report sharp

[3] Countries in SSA have collectively signed onto the African Continental Free Trade Area.

deficits in digital and physical infrastructure such as ICT, electricity, transportation, as well as water and sanitation compared to their North African counterparts. Finally, countries in SSA are substantially similar in terms of growth trajectories, level of development, and lingering concerns of inability to build sustained momentum.

The rest of the paper is organised as follows: the next section presents a brief survey of economics-related studies applying machine learning. The data and empirical models are also presented in Section 3. The results and discussions are presented in Section 4 while Section 5 concludes with some policy recommendations.

2.0 Literature survey on empirical works using machine learning

The literature on economic growth is vast and an attempt to present all of them will be a daunting one. Therefore, attention is paid to the recent advances and applications of machine learning regularization techniques by researchers in aiding policymakers plan and target growth. For instance, this study is similar to Schneider and Wagner (2012; 2009) who focus solely on the machine learning algorithm of lasso (Least Absolute Shrinkage and Selection Operator) in determining key drivers of growth in the NUTS2 region[4] of the European Union over the period 1995 – 2005. The results indicate that covariates such as initial GDP per capita, human capital, and initial unemployment rate matter for economic growth.

Similarly, in identifying which income distribution measure matter for development outcomes, Dutt and Tsetlin (2016) applied the elasticnet and lasso techniques to select from 37 potential covariates of development. The authors find that the poverty headcount indicator matters most in predicting three development outcomes (i.e., per capita income, schooling, and institutional quality) compared to 36 other distributional measures. A similar work is Tkacz (2001), which in forecasting Canadian GDP growth, applied the neural network algorithms. The study finds that, relative to traditional methods such as linear and univariate forecasting methods, neural network techniques yield lower forecast errors on annual growth rate. The author goes further to indicate that neural techniques perform better in forecasting long-term growth than short-term growth. Further, Richardson *et al.* (2021) explore the power of several machine learning techniques[5] relative to classical methods in forecasting real GDP growth in New Zealand. The authors find that machine learning algorithms outperform classical statistical methods in prediction. Jung *et al.* (2018) also employ machine learning algorithms of lasso, ridge, elasticnet, neural networks, and super learner to examine

[4]Nomenclature of Territorial Units for Statistics.
[5] Support-vector machine, neural network, lasso, boosted tree, regularized generalized linear model and ridge,

the GDP growth of the G7 countries. The authors provide strong evidence to conclude that machine learning algorithms outperformed standard prediction techniques.

In the case of SSA, however, the literature shows that researchers have not explored how relevant these techniques can be in aiding policymakers plan and target growth. The results we provide could prove invaluable in helping policymakers turnaround the slow growth (real GDP per capita) trajectories of the SSA region as presented in **Fig.**1 and **Fig.**1.1 in the Appendices section.

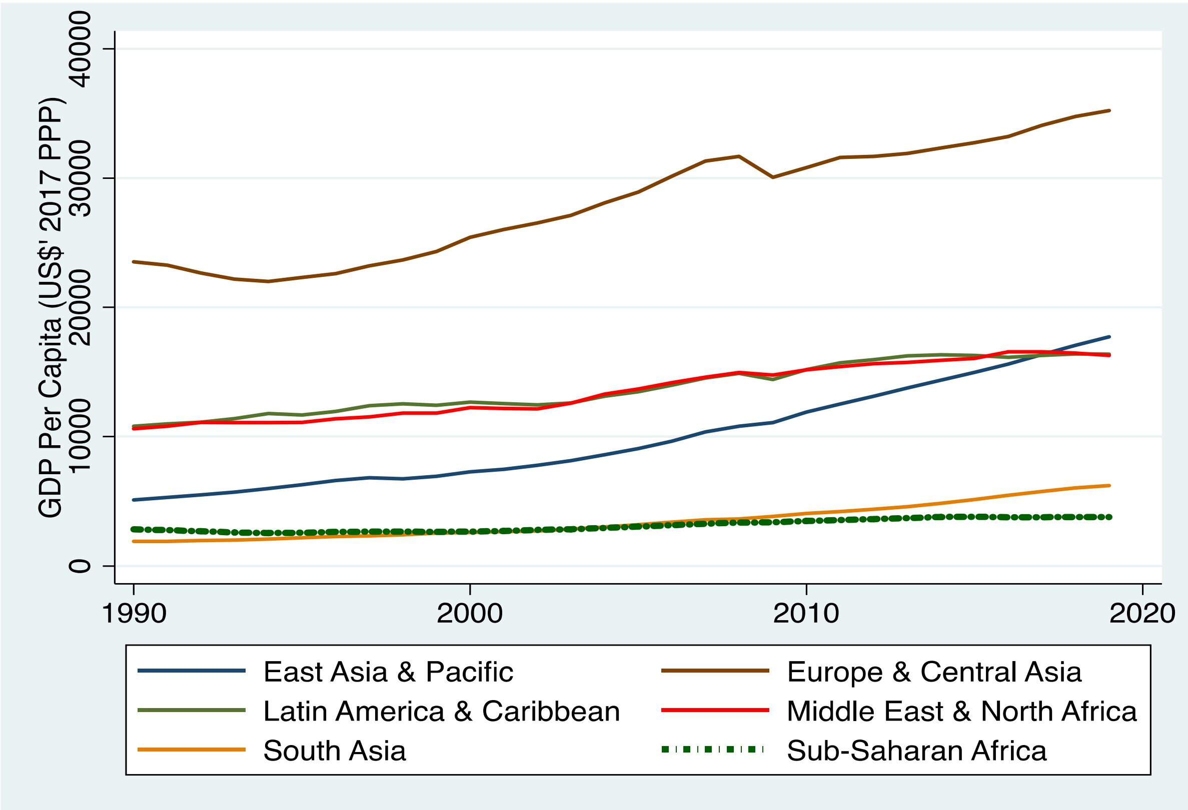

Fig.*1: Trend of GDP Per Capita Across Regions, 1990 – 2019*
Source: Authors' construct using data from World Development Indicators

2.1 Literature survey on drivers of economic growth using non-regularization techniques

In this section, we present a survey of the literature on the effect of several covariates included in this study on economic growth. Using a dataset on 21 SSA countries over period 2000-2014, Ngongang (2015) employed the dynamic panel GMM technique to examine the relationship between financial development and economic growth. The author finds a significant positive relationship between financial development and economic growth. In the same way, Ibrahim and Alagidede (2018) use panel data spanning 1980 – 2014 on 29 SSA countries to examine the conditional and unconditional effects of financial development in economic growth. The results suggest that while financial development has a positive impact on economic growth, the joint effect of financial development and investment are rather remarkable growth. Kodongo and Ojah (2016) also explore the link between infrastructure and economic growth and development in SSA countries. The results which are based on system GMM estimator and a dataset on 45 SSA countries for the period 2000 – 2011 shows that relative to middle income countries, infrastructure plays a salient role in least developed countries.

Omoteso and Mobolaji (2015) also apply the panel fixed effect, random effect and the maximum likelihood estimation techniques to test the linear relationship governance and economic growth in some selected SSA countries for the period 2002 to 2009. The authors find strong evidence to conclude that while political stability and regulatory quality enhance growth, government effectiveness adversely affect economic growth. Using a panel of 27 countries in SSA, Kebede and Takyi (2017) employed the panel causality and system GMM estimation techniques to examine the relationship between institutional quality and economic growth. While the authors find a unidirectional causality from economic growth to institutional quality, reverse does not hold. The results further show that institutional quality, trade openness, financial development, and debt positively affect economic growth.

In exploring the link between government expenditure and economic growth, Olaoye et al. (2018) apply the system GMM, and the Driscoll and Kraay nonparametric technique to examine the asymmetrical phenomenon in government spending and growth relationship in 15 ECOWAS countries for the period 2005 – 2017. Aside from confirming the asymmetric link between government spending and economic growth, the authors find evidence of an inverted U-shaped connection between government spending and economic growth. Also. Adams and Opoku (2015) investigate the effect of FDI on economic growth using a panel of 22 SSA for the period 1980 – 2011. The authors find evidence from the GMM estimation technique to show though unconditionally both FDI does not drive economic growth, the joint effect of FDI and regulations is positive and statistically significantly. Adams et al. (2016) also examine the link between energy consumption and economic growth, and the modulating role of democracy using a panel data of 16 SSA countries from 1971-2013. The study provides evidence from the panel vector autoregressive model to show that energy consumption enhances economic growth in the region. The study further finds that the joint effect of democracy and energy consumption on economic growth is positive and significant.

In addition, Adams and Klobodu (2016) assess the effect of remittances and regime change on economic growth for 33 SSA counties over the period 1970-2012. Their results from the system GMM estimation technique show while remittances do not significantly affect growth, regime change suppressing effect on growth. The study concludes that the growth-enhancing effect of remittances is amplified in the presence of a democratic and stable government. Appiah-Otoo and Song (2021) use a panel of 123 countries composed of 45 high-income countries, 58 middle-income countries, and 20 low-income countries for the period 2002-2017 to examine the impact of ICT on economic growth. The authors provide strong evidence to show that the effect of ICT diffusion on growth across rich and poor countries are significantly different and that poor countries tend to gain more from the ICT revolution. Employing a panel dataset on 20 African countries, Akadiri and Akadiri (2018) applied the fixed effect estimator to test the relationship between growth and income inequality on the one-hand, and the pathway through which growth determinants influences income inequality for the period 1991 to 2015. The study finds evidence of positive long-run relationship between income inequality and growth. The study further reveals that population growth, mortality rate, government consumption expenditure and foreign direct investment are principal determinants of the long-run growth and income inequality in the sampled countries.

In the same vein, Mavikela et al. (2019) examined the effect of inflation on economic growth for South Africa and Ghana with data over the period 2001 to 2016. Evidence from the quantile regression shows that while high inflation is positively related with growth in Ghana, it is the opposite in the case of South Africa. The study further shows an adverse effect of inflation at all threshold levels on growth in the post 2008/09 global financial crisis.

3. Data and methodology

3.1 Data

The dataset used for the analysis is a large balanced panel spanning 1980 – 2019. The study covers 42 SSA countries[6] on grounds of data availability. The outcome variable, economic growth, is the annual real GDP growth rate and is drawn from the World Development Indicators (World Bank, 2021b). Data on 113 potential drivers of growth are considered based on extant scholarship on economic growth. Taking into consideration the real sector of the economies under consideration, variables such as vulnerable employment, inflation and self-employment are considered (Bittencourt *et al.,* 2015; Barro 2103). Likewise, we include

[6]Angola, Benin, Botswana, Burkina Faso, Burundi, Cabo Verde, Cameroon, Central African Republic, Chad, Comoros, Congo, Dem. Rep., Congo, Rep., Cote d'Ivoire, Ethiopia, Gabon, Gambia (The), Ghana, Guinea, Guinea-Bissau, Kenya, Lesotho, Liberia, Madagascar, Malawi, Mali, Mauritania, Mauritius, Mozambique, Namibia, Niger, Nigeria, Rwanda, Sao Tome and Principe, Senegal, Seychelles, Sierra Leone, South Africa, Sudan, Tanzania, Togo, Uganda, Zambia.

variables such as globalisation index, trade openness, and tariff considering the rise in economic globalisation of SSA following the implementation of the AfCFTA and the projected rise in FDI inflow to the region in 2022 (UNCTAD, 2021; OECD/ACET, 2020). The essence of economic integration for growth in marginalised settings like SSA rests in the classical trade argument that it can foster social progress, and the contemporary view that trade is essential for innovation diffusion, technological transfer, global value chain participation and export diversification (Asongu & Odhiambo, 2019; Asongu & Nwchukwu, 2016; Sakyi *et al.*, 2015; Adams & Opoku., 2015).

Variables such as financial development and infrastructure are also considered due to their contribution to growth through resource allocation and the facilitation of economic activities (Koomson *et al.*, 2020; Opoku *et al.*, 2019; Peprah *et al.*, 2019, African Development Bank, 2018). We source data on financial development from the World Bank's Global Financial Development Database (Čihák *et al.*,2013), and the International Monetary Fund's financial development index (Svirydzenka 2016). Also, the study includes welfare variables of poverty and inequality due to their prevalence in the selected countries despite gains chalked in recent years and the fact that such developments waste human capital, consequently dragging growth down.

Data on poverty and inequality are sourced from the World Bank's Poverty and Equity Database, and the Global Consumption and Poverty Project (Lahoti *et al.*, 2016), while that of globalisation[7] is drawn from the Konjunkturforschungsstelle (KOF) index (Gygli *et al.*, 2019). Per empirical evidence on the contribution of institutions and policy to growth, we consider country policy and institutional scores on macroeconomic management, trade policy, social protection, social inclusion, and financial sector management (Asongu & Nwachukwu, 2017; Akobeng, 2016; Asongu & Gupta, 2015; Fosu 2012; Anyanwu, 2003). Also, we consider ICT skills, access, and usage given the momentous rise in the digital infrastructure of the region (Appiah-Otoo & Song, 2021; Tchamyou *et al.*, 2019; Adeleye *et al.*, 2019). The variable definitions and sources are reported in Table A1 in the Appendices section.

3.2 Estimation strategy

The empirical focus of this paper is in two parts. The first part is dedicated to the specification of the regularization and variable selection techniques while the inferential

[7] The KOF globalisation index is a cross-country composite index comprising key dimensions of global interrelationships (i.e., economic, social, trade, financial, culture and political)

models are presented in the second part. In line with the objectives of the study, we do not employ traditional panel data estimation techniques for the analysis. For instance, the panel least squares estimation technique is inappropriate as it cannot explicitly perform variable selection out of the 113-potential drivers of growth. Second, traditional estimation techniques such as the panel corrected standard errors and generalised method of moment cannot be relied upon as the presence of more predictors can cause the required matrix ($X'X$) to be invertible. Even if it is possible, the presence of too many covariates underpinning this study may cause overfitting. Overfitting is the inclusion of extra parameters the attendant estimates are not biased, they are less efficient[8] (James *et al.,* 2013). This is due to the fact that as the covariates become large, least squares assumptions of no multicollinearity, homoscedasticity and exogeneity typically breakdown therefore overfitting the model. This causes the out of sample error to increase, making inference and predictions flawed (James *et al.*, 2013).

Addressing this econometric concern can be through the application of machine learning regularization techniques, which are effective for variable selection regardless of the number of covariates, model specification, nonlinearity and time (Tibshirani, 1996). In this study, therefore, we train recent machine learning regularization algorithms to learn patterns in the underlying dataset to yield sparse drivers of economic growth. Regularization is done by utilising the bias-variance trade-off, where a tuning parameter (i.e., the bias) is introduced to reduce the variance associated with large datasets and consequently yield sparse estimates. In specifics, we train four alternative shrinkage models— the first three from the lasso family (i.e., the Standard lasso, the Minimum Schwarz Bayesian information criterion lasso, and Adaptive lasso), and the Elasticnet to achieve the first objective[9]. Next, we perform causal inference on the selected covariates in Objective 1 by running the lasso inferential models of double-selection linear lasso, partialing-out lasso linear regression, and partialing-out lasso instrumental variable regression to address Objective 2. To this end, the STATA (Version 16) and R (Version 3.6) software are employed. The latter is employed primarily for data engineering and descriptive purposes while the data partitioning, regularization and inferential estimates are carried out using the former.

[8] Inefficiency due to model complexity, specification problems and/ or overfitting. Further, the traditional least squares estimator is not only less sparse but also, more susceptible and sensitive to problems like multicollinearity and outliers.

[9] Since the ordinary least squares technique and Ridge regression cannot yield variable selection, their estimations are relaxed

3.2.1 Specification of regularization models

3.2.1.1 Specification of Standard lasso and Minimum BIC lasso models

To address the ineffectiveness of traditional regression techniques in variable selection in large datasets, Tibshirani (1996) introduced the Standard lasso. Like other shrinkage techniques, the main advantages of the Standard lasso are that it: (1) enhances the model interpretability by eliminating irrelevant variables that are not associated with the response variable; (2) enhances prediction accuracy, because shrinking and removing irrelevant predictors can reduce variance without a substantial increase in the bias; and (3) is limitless to data dimensionality.

In line with Objective 1 of this study, the *Standard lasso* is applied to select the key drivers of economic growth by penalising the model coefficients through a tuning parameter (λ) (Tibshirani, 1996; Belloni & Chernozhukov, 2013). Following Tibshiran (1996), we specify the objective function of the Standard lasso as shown in Equation (1). For the Standard lasso algorithms to detect the key predicators of economic growth from a pool of several possible predictors, the penalty ($\lambda \sum_{j=1}^{\rho} |\beta_j|$), also referred to ℓ_1-norm, is introduced to obtain $\widehat{\beta}_{lasso}$ defined in Equation (2)

$$Q_L = \frac{1}{N} \sum_{i=1}^{N} \omega_i \, f(y_{it}, \beta_0 + X_{it}\beta') + \lambda \sum_{j=1}^{p} k_j \, |\beta_j| \quad (1)$$

$$\hat{\beta}_{lasso} = min\left\{SSE + \lambda \sum_{j=1}^{\rho} |\beta_j|\right\}, \quad (2)$$

where y_{it} is economic growth in country ***i*** in year ***t***, X_{it} is a vector of all possible predictors of economic growth. The objective, therefore, is the minimisation of the model sum of square errors with a given ℓ_1-norm. It is imperative to point out that if the tuning parameter, $\lambda = 0$, then we have a full model as in the least square estimator, while $\lambda \to \infty$ is an intercept-only model. For brevity, we indicate that the specification of the *Minimum BIC* lasso follows that of the *Standard lasso* with the same penalty and objective function but subset selection is based on the model with the least BIC (Schwarz, 1978). Some known drawbacks of the Standard lasso and Minimum BIC lasso techniques are that they: (1) may become inconsistent as features grow rapidly, and (2) are unable to perform hypothesis tests.

3.2.1.2 Specification of Adaptive lasso model

To enhance the consistency of regularization, Zou (2006) introduced the Adaptive lasso technique, which in addition to the ℓ_1-norm penalty, adds the '*oracle property'* (z_j). Relative to the Standard lasso, the '*oracle property*' enhances shrinkage or subset selection even when data attributes grow faster than the number of observations. In this study, we employ the Adaptive lasso technique as an alternative to the Standard lasso and Minimum BIC lasso in addressing Objective 1. Following Zou (2006), we minimise the objective function in (3) by applying the Adaptive lasso estimator ($\hat{\beta}_{AdaptiveLasso}$) specified in Equation (4),

$$Q_L = \frac{1}{N}\sum_{i=1}^{N} \omega_i \, f(y_{it}, \beta_0 + X_{it}\beta') + \lambda \sum_{j=1}^{p} k_j \, |\beta_j| \quad (3)$$

$$\hat{\beta}_{AdaptiveLasso} = min\left\{SSE + \lambda \sum_{j=1}^{\rho} z_j |\beta_j|\right\} \quad (4)$$

Where y_{it} is the outcome variable (economic growth) in country ***i*** in year ***t***, X_{it} is a vector of all 113 covariates of economic growth and β' are the attendant parameters.

3.2.1.3 Specification of Elasticnet model

The Elasticnet method draws on the strengths of the Standard lasso and Ridge regression by applying the ℓ_1 and ℓ_2 penalisation norms. The strength of the Elasticnet is that in highly correlated covariates, it can produce sparse and consistent regularization than the lasso family algorithms (Zou & Hastie, 2005). Also, with the application of the ℓ_1 and ℓ_2 penalization norms, the Elasticnet becomes flexible in subset selection. To perform variable selection, the Elasticnet estimator minimizes the objective function:

$$Q_{en} = \frac{1}{N}\sum_{i=1}^{N} \omega_i \, f(y_{it}, \beta_0 + X_i\beta') + \lambda \sum_{j=1}^{p} k_j \left\{\frac{1-\alpha}{2}\beta_j^2 + |\beta_j|\right\} \quad (5)$$

Where y_{it}, X_i, and β' are as defined in previous sections, and α is an additional Elasticnet penalty parameter[10], which takes on values only in [0,1]. That is, sparsity occurs when $0 < \alpha < 1$ and $\lambda > 0$. This implies that in special cases, the Elasticnet plunges into either the Ridge estimator (i.e., when λ=0) or the Standard lasso estimator (i.e., when λ=1).

3.2.2 Choice of tuning parameter

[10] This adds to the regular λ penalty.

A fundamental concern regarding variable selection is the choice of the tunning parameter (**λ**). A good value of **λ** is essential for the overall performance of regularization models as it controls the strength of shrinkage and the concomitant prediction and inference (Schneider and Wagner, 2012). Among the widely used methods for choosing an efficient **λ** are Cross validation (CV), Bayesian Information Criterion (BIC), and Akaike Information Criterion (AIC) (Tibshirani and Taylor, 2012). But it needs to be pointed out that if regularization becomes too strong, relevant variables may be omitted and coefficients may be shrunk excessively. Therefore, information criteria such as the BIC and AIC might be preferable to CV, since they are faster to compute and are less volatile in small samples (Zou *et al.*, 2007). However, to the extent that setting **λ** under a researcher's discretion can yield 'target sparsity' and harm both predictive capacity and inferencing (Hastie *et al.*, 2019), we rely on both information criterion and CV[11]in determining **λ**.

3.2.3 Specification of lasso inferential models

Since the aforementioned variable selection techniques do not provide estimates and confidence intervals essential for inference[12], we apply the lasso inferential techniques to provide robust estimates on the selected predictors of economic growth. In specifics, we run the double-selection lasso linear model (DSL), the partialing-out lasso linear regression (POLR), and the partialing-out lasso instrumental-variables regression (POIVLR) using the selected covariates in Objective 1 as the variables of interest, and all the unselected (redundant) variables as controls.

It is imperative to note that due to the sparsity of the regularization techniques, the control variables are usually many. In view of this, the lasso inferential models consider these controls as irrelevant and therefore, their inferential statistics are not reported. However, the number of relevant controls and instruments are indicated as part of the general regression statistics. Further, unlike the variables of interest, which the researcher has no flexibility of adding or excluding from model, the researcher can indicate the number of controls in the model[13].The strength of these models is that they are built to produce unbiased and efficient estimates irrespective of data dimensionality, model specification, and multicollinearity.

[11] In this study, we invoke the 10-fold cross-validation.

[12] Traditional estimation techniques such as the OLS cannot be employed either as the new variability introduced in the dataset by the regularization techniques are not captured by such techniques.

[13] We include 56 out of the remaining 106 covariates as control against the backdrop that several alternative measures of globalisation, institutional quality and welfare are used.

3.2.3.1 Double-selection lasso linear model

In line with the second objective, we follow Belloni *et al.* (2014) by specifying the double-selection lasso (DSL) linear model as:

$$E[Y|d,x] = \psi\alpha' + \phi\beta', \tag{6}$$

where $\mathbf{y}$ is economic growth and is modelled to depend on $\boldsymbol{\psi}$, containing $\boldsymbol{J}$ covariates of interest (i.e., the elasticnet or lasso selected key drivers of economic growth) and $\boldsymbol{\phi}$, which contains $\boldsymbol{p}$ controls (i.e., the weak drivers of economic growth). The DSL estimator produces estimates on $\boldsymbol{J}$ while relaxing the estimates for $\boldsymbol{p}$.

3.2.3.2 Partialing-out lasso linear regression

In reference to the DSL, an added advantage of the partialing-out lasso linear regression (POLR) is that it enhances the efficacy of inference as the model becomes too complex. Following Belloni *et al.* (2012) and Chernozhukov *et al.* (2015), we specify the POLR estimator as:

$$E[Y|d,x] = d\alpha' + X\beta', \tag{7}$$

Where $\boldsymbol{y}$ is outcome variable (economic growth), $\boldsymbol{d}$ is a vector containing the $\boldsymbol{J}$ predictors of interest (i.e., the non-zero selected covariates of economic growth), and $\boldsymbol{X}$ contains the $\boldsymbol{p}$ controls (i.e., the unselected predictors of economic growth). Like the DSL, the POLR yields estimates, standard errors and confidence intervals on the $\boldsymbol{J}$ covariates while relaxing that of the $\boldsymbol{p}$ controls.

3.2.3.3 Partialing-out lasso instrumental-variables regression

In large data regression problems like this study, sources of endogeneity abound largely due to bi-causality. For example, endogeneity can arise from the argument espoused in the supply-leading and demand-following hypotheses of financial development and economic growth (King & Levine, 1993). To address this, we follow Chernozhukov *et al*. (2015) by performing a partialing-out lasso instrumental variable regression (POIVLR). The POIVLR is specified as:

$$y = \Psi\alpha_d' + \Phi\alpha_f' + X\beta' + \varepsilon, \tag{8}$$

where y is economic growth; $\boldsymbol{\Psi}$ comprises $\boldsymbol{J_d}$ endogenous covariates of interest; $\boldsymbol{f}$ contains the $\boldsymbol{J_f}$ exogenous covariates of interest; and $\boldsymbol{X}$ contains $\boldsymbol{\wp_x}$ controls. Allowing for potential endogeneity primarily due to simultaneity, $\boldsymbol{\wp_z}$outside instrumental variables denoted by $\boldsymbol{z}$ that are correlated with $\boldsymbol{d}$ but not with $\boldsymbol{\varepsilon}$are introduced. As aforesaid, the simultaneity between financial development and economic growth presents endogeneity concerns which is addressed using the $\boldsymbol{z}$ instruments[14].Theoretically, the controls and instrument can grow with the sample size; however, $\boldsymbol{\beta}$ and non-zero coefficients in $\boldsymbol{z}$ must be sparse.

3.4 Data engineering and partitioning

One of the key requirements of effective regularization is that the underlying dataset is strongly balanced. To this end, we employ the *K-Nearest Neighbour* (KNN) data imputation technique to address missing observations, particularly for variables such as the policy and institutional indicators (see, **Fig.**1.3). The KNN follows the principle that developments regarding variables drawn from a similar population exhibit similar properties (Van Hulse and Khoshgoftaar 2014). In principle, the KNN selects the nearby neighbours based on a distance metric, and estimates the missing observation with the attendant mean or mode. It is worth noting that while the mean rule is used to address missing observations in numerical variables, the latter is employed to address missing observations in categorical variables (Pan *et al.* 2015). Per this principle, this study relies on the mean rule, which uses the Minkowski distance as specified in equation (9) in addressing the missing observations.

$$d(i,j) = (|x_{i1} - x_{j1}|^{q} + |x_{i2} - x_{j2}|^{q} + \cdots + |x_{i\rho} - x_{j\rho}|)^{q^{1/q}}, \tag{9}$$

where q is the Minkowski coefficient, $d(i,j)$ is the Minkowski distance for observations i and j, and x are the variables. That said, we take cues from Ofori et al. (2022) by partitioning the dataset into two parts— the training set (70%) and testing test (30%) samples. We do this by applying the BASE R and Stratified data splitting techniques. In line with Ofori et al. (2022), we take cues from James et al. (2013) that among all other possible sets, the 70-30,

[14] List of instruments in POIVLR: transparency score, trade score, public management score, macroeconomic management score, gender equality score, financial sector management score, internet access (per 1 million of the population), mobile cellular subscription (per 100 of the population), fixed telephone subscription (per 100 of the population), fixed broadband subscription (per 100 of the population).

and 80-20 splits are the data partitioning sets allowing reasonable representation of all variables in both the training and testing samples.

4. Presentation and discussion of results

4.1 Exploratory data analysis

For brevity, the exploratory data analysis is limited to the data partitioning results[15], the distribution of economic growth, and the summary statistics. Information gleaned from the summary statistics in Table A2[16] shows an average economic growth (i.e., real GDP growth rate) value of 3.58 per cent in the training set as compared to 3.95 per cent in the testing set. Also, the average trade openness value as a percentage of GDP is 67.48 in the training set compared to 66.85 per cent in the testing set. Additionally, we observe a mean unemployment rate of 7.58 per cent in the training set compared to 7.68 per cent in the testing set. It is also evident from Table A2 in the Appendices section that the average transparency, accountability, and corruption score of 2.81 and 2.8 in the training and testing sets respectively. Finally, **Fig.**1.2 shows that 99.9 per cent observations were present in the dataset before the data imputation.

4.1.1 Data partitioning and distribution of economic growth results

A major decision regarding regularization is the form the outcome variable takes— either level or log transformed. On the latter, the distribution of economic growth as we show in Fig.2 (right) is right-skewed. However, at level, as shown in Fig.2 (left), economic growth is more symmetric and less heavy-tailed. At the backdrop that skewed distribution can have dire implications for regularization and the attendant inferential statistics, we run our shrinkage models using economic growth at level. Further, though non-standardization of covariates of economic growth does not constrain regularization, it is essential for ensuring the internal consistency of the data and comparability of the covariates. In view of this, the standardise option is invoked.

[15] That is the distribution of economic growth in the training and testing sets.
[16] See appendix section.

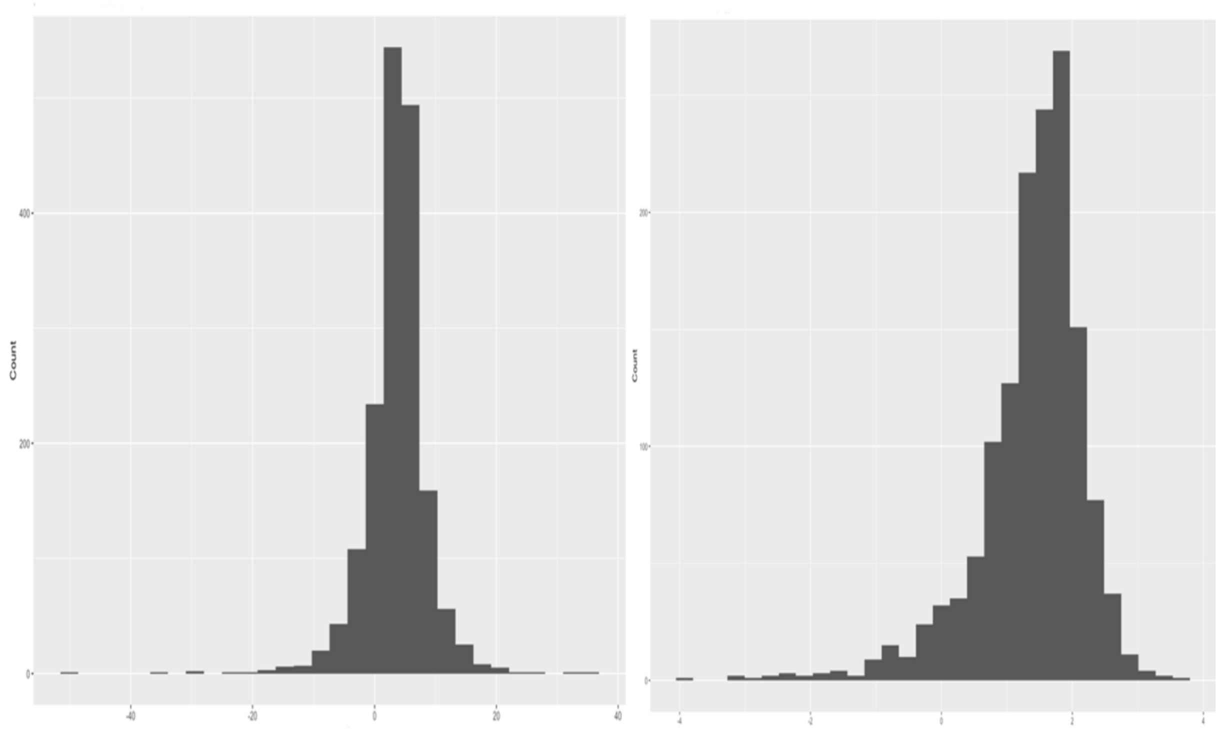

Fig.2*: Distribution of economic growth at level (left) and its log-transformation (right)*

On data partitioning, we perform a 70-30 split of the dataset using the *Stratified method* (see, Fig.3 (left)). Additionally, in checking the reliability or consistency of the stratified split, we run the *Rsample* technique, which yields similar results (Fig.3 (right)).

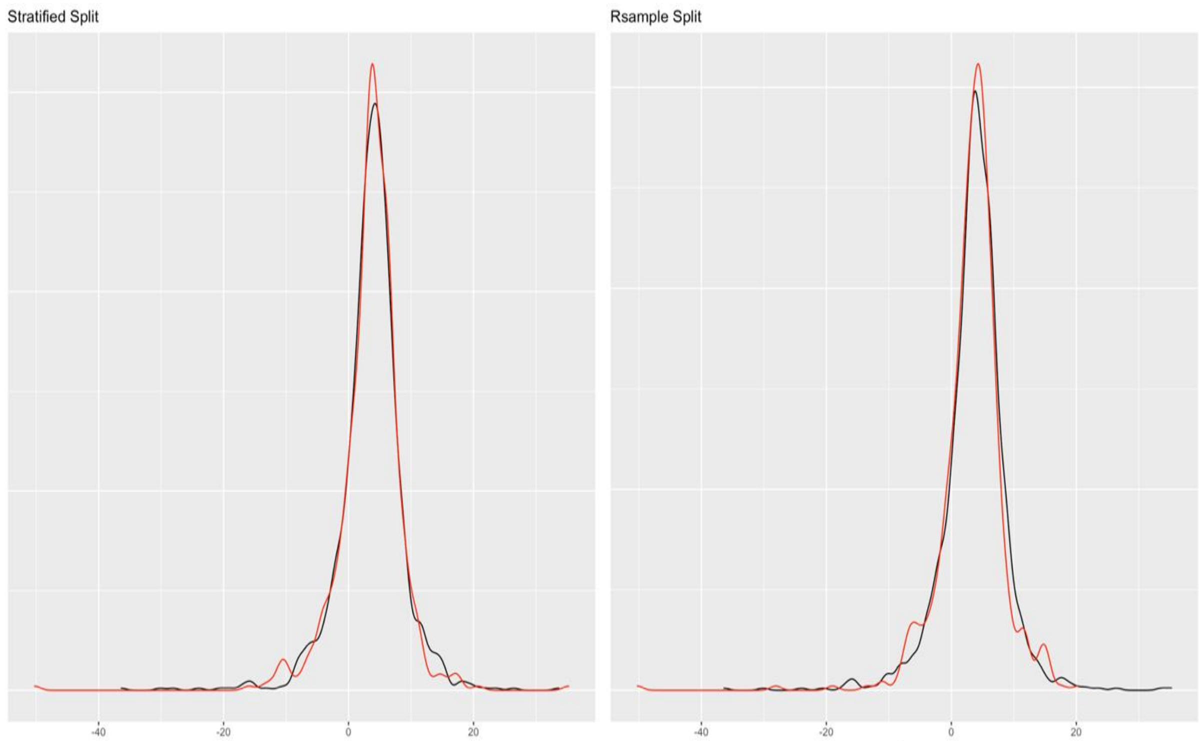

Fig. 3*: Data partitioning plot (Base R method), Training (Black) and Test (Red)*

4.2 Regularization results on drivers of economic growth in Sub-Saharan Africa

In this section, results on the first objective are presented. As we show from Figures 4 – 7, the lassos and Elasticnet algorithms select different non-zero coefficients (i.e., predictors) of economic growth. We find that the *Standard lasso* selects 12 covariates as key drivers of economic growth under a 10-fold cross-validation tuning parameter (λ) value of 0.33 (see Fig.4). Further, the A*daptive lasso* selects only 10 covariates from the total 113 as chief drivers of economic growth in SSA with a tuning parameter (λ) value of 0.24. Similarly, we find a special case for the *Elasticnet* regularization as it selects covariates based on a minimum cross-validation lambda of 0.33, and a minimum cross-validation alpha of 1. While the *Elasticnet* plunges into the *Standard lasso* (i.e., selects 12 non-zero predictors), we find a sparser regularization in the *Minimum BIC lasso* as it selects only 7 covariates of the total 113.

In Table A3 in the Appendices section, a detailed output of how covariates enter and leave the respective shrinkage models is presented. The results from the Minimum BIC lasso, which yields the best regularization indicates that the key drivers of economic growth in SSA are: *manufacturing (value addition), population, financial development, government spending, macroeconomic management, globalisation, and social inclusion.* The appropriateness of the results is evident in the postestimation tests of cross-validation and coefficient path plots associated with each model (Figures 4 – 7).

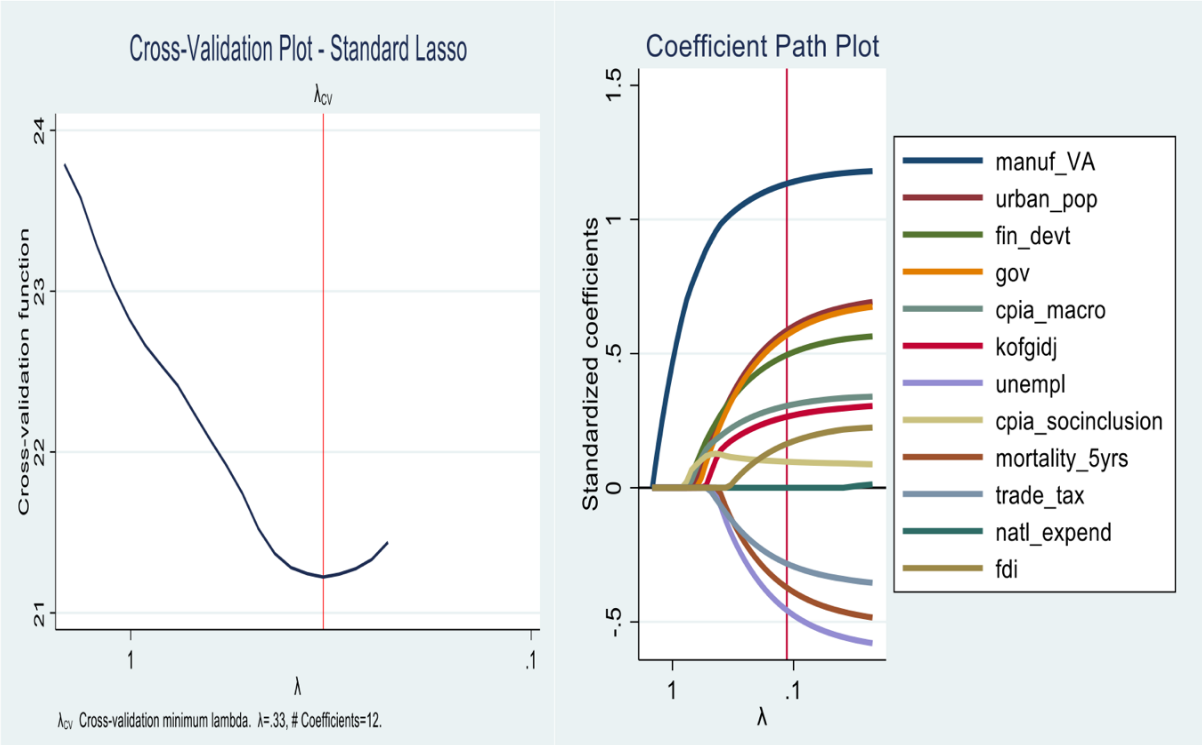

Fig. 4*: Cross-validation plot (left) and coefficient path plot (right) for Standard lasso*

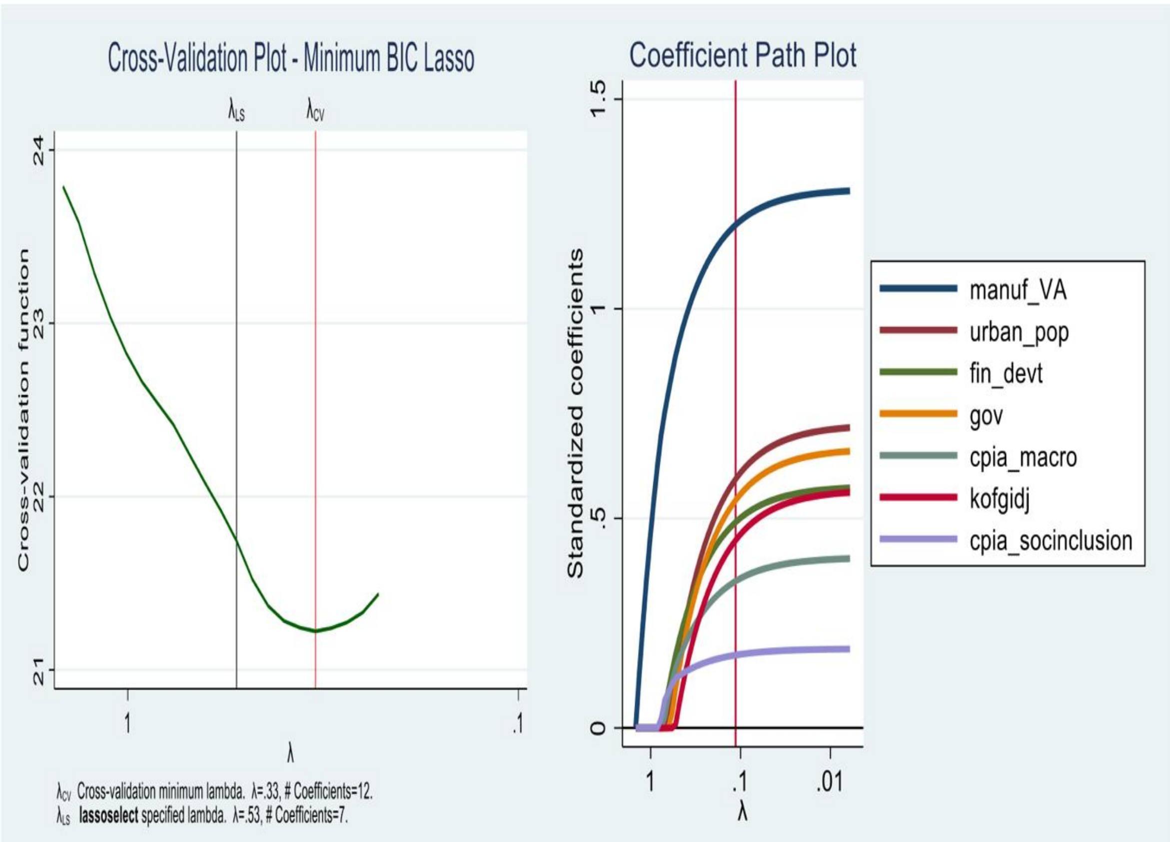

Fig. 5*: Cross-validation plot (left), and coefficient path plot (right) for Minimum BIC lasso*

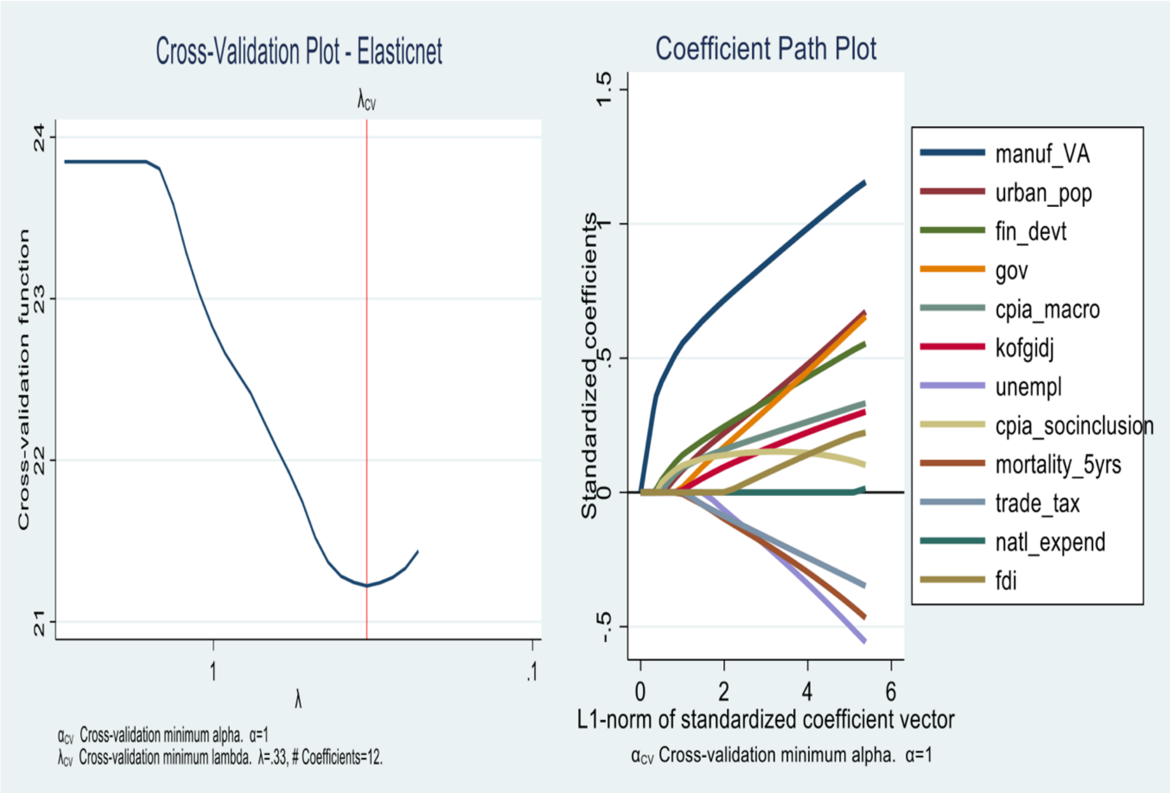

Fig. 6: *Cross-validation plot (left), and coefficient path plot (right) for Elasticnet*

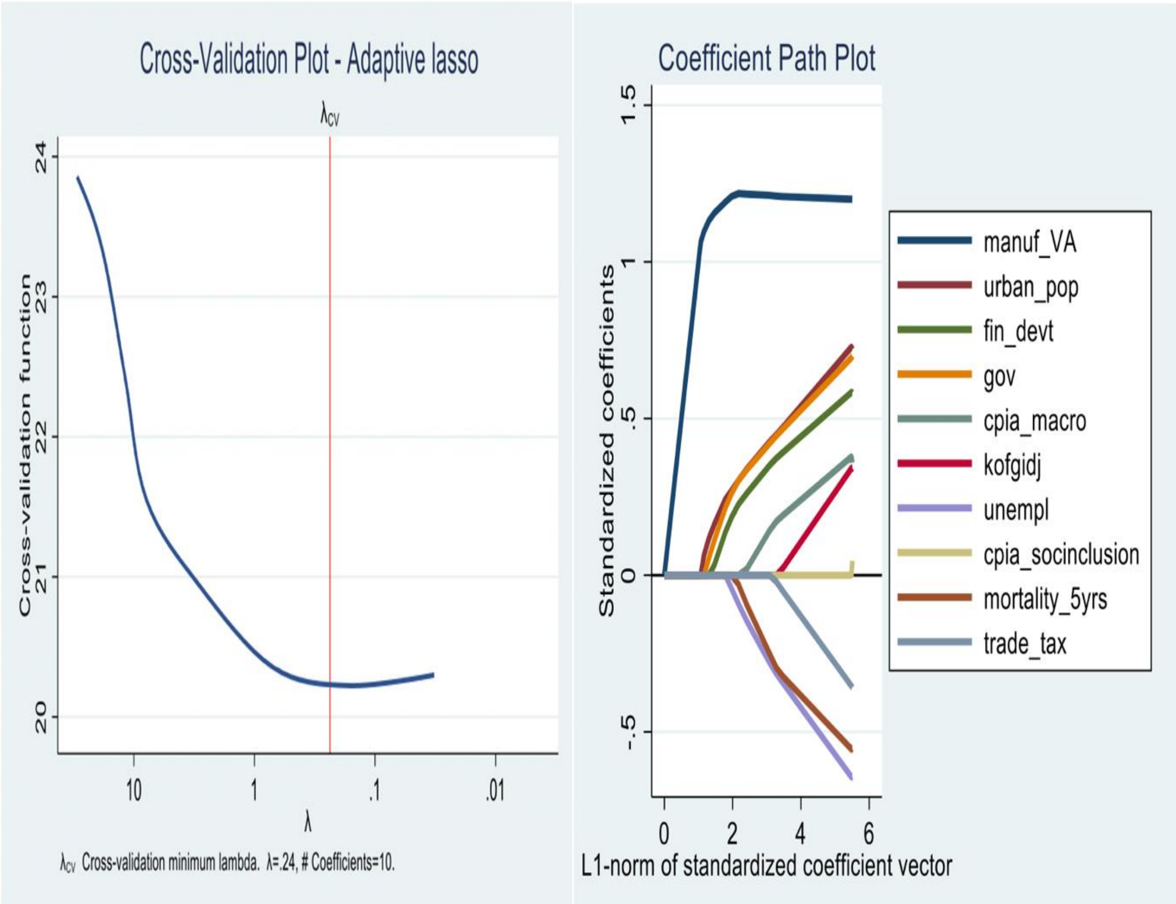

Fig. 7: *Cross-validation plot (left), and coefficient path plot (right) for Adaptive lasso*

4.3 Inferential results on drivers of economic growth in Sub-Saharan Africa

In this section, estimates on the 7 covariates of growth identified in Objective 1 are provided. The results, which are reported in Table 1 are based on the DSL, POLR and POIVLR estimation techniques, meaning that they are robust to heteroskedasticity, endogeneity, and model misspecification. To inform policy appropriately, we run three separate results for the: (i) full sample, (ii) low-income countries and (iii) middle- and high-income countries.

Table 1: Lasso estimates on key drivers of economic growth in Sub-Saharan Africa

	All Countries			Low Income	Countries		Middle and	High Income	Countries
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Variables	**DSL**	**POLR**	**POIVLR**	**DSL**	**POLR**	**POIVLR**	**DSL**	**POLR**	**POIVLR**
Manufacturing Value Addition	0.062*	0.066*	0.066*	0.125***	0.114***	0.114***	0.008	0.010	0.010
	(0.034)	(0.034)	(0.034)	(0.028)	(0.024)	(0.024)	(0.007)	(0.008)	(0.008)
Population (urban)	0.273***	0.309***	0.255**	0.318**	0.349**	0.349**	0.110	0.152	0.152
	(0.089)	(0.104)	(0.114)	(0.155)	(0.170)	(0.170)	(0.300)	(0.264)	(0.264)
Financial development	0.142***	0.133***	0.144***	0.165***	0.142***	0.142***	0.040	0.047	0.047
	(0.031)	(0.027)	(0.030)	(0.047)	(0.045)	(0.045)	(0.034)	(0.030)	(0.030)
Government expenditure	0.032***	0.034***	0.065**	0.058***	0.059***	0.059***	0.036*	0.023	0.023
	(0.011)	(0.011)	(0.029)	(0.020)	(0.018)	(0.018)	(0.020)	(0.022)	(0.022)
Macroeconomic management	0.757***	0.654**	0.737***	1.493***	0.825**	0.825**	0.885*	0.602	0.602
	(0.274)	(0.261)	(0.269)	(0.511)	(0.371)	(0.371)	(0.497)	(0.438)	(0.438)
Globalisation	0.090**	0.078*	0.063	0.002	0.019	0.019	-0.002	0.036***	0.036***
	(0.041)	(0.041)	(0.042)	(0.031)	(0.028)	(0.028)	(0.023)	(0.013)	(0.013)
Social inclusion	3.308***	2.497***	3.007***	-1.502	-1.017	-1.017	3.231**	1.352	1.352
	(0.904)	(0.825)	(1.006)	(1.447)	(1.222)	(1.222)	(1.425)	(1.045)	(1.045)
Observations	1,720	1,720	1,720	798	798	798	350	350	350
Variables of Interest	7	7	7	7	7	7	7	7	7
Controls	56	56	56	66	66	66	66	66	66
Controls selected	39	39	46	25	25	25	27	27	27
Instruments	–	–	12	–	–	7	–	–	7
Wald Statistics	89.43***	81.72***	78.11***	38.22***	49.72***	49.72***	16.28**	24.29***	24.29***
Wald P-value	0.000	0.000	0.000	0.000	0.000	0.000	0.022	0.001	0.001
Countries	42	42	42	20	20	20	22	22	22

Note: Double selection lasso; POLR is Partialing out linear lasso regression; POIVLR is Partialing out instrumental variable linear regression
Robust standard errors in parentheses
**** p<0.01, ** p<0.05, * p<0.1*

To begin with, we find that manufacturing value addition matters for economic growth in SSA. The results show that a 1 per cent increase in manufacturing value addition boosts economic growth by 0.06 per cent. Across the low-income and middle-income divide, however, we find that manufacturing value addition is significant only in the case of the former. The evidence suggests that with appropriate economic governance, it is low-income countries that can make remarkable strides in economic growth through enhanced manufacturing value addition considering the implementation of the AfCFTA. This is more so as improvement in manufacturing can spur forward and backward linkages as well as global value chain participation.

Further, the results show that though financial development is directly related to economic growth in both low-income and middle-income countries, it is statistically significant only in the former. In terms of magnitudes, the results suggest that for every 1-point increase in financial development, economic growth rises by 0.14 per cent in low-income countries. The effect of financial development is remarkable, suggesting that access to financial products and services can propel the huge informal sector of low-income countries realised their innovative and entrepreneurial objectives. This is more so considering the fact lags in financial access are glaring in the low-income countries compared to middle-income countries.

Additionally, we find that economic globalisation drives economic growth in SSA. In the remit of low-income and middle-income countries, however, we find that economic integration matters for growth only in the case of the latter. The plausible explanation for this tis that, relative to low-income countries, middle-income countries have made remarkable strides in developing their manufacturing base, coupled with a good absorptive capacity that can enable them to gain significantly from economic globalisation. Albeit statistically insignificant, the positive relationship between growth and economic optimism also suggests that with the implementation of the AfCFTA and the expected of FDI to Africa from 2022, low-income countries can gain from economic integration.

The result on economic globalisation is linked to the remarkable finding on macroeconomic management. There is strong empirical evidence to show that every 1-point increase in the score of macroeconomic management of the SSA boosts economic growth by 0.73 per cent (Column 3). This result is even strong (i.e., 0.82%) in the case of low-income countries (Column 6). Indeed, one of the major problems of the region has been poor macroeconomic management often resulting in bailouts by foreign institutions[17]. Though these bailouts have proved effective in propelling beneficiary countries toward prudent macroeconomic management paths, gains are mostly disrupted following exist, signifying the need for sustained commitment to fiscal and monetary discipline.

Also, we find that government expenditure is instrumental for economic growth in SSA. The result shows that a 1 per cent increase in government expenditure boosts economic growth by 0.06 per cent. However, this evidence is only significant in low-income SSA countries. A possible explanation for this is that in middle-income countries, a high percentage of government expenditure goes into the recurrent expenditure compared to capital expenditure.

Moreover, we find that urban population matters for economic growth in the SSA. Additionally, the result reveal that urbanisation is effective in fostering growth in the low-income countries compared to middle-income countries. This evidence appeals to logic in that economic activities driving growth in low-income SSA countries are mostly concentrated in urban centres. The result is in line with a World Bank (2009) report which argues that urban concentration is crucial in fostering growth in economies at the early stages of development. There is also the supporting evidence of gains from urbanization in that it reduces poverty and inequalities in opportunities, services and assets (see, Sekkat, 2017), and inequality (see, Oyvat, 2016).

Also, we provide strong empirical evidence to show that improving the coverage of social inclusion polices boosts economic growth in SSA by 3 per cent (Column 3). The result suggests that rolling out social interventionary programs can propel SSA countries towards sustained growth trajectories. This is more so as social inclusion policies can build private sector capacity to withstand socioeconomic shocks. This is however not effective for growth in the low-income countries. This is also possible since institutions for developing human capital in these settings are weak thereby providing little or no growth gains for such expenditure.

5. Conclusion and policy implications

The study contributes to the economic growth literature on SSA by employing recent advances in machine learning to identify the key drivers of growth. In doing so, we train four machine learning regularization models— the Standard lasso, the Minimum BIC lasso, Adaptive lasso, and Elasticnet based on a dataset spanning 1980 – 2019 for 42 African

[17] For example, the case of the IMF and Ghana in 2015/16, and the IMF and South Africa in 2021.

countries for the analysis. Our results show that machine learning techniques are powerful and effective in reducing model complexities associated with large-data time series regression problems. In this study, while both the Standard lasso and Elasticnet techniques select 12 covariates as key determinants of economic growth, the Minimum BIC lasso selects only 7 out of the total 113 possible predictors. The uniqueness of the study is that it presents policymakers interested in the SSA growth agenda, variables to target to foster and sustain growth. These variables are: manufacturing (value addition), urban population, financial development, government spending, macroeconomic management, economic globalisation, and social inclusion. Further, we find that only economic globalisation and social matter for growth in middle-income countries. Policy recommendations are therefore provided in this regard.

For middle-income SSA countries, the suggest the following recommendations. First, in line with the implementation of the AfCFTA and the green growth agenda of the SSA, it is recommended that policymakers invest strategically in the manufacturing sectors of their economies. This can prove crucial in turning around the slow growth trajectories of the region as economic globalisation can spur the industrial drive of the region through forward and backward, innovation diffusion and global value chain participation. Policymakers are therefore advised to build the technical workforce of their economies are to make sense of the knowledge and innovation transfers associated with economic integration. Second, to improve the ability, opportunities, and dignity of the marginalised to contribute meaningfully to national development, policymakers are to invest strategically in areas such as health, education, and vocational training. This is more so as ICT diffusion can reduce inequalities in accessing information, and high cost of accessing opportunities due to polarisation of administrative procedures in the SSA.

For low-income countries, efforts should be made to develop the financial sector. This could prove crucial for efficient resource allocation, which can be a gamechanger in spurring the industrialisation drive of the region thorough competition, innovation, dynamism, and enhanced global value chain participation. Resources should thus be channelled towards the development of payment system platforms and services, financial innovation, and information flow on consumers. In this regard, institutions interested in the SSA agenda such as the African Development Bank, IMF and World Bank should provide technical and logistical support to aid the transformation of the region’s predominantly low productive informal sector to a more dynamic, highly competitive and export-oriented one.

Additionally, we recommend that policymakers commit to prudent macroeconomic management. We reckon that in a setting like SSA where vulnerabilities are widespread, sound macroeconomic management will prove momentous in mitigating the welfare setbacks imposed by the coronavirus pandemic while lessening the impact of future economic and health crisis. This calls for need to channel resources into productive expenditure like infrastructure and energy supply, which could contribute to ensuring that economic globalisation propel these countries sustained growth trajectories.

The study leaves room for future works. First, considering the contributions this study makes through machine learning techniques, the academic community can also draw on similar techniques, for instance, to identify factors key for analysing poverty and inequality. Second, these techniques can be employed to examine whether the growth-globalization relationship we find differs between landlocked and non- landlocked countries. Finally, considering the green growth agenda of the continent, regularization techniques can be employed to determinant whether durable shared growth is driven largely by environmental factors or income growth and distributions.

References

Adams, S., & Akobeng, E. (2021). ICT, governance and inequality in Africa. *Telecommunications Policy*, *45*(10), 102198.

Adams, S., Klobodu, E. K. M., & Opoku, E. E. O. (2016). Energy consumption, political regime and economic growth in sub-Saharan Africa. *Energy Policy*, *96*, 36-44.

Adams, S., & Klobodu, E. K. M. (2016). Remittances, regime durability and economic growth in Sub-Saharan Africa (SSA). *Economic Analysis and Policy*, *50*, 1-8.

Adams, S., & Opoku, E. E. O. (2015). Foreign direct investment, regulations and growth in sub-Saharan Africa. *Economic Analysis and Policy*, *47*, 48-56.

Adeleye, B. N., & Eboagu, C. (2019). Evaluation of ICT development and economic growth in Africa. *Netnomics*, 1-25.

African Development Bank (2018) Africa's Infrastructure: Great Potential but Little Impact on Inclusive Growth, Africa Development Outlook.

Agbloyor, E. K., Abor, J. Y., Adjasi, C. K. D., & Yawson, A. (2014). Private capital flows and economic growth in Africa: The role of domestic financial markets. *Journal of International Financial Markets, Institutions and Money*, *30*, 137-152.

Aghion, P., & Howitt, P. (1990). A model of growth through creative destruction (No. w3223). National Bureau of Economic Research.

Akbari, A., Ng, L., & Solnik, B. (2021). Drivers of economic and financial integration: A machine learning approach. *Journal of Empirical Finance*, *61*, 82-102.

Akobeng, E. (2016). Growth and institutions: A potential medicine for the poor in Sub-Saharan Africa. *African Development Review*, *28*(1), 1-17.

Alagidede, P., & Ibrahim, M. (2017). On the causes and effects of exchange rate volatility on economic growth: Evidence from Ghana. *Journal of African Business*, *18*(2), 169-193.

Anyanwu, J. C. (2014). Factors affecting economic growth in Africa: are there any lessons from China? *African Development Review*, *26*(3), 468-493.

Anyanwu, J. C. (2003). Estimating the macroeconomic effects of monetary unions: the case of trade and output. *African Development Review*, *15*(2-3), 126-145.

Appiah-Otoo, I., & Song, N. (2021). The impact of ICT on economic growth-Comparing rich and poor countries. *Telecommunications Policy*, *45*(2), 102082.

Asongu, S., & Gupta, R. (2015). Trust and quality of growth: a note. *Economics Bulletin*, *36*(3), A181.

Asongu, S. A., & Nwachukwu, J. C. (2017). Quality of growth empirics: comparative gaps, benchmarking and policy syndromes. *Journal of Policy Modeling, 39(5)*, 861-882.

Asongu, S. A., & Nwachukwu, J. C. (2016). Welfare spending and quality of growth in developing countries: A note on evidence from Hopefuls, Contenders and Best Performers. *The Social Science Journal*, *53*(4), 495-500.

Asongu, S. A., & Odhiambo, N. M. (2019). Mobile banking usage, quality of growth, inequality and poverty in developing countries. *Information Development*, *35*(2), 303-318.

Barro, R. J. (2013). Inflation and economic growth. Annals of Economics & Finance, 14(1).

Bazarbash, M. (2019) Fintech in Financial Inclusion: Machine Learning Applications in Assessing Credit Risk (May 2019). IMF Working Paper No. 19/109

Beam, A. L., & Kohane, I. S. (2018). Big data and machine learning in health care. *Jama*, *319*(13), 1317-1318.

Belloni, A., Chernozhukov, V., & Hansen, C. (2014). High-dimensional methods and inference on structural and treatment effects. *Journal of Economic Perspectives*, *28*(2), 29-50.

Belloni, A., D. Chen, V. Chernozhukov, and C. Hansen (2012). Sparse Models and Methods

for Optimal Instruments with an Application to Eminent Domain. *Econometrica (80),* 2369–2429, Arxiv, 2010.
Bergh, A., & Nilsson, T. (2010). Do liberalization and globalization increase income inequality? *European Journal of political economy*, *26*(4), 488-505.
Berhane, K. (2018). The role of financial development and institutional quality in economic growth in Africa in the era of globalization. In *Determinants of economic growth in Africa* (pp. 149-196). Palgrave Macmillan, Cham.
Bhavsar, P., Safro, I., Bouaynaya, N., Polikar, R., & Dera, D. (2017). Machine learning in transportation data analytics. In *Data analytics for intelligent transportation systems*, 283-307, Elsevier.
Bittencourt, M., Van Eyden, R., & Seleteng, M. (2015). Inflation and Economic Growth: Evidence from the Southern African Development Community. *South African Journal of Economics, 83*(3), 411-424.
Bredt, S. (2019). Artificial Intelligence (AI) in the financial sector—Potential and public strategies. *Frontiers in Artificial Intelligence*, *2* – 16.
Brown, C. S., Ravallion, M., & Van De Walle, D. (2020). Can the World's Poor Protect Themselves from the New Coronavirus? (No. w27200). *National Bureau of Economic Research*.
Chakamera, C., & Alagidede, P. (2018). The nexus between infrastructure (quantity and quality) and economic growth in Sub Saharan Africa. *International Review of Applied Economics*, *32*(5), 641–672.
Čihák, M., Demirgüč-Kunt, A., Feyen, E., & Levine, R. (2013). Financial development in 205 economies, 1960 to 2010 (No. w18946). *National Bureau of Economic Research*.
Chernozhukov, V., Hansen, C. & Spindler M. (2015). Valid Post-Selection and Post-Regularization Inference: An Elementary. General Approach. *Annual Review of Economics*, *7*(1), 649–688.
De Almeida Rocha, D., & Duarte, J. C. (2019). Simulating human behaviour in games using machine learning. In *2019 18th Brazilian Symposium on Computer Games and Digital Entertainment*, 163–172. IEEE.
Domar, E. (1947). Expansion and Employment. *American Economic Review*, 14(1), 34 – 55.
Doupe, P., Faghmous, J., & Basu, S. (2019). Machine learning for health services researchers. *Value in Health*, *22*(7), 808-815.
Dutt, P., & Tsetlin, I. (2016). Income distribution and economic development: Insights from machine learning. *Economics & Politics*.
Emmanuel, O. G., & Kehinde, A. J. O. S. E. (2018). Domestic investment and economy growth in Nigeria: An empirical investigation. *International Journal of Business and Social Science, 9(2)*, 130-138.
Fosu, A. K. (2012). The African economic growth record, and the roles of policy syndromes and governance. Good growth and governance in Africa: Rethinking development strategies, 175-218.
Grossman, G. M., & Helpman, E. (1991). *Innovation and growth in the global economy*. MIT press.
Gygli, S., Haelg, F., Potrafke, N., & Sturm, J. E. (2019). The KOF Globalisation Index Revisited. *Review of International Organizations*, *14*(3), 543–574.
Gyimah-Brempong, K., Paddison, O., & Mitiku, W. (2006). Higher education and economic growth in Africa. *The Journal of Development Studies*, *42*(3), 509-529.
Harrod, R. (1939). An Essay in Dynamic Theory. *The Economic Journal*, 44(1), 14 – 31.
Hastie, T., Tibshirani, R., & Wainwright, M. (2019). Statistical learning with sparsity: the lasso and generalizations. Chapman and Hall/CRC.
Ibrahim, M., & Alagidede, P. (2018). Effect of financial development on economic growth in

sub-Saharan Africa. *Journal of Policy Modeling*, *40*(6), 1104–1125.
ILO (2020). COVID-19 Cruelly Highlights Inequalities and Threatens to Deepen Them. ILO Newsroom
ILO (2020b). The Impact of The Coronavirus (COVID-19) Crisis on Development Finance, June, 2020.
IMF (2020). World Economic Outlook: A Long and Difficult Ascent. October. *Washington, DC,*
IMF & World Bank. (2020). Enhancing Access to Opportunities. International Monetary Fund and World Bank, *Washington, DC*: World Bank.
James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An Introduction to Statistical Learning (Vol. 112). New York: Springer.
Jung, J.-K., Patnam, M., & Ter-Martirosyan, A. (2018). An algorithmic crystal ball: Forecasts-based on machine learning. *IMF Working Papers*. Washington, D.C.: International Monetary Fund
Kebede, J. G., & Takyi, P. O. (2017). Causality between institutional quality and economic growth: Evidence from sub-Saharan Africa. European Journal of Economic and Financial Research.
King, R. G., & Levine, R. (1993). Finance and growth: Schumpeter might be right. *The quarterly journal of economics*, *108*(3), 717–737.
Kodongo, O., & Ojah, K. (2016). Does infrastructure really explain economic growth in Sub-Saharan Africa? Review of Development Finance, 6(2), 105-125.
Koomson, I., Villano, R. A., & Hadley, D. (2020). Effect of financial inclusion on poverty and vulnerability to poverty: Evidence using a multidimensional measure of financial inclusion. Social Indicators Research, 149(2), 613-639.
Lahoti, R., Jayadev, A., & Reddy, S. (2016). The global consumption and income project (GCIP): An overview. *Journal of Globalization and Development*, *7*(1), 61-108.
Luxton, D. D. (2016). An introduction to artificial intelligence in behavioral and mental health care. In Artificial intelligence in behavioral and mental health care, *Academic Press*,1-26
Mateen, B. A., Liley, J., Denniston, A. K., Holmes, C. C., & Vollmer, S. J. (2020). Improving the quality of machine learning in health applications and clinical research. *Nature Machine Intelligence*, *2*(10), 554-556.
Mavikela, N., Mhaka, S., & Phiri, A. (2019). The Inflation-Growth Relationship In SSA Inflation-Targeting Countries. *Studia Universitatis Babes-Bolyai, Oeconomica, 64(2),* 84-102.
Morik K. (2011) Medicine: Applications of Machine Learning. In: Sammut C., Webb G.I. (eds) *Encyclopedia of Machine Learning*. Springer, Boston, MA.
Ngongang, E. (2015). Financial development and economic growth in Sub-Saharan Africa: A dynamic panel data analysis. *European Journal of Sustainable Development, 4(2),* 369-369.
OECD/ACET (2020). Quality Infrastructure in 21st Century Africa: Prioritising, Accelerating and Scaling up in the Context of Pida (2021-30).
OECD/ILO (2019), Tackling Vulnerability in the Informal Economy, Development Centre Studies, OECD Publishing, Paris.
Ofori, I. K., Quaidoo, C., & Ofori, P. E. (2022). What drives financial sector development in Africa? Insights from machine learning. Applied Artificial Intelligence, 1-33.
Olaoye, O. O., Eluwole, O. O., Ayesha, A., & Afolabi, O. O. (2020). Government spending and economic growth in ECOWAS: An asymmetric analysis. The Journal of Economic Asymmetries, 22, e00180.
Omoteso, K., & Mobolaji, H. I. (2014). Corruption, governance and economic growth in Sub-

Saharan Africa: a need for the prioritisation of reform policies. Social Responsibility Journal.
Opoku, E. E. O., Ibrahim, M., & Sare, Y. A. (2019). The causal relationship between financial development and economic growth in Africa. *International Review of Applied Economics*, 33(6), 789-812.
Oyvat, C. (2016). Agrarian structures, urbanization, and inequality. *World Development*, *83*, 207-230.
Pan, R., Yang, T., Cao, J., Lu, K., & Zhang, Z. (2015). Missing data imputation by K nearest neighbours based on grey relational structure and mutual information. *Applied Intelligence*, *43*(3), 614-632.
Peprah, J. A., Kwesi Ofori, I., & Asomani, A. N. (2019). Financial development, remittances and economic growth: A threshold analysis. *Cogent Economics & Finance*, 7(1), 1625107.
Romer, P. M. (1990). Endogenous technological change. *Journal of political Economy*, *98*(5, Part 2), S71–S102.
Richardson, A., van Florenstein Mulder, T., & Vehbi, T. (2021). Nowcasting GDP using machine-learning algorithms: A real-time assessment. *International Journal of Forecasting*, *37*(2), 941–948.
Saint Akadiri, S., & Akadiri, A. C. (2018). Growth and inequality in Africa: Reconsideration. *Academic Journal of Economic Studies*, *4(3)*, 76-86.
Sakyi, D., Commodore, R., & Opoku, E. E. O. (2015). Foreign direct investment, trade openness and economic growth in Ghana: An empirical investigation. *Journal of African Business*, *16*(1-2), 1-15.
Sandeep, S., Shelton, C. R., Pahor, A., Jaeggi, S. M., & Seitz, A. R. (2020). Application of Machine Learning Models for Tracking Participant Skills in Cognitive Training, *Frontiers in Psychology*, *11*, 15 – 32.
Schneider, U., & Wagner, M. (2012). Catching growth determinants with the adaptive lasso. *German Economic Review*, *13*(1), 71-85.
Schowe, B., & Morik, K. (2011). Fast-ensembles of minimum redundancy feature selection. In *Ensembles in Machine Learning Applications*, Springer, Berlin, Heidelberg, 75-95.
Schwarz, G. (1978). Estimating the dimension of a model. *Annals of statistics*, *6*(2), 461-464.
Sekkat, K. (2017). Urban concentration and poverty in developing countries. *Growth and Change*, *48*(3), 435-458.
Svirydzenka, K. (2016). Introducing a New Broad-Based Index of Financial Development. *IMF Working Paper No. 16/5*.
Solow, R. M. (1956). A Contribution to the Theory of Economic Growth. *The Quarterly Journal of Economics*, *70*(1), 65-94.
Tchamyou, V. S., Asongu, S. A., & Odhiambo, N. M. (2019). The role of ICT in modulating the effect of education and lifelong learning on income inequality and economic growth in Africa. *African Development Review*, *31*(3), 261-274.
Tkacz, G. (2001). Neural network forecasting of Canadian GDP growth. *International Journal of Forecasting*, *17*(1), 57-69.
Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, *58*(1), 267-288.
Tibshirani, R. J., & Taylor, J. (2012). Degrees of freedom in lasso problems. *The Annals of Statistics*, 40(2), 1198-1232.
Tizghadam, A., Khazaei, H., Moghaddam, M. H., & Hassan, Y. (2019). Machine Learning In Transportation. *Journal of Advanced Transportation*, Special Issue, Volume (2019), 1 – 3.
UNCTAD (United Nations Conference on Trade and Development) (2021). Investment

Trends Monitor. Issue 38. Geneva: UNCTAD.
Van Hulse, J., & Khoshgoftaar, T. M. (2014). Incomplete-case nearest neighbor imputation in software measurement data. *Information Sciences*, *259*, 596-610.
World Bank. (2021a). COVID-19 and Food Security. Brief, World Bank, Washington, DC. https://www.worldbank.org/en/topic/agriculture/brief/foodsecurity-and-covid-19.
World Bank (2021b). World Development Indicators. Washington, DC: World Bank.
World Bank. (2020a). Global Economic Prospects, June 2020. Washington, DC: World Bank.
World Bank (2020b). Poverty and Shared Prosperity 2020: Reversals of Fortunes. October. Washington, DC, World Bank.
World Bank (2009), Reshaping Economic Geography; World Development Report 2009, Washington DC, World Bank.
Zahonogo, P. (2016). Trade and economic growth in developing countries: Evidence from sub-Saharan Africa. Journal of African Trade, 3(1-2), 41-56.
Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of The American Statistical Association*, *101*(476), 1418-1429.
Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of The Royal Statistical Society*: series B (statistical methodology), 67(2), 301-320.
Zou, H., Hastie, T., & Tibshirani, R. (2007). On the “degrees of freedom” of the lasso. *The Annals of Statistics*, *35*(5), 2173-2192.

APPENDICES

Table A1: Variable definition and data sources

Variable	Description	Source
women	Proportion of seats held by women in national parliament (%)	WDI
Women_business	Women businesses and law index score (scale 1 – 100)	WDI
wagessalary	Wage and salaried workers, total (% of total employment)	WDI
vul_tot	Vulnerable employment is contributing family workers and own-account workers as a percentage of total employment.	WDI
lossesdue_power	Productivity losses due to power	WDI
urban_pop	Urban population growth (annual %)	WDI
unempl	Unemployment, total (% of total labour force)	WDI
trade	Trade is the sum of exports and imports of goods and services measured as a share of gross domestic product.	WDI
trade_tax	Taxes on international trade include import duties, export duties, profits of export or import monopolies, exchange profits, and exchange taxes.	WDI
taxrev	Tax revenue (% of GDP)	WDI
tariffwm	Tariff rate, applied, weighted mean, all products (%)	WDI
self_employ	Self-employed, total (% of total employment)	WDI
rur_popgrof	Rural population growth (annual %)	WDI
rd	Research and development expenditure (% of GDP)	WDI
rer	Real effective exchange rate index (2010 = 100)	WDI
HIV_preva	Prevalence of HIV, total (% of population ages 15-49)	WDI
prenatal	Pregnant women receiving prenatal care (%)	WDI
povert_hc	Poverty headcount ratio at national poverty lines (% of population)	PED
povertyhc_mid	Poverty headcount ratio at $3.20 a day (2011 PPP) (% of population)	PED
povertyhc_low	Poverty headcount ratio at $1.90 a day (2011 PPP) (% of population)	PED
povmidd	Poverty gap at $3.20 a day (2011 PPP) (%)	PED
povint	Poverty gap at $1.90 a day (2011 PPP) (%)	PED
urbanization	Urban population (% of total population)	WDI
popgrof	Population growth (annual %)	WDI
finan_insti	Deposit money banks' assets to GDP (%)	WDI
sanitation_pop	People using at least basic sanitation services (% of population)	WDI
opendefeca_pop	People practicing open defecation (% of population)	WDI
exr	Nominal exchange rate (US dollar)	WDI
unfpa_aid	Net official flows from UN agencies, UNFPA (current US$)	WDI
unicef_aid	Net official flows from UN agencies, UNICEF (current US$)	WDI
undp_aid	Net official flows from UN agencies, UNDP (current US$)	WDI
noda	Net Official Development Assistance received (% of GNI)	WDI
netmigration	5-year estimates on net migration (immigrants less emigrants)	WDI
mortality_5yrs	Number of under-five deaths	WDI
manuf_VA	Manufacturing, value added (annual % growth)	WDI
logisticqua_overal	Logistics performance index: Overall (1=low to 5=high)	WDI
logisticqua_TT	Quality of trade and transport-related infrastructure (1=low to 5=high)	WDI
logisticqua_ship	Logistics performance index: Frequency with which shipments reach consignee within scheduled or expected time (1=low to 5=high)	WDI
logisticqua_custom	Logistics performance index: Efficiency of customs clearance process (1=low to 5=high)	WDI
literacy_adult	Literacy rate, adult total (% of people ages 15 and above)	WDI
labforce_pr	Labour force participation rate, total (% of total population ages 15-64)	WDI
transport_invest	Investment in transport with private participation (current US$)	WDI
inflation	Inflation, consumer prices (annual %)	WDI
informalemp_Tot	Informal employment (total)	WDI
industry_VA	Industry (including construction), value added (% of GDP)	WDI

hci	Human Capital Index (HCI) (scale 0-1)	WDI
house_spend	Households and NPISHs final consumption expenditure (% of GDP)	WDI
grossavings	Adjusted annual gross savings (% of GNI)	WDI
natl_expend	National expenditure (% of GNI)	WDI
gfcf	Gross fixed capital formation (annual % growth)	WDI
domesticinvest	Gross fixed capital formation, private sector (% of GDP)	WDI
gov_educ	Expenditure on secondary education (% of government expenditure on education)	WDI
gov	General government final consumption expenditure (% of GDP)	WDI
gov_gdp	Government recurrent expenditure (%GDP)	WDI
gpc_grof	GDP per capita growth (annual %)	WDI
gpc	GDP per capita, PPP (constant 2017 international $)	WDI
gdpg	GDP growth (annual %)	WDI
fdi	Foreign direct investment, net inflows (% of GDP)	WDI
emp_ind	Employment in Employment in industry (% of total employment)	WDI
emp_agric	Employment in agriculture (% of total employment)	WDI
ease	Ease of doing business index (1=most business-friendly regulations)	WDI
health_exp	Government health expenditure (%GDP)	WDI
cps	Private credit by deposit money banks and other financial institutions to GDP (%)	GFDD
cpia_transparency	CPIA transparency, accountability, and corruption in the public sector rating (1=low to 6=high)	CPIA
cpia_trade	CPIA trade rating (1=low to 6=high)	CPIA
cpia_socprotect	CPIA social protection rating (1=low to 6=high)	CPIA
cpia_publicmgt	CPIA public sector management and institutions cluster average (1=low to 6=high)	CPIA
cpia_socinclusion	CPIA policies for social inclusion/equity cluster average (1=low to 6=high)	CPIA
cpia_macro	CPIA macroeconomic management rating (1=low to 6=high)	CPIA
cpia_gender	CPIA gender equality rating (1=low to 6=high)	CPIA
cpia_finsector	CPIA financial sector rating (1=low to 6=high)	CPIA
debt	Public debt stock (%GDP)	WDI
moneyg	Broad money growth (annual %)	GFDD
agric_VA	Agriculture, forestry, and fishing, value added (% of GDP)	WDI
electricaccess_pop	Access to electricity (% of population)	WDI
electricaccess_rur	Access to electricity, rural (% of rural population)	WDI
importburden	Cost to import, documentary compliance (US$)	WDI
exportburden	Cost to export, documentary compliance (US$)	WDI
natresourcerent	Natural resource rent %GDP)	WDI
kofgidj	KOF. overall globalisation index (de jure)	KOF. Index
kofecgj	KOF. economic globalisation index (de jure)	KOF. Index
koftrgj	KOF. trade globalisation index (de jure)	KOF. Index
koffindj	KOF. financial globalisation index (de jure)	KOF. Index
kofsodj	KOF. social globalisation index (de jure)	KOF. Index
gini	Gini index inequality indicators	GCIP
fin_devt	Financial development index	Findex
fi	Financial institutions index	Findex
fm	Financial markets index	Findex
fid	Financial institutions depth index	Findex
fia	Financial institutions access index	Findex
fie	Financial institutions efficiency index	Findex
fmd	Financial markets depth index	Findex
fma	Financial markets access index	Findex
fme	Financial markets efficiency index	Findex
npl	Bank non-performing loans to gross loans (%)	GFDD

bankOHcost	Bank overhead costs to total assets (%)	GFDD
roa_net	Bank return on assets (%, after tax)	GFDD
roe_net	Bank return on equity (%, after tax)	GFDD
bankCrisis	Banking crisis dummy (1=banking crisis, 0=none)	GFDD
boone	Boone indicator. A measure of degree of competition based on profit-efficiency in the banking market	GFDD
onlinepayment	Electronic payments used to make payments (% age 15+)	GFDD
insurancePrem	Life insurance premium volume to GDP (%)	GFDD
phonePayment	Mobile phone for paying bills online	GFDD
phoneMomo	Mobile phone penetration (Able to perform mobile money transaction)	GFDD
remit	Remittance inflows (%GDP)	GFDD
stockPxVol	Stock price volatility index	GFDD
infrastr_qua	Infrastructure quality score	WDI
sse_gp	School enrolment, secondary (gross), gender parity index (GPI)	WDI
sis_m	Secure Internet servers (per 1 million people)	WDI
int_pop	Individuals using the Internet (% of population)	WDI
mcs_hd	Mobile cellular subscriptions (per 100 people)	WDI
fts_hd	Fixed telephone subscriptions (per 100 people)	WDI
fbs_hd	Fixed broadband subscriptions (per 100 people)	WDI
fd2	Square of financial development index	Generated
ps	Severity of poverty	Generated

Note: FD Index is Financial Development (International Monetary Fund); GFDD is Global Financial Development Database (Word Bank); KOF. Index is the Konjunkturforschungsstelle (KOF) index; GCIP is Global Consumption and Income Project; CPIA is Country Policy and Institutional Assessment (World Bank); and WDI is World Development Indicators (World Bank).

Source: Author's construct, 2021

Table A2: Summary Statistics for Training and Testing sets

Variable	Obs	Mean (Training set)	Std. Dev. (Training set)	Min (Training set)	Max (Training set)	Mean(Testing set)	Std. Dev. (Testing set)	Min(Testing set)	Max (Testing set)
wagessalary	1204(516)	26.779	21.77	5.049	85.412	27.151	22.537	5.106	85.871
vul tot	1204(516)	71.106	22.471	9.429	94.75	70.747	23.268	8.826	94.759
lossesdue power	1204(516)	6.804	4.938	.7	25.1	6.427	4.718	0	25.1
urban pop	1204(516)	4.268	1.885	-6.879	17.499	4.195	1.888	-7.182	15.714
unempl	1204(516)	7.58	7.48	.3	37.976	7.684	7.513	.3	37.94
trade	1204(516)	67.487	36.253	9.136	290.499	66.853	34.662	0	311.354
trade tax	1204(516)	16.15	13.095	.107	63.451	15.404	12.787	0	63.451
taxrev	1204(516)	14.198	6.339	4.204	39.258	14.182	6.287	0	37.353
tariffwm	1204(516)	12.301	5.381	.84	32.6	12.771	7.106	0	91.27
self_employ	1204(516)	73.221	21.77	14.588	94.951	72.849	22.537	14.129	94.894
rurpopgrof	1204(516)	1.735	1.335	-6.707	10.906	1.656	1.477	-7.866	7.297
rd	1204(516)	.275	.18	.005	.888	.28	.18	0	.898
rer	1204(516)	198.423	168.865	46.021	3520.534	200.184	122.564	0	2182.799
hiv_preva	1204(516)	2.664	3.554	.1	24.1	2.584	3.74	0	24.2
prenatal	1204(516)	77.749	17.935	23.4	99.4	75.969	19.565	0	99.3
poverthc	1204(516)	48.929	13.849	7.9	73.2	47.661	14.206	7.9	73.2
povertyhc mid	1204(516)	69.272	23.899	2.2	98.5	68.718	24.645	3.1	98.5
povertyhc low	1204(516)	49.48	24.837	.2	94.3	49.342	25.332	.4	94.3
povmidd	1204(516)	38.282	18.955	.4	86.7	38.317	19.581	.7	86.7
povint	1204(516)	23.018	16.497	0	86.7	23.342	17.313	.1	86.7
urbanization	1204(516)	39.323	14.338	10.838	92.697	38.745	13.696	0	100
popgrof	1204(516)	2.614	.875	-1.305	7.449	2.532	1.101	-6.766	8.118
finaninsti	1204(516)	2.623e+08	9.764e+08	-2.027e+09	8.594e+09	2.503e+08	9.760e+08	-7.162e+09	6.988e+09
exr	1204(516)	408.351	1250.663	0	19068.417	407.802	1462.092	0	18498.601
noda	1204(516)	11.632	11.713	-.188	79.827	11.091	11.397	-.251	94.946
netmigration	1204(516)	-21467.614	262740	-1374270	1457943	-19246.359	295523.94	-1374270	1457943
mortality 5yrs	1204(516)	127.098	63.887	13.7	329.3	127.848	68.421	0	337.4
manuf VA	1204(516)	2.456	9.168	-34.921	97.709	3.636	16.181	-43.84	375.158
logisticqua TT	1204(516)	2.167	.335	1.27	3.79	2.132	.349	0	3.79
logisticquaoveral	1204(516)	2.398	.302	1.61	3.775	2.374	.321	0	3.67
logisticqua ship	1204(516)	2.845	.431	1.67	4.03	2.839	.458	0	4.03
literacy adult	1204(516)	56.851	21.081	10.895	95.868	56.709	21.682	0	95.868
labforce pr	1204(516)	69.802	11.312	42.381	92.453	69.87	11.536	42.409	92.453
transport invest	1204(516)	3.050e+08	5.834e+08	0	3.483e+09	3.543e+08	6.454e+08	0	3.483e+09
inflation	1204(516)	45.776	822.219	-9.809	23773.132	17.538	102.741	-13.057	2154.437
industry VA	1204(516)	22.678	12.112	1.305	72.153	23.735	12.355	0	72.717
hci	1204(516)	.393	.069	0	.678	.396	.077	0	.678
house spend	1204(516)	.713	8.947	-45.41	65.181	1.032	8.214	-46.068	87.014
grossavings	1204(516)	15.406	18.256	-69.534	84.49	17.083	16.674	-70.263	87.096
natl expend	1204(516)	109.86	18.643	57.699	255.256	108.993	16.652	0	261.428
gfcf	1204(516)	21.085	10.653	-2.424	85.941	21.436	10.946	0	93.547
domesticinvest	1204(516)	11.914	20.269	-133.979	85.541	13.388	18.583	-141.974	88.389
gov_educ	1204(516)	15.581	5.659	4.673	37.521	15.445	5.567	0	34.309
gov	1204(516)	6.38	31.534	-71.464	565.539	4.477	18.399	-68.238	165.168
gov_gdp	1204(516)	14.813	6.874	0	51.975	14.894	6.773	0	51.975
gpc	1204(516)	3783.244	4347.165	464.018	29223.465	3855.974	4458.085	0	27242.656
gdpg	1204(516)	3.588	4.881	-30.145	33.629	3.592	5.522	-50.248	35.224
fdi	1204(516)	3.353	7.973	-8.703	103.337	2.435	4.215	-28.624	40.167
emp ind	1204(516)	12.601	8.19	1.43	42.939	12.711	8.62	1.539	43.114
emp agric	1204(516)	54.751	21.465	4.6	92.298	54.951	22.264	4.65	92.303
electricity	1204(516)	545.412	953.395	0	4665.176	533.118	958.839	0	4851.693
ease	1204(516)	136.628	38.692	13	184	134.261	41.851	0	184
health exp	1204(516)	1.661	1.098	.062	5.496	1.631	1.102	0	6.049
cps	1204(516)	17.836	20.968	.403	149.234	18.695	21.393	0	160.125
cpia transparency	1204(516)	2.812	.592	1.5	4.5	2.808	.6	0	4.5

cpia trade	1204(516)	3.728	.522	2	4.5	3.743	.511	0	4.5
cpiapublicmgt	1204(516)	3.006	.455	2	4.1	3.02	.478	0	4
cpiasocinclusion	1204(516)	3.157	.471	2.2	4.3	3.166	.477	0	4.3
cpia macro	1204(516)	3.641	.66	1.5	5	3.671	.645	0	5
cpia gender	1204(516)	3.195	.535	2	4.5	3.196	.555	0	4.5
cpiafinsector	1204(516)	2.954	.428	2	4	2.947	.437	0	4
debt	1204(516)	104.141	103.919	0	289.845	109.07	105.418	0	289.845
moneyg	1204(516)	66.543	455.289	-99.864	6968.922	76.271	457.558	-29.245	4105.573
agric VA	1204(516)	26.747	16.109	1.828	76.534	25.563	15.59	0	79.042
importburden	1204(516)	137.893	100.549	3.5	420	136.994	107.926	0	588
exportburden	1204(516)	108.208	71.445	4	347	107.994	77.735	0	515
natresourcerent	1204(516)	10.836	9.99	0	56.939	10.861	10.305	0	59.604
kofgi	1204(516)	40.353	10.119	16.922	72.354	39.856	10.381	0	72.269
kofgidj	1204(516)	40.995	11.306	13.308	80.993	40.575	11.51	0	81.288
kofecgj	1204(516)	34.676	10.965	10.514	81.49	34.073	11.136	0	79.549
koftrgj	1204(516)	28.778	14.291	6.494	88.014	28.567	14.566	0	88.497
koffindj	1204(516)	40.481	14.338	6.099	80.37	39.63	13.921	0	81.357
kofsodj	1204(516)	35.241	16.046	4.289	84.779	34.64	16.833	0	85.141
gini	1204(516)	51.994	21.443	0	86.276	51.909	21.048	0	86.832
fin_devt	1204(516)	.124	.089	0	.641	.124	.088	0	.648
fi	1204(516)	.209	.125	0	.74	.211	.126	0	.73
fm	1204(516)	.034	.072	0	.52	.031	.068	0	.54
fid	1204(516)	.099	.149	0	.88	.095	.144	0	.88
fia	1204(516)	.075	.125	0	.88	.078	.132	0	.86
fie	1204(516)	.49	.197	0	.99	.499	.208	0	.98
fmd	1204(516)	.052	.1	0	.83	.045	.09	0	.75
fma	1204(516)	.031	.11	0	.89	.032	.112	0	.58
fme	1204(516)	.016	.058	0	.96	.014	.046	0	.42
npl	1204(516)	13.187	12.73	0	74.1	13.844	13.578	0	74.1
bankOHcost	1204(516)	6.471	5.082	0	89.423	6.159	4.073	0	28.192
roe net	1204(516)	17.975	24.454	-93.62	146.913	19.74	24.961	-93.62	160.344
roa net	1204(516)	1.675	2.837	-15.047	9.182	1.814	2.8	-15.047	12.106
Boone	1204(516)	-.048	.233	-2.541	1.607	-.031	.235	-.896	1.607
onlinepayment	1204(516)	21.345	18.123	0	76.411	20.576	17.807	0	76.411
insurancePrem	1204(516)	.633	1.79	0	14.52	.622	1.72	0	15.381
phonePayment	1204(516)	3.801	5.157	0	37.105	3.626	5.141	0	37.105
phoneMomo	1204(516)	10.532	13.179	0	50.122	10.009	12.84	0	50.122
remit	1204(516)	4.666	19.582	0	235.924	4.057	15.585	0	232.217
stockPxVol	1204(516)	11.08	5.729	0	43.1	10.949	5.513	0	34.376
infrastr qua	1204(516)	3.467	.749	1.8	5.417	3.432	.784	0	5.641
ssegp	1204(516)	.737	.281	0	1.527	.728	.285	0	1.504
int pop	1204(516)	4.942	10.76	0	62	4.937	10.667	0	64
sis m	1204(516)	495.477	9655.395	0	264256.63	302.971	5725.258	0	155191.3
mcshd	1204(516)	23.577	38.438	0	198.152	23.21	38.546	0	173.811
fts hd	1204(516)	2.019	4.51	0	32.669	2.256	5.074	0	34.273
fbshd	1204(516)	.299	1.767	0	27.603	.322	1.717	0	21.639
fd2	1204(516)	.023	.044	0	.41	.023	.042	0	.42
ps	1204(516)	16.016	21.48	0	169.299	17.132	23.363	.001	169.299

Note: 1204 is number of observations in Training set; 516 is observations in Testing set

Source: Author's construct, 2021

Table A3: Variable selection in regularization models

	Standard_lasso	Minimum_BIC_lasso	Elastic_Net	Adaptive_lasso
manuf_VA	x	x	x	x
urban_pop	x	x	x	x
gov	x	x	x	x
house_spend	x	x	x	x
cpia_macro	x	x	x	x
kofecdj	x	x	x	x
unempl	x		x	x
cpia_socinclusion	x	x	x	x
mortality_5yrs	x		x	x
trade_tax	x		x	x
natresourcerent	x		x	
fdi	x		x	
_cons	x	x	x	x

Legend: o – omitted; x – estimated

Source: Authors' construct, 2021

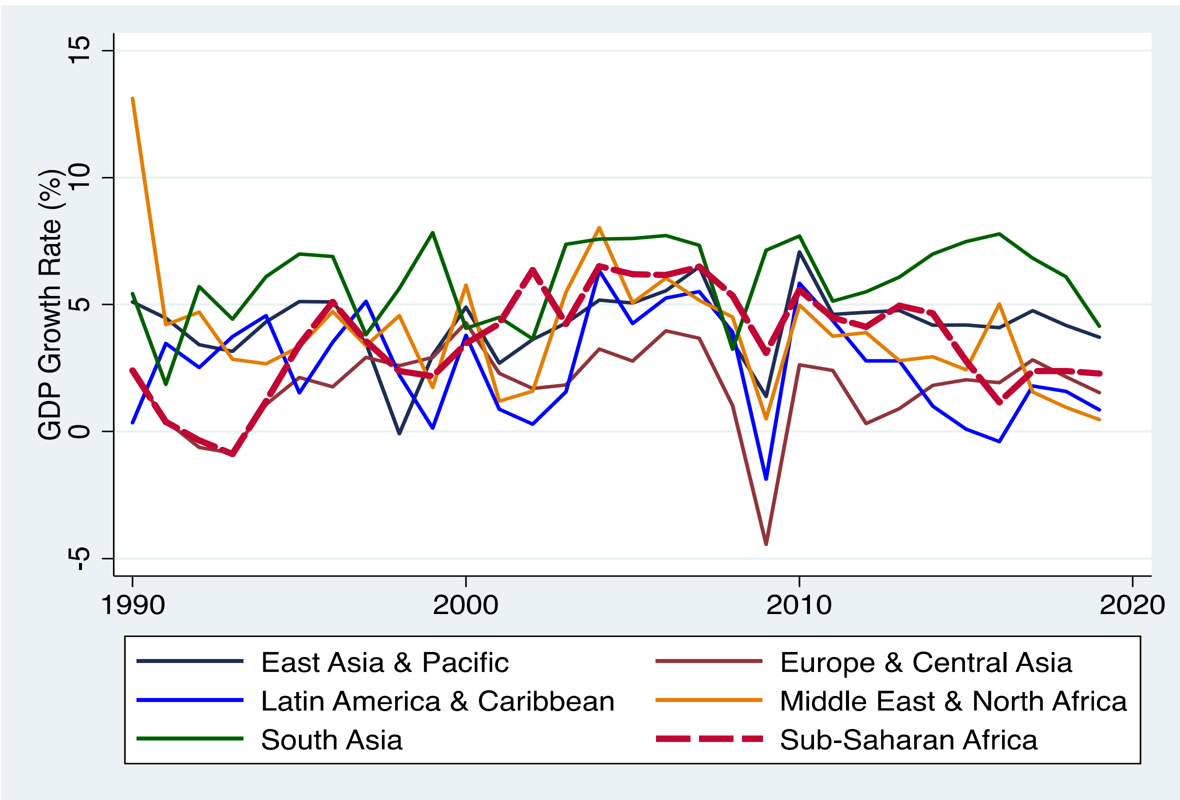

***Fig.**1.1: Trend of GDP Growth Across Regions, 1990 – 2019*
Source: Authors' construct, 2021

proportion of seats held by women in national parliaments (%)

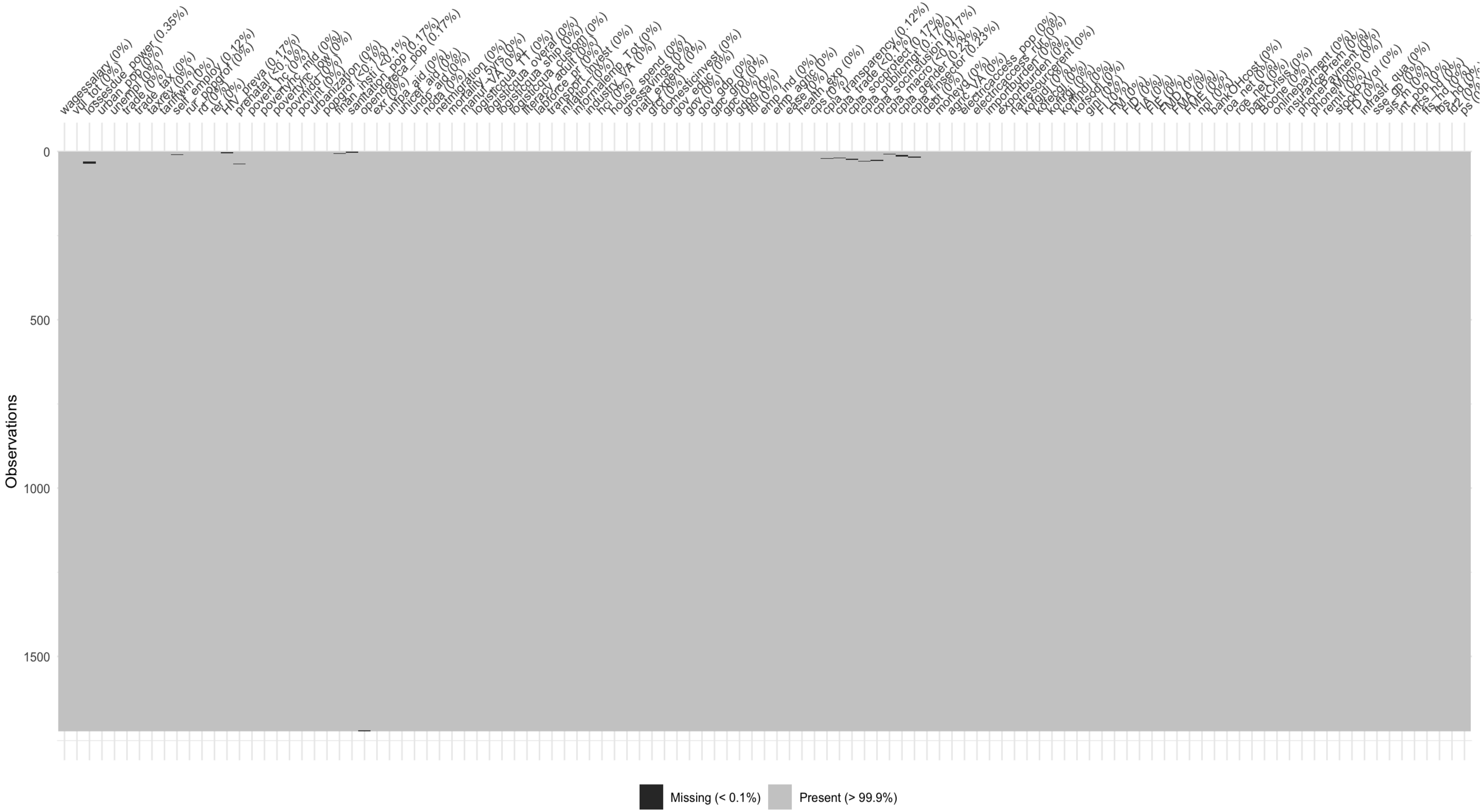

Fig.1.2: Overview of the dataset before data engineering
Source: Authors' construct (2021)

Fig.1.3: Overview of the dataset after data engineering
Source: Authors' construct (2021)