A G D I Working Paper

WP/21/074

What Drives Financial Sector Development in Africa? Insights from Machine Learning

Forthcoming: Applied Artificial Intelligence

Isaac K. Ofori[*a]
(Corresponding Author)
Department of Economics,
University of Insubria, Varese, Italy.
E-mail: ikofori@oulook.com

Christopher Quaidoo[b]
Department of Banking and Finance,
University of Professional Studies,
Legon, Accra, Ghana
E-mail: christopher.quaidoo@upsamail.edu.gh

Pamela E. Ofori[*a]
Department of Economics,
University of Insubria, Varese, Italy.
E-mail: peofori@outlook.com

2021 African Governance and Development Institute WP/21/074

Research Department

What Drives Financial Sector Development in Africa? Insights from Machine Learning

Isaac K. Ofori, Christopher Quaidoo & Pamela E. Ofori

January 2021

Abstract

This study uses machine learning techniques to identify the key drivers of financial development in Africa. To this end, four regularization techniques— *the Standard lasso, Adaptive lasso, the minimum Schwarz Bayesian information criterion lasso*, and the *Elasticnet* are trained based on a dataset containing 86 covariates of financial development for the period 1990 – 2019. The results show that variables such as cell phones, economic globalisation, institutional effectiveness, and literacy are crucial for financial sector development in Africa. Evidence from the *Partialing-out lasso instrumental variable regression* reveals that while inflation and agricultural sector employment suppress financial sector development, cell phones and institutional effectiveness are remarkable in spurring financial sector development in Africa. Policy recommendations are provided in line with the rise in globalisation, and technological progress in Africa.

Keywords: Africa, Elasticnet, Financial Development, Financial Inclusion, Lasso, Regularization, Variable Selection

JEL Codes: *C01; C14; C52; C53; C55; E5; O55*

1.0 Introduction

The slump in global economic activity in the last two years is primarily due to the loss of routine engagements imposed implicitly by the emergence of the coronavirus disease (COVID-19). The concern of policymakers is not only on the welfare implications of the pandemic but how economic activity can be sustained even in future health and economic turmoil. Indeed, such a breakthrough will lessen the impact of future pandemics on jobs, welfare, and the resources of policymakers. Crucially, in the developing world, the high physical contact in transactions coupled with the relatively low financial inclusion means that the progress towards shared prosperity is likely to be derailed in the event of future economic or health uncertainties. Per the long-term growth aspirations of Africa as spelt out in the Africa Agenda 2063, the development of the continent's financial system should be a key policy consideration. This stems from the argument that at the heart of robust and equitable growth is a sound, efficient, dynamic and innovative financial sector crucial for resource allocation, reduction in transaction cost, and creation of opportunities (World Bank, 2019; Beck, 2012; Mckinnon, 1973; Shaw, 1973).

While a burgeoning financial sector can be growth-enhancing, Peprah *et al.* (2019), Law and Singh (2014) and Arcand *et al.* (2015) warn that, in the developing world, excessive financial development can cause a heating-up in the economy, dragging down growth in the process. Particularly, Peprah *et al.* (2019) put a 70 per cent cap on the financial sector development-growth nexus in the case of Ghana while Law and Singh report 88 per cent for a panel of 87 developed and developing economies. The foregoing arguments imply that, realising the lubricating effects of the financial sector while keeping it in check rests on the identification of key variables shaping the sector. The relevance of this is enshrined in the World Bank's Reference Framework for Financial Inclusion Strategies[1], which comprises a set of programmes, knowledge and tools aimed at broadening financial inclusion especially in the developing world (World Bank 2018).

Indeed, the literature on the drivers of financial sector development in Africa is growing. Among others, the literature shows that financial development is driven by institutions, particularly, those for financial sector regulation and supervision, the macroeconomy, bank-specific factors and technology (see e.g., Ibrahim and Sare, 2018; Aluko and Ajayi 2018). Notwithstanding these contributions, conspicuous gaps in the financial development literature, particularly, on Africa are that: (1) proxies are used to

[1] The RFFIS has been adopted by African countries such as Burundi, Ghana, Liberia, Nigeria, Tanzania, Cote, Sierra Leone, Niger, Mauritius, Mauritania, Swaziland, Madagascar, Zambia, and Zimbabwe

capture financial development[2], and (2) prior contributions are inconclusive as to which variables are key for financial sector development in Africa (see e.g., Madsen *et al.*, 2018; Aluko and Ajayi, 2018; Almarzoqi *et al.*, 2015; Jedidia *et al.*, 2014; Arcand *et al.*, 2015). Though the first issue has been addressed to some extent by Čihák *et al.* (2013) and Svirydzenka (2016) who on recognising that a country's financial sector comprises a variety of financial institutions, markets and products, developed the Global Financial Development Database and Global Financial Development Index (FD Index)[3], respectively, comprehensive empirical work(s) responding to the latter is(are) hard to find.

A survey of the literature shows that studies attempting such a contribution are plagued with some methodological flaws due to: (1) the application of techniques that lack regularization powers for inference even in large datasets, and (2) the preferential/subjective selection of covariates in regression problems (see e.g., Nguyen, 2020, Ibrahim and Sare, 2018; Aluko and Ajayi, 2018; Adu *et al.*, 2013). The concern with these empirical works is that even tenuous variables may be deemed relevant for driving financial development under some modelling assumptions, specifications and data transformation. Addressing this challenge and thus informing policy appropriately can be through the use of machine learning[4] (artificial intelligence) algorithms for regularization, prediction, and inference (see, Tibshirani, 1996, Zou, 2006; Saura 2020). This forms the contribution of this paper where two objectives are introduced to extend the financial development literature. First, we train algorithms for the *Standard lasso, Adaptive lasso, the minimum Schwarz Bayesian information criterion lasso (Minimum BIC lasso)*, and *Elasticnet* to study patterns underlying a dataset on 42 African countries to identify the main determinants of financial development. Second, to provide inferences robust to potential endogeneity concerns, model misspecification and the underlying data complexity on the selected drivers of financial development, we apply the *double-selection linear lasso regression, partialing-out lasso linear regression,* and *partialing-out lasso instrumental variable regression.*

The relevance of our contribution is that it can prove crucial in informing policy actions in Africa on the key variables to target if monetary policy propositions, resource allocation, and the overall effectiveness of the financial sector in fostering shared prosperity

---

[2] For instance, variables such as the ratio of financial institutions' assets to GDP, the ratio of liquid liabilities to GDP, and the ratio of deposits to GDP, are often chosen as proxies/indicators for financial sector development (see e.g., Mtar and Belazreg, 2020; Barajas *et al.*, 2013; Adu *et al.*, 2013)

[3] The FD index provides comprehensive information on the degree of access, depth, efficiency and stability of the financial institutions and markets of a country's financial sector[3] (see, Svirydzenka, 2016).

[4] Machine learning has gained attention in recent times due to its ability to detect relevant patterns in big data for prediction and analysis.

is to be achieved. It could also prove invaluable to various African governments in their bid to broadening access to formal financial services especially for the financially excluded as well as the efficient allocation of resources to transform the continent's highly informal structure to a formal one. Additionally, the study could aid stakeholders interested in Africa's financial sector development, plan, strategize and possibly initiate necessary reforms to spur a sound, responsible, and innovative financial sector.

The rest of the chapter is organised as follows. The next section presents an overview of Africa's financial sector and a literature review on drivers of financial development. Section 3 also presents the methods and data underpinning the analysis. In Section 4, we present our results while the conclusion and policy recommendations are provided in Section 5.

2.0 Literature survey

2.1 Financial sector development in Africa: current and historical perspectives

In 2017, the World Bank reported that an astounding 1.7 billion people were financially excluded, down from 3 billion in 2014 (Demirgüç-Kunt *et al.*, 2018). The report further indicates that at least 300 million adults in Africa do not have accounts with banks or any form of financial institution. Indeed, compared to regions such as Europe, and the Americas, the financial sector of Africa lags behind. In the 1960s to 1990s, Africa's financial sector was highly repressed or polarised for protectionist motives of various governments (e.g., Ghana, Nigeria, and Guinea), resulting in inefficient resource allocation. It was until the last decade that financial openness and repression has eased in the region. Albeit not surprising, it is worrying to note that no African country has attained the average financial development threshold of 0.5 per IMF's classification as apparent in the upper panel of Figure 1. Further, information gleaned from the upper panel of Figure 1 shows that, though the likes of South Africa, Mauritius. Seychelles, Botswana, and Nigeria have made significant strides in financial sector development, that of Cameroon, Comoros, Congo DR., Guineas-Bissau, Sierra Leone and the Central African Republic remain significantly underdeveloped.

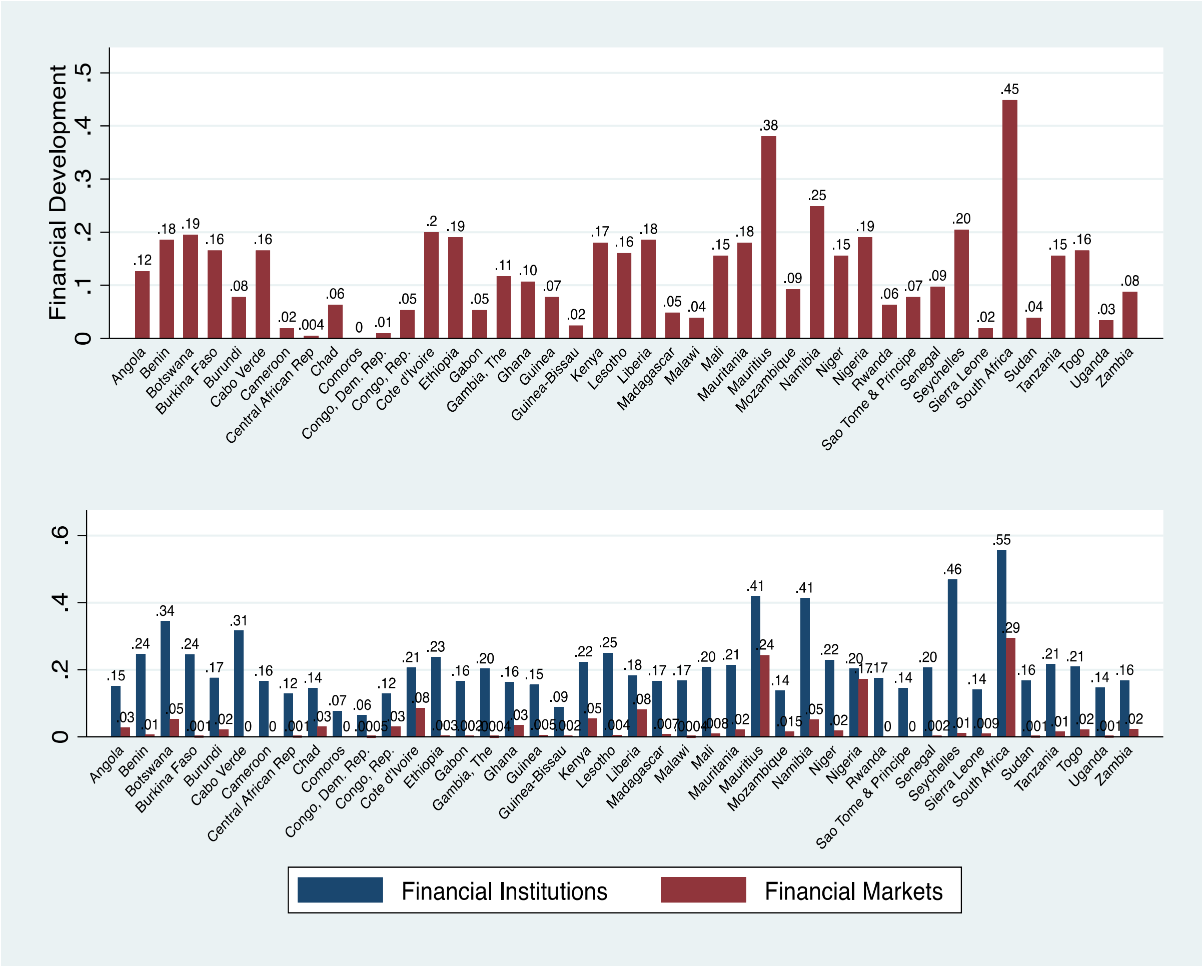

Figure 1: Average Financial Development (Upper Panel), And Financial Markets and Institutions (Lower Panel) in Africa, 1980 – 2019, IMF Findex data.

In particular, information garnered from the lower panel of Figure 1 shows that vis-à-vis financial institutions, Africa's financial market is significantly underdeveloped. Also conspicuous is the striking within-country experiences in Figure 1 (lower panel), which also reveal that countries such as South Africa, Nigeria, Mauritius, Botswana, Cote d'Ivoire, and Kenya have made significant progress in the development of their financial institutions. The overview of Africa's financial sector development in Figure 1 underscores the need to strengthen the continent's financial sector. Achieving this objective will among other rest chiefly on identifying variables that are crucial for financial sector development to aid decision-makers plan, reform or re-strategize— one reason why this study is relevant.

2.2 Theoretical and empirical literature review

In this section, we present some theoretical and empirical evidence on the drivers of financial development.

2.2.1 Endowment theory (settler mortality hypothesis)

The endowment theory as put forward by Acemoglu *et al.* (2001) points to the relevance of institutions, resource endowment, and geography for financial sector development. The authors indicate that, in the 1960s and 1970s, institutions were established to offer protection for private property; protection against government power of expropriation; guarantee the transfer of resources from colonies to the colonisers with little or no investment (Acemoglu *et al.,* 2001). Broadening the import of this theory, Beck *et al.* (2003) also argue that initial endowments are rather germane in explaining international differences in financial sector development than legal origins and that countries with poor geographical endowments are likely to have less developed financial sector.

2.2.2 Law and finance theory

La Porta *et al.* (1998) championed this theory with the fundamental proposition that a country's legal framework matters for financial sector development. The theory comes in two forms— a part that recognises the relevance of robust legal systems in financial sector development (Beck *et al*. 2003), and another part that identifies legal traditions[5] as the driving force behind cross-country differences in financial sector development. Empirical evidence for this theory is found in Djankov *et al.* (2007) who argue that civil law countries realise lesser bureaucracy, corruption, enhanced government credibility and greater financial development. In the context of Africa, however, Fowowe (2014) does not find empirical support for this theory.

2.2.3 Financial liberalisation theory

This is the McKinnon-Shaw hypothesis theorising growth in a country's financial sector following financial liberalisation (McKinnon, 1973; Shaw, 1973). The theory indicates that both domestic savings and credit to the private sector increases if there is a moderately high and positive interest rate. They argue that financial repression results in market disequilibrium, consequently limiting allocative efficiency. The authors further suggest that in developing countries like Africa, financial repression can lead to firms facing financing constraints due to limited access to external finance and credit controls. In line with this theory is empirical evidence by Baltagi *et al.* (2009) who find that financial sector

[5] La Porta *et al*. (1998) argue that, common law countries provide stronger legal protection for investors than civil law countries

development grows even faster if financial liberalisation is accompanied by greater trade and financial openness.

2.2.4 Inflation and finance theory

This theory was put forward by Huybens and Smith (1999) with the fundamental proposition that high inflation levels suppress financial development. Furthering this argument, Rousseau and Wachtel (2002) argue that macroeconomic instability causes financial institutions to ration credit, reducing financial market activity and profitability in the process. The authors further indicate that high inflation can discourage long-term loans, resulting in inefficient allocation of resources. In related empirical work, Boyd *et al*. (2001) and Kim and Lin (2010) find evidence that the inflation-finance nexus is nonlinear and exists only up to a certain point.

2.2.5 Demand-following* (growth-led) *hypothesis

The demand-following hypothesis is the well-known argument by Robinson (1952) that growing economic activity leads to greater demand for financial services by the real sector, enhancing the utilisation of financial products and services. Thus, increasing economic growth reflects rising living standards and the likely participation of the populace in the country's financial sector. This theory has been enhanced significantly by empirical evidence from authors such as Akinlo and Egbetunde (2010), who argue that economic growth is crucial for driving both financial inclusion and financial development.

2.3 Empirical literature survey

The literature shows that variables such as inflation and public debt impede financial development (Ayadi *et al.* 2015; Elsherif 2015; Sanusi, Meyer and Ślusarczyk 2017; Aluko and Ibrahim 2019). Particularly, Ayadi *et al.* (2015) argue that growth in government debt deteriorates the growth of credit and crowds out private lending and investment. Boyd *et al.* (2001) also provide convincing evidence to conclude that high inflated economies are more likely to have banks and equity markets that are less robust and inefficient. Specifically, in inflation targeting economies like Ghana, information asymmetry can bid inflation up, creating frictions in the credit market, leading to financial sector deterioration in the process (Padachi *et al.* 2008). Similar evidence is found in Bittencourt (2011) who examined the relationship between inflation and finance in Brazil from 1995 to 2002.

There is also the evidence that financial sector development thrives on conducive economic, financial and institutional settings. Indeed, evidence gleaned from Khalfaoui (2015) and Shabbir *et al.* (2018) indicate that fiscal discipline, economic growth and transparent monetary regime are crucial for enhancing the access, depth and efficiency of financial systems. In a related study by Beck and Levine (2005), regulatory quality in the form of prudential supervision has been identified to enhance financial development and stability. In line with this evidence is finding by Naqvi *et al.* (2017) that geopolitical fragilities peculiar of the developing world tend to hinder financial sector development. Similarly, authors such as Ayadi *et al*. (2015) and Cherif and Dreger (2016) report that legal institutions, good democratic governance and adequate implementation of financial reforms are necessary for spurring financial sector development. Particularly, while authors such as Voghouei *et al.* (2011) and Khalfaoui (2015) point to the crucial implications of institutions, financial markets, legal tradition, and political economy as factors driving financial sector development, Raza *et al.* (2014), and Cherif and Dreger (2016) identify corruption and rule of law as fundamental ingredients for achieving a robust and burgeoning financial sector.

A recent study by Aluko and Ajayi (2018) find that variables such as population density, trade openness, and capital investment are significant drivers of financial development in Africa. Also, there is evidence that government expenditure boosts financial sector development either through competition or infrastructural development (Naceur *et al.* 2014). Further, studies such as Peprah *et al*. (2019) and Aggarwal *et al.* (2011) find that remittances increase the volume of bank deposits, financial intermediation, and financial sector development. Last but not the least, the literature shows that human capital mattes for financial development (Kodila-Tedika and Asongu, 2015).

3.0 Data and methodology

3.1 Data

The dataset underpinning the analysis is entirely macro and spans 1980 – 2019 for 42 African countries[6]. The variable of interest in this study is financial development and is drawn from

[6] Angola, Benin, Botswana, Burkina Faso, Burundi, Cabo Verde, Cameroon, Central African Republic, Chad, Comoros, Congo, D.R., 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.

the International Monetary Fund's global Financial Development Index (Svirydzenka, 2016). Data on its potential bank-specific, institutional/regulatory, and socioeconomic drivers as elaborated in Section 2 are also taken from the World Bank's Global Financial Development Database (Čihák *et al.*, 2013). Variables such as interest rate spread, lending rate, deposit rate, non-performing loans, Boone indicator, net interest margin, return on asset and stock market capitalization are found in the dataset. Our welfare distribution variables such as the poverty headcount, poverty gap (US$1.90), Gini index, Palma ratio and the Atkinson index are also taken Global Consumption and Income Project (Lahoti *et al.*, 2016). Taking cues from Aluko and Ajayi (2018), we capture the potential relevance of the rise in global interconnectedness, driven chiefly by information technology (Ofori and Asongu, 2021), for financial sector development in Africa. Our globalisation variables such as economic globalisation, social globalisation, political globalisation, financial globalisation and trade globalisation are sourced from the Konjunkturforschungsstelle (KOF) globalisation index (Gygli *et al.*, 2019). Additionally, institutional, structural and macroeconomic variables such as agricultural sector employment, the ease of doing business, financial sector regulation, inflation, government expenditure and unemployment are drawn from the World Bank's World Development Indicators (World Bank, 2021). The definitions of the variables are presented in Table A2 in the Appendices section.

3.2 Estimation strategy

Taking cues from Saura *et al.* (2021), we elaborate the theoretical and empirical foundation of the study in this section. In the first part of this section, we pay attention to the relevance and specifications of the variable selection techniques. The second part also deals with the inferential results. In particular, the first part is in response to growing debate among researchers as to whether it is appropriate to apply classical estimation techniques such as the ordinary least squares (OLS) for inference even in large datasets or resort to machine learning techniques for variable selection and inference. The argument for the former centres on the fact that with appropriate theories researchers can choose the right covariates in regression problems or resort to systematic reviews to identify the salient determinants of the outcome variable (see e.g., Ribeiro-Navarrete *et al.* 2021). However, this may not be feasible if there are more predictors than observations as the required matrix $(X'X)$ becomes invertible. Even if it is feasible, the presence of several predictors, for example, 86 in the case of this study, may cause overfitting of the model.

Overfitting is the inclusion of extra parameters that improve the in-sample fit but increases the out-of-sample prediction error. In the presence of overfitting, even though the attendant estimates are not biased, they are less efficient[7](James *et al.*, 2013). This is because as the variables/features become large, least squares assumptions of no multicollinearity, homoscedasticity and exogeneity typically break down, causing the out of sample error to increase and thus making inference and predictions flawed (James *et al.*, 2013). This partly explains the inconclusive results on variables deemed crucial for driving/predicting financial development. Navigating this econometric blunder requires the use of reliable techniques for variable selection, inferences and prediction.

Such techniques as Tibshirani (1996) argue are efficient regardless of the number of covariates, model specification, nonlinearity and time (Tibshirani, 1996). The relevance of our machine learning techniques in reducing data complexity and aiding decision-making is seen in its application in policy relevant areas such as financial risk analysis (Kou *et al.* 2014), health (Mateen *et al.*, 2020), transportation (Tizghadam *et al.*,2019), games and psychology (Sandeep *et al.*, 2020), bankruptcy prediction (Kou *et al*. 2021) and Large-scale group decision-making (Chao *et al.* 2021). In this study, therefore, we train four alternative shrinkage models— the first three from the lasso family (i.e., the Standard lasso, the Minimum BIC lasso, and Adaptive lasso), and the Elasticnet to achieve the first objective[8]. 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. Next, we perform causal inference on the selected covariates in Objective 1 by running the lasso inferential models: the double-selection linear lasso regression, the partialing-out lasso linear regression, and the partialing-out lasso instrumental variable regression to address Objective 2.

3.2.1 Specification of regularization models

3.2.1.1 Specification of Standard lasso and Minimum BIC lasso models

The Standard lasso variable selection technique was introduced by Tibshirani (1996) to address the poor prediction and inference arising due to discretional selection of covariates in large dataset problems. The key advantages of the Standard lasso over traditional techniques

---

[7] 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.

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

are that it can: (i) enhance model interpretability by eliminating irrelevant predictors; (ii) enhance prediction accuracy, as the elimination of irrelevant predictors reduces model variance without a substantial increase in the bias; and (3) be applied regardless of data dimensionality.

It is imperative to note that the Standard lasso technique yields sound regularization based on a given tuning parameter (λ), which determines the extent of the shrinkage (Tibshirani, 1996; Belloni and Chernozhukov, 2013). In this study, we follow Tibshiran (1996) by specifying the Standard lasso objective function as apparent in Equation (1). This approach runs on the penalty ($\lambda \sum_{j=1}^{\rho} |\beta_j|$), also referred to as the ℓ_1-norm, to obtain $\widehat{\beta}_{slasso}$ defined in Equation (2)

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

$$\hat{\beta}_{slasso} = min\left\{SSE + \lambda \sum_{j=1}^{\rho} |\beta_j|\right\} \tag{2}$$

Where y_{it} is financial development in country *i* in year *t*, X_{it} is a matrix of 86 potential key predictors of financial development. Effective regularization is done by minimising the model sum of square errors with the given ($\lambda \sum_{j=1}^{\rho} |\beta_j|$) or ℓ_1-norm. Therefore, if $\ell_1 = 0$, then $\hat{\beta}_{slasso}$ plunges into the least square estimator[9]. Accordingly, if $\lambda \to \infty$ then all the predictors are eliminated from our model.

For brevity, we point that the specification of the Minimum BIC lasso follows that of the Standard lasso as elaborated above with the same penalty (ℓ_1). It is worth noting, however, that, unlike the Standard lasso, variable selection under the Minimum BIC is based on the model with the least BIC (Schwarz, 1978). Despite the regularization powers of the Standard lasso and Minimum BIC lasso techniques, two key drawbacks have been identified. First, both techniques can be inconsistent as features grow rapidly, and second, the techniques are unable to perform hypothesis tests and confidence intervals.

3.2.1.2 Specification of Adaptive lasso model

The Adaptive lasso technique was introduced by Zou (2006) to address the first regularization shortfall of the Standard lasso and Minimum BIC lasso techniques. Thus, the key contribution of the Adaptive lasso is that it aids sound variable selection even when data attributes grow faster than the number of observations. This is done by adding another

[9] That is no variable is shrank to zero.

property called the '*oracle property*' (z_j) to the ℓ_1-norm. In this study, we apply the Adaptive lasso technique as an alternative to the Standard lasso and Minimum BIC lasso to address Objective 1. To this end, we follow Zou (2006) by minimising the objective function in Equation (3) to obtain ($\hat{\beta}_{Alasso}$) as specified in Equation (4),

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

$$\hat{\beta}_{Alasso} = min\left\{SSE + \lambda \sum_{j=1}^{\rho} z_j |\beta_j|\right\} \tag{4}$$

Where y_{it} is financial development in country *i* in year *t*, X_{it} is a vector of the 86 covariates of financial development and β' are the attendant parameters.

3.2.1.3 Specification of Elasticnet model

The Elasticnet technique draws on the strengths of the Standard lasso and Ridge regression for effective variable selection. The technique is thus built to apply the ℓ_1 and ℓ_2 penalisation norms in variable selection. The strength of the Elasticnet is that in highly correlated covariates, it can produce sparse and consistent regularization than the lasso family algorithms (Zou and Hastie, 2005). Also, with the application of the ℓ_1 and ℓ_2 penalization norms, the Elasticnet becomes flexible in variable selection. The Elasticnet estimator minimises the objective function:

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

Where y_{it}, X_i, and β' in Equation (5) are as defined in Equation (4), and α is an additional Elasticnet penalty parameter[10], which takes on values only in [0,1]. This implies that sparsity occurs only when $0 < \alpha < 1$ and $\lambda > 0$. It is important to point out that in some special cases, the Elasticnet plunges into either the Standard lasso estimator (i.e., when λ=1) or the Ridge estimator (i.e., when λ=0)

3.2.2 Choice of tuning parameter

[10] This adds to the regular λ penalty.

A key concern in regularization is the choice of the tuning parameter (**λ**), which controls the degree of shrinkage. Accordingly, a good value of **λ** is essential for the overall performance of regularization techniques and the attendant prediction results (Schneider and Wagner, 2012). For instance, if **λ** becomes too large, regularization becomes too strong and this can shrink relevant variables to zero. Additionally, if **λ** is set under a researcher's discretion, it can yield '*target sparsity*[11]*'* (Hastie *et al.,* 2019). Therefore, information criteria such as the Cross-validation (CV), Bayesian Information Criterion (BIC), and Akaike Information Criterion (AIC) are usually relied upon to select appropriate **λ** (Tibshirani and Taylor, 2012). For instance, the BIC and AIC are sometimes preferred to CV as they are faster to compute and are less volatile in small samples (Zou *et al.,* 2007). In this study, we rely on the BIC information criterion and CV[12] in determining **λ**.

3.2.3 Specification of lasso inferential models

To provide estimates and confidence intervals on the selected drivers of financial development[13], we apply the lasso inferential techniques. In specifics, we run the double-selection lasso linear regression (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 redundant (weak) covariates as controls (see Chernozhukov *et al*. 2015b). It is worth noting that the lasso inferential techniques consider these controls as irrelevant and therefore, their inferential statistics are not reported (see, Belloni *et al.* 2016).

However, the number of relevant controls selected and the instruments used in cases where there is endogeneity are reported as part of the general regression statistics (Chernozhukov *et al*. 2015a). Further, unlike the variables of interest, which the researcher has no flexibility of adding to or excluding from the model, one can determine the number of controls in the model[14]. The strength of these models is that they are built to produce unbiased and efficient estimates irrespective of data dimensionality, model specification, and multicollinearity.

3.2.3.1 Double-selection lasso linear model

[11] A situation where covariates are selected when a researcher determines the value of **λ**
[12] In this study, we invoke the 10-fold cross-validation.
[13] 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.
[14] 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.

In line with Objective 2 of this study, we follow Belloni *et al.* (2016) and 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 financial development, which is modelled to depend on $\boldsymbol{\psi}$, containing $\boldsymbol{J}$ covariates of interest (i.e., the Elasticnet or lasso selected key drivers of financial development) and $\boldsymbol{\phi}$, which contains $\boldsymbol{p}$ controls (i.e., the redundant predictors of financial development). As indicated in Section 3.2.3, the DSL estimator produces estimates on $\boldsymbol{J}$ while relaxing the estimates for $\boldsymbol{p}$.

3.2.3.2 Partialing-out lasso linear regression

Vis-à-vis the DSL, an added advantage of the partialing-out lasso linear regression (POLR) is that it enhances the efficacy of estimation as the model becomes too complex. Following Belloni *et al.* (2012) and Chernozhukov *et al.* (2015a; 2015b), we specify the POLR estimator as:

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

Where $\boldsymbol{y}$ is financial development, $\boldsymbol{d}$ 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 weak predictors of financial development). Like the DSL, the POLR yields inferential statistics only on the $\boldsymbol{J}$ covariates while relaxing that of the $\boldsymbol{p}$ controls.

3.2.3.3 Partialing-out lasso instrumental-variables regression

We employ the partialing-out lasso instrumental variable regression (POIVLR) to address potential endogeneity concerns in this study. In particular, endogeneity is apparent taking cues from the supply-leading and demand-following hypotheses where financial development and economic growth are considered simultaneous. To address this, we follow Chernozhukov *et al*. (2015a) by specifying our POIVLR model as:

$$y = \Psi\alpha'_d + \Phi\alpha'_f + X\beta' + \varepsilon \tag{8}$$

Where $\boldsymbol{y}$ is financial development; $\boldsymbol{\Psi}$ comprises $\boldsymbol{J_d}$ endogenous covariates of interest; $\boldsymbol{f}$ contains the $\boldsymbol{J_f}$ exogenous covariates of interest; and $\boldsymbol{X}$ contains $\boldsymbol{p_x}$ controls. Allowing for potential endogeneity primarily due to the simultaneity between financial development and economic growth, $\boldsymbol{p_z}$ outside instrumental variables[15] denoted by $\boldsymbol{z}$ that are correlated with $\boldsymbol{d}$ but not with $\boldsymbol{\varepsilon}$ are introduced. Theoretically, the controls and instrument can grow with the sample size; however, $\boldsymbol{\beta}$ and non-zero coefficients in $\boldsymbol{z}$ must be sparse.

3.3 Data engineering and partitioning procedure

Figure A1 shows that 98.8 per cent of the observations are present in our dataset. Mindful of a strongly balanced panel for training algorithms, the *K-nearest* neighbour (KNN) data engineering technique is applied, particularly for variables such as the policy and institutional indicators[16], insurance premium, stock market volatility, and infrastructure quality (see, Figure A2). The KNN is based on the principle that developments in a dataset generally exist in close proximity with other cases that have similar properties (Van Hulse and Khoshgoftaar 2014). The KNN is mostly used when one has no prior knowledge about the distribution of the data. The KNN then selects closest neighbours according to a distance metric, and estimates missing data with the corresponding mean or mode. The mean rule is used to predict missing numerical features while that of missing categorical features is addressed using the mode rule (Pan *et al.* 2015). In this study, therefore, the mean rule is used based on the Minskowski distance as specified in equation (9)

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

Where q is the called the Minkowski coefficient. The Minskowski distance reduces to the Manhattan distances if q =1 and as the Euclidean distance if q =2. Finally, we split the dataset into two parts— the training (70%) and testing (30%) samples by applying the stratified data partitioning method, taking into account the skewed distribution of financial development as apparent in Figure 2.

---

[15] 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).

[16] These are data on net migration, and country policy and institutional scores for macroeconomic management, pubic administration, and financial sector management.

4.0 Presentation and discussion of results

4.1 Exploratory data analysis

For brevity, the exploratory data analysis is limited to the data partitioning results[17], the distribution of financial development, and the summary statistics. Information gleaned from the summary statistics in Table A2[18] shows an average financial development figure of 0.128 in the training set as compared to 0.121 in the testing set. Also, the average remittance inflow into Africa is 4.75 per cent in the training set as compared to 4.02 per cent in the testing set. Additionally, the data shows a mean institutional effectiveness score of 2.967 in the training set compared to 2.938 in the testing set, both shy of the average 3.0. Further, the data shows an average income per capita of US$3730.3 and US$3938.6 in the training and testing sets, respectively.

4.1.1 Data partitioning and distribution of financial development results

Figure 2 shows the 70-30 split of the dataset. It is clear from Figure 2 that financial development follows similar distribution in both the training and testing samples.

[17] That is the distribution of financial development in the training and testing sets.
[18] See Appendices section

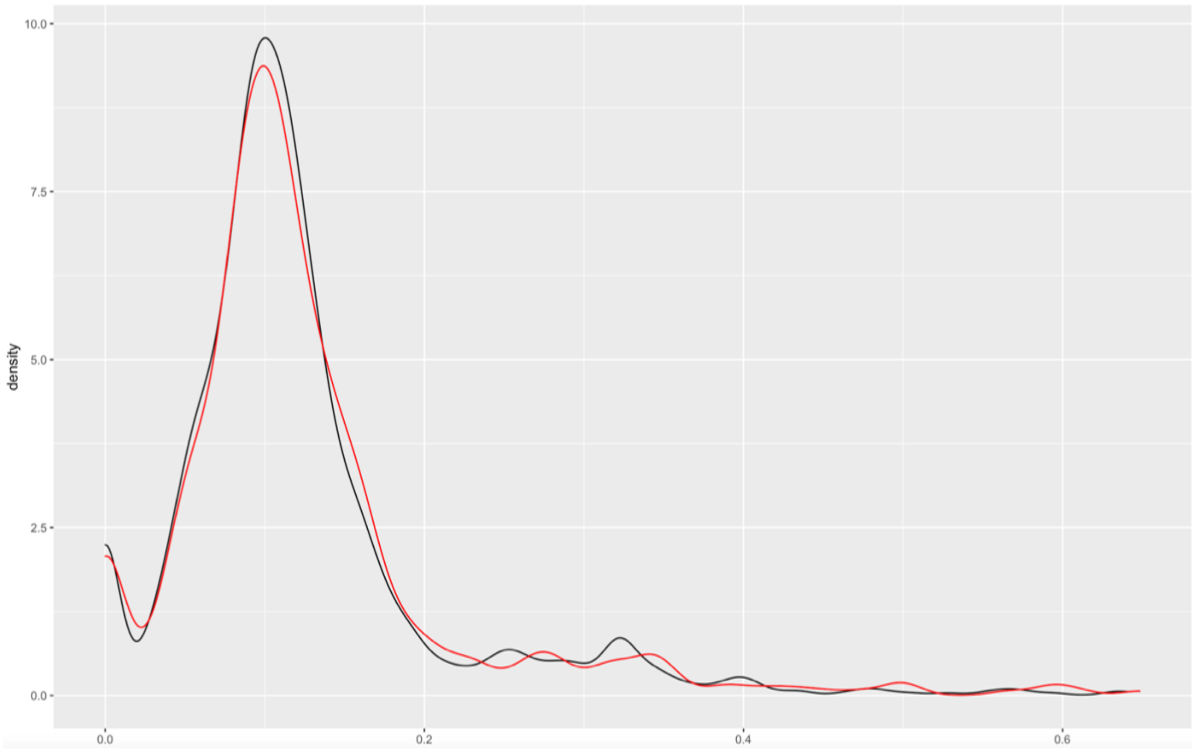

Figure 2: Data partitioning plot, Training (Black) and Test (Red)

The distribution of financial development in Figure 2 as emphasized in Figure 3 (left) is left-skewed. Since skewed distributions can have adverse implications for regularization, financial development is normalised by taking a logarithmic transformation of the series. Figure 3 (right) shows that financial development is more symmetric and less heavy-tailed after the normalisation.

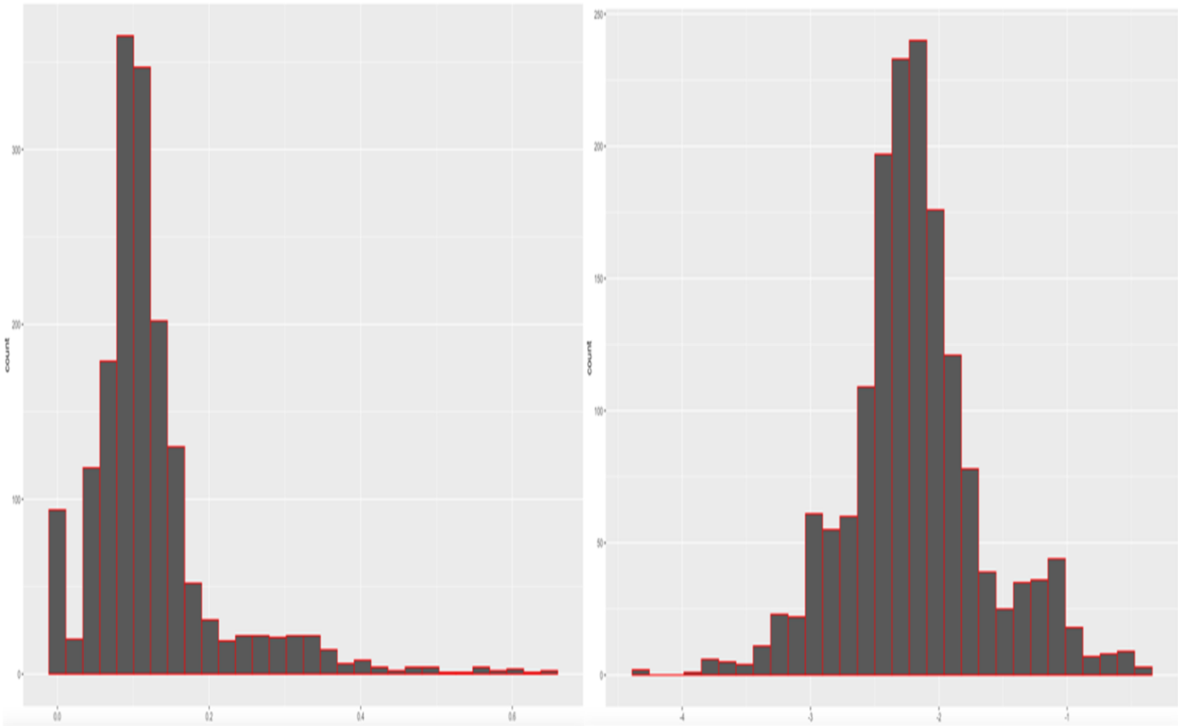

Figure 3: Distribution of financial development at level (left) and its log-transformation (right)

4.2 Regularization results on drivers of financial development in Africa

In this section, results on the first objective are presented. It is evident from Figures 4 – 7 that, the lasso and Elasticnet algorithms select different tuning parameters but a similar number of covariates (i.e., non-zero coefficients) as drivers of financial development. Interestingly, we find that the *Standard lasso* ($\lambda = 0.07$), *Adaptive lasso* ($\lambda = 0.0019$), and *Elasticnet* ($\lambda = 0.07$ and $\alpha = 1$) algorithms select the same number of covariates (17) as drivers of financial development in Africa. A more parsimonious regularization is, however, found in the *Minimum BIC lasso* model, which selects 10 variables out of the 86 covariates. These key covariates are literacy, cell phones, economic growth, economic globalisation, employment, inflation, government expenditure, Z-score, bank overhead cost, and institutional effectiveness (see Table A3 and Figure 5 (right)). For brevity, we present the post estimation tests of cross-validation and coefficient path plots to show the behaviour of the covariates across the four models.

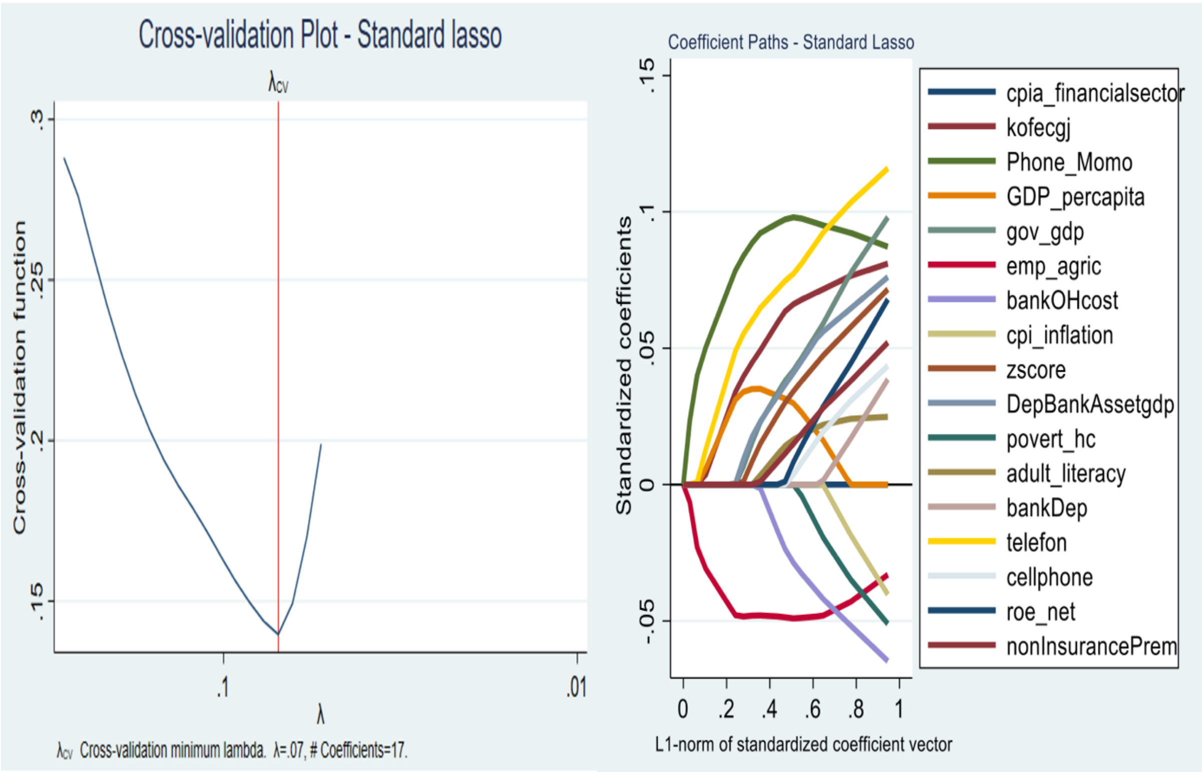

Figure 4: Cross-validation plot (left) and coefficient path plot (right) for Standard lasso

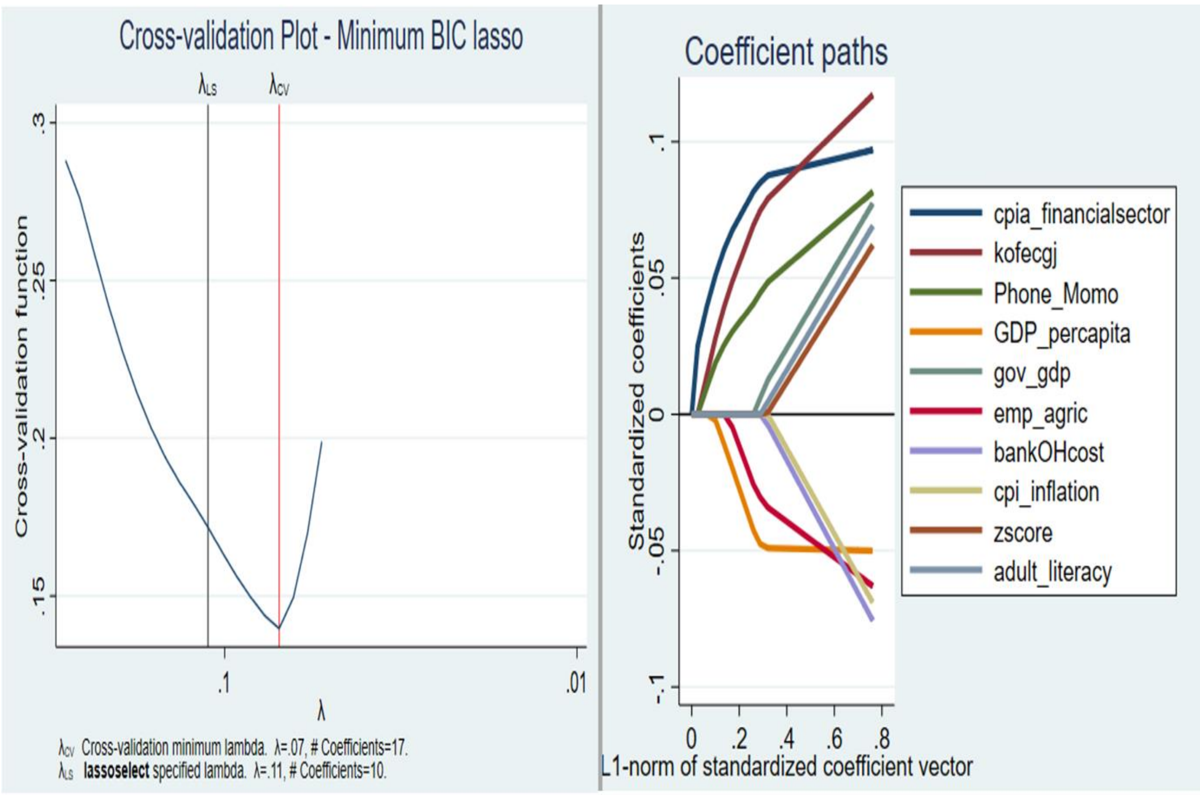

Figure 5: Cross-validation plot (left), and coefficient path plot (right) for Minimum BIC lasso

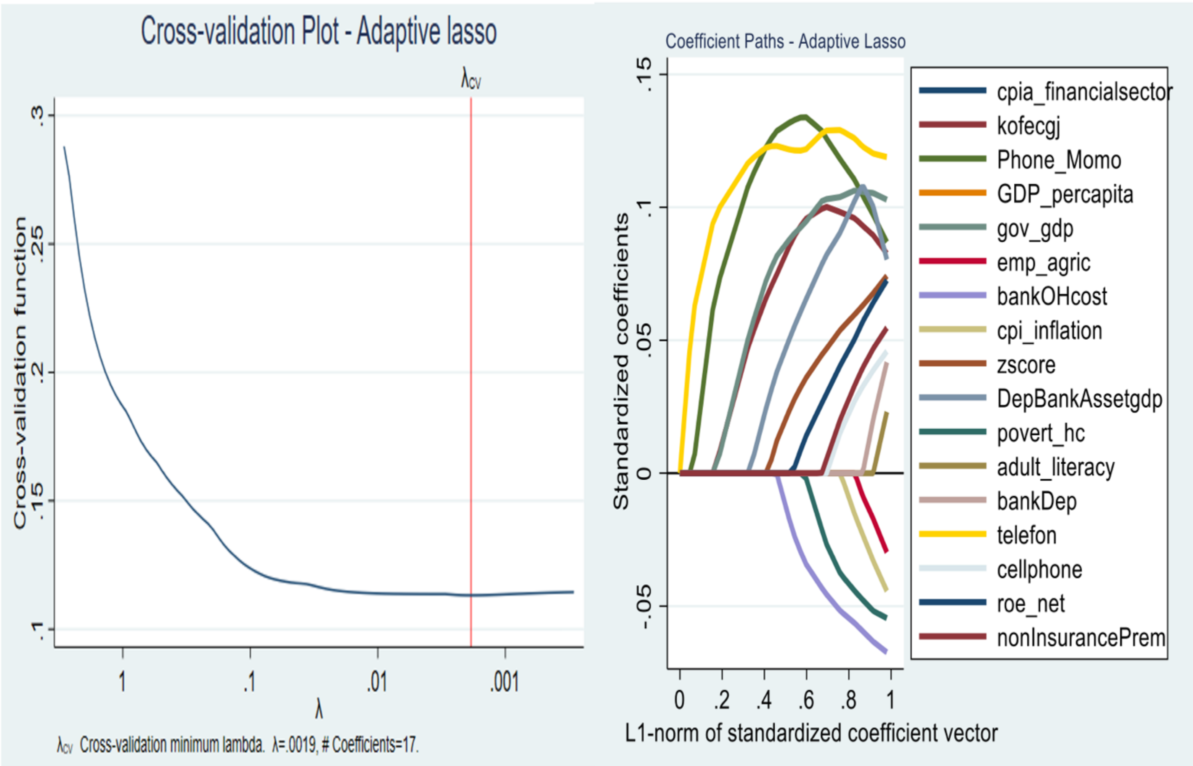

Figure 6: Cross-validation plot (left), and coefficient path plot (right) for Adaptive lasso

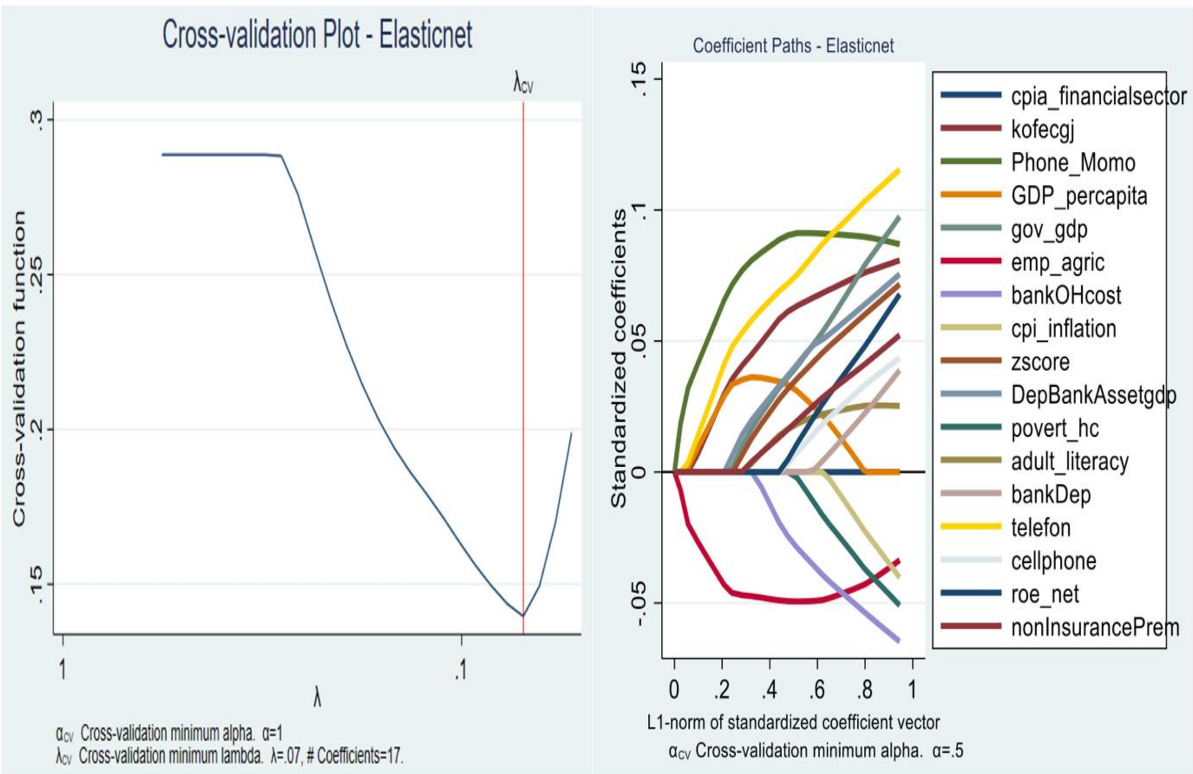

Figure 7: Cross-validation plot (left), and coefficient path plot (right) for Elasticnet

4.3 Inferential results on drivers of financial development in Africa

Using the 10 key predictors of financial development as the variables of interest, we apply the DSL, POLR, and POIVLR estimation techniques to address Objective 2 of the study. The attendant estimates are presented in Table 1. We point out that we rely on the estimates in Column 3 due to its added advantage of addressing endogeneity. Further, aside from the joint significance of the 10 predictors in explaining variations in financial development, the reliability of the estimates is seen in the robustness of the POIVLR to heteroskedasticity, endogeneity, and misspecification.

Table 1: Lasso estimates on the key drivers of financial development in Africa

Variables	(1) DSL lasso	(2) POLR lasso	(3) POIVLR lasso
Institutional effectiveness	0.047***	0.046***	0.082***
	(0.007)	(0.007)	(0.013)
Economic globalisation	0.007***	0.007***	0.009***
	(0.001)	(0.001)	(0.002)
Cell phones	0.007***	0.007***	0.027***
	(0.002)	(0.002)	(0.005)
GDP per capita	0.001	0.001	0.033***
	(0.001)	(0.001)	(0.006)
Government expenditure	0.009***	0.009***	0.017**
	(0.001)	(0.001)	(0.008)
Employment (agriculture)	-0.002***	-0.002***	-0.001
	(0.000)	(0.000)	(0.001)
Overhead cost	-0.015***	-0.015***	-0.022***
	(0.005)	(0.005)	(0.006)
Inflation	-0.005***	-0.005***	-0.003
	(0.001)	(0.001)	(0.002)
Z-score	0.011***	0.011***	0.018***
	(0.001)	(0.001)	(0.003)
Literacy	0.008***	0.008***	0.027***
	(0.001)	(0.001)	(0.004)
Observations	1,628	1,628	1,628
Wald X^2 Statistic	407.14	395	166.75
Wald P-value	0.000	0.00	0.00

CPIA is country and institutional policy assessment score for the financial sector;
*Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1*

We find strong empirical evidence that *literacy* matters for financial sector development in Africa. The result shows that a 1 per cent increase in literacy is associated with a boost in financial development by 0.02 per cent. The significance of literacy (human capital) for financial development follows the proposition that the educated are more likely to invest and/or consume financial products and services. Additionally, as Boopen *et al.* (2021), Kodila-Tedika and Asongu (2015) and Elsherif (2015) point out, the literates are most financially included and are more likely to comprehend financial sector reforms compared to their illiterate counterparts.

Also, we find that cell phones (ICT usage) is also statistically significant in driving Africa's financial sector development[19] agenda. The rise in ICT diffusion has made cell phones a viable and youth-friendly channel for fostering financial development, especially for capturing the financially excluded into the financial sector and achieving a cashless system. Indeed, empirical evidence in Asongu *et al*. (2019) and Asongu (2013) show that cell phone penetration offers cheaper means of achieving financial inclusion, the consumption of financial services and products, and financial sector development. This result also amplifies the finding on literacy as the educated are more likely to use mobile phones and internet banking services. Our result provides optimism particularly, regarding empirical evidence by Jacolin *et al.* (2021) that mobile financial services reduce informality in the developing world.

Further, we find strong evidence that economic globalisation[20] is crucial for Africa's financial sector development. The magnitude of the coefficient indicates that for every 1 per cent improvement in economic globalisation, there is a surge in financial development by 0.009 per cent. This finding corroborates that of Aluko and Ibrahim (2019) and Boopen *et al.* (2011), who provide empirical support that opening up Africa to trade, investment and capital flows can boost financial development. The concern with economic globalisation, however, as Aluko and Ajayi (2018), Mahawiya (2015) and Asongu (2012) argue, is that, it leaves the financial sector more susceptible to cybercrime, money laundering, Ponzi schemes, and global financial crisis spillover.

Also, we find strong empirical evidence that Africa's financial sector grows by 0.017 for every 1 per cent increase in government expenditure. Indeed, in the developing world, empirical contributions such as Filippidis and Katrakilidis (2014) and Aluko and Ibrahim (2019) indicate that government expenditure can boost financial sector performance if the expenditure results in a more lubricated economy. However, excessive government borrowing from the financial system, which is ubiquitous in Africa can result in the crowding-out private investment or inefficient resource allocation (Naceur *et al.*, 2014;

[19] The internet, can, in this case, be a good medium to offer the public a broad range of affordable and quality financial products, services

[20] Economic globalisation comprises tariff, foreign direct investment, trade openness and capital flows across borders.

Cooray. 2011). This means that government expenditure should thus enhance financial infrastructure, especially the development of payment system platforms and services; support for financial innovation and the enhancement of information flow on consumers[21]. Our results also suggest that the highly informal nature of Africa (proxied by agricultural sector employment) hinders financial sector development. Our finding is in line with that of Elgin and Uras (2013). Indeed, in Africa, individuals employed[22] in the agricultural sector are less likely to consume financial services and products continuously due to unsustainable income growth. Particularly, the vulnerabilities in economic activities can be a barrier to financial inclusion and more especially the utilisation of financial market services and products.

The results also show that financial strength/stability (Z-score), which has a marginal effect of 0.01 per cent, and financial institutions' overhead cost ($\beta = 0.02$) are also germane for financial sector development. The significance of the former signifies that building a robust system for reducing risk, improving intra-firm information flow while breeding competition in the financial system could prove crucial. The latter, as Beck and Levine (2005) argue, also signifies the relevance of prudent macroeconomic management and financial system supervision/regulation, which can ultimately lead to a reduction in accounting fees, advertising, insurance fees, cost of borrowing, legal fees, rent, supplies, taxes, and utilities. In line with this finding is the statistically significant effect of institutional effectiveness for financial development. The result is remarkable (0.08). Considering the underdeveloped nature of Africa's financial system, the revision of prudential standards as well as improvement in on-site and off-site supervision is worthwhile. Additionally, a sound legal and regulatory framework for financial consumer protection as Cherif and Dreger (2016) and Ayadi *et al*. (2015) argue could prove crucial for boosting consumer confidence in the financial system.

5.0 Conclusion

The study employs machine learning techniques for identifying the key drivers of financial development in 42 African countries. Using a dataset containing 86 potential predictors of financial development for the period 1980 – 2019, we run four machine learning regularization models— the *standard lasso,* the *Minimum BIC lasso,* the *Adaptive lasso,* and the *Elasticnet* to show that literacy, cell phones, economic growth, economic globalisation,

[21] Tightening the national identification system

[22] Even the few who are financially included are more likely to default on loans plausible due to vulnerabilities in employment.

employment (agriculture), inflation, government expenditure, Z-score, bank overhead cost, and institutional effectiveness are crucial for driving Africa's financial sector development. Evidence from the lasso inferential estimation techniques also show that, but for inflation and employment, all the selected covariates are statistically significant in driving Africa's financial sector development. Our findings show that machine learning techniques can be applied to reduce data complexity and aid sound decision making. In particular, the approach solves the problem of selection bias and inconclusive results by eliminating researcher discretion in the selection of variables in large data regression problems

For policy, we recommend that strategic government expenditure, preferably one that supplements the private sector's effort in human capital development, financial infrastructure, and economic growth be enhanced to foster greater financial activity, inclusion, and development. Also, in line with the youthful nature of Africa's population and the technological progress, government intervention is required in reducing the cost of internet access while broadening telecommunication network access for the rural folks who are more likely to use mobile money services. Various governments should thus liaise with financial institutions, markets, and telecommunication service providers to make financial products and services accessible via mobile phones. Additionally, it is recommended that financial institutions and markets provide greater incentives, for example, through low charges or discounts for clients using cell phones for transactions. Finally, we recommend that regulation and supervision institutions be strengthened to enhance information flow, consumer protection, and confidence in the financial system considering the rise in the economic integration of Africa following the implementation of the Africa Continental Free Trade Area. This can be enhanced if international bodies such as the World Bank and African Development Bank support Africa's monetary authorities to strengthen the secured transactions and collateral frameworks; and the insolvency regimes.

For the academic community, researchers can draw on our contribution to identify which variables matter for addressing poverty and inequality. This could prove crucial for making resources count considering the huge investment made by African governments and development partners such as the World Bank and African Development Bank in their quest to alleviate poverty and income inequality. Additionally, considering the underdeveloped nature of the region's financial market, researcher can draw on our contribution to narrow the scope and inform policy as to which the key drivers of financial market development are. Also, following the implementation of the African Continental Free Trade Area agreement, other researcher can employ the techniques used in this study to inform policy as to which goods/products the African countries should produce to diversify export.

A conspicuous drawback to this study is that we do not consider all African countries on grounds of data limitation. For future research, this study could be executed at the regional level, for instance, in the West African Monetary Zone, to guide policy actions.

References:

Acemoglu, D., Johnson, S., & Robinson, J. A. (2001). The colonial origins of comparative development: An empirical investigation. *American economic review*, *91*(5), 1369-1401.

Almarzoqi, R., Naceur, M.S.B. & Kotak, A. (2015). What matters for financial development and stability? (No. 15-173). *International Monetary Fund*.

Adu, G., Marbuah, G., & Mensah, J. T. (2013). Financial development and economic growth in Ghana: Does the measure of financial development matter? *Review of Development Finance*, *3*(4), 192–203.

Akinlo, A. E., & Egbetunde, T. (2010). Financial development and economic growth: The experience of 10 sub-Saharan African countries revisited. *The review of finance and banking*, *2*(1).

Aluko, O. A., & Ibrahim, M. (2020). On the macroeconomic determinants of financial institutions development in sub-Saharan Africa. International Review of Economics, 67(1), 69-85

Aluko, O. A., & Ajayi, M. A. (2018). Determinants of banking sector development: Evidence from Sub-Saharan African countries. *Borsa Istanbul Review*, *18*(2), 122-139.

Aggarwal, R., Demirgu ̈ç-Kunt, A., & Pería, M. S. M. (2011). Do remittances promote financial development. *Journal of Development Economics*, 96(2), 255-264.

Arcand, J., Berkes, E., & Panizza, U. (2015). Too much finance? IMF Working paper WP/12/161. International Monetary Fund.

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.

Asongu, S. A. (2012). Bank efficiency and openness in Africa: Do income levels matter? *The Review of Finance and Banking*, 4(2), 115–122

Asongu, S. A. (2013). How has mobile phone penetration stimulated financial development in Africa? *Journal of African Business*, 14(1), 7-18.

Ayadi, R., Arbak, E., Naceur, S. B., & De Groen, W. P. (2015). Determinants of financial development across the Mediterranean. In Economic and Social Development of the Southern and Eastern Mediterranean Countries (159–181). Springer

Barajas, M.A., Beck, T., Dabla-Norris, M.E. & Yousefi, M.R. (2013). Too cold, too hot, or just right? Assessing financial sector development across the globe (No. 13-81). *International Monetary Fund.*

Baltagi, B. H., Demetriades, P. O., & Law, S. H. (2009). Financial development and openness: Evidence from panel data. *Journal of Development Economics*, 89(2), 285-296.

Beck, T. (2012). The role of finance in economic development–benefits, risks, and politics. *Oxford Handbook of Capitalism,161-203.*

Beck, T., Demirgu ̈ç-Kunt, A., & Levine, R. (2003). Law, endowments and finance. *Journal of Financial Economics*, 70(2), 137-181.
Beck, T., & Levine, R. (2005). Legal institutions and financial development. In Handbook of new institutional economics (pp. 251-278). Springer, Boston, MA.
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., Chernozhukov, V., & Wei, Y. (2016). Post-selection inference for generalized linear models with many controls. *Journal of Business & Economic Statistics*, *34*(4), 606-619.
Belloni, A., & Chernozhukov, V. (2013). Least squares after model selection in high-dimensional sparse models. Bernoulli, 19(2), 521-547.
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.
Bittencourt, M. (2011). Inflation and financial development: Evidence from Brazil. Economic Modelling, 28(1–2), 91–99.
Boyd, J. H., Levine, R., & Smith, B. D. (2001). The impact of inflation on financial sector performance. *Journal of monetary Economics*, *47*(2), 221-248.
Boopen, S., Kesseven, P., Jashveer, H., & Binesh, S. (2011). Determinants of financial development: The case of Mauritius. *Finance and Corporate Governance Conference*.
Chao, X., Kou, G., Peng, Y., & Viedma, E. H. (2021). Large-scale group decision-making with non-cooperative behaviors and heterogeneous preferences: an application in financial inclusion. *European Journal of Operational Research*, *288*(1), 271-293.
Cherif, M., & Dreger, C. (2016). Institutional determinants of financial development in MENA countries. *Review of Development Economics*, *20*(3), 670–680.
Chernozhukov, V., Hansen, C. & Spindler M. (2015a). Valid Post-Selection and Post-Regularization Inference: An Elementary. General Approach. *Annual Review of Economics*, 7(1), 649–688.
Chernozhukov, V., Hansen, C., & Spindler, M. (2015b). Post-selection and post-regularization inference in linear models with many controls and instruments. *American Economic Review, 105*(5), 486-90
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*.
Cooray, A. (2011). The role of the government in financial sector development. *Economic Modelling*, *28*(3), 928–938.
Demirgüç-Kunt, Asli, Leora Klapper, Dorothe Singer, Saniya Ansar, and Jake Hess. (2018). The Global Findex Database 2017: Measuring Financial Inclusion and the Fintech Revolution. *Washington, DC: World Bank*.
Djankov, S., McLiesh, C., & Shleifer, A. (2007). Private credit in 129 countries. *Journal of Financial Economics*, 84(2), 299-329.
Elgin, C., Uras, B.R. Is informality a barrier to financial development? SERIEs 4, 309–331
Elsherif, M. A. (2015). The determinants of financial development: Empirical evidence from Egypt. *The Macrotheme Review*, *4*(3), 69–87.
Fowowe, B. (2014). Law and finance revisited: Evidence from African countries. *South African Journal of Economics*, 82(2), 193-208.
Filippidis, I., & Katrakilidis, C. (2014). Institutions, policy and banking sector development: A reassessment. *Finance a Uver*, *64*(6), 501.
Gygli, S., Haelg, F., Potrafke, N., & Sturm, J. E. (2019). The KOF Globalisation Index –

Revisited. *Review of International Organizations*, *14*(3), 543–574.
Hastie, T., Tibshirani, R., & Wainwright, M. (2019). *Statistical learning with sparsity: the lasso and generalizations*. Chapman and Hall/CRC.
Huybens, E., & Smith, B. D. (1999). Inflation, financial markets and long-run real activity. *Journal of Monetary Economics*, 43(2), 283-315.
Madsen, J. B., Islam, M. R., & Doucouliagos, H. (2018). Inequality, financial development and economic growth in the OECD, 1870–2011. *European Economic Review*, 101, 605-624.
Ibrahim, M., & Sare, Y. A. (2018). Determinants of financial development in Africa: How robust is the interactive effect of trade openness and human capital? *Economic analysis and policy*, *60*, 18-26.
Jacolin, L., Keneck Massil, J., & Noah, A. (2021). Informal sector and mobile financial services in emerging and developing countries: Does financial innovation matter? World Economy; 00: 1–35. https://doi.org/10.1111/twec.13093
James, G., Witten, D., Hastie, T. & Tibshirani, R. (2013). Linear model selection and regularization. In *An introduction to statistical learning* (203-264). Springer, New York, NY.
Jedidia, K. B., Boujelbène T., & Helali, K. (2014). Financial Development and Economic Growth: New Evidence from Tunisia. *Journal of Policy Modeling*, 36 (5): 883–898.
Khalfaoui, H. (2015). The determinants of financial development: Empirical evidence from developed and developing countries. *Applied Economics and Finance*, 2(4), 1–9.
Kim, D. H., & Lin, S. C. (2010). Dynamic relationship between inflation and financial development. *Macroeconomic Dynamics*, 14(3), 343-364.
Kodila-Tedika, O., & Asongu, S. A. (2015). The effect of intelligence on financial development: A cross-country comparison. *Intelligence*, 51, 1-9.
Kou, G., Peng, Y., & Wang, G. (2014). Evaluation of clustering algorithms for financial risk analysis using MCDM methods. *Information Sciences*, *275*, 1-12.
Kou, G., Xu, Y., Peng, Y., Shen, F., Chen, Y., Chang, K., & Kou, S. (2021). Bankruptcy prediction for SMEs using transactional data and two-stage multiobjective feature selection. *Decision Support Systems*, 140, 113429.
Kou, Gang, Özlem Olgu Akdeniz, Hasan Dinçer, and Serhat Yüksel. "Fintech investments in European banks: a hybrid IT2 fuzzy multidimensional decision-making approach." *Financial Innovation* 7, no. 1 (2021): 1-28.
La Porta, R., Lopez-de-Silanes, F., Shleifer, A., & Vishny, R. W. (1998). Law and finance. *Journal of political economy*, *106*(6), 1113-1155.
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.
Law, S. H., & Singh, N. (2014). Does too much finance harm economic growth? Journal of Banking & Finance, 41, 36-44.
Mahawiya, S. (2015). Financial sector development, inflation and openness: A comparative panel study of ECOWAS and SADC. *Economic Research Southern Africa Working Paper* (528) 1-37.
Marcelin, I. & Mathur, I. (2014). Financial development, institutions and banks. *International Review of Financial Analysis*, *31*, 25-33.
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.
McKinnon, R.I., (1973). Money and capital in economic development. *Brookmgs Instttutton, Washmgton, DC.*
Mtar, K., & Belazreg, W. (2020). Causal Nexus Between Innovation, Financial Development,

and Economic Growth: The Case of OECD Countries. *Journal of the Knowledge Economy*, 1-32.
Naceur, S. B., Cherif, M., & Kandil, M. (2014). What drives the development of the MENA financial sector? *Borsa Istanbul Review*, 14(4), 212-223.
Naqvi, T., Waheed, A., Mahmood, H., & Rafique, M. (2017). Impact of Political Instability on Financial Development of Pakistan. *International Journal of Management Sciences and Business Research,* 6(4), 1 – 13
Nguyen, C. P., Su, T. D., & Doytch, N. (2020). The drivers of financial development: Global evidence from internet and mobile usage. Information Economics and Policy, 53, 100892.
Ofori I. K., & Asongu S. A. (2021b). ICT Diffusion, Foreign Direct Investment and Inclusive Growth in Sub-Saharan Africa. Telematics and Informatics, 65, 101718.
Padachi, K., Rojid, S., & Seetanah, B. (2008). Investigating into the factors that influence the adoption of internet banking in Mauritius. *Journal of Internet Business,* (5), 98 – 120.
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., Ofori, I.K., & Asomani, A.N. (2019). Financial development, remittances and economic growth: A threshold analysis. *Cogent Economics & Finance*, *7*(1),1625107.
Government of Ghana (2018). National financial inclusion and development strategy, 2018 – 2023. *Ministry of Finance*, Accra, Ghana.
Raza, S. H., Shahzadi, H., & Akram, M. (2014). Exploring the determinants of financial development (using panel data on developed and developing countries). *Journal of Finance and Economics,* 2(5), 166–172.
Ribeiro-Navarrete, S., Saura, J. R., & Palacios-Marqués, D. (2021). Towards a new era of mass data collection: Assessing pandemic surveillance technologies to preserve user privacy. *Technological Forecasting and Social Change*, *167*, 120681.
Robinson, J. C. (1952). The generalization of the general theory. In The rate of interest and other essays. London: Macmillan.
Rousseau, P. L., & Wachtel, P. (2002). Inflation thresholds and the finance-growth nexus. *Journal of International Money and Finance*, 21, 777-795.
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.
Sanusi, K. A., Meyer, D., & Ślusarczyk, B. (2017). The relationship between changes in inflation and financial development. *Polish Journal of Management Studies*, 16(2), 253–265.
Saura, J. R. (2021). Using data sciences in digital marketing: Framework, methods, and performance metrics. *Journal of Innovation & Knowledge*, *6*(2), 92-102.
Saura, J. R., Ribeiro-Soriano, D., & Palacios-Marqués, D. (2021). From user-generated data to data-driven innovation: A research agenda to understand user privacy in digital markets. *International Journal of Information Management*, 60, 102331.
Schneider, U., & Wagner, M. (2012). Catching growth determinants with the adaptive lasso. *German Economic Review*, 13(1), 71-85.
Schwarz, G. (1978). Estimating the dimension of a model. *Annals of statistics*, *6*(2), 461-464.
Shabbir, B., Jamil, L., Bashir, S., Aslam, N., & Hussain, M. (2018). Determinants of Financial Development. A Case Study Of Pakistan. A Case Study of Pakistan.
Shaw, E. S. (1973). Financial deepening and economic development. *Oxford University Press, New York.*
Svirydzenka, K. (2016). Introducing a New Broad-Based Index of Financial Development.

IMF Working Paper No. 16/5.
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.
Van Hulse, J., & Khoshgoftaar, T. M. (2014). Incomplete-case nearest neighbor imputation in software measurement data. *Information Sciences*, *259*, 596-610.
Voghouei, H., Azali, M., & Jamali, M. A. (2011). A survey of the determinants of financial development. Asian-Pacific *Economic Literature*, 25(2), 1–20.
World Bank (2020). World Development Indicators. *Washington, DC*: World Bank.
World Bank (2019). Global Financial Development Report 2019/2020: Bank Regulation and Supervision a Decade after the Global Financial Crisis. *Washington, DC*: World Bank.
World Bank (2018). Developing and Operationalizing a National Financial Inclusion Strategy: Toolkit. *Washington, DC.* World Bank,
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

Variables	Definition	Source
unempl	Unemployment, total (% of total labour force)	WDI
rer	Real effective exchange rate index (2010 = 100)	WDI
povert hc	Poverty headcount ratio at national poverty lines (% of population)	WDI
Povertyhc_mid	Poverty headcount ratio at $3.20 a day (2011 PPP) (% of population)	WDI
Povertyhc_low	Poverty headcount ratio at $1.90 a day (2011 PPP) (% of population)	WDI
urbanization	Annual population growth rate in urban centres (% population)	WDI
popgrof	Annual population growth rate in rural centres (% population)	WDI
exr	Nominal exchange rate, dollar-local currency rate	WDI
noda	Net official development assistance (%GNI)	WDI
cellphone	Active mobile phone subscription (mobile money enabled)	WDI
logisticqua overal	Logistics performance index: Overall (1=low to 5=high)	WDI
literacy adult	Literacy rate, adult total (% of people ages 15 and above)	WDI
labforce pr	Labour force participation rate, total (% of total population ages 15-64)	WDI
transport invest	Investment in transport with private participation (current US$)	WDI
inflation	End-of-period inflation (%)	WDI
hci	Human Capital Index (HCI) (scale 0=lowest; 1=Highest)	WDI
house spend	Household final consumption expenditure (annual % growth)	WDI
grossavings	Adjusted annual gross savings (% of GNI)	WDI
Firmsbank_invest	Firms using banks to finance investments (%)	GFDD
gfcf	Gross fixed capital formation	WDI
gov gdp	Government consumption expenditure (%GDP)	GFDD
internet	Secure internet servers (per 1 million people)	WDI
gpc	GDP per capita, US$ 2017 (constant)	WDI
gdpg	GDP growth (annual %)	WDI
fdi	Foreign direct investment, net inflows (%GDP)	WDI
telefon	Telephone subscription per 1000 people	GFDD
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
cpia publicmgt	Public sector management and institutions cluster average (1=low to 6=high)	CPIA
cpia macro	Macroeconomic management rating (1=low to 6=high)	CPIA
cpia finsector	Financial sector management rating (1=low to 6=high)	CPIA
debt	Overall national debt (%GDP)	WDI
moneyg	Money supply growth (M2+)	GFDD
kofgidj	KOF. overall globalisation index (de jure)	KOF. Index
kofecgj	KOF. economic globalisation index (de jure)	KOF. Index
koffindj	KOF. financial globalisation index (de jure)	KOF. Index
palma	Palma ratio, inequality indicator	GCIP
theil	Theil index, inequality indicator	GCIP
gini	Gini index, inequality indicator	GCIP
bank5	5-bank asset concentration	GFDD
formalAcc	Account at a formal financial institution (% age 15+)	GFDD
atm	Automated Teller Machines per 100,000 adults	GFDD
bankAcc	Bank accounts per 1,000 adults	GFDD
bankBran	Bank branches per 100,000 adults	GFDD
bankCaptAsset	Bank capital to total assets (%)	GFDD
bankConcent	Bank concentration (%)	GFDD

bankCostInc	Bank cost to income ratio (%)	GFDD
bankCreditDep	Bank credit to bank deposits (%)	GFDD
bankDep	Bank deposits to GDP (%)	GFDD
irs	Bank lending-deposit spread calculated as difference between lending and deposit interest rates	GFDD
nim	Bank net interest margin (%)	GFDD
banknonIntInc	Bank noninterest income to total income (%)	GFDD
npl	Bank non-performing loans to gross loans (%)	GFDD
bankOHcost	Bank overhead costs to total assets (%)	GFDD
bankRegCap	Bank regulatory capital to risk-weighted assets (%)	GFDD
roa net	Bank return on assets (%, after tax)	GFDD
roe net	Bank return on equity (%, after tax)	GFDD
zscore	Bank Z-score, financial system stability	GFDD
bankCrisis	Banking crisis dummy (1=banking crisis, 0=none)	GFDD
Boone	Boone indicator (Banking efficiency)	GFDD
Cpi	Inflation (consumer price index, 2005 = 100)	GFDD
GovStateCredit	Credit to government and state-owned enterprises to GDP (%)	GFDD
DepBankAsset	Deposit money bank assets to deposit money bank assets and central bank assets (%)	GFDD
DepBankAssetgdp	Deposit money banks' assets to GDP (%)	GFDD
credit	Private credit by deposit money banks and other financial institutions to GDP (%)	GFDD
onlinepayment	Electronic payments used to make payments (% age 15+)	GFDD
finsystemDep	Financial system deposits to GDP (%)	GFDD
foreignBankAsset	Foreign bank assets among total bank assets (%)	GFDD
foreignBanks	Foreign banks among total banks (%)	GFDD
Hstats	H-statistics, Banking sector competition	GFDD
insuranceAsset	Insurance company assets to GDP (%)	GFDD
lerner	Lerner index, market power of financial institutions	GFDD
insurancePrem	Life insurance premium volume to GDP (%)	GFDD
phonePayment	Mobile phone for paying bills online	GFDD
phoneMomo	Mobile phone (mobile money capable)	GFDD
nonBankFinsInsti	Nonbank financial institutions' assets to GDP (%)	GFDD
nonInsurancePrem	Non-life insurance premium volume to GDP (%)	GFDD
remit	Remittance inflows to GDP (%)	GFDD
stockMktcap	Stock market capitalization to GDP (%)	GFDD
stockMktreturn	Stock market return (%, year-on-year)	GFDD
stockMktValue	Stock market total value traded to GDP (%)	GFDD
stockMktTurnover	Stock market turnover ratio (%)	GFDD
stockPxVol	Stock price volatility index	GFDD
FD	Financial development index	FD Index
infrastr qua	Infrastructure quality score	WDI

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
Source: Author's construct (2021)

Table A2: Summary Statistics for Training and Testing sets

Variables	Obs	Mean Training Set	Mean Testing Set	Std. Dev. Training Set	Std. Dev Testing Set	Min Training Set	Min Testing Set	Max Training Set	Max Testing Set
unempl	1680	7.842	7.442	7.799	7.218	.3	.3	37.94	37.976
rer	1680	194.541	200.538	153.382	125.636	49.296	46.021	3520.534	2182.799
povert hc	1680	48.656	47.775	14.233	13.87	7.9	7.9	73.2	73.2
povertyhc mid	1680	68.978	68.929	23.951	24.58	2.2	3.1	98.5	98.5
povertyhc low	1680	49.22	49.445	24.844	25.288	.2	.4	94.3	94.3
urbanization	1680	39.102	38.895	14.19	13.695	10.884	10.838	92.697	100
popgrof	1680	2.554	2.591	.984	1.014	-6.766	-5.539	7.902	8.118
exr	1680	401.785	431.028	1170.927	1554.637	0	0	19068.417	18498.601
noda	1680	11.301	11.39	11.518	11.544	-.251	-.188	94.946	78.707
cellphone	1680	24.473	23.483	39.436	38.128	0	0	198.152	163.875
logisticqua overal	1680	2.395	2.38	.323	.29	0	1.61	3.775	3.67
literacy adult	1680	57.12	56.824	21.396	21.259	0	10.895	95.868	95.868
labforce pr	1680	70.1	69.649	11.54	11.338	42.388	42.381	92.453	92.453
transport invest	1680	3.140e+08	3.480e+08	6.040e+08	6.270e+08	0	0	3.483e+09	3.483e+09
inflation	1680	43.272	20.781	823.613	159.942	-13.057	-11.686	23773.132	4129.17
hci	1680	.394	.395	.069	.076	0	0	.678	.678
house spend	1680	1.147	.641	8.03	9.041	-46.068	-35.333	62.472	87.014
grossavings	1680	16.052	16.527	17.188	17.52	-70.263	-69.534	87.096	87.096
national_expend	1680	109.561	109.306	18.184	16.768	63.638	51.452	261.428	212.246
gfcf	1680	20.937	21.505	9.847	11.413	0	-2.424	89.386	93.547
gov gdp	1680	14.74	14.95	6.373	7.105	0	0	44.486	51.975
internet	1680	.855	1.146	5.396	4.688	-47.503	-26.412	37.536	28.676
gpc	1680	3730.328	3938.626	4183.257	4638.931	436.72	469.189	29223.465	26421.941
gdpg	1680	3.453	3.818	5.532	4.785	-50.248	-23.983	35.224	33.629
fdi	1680	2.753	3.124	5.538	7.259	-8.703	-28.624	103.337	86.989
telefon	1680	210000	155000	730000	487000	0	0	5492840	4961740
emp ind	1680	12.834	12.447	8.586	8.245	1.505	1.465	43.114	42.903
emp agric	1680	54.7	55.437	22.366	21.275	4.6	4.838	92.298	92.303
ease	1680	134.035	137.061	40.567	39.772	13	13	184	184
cpia publicmgt	1680	3.028	3	.468	.454	2	2	4.1	4
cpia macro	1680	3.668	3.647	.642	.65	2	1.5	5	5
cpia finsector	1680	2.967	2.938	.426	.428	2	2	4	4
debt	1680	104.458	108.634	104.063	105.237	0	0	289.845	289.845
moneyg	1680	67.386	77.744	451.498	472.302	-29.245	-99.864	6968.922	4105.573
kofgidj	1680	41.341	40.806	11.407	11.25	0	13.308	80.993	81.288
kofecgj	1680	34.639	34.386	10.939	11.164	0	10.514	78.365	81.49
koffindj	1680	40.369	40.101	13.957	14.232	0	6.073	80.37	81.357
palma	1680	7.358	7.179	3.696	3.694	0	0	30.065	30.065
theil	1680	.686	.676	.13	.127	0	0	1.164	1.165
gini	1680	53.768	52.733	19.511	20.142	0	0	86.276	86.832
bank5	1680	92.792	92.231	12.762	13.224	0	0	100	100
formalAcc	1680	22.783	22.382	17.066	16.687	0	0	89.495	89.495
atm	1680	6.391	6.617	11.112	11.509	0	0	79.164	71.801
bankAcc	1680	191.574	191.552	371.524	363.474	0	0	2084.59	2019.34
bankBran	1680	4.377	4.921	7.117	8.529	0	0	52.329	53.348
bankCaptAsset	1680	10.664	10.47	3.423	3.215	0	0	23.677	22.33
bankConcent	1680	83.454	83.51	18.318	18.823	0	0	100	100

bankCostInc	1680	62.831	63.955	28.114	31.057	0	0	218.087	218.087
bankCreditDep	1680	87.669	88.705	51.085	55.059	0	0	397.115	388.545
bankDep	1680	24.866	28.277	66.217	89.83	0	0	883.404	972.186
irs	1680	9.513	9.824	9.606	10.597	0	0	80.333	80.333
nim	1680	8.336	8.43	5.706	5.996	0	0	39.21	28.982
banknonIntInc	1680	43.509	44.603	16.927	16.708	0	0	95.34	90.123
npl	1680	13.538	13.427	13.27	12.988	0	0	74.1	74.1
bankOHcost	1680	6.361	6.267	5.019	4.139	0	0	89.423	28.639
bankRegCap	1680	17.205	16.831	7.199	7.118	0	0	43.4	42.203
roa net	1680	1.784	1.705	2.689	2.935	-15.047	-15.047	12.106	9.569
roe net	1680	18.921	18.762	23.897	25.463	-93.62	-93.62	160.344	126.138
zscore	1680	10.923	10.784	8.057	7.66	0	0	96.68	47.341
bankCrisis	1680	.049	.074	.216	.262	0	0	1	1
Boone	1680	-.048	-.032	.184	.274	-1.022	-2.541	1.607	1.607
cpi	1680	60.866	57.863	46.862	46.141	0	0	410.94	349.819
GovStateCredit	1680	5.206	5.57	7.238	8.309	0	0	71.28	60.47
DepBankAsset	1680	67.289	66.643	25.401	25.646	0	0	100	100
DepBankAssetgdp	1680	22.492	23.829	37.339	54.799	0	0	661.731	892.896
credit	1680	22.177	22.605	38.316	43.439	0	0	328.493	361.763
onlinepayment	1680	21.328	20.645	18.195	17.731	0	0	76.411	76.411
finsystemDep	1680	24.97	28.367	66.197	89.816	0	0	883.404	972.186
foreignBankAsset	1680	55.137	55.098	27.926	27.779	0	0	100	100
foreignBanks	1680	45.894	45.573	23.966	23.678	0	0	100	100
Hstats	1680	.504	.502	.232	.233	-.036	-.107	1.431	1.431
insuranceAsset	1680	6.59	6.26	11.507	10.539	0	0	69.049	60.193
lerner	1680	.295	.295	.175	.172	-.386	-.212	.64	.599
insurancePrem	1680	.726	.536	1.983	1.503	0	0	14.52	15.381
phonePayment	1680	3.813	3.624	5.217	5.092	0	0	37.105	37.105
phoneMomo	1680	10.471	10.09	12.982	13.046	0	0	50.122	50.122
nonBankFinsInsti	1680	7.868	6.488	14.663	9.151	0	0	119.855	112.484
nonInsurancePrem	1680	.891	.882	1.439	1.529	0	0	14.723	14.013
remit	1680	4.75	4.02	20.082	15.28	0	0	232.217	235.924
stockMktcap	1680	17.466	15.388	34.745	26.907	0	0	270.278	328.361
stockMktreturn	1680	12.249	11.061	18.889	18.356	-30.365	-55.016	81.91	71.516
stockMktValue	1680	2.438	1.911	8.728	6.159	0	0	102.462	123.245
stockMktTurnover	1680	4.73	4.537	5.351	4.6	0	0	50.346	35.766
stockPxVol	1680	11.073	10.986	5.769	5.483	0	0	43.1	31.877
FD	1680	.128	.121	.095	.082	0	0	.648	.641
infrastr qua	1680	3.454	3.448	.757	.77	1.372	1.8	5.417	5.641

Source: Author's construct (2021)

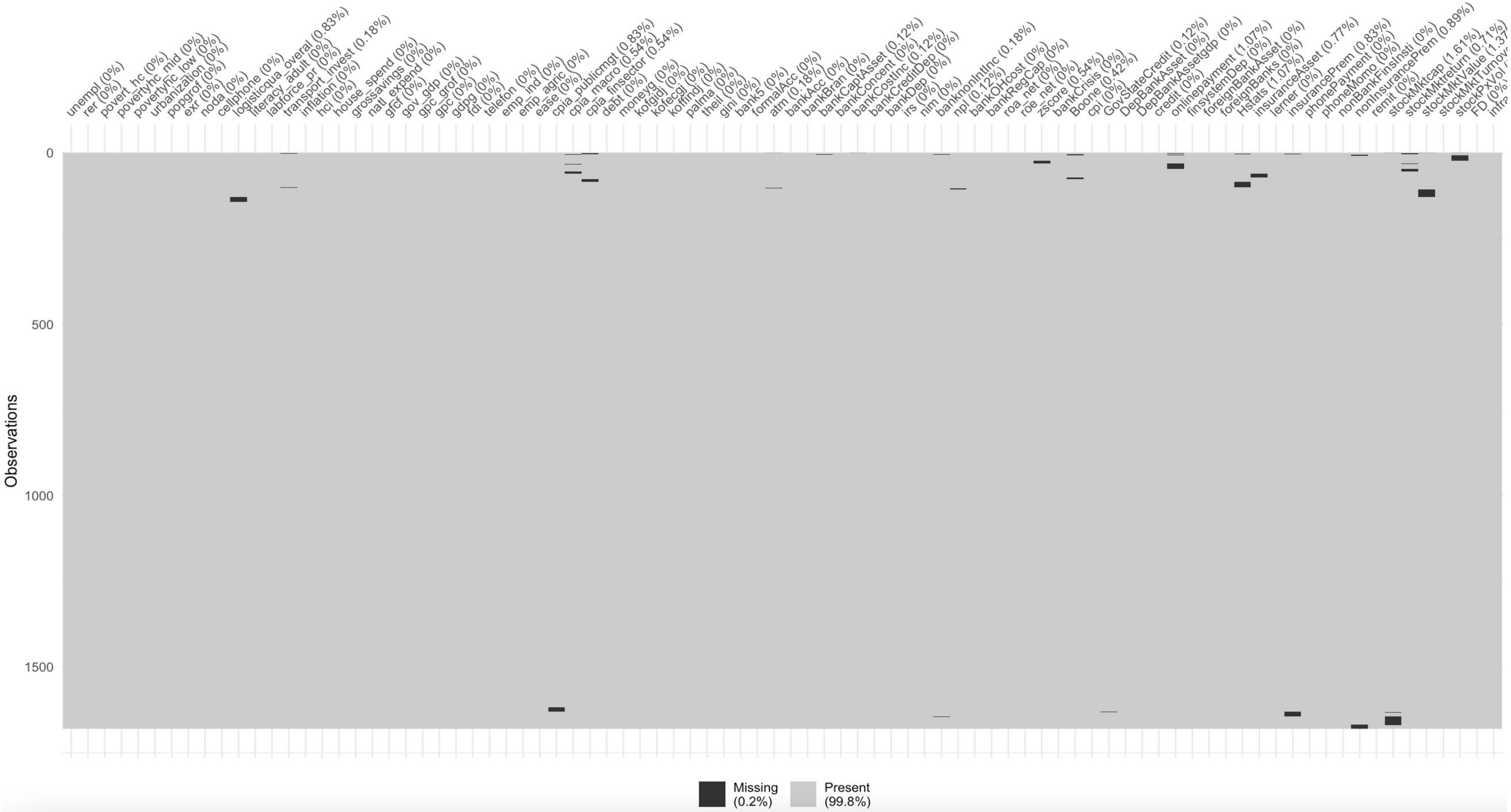

Figure A2: ***Overview of the dataset before data engineering***
Source: Author's construct (2021)

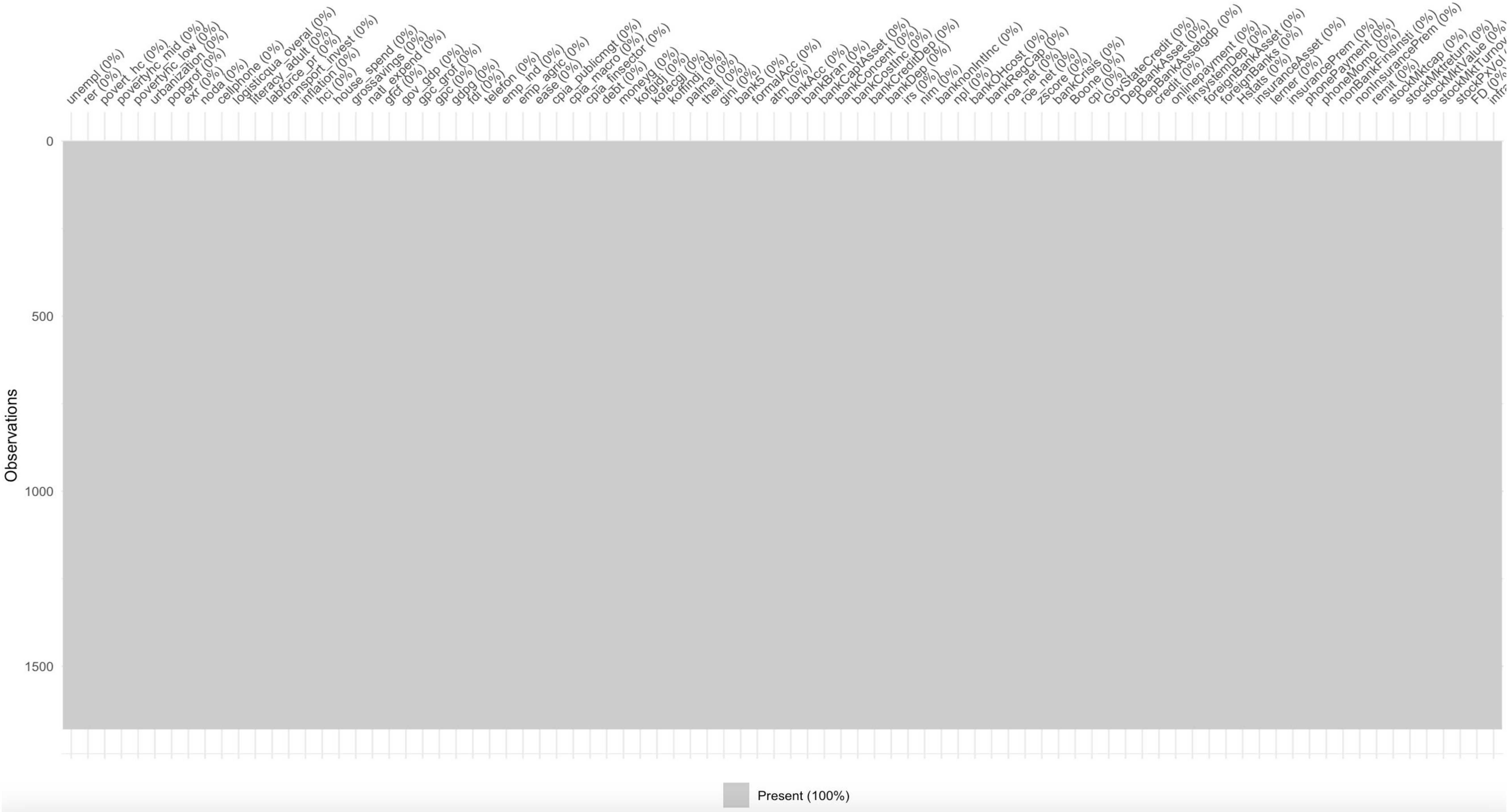

Figure A2: ***Overview of the dataset after data engineering***
Source: Author's construct (2021)

Table A3: Variable selection in regularization models

	standard_lasso	Minimum_BIC_lasso	Adaptive_lasso	Elastic_Net
cpia_financialsector	x	x	x	x
kofecgj	x	x	x	x
Phone_Momo	x	x	x	x
GDP_percapita	x	x	x	x
gov_gdp	x	x	x	x
emp_agric	x	x	x	x
bankOHcost	x	x	x	x
cpi_inflation	x	x	x	x
zscore	x	x	x	x
DepBankAssetgdp	x		x	x
povert_hc	x		x	x
adult_literacy	x	x	x	x
bankDep	x		x	x
telefon	x		x	x
cellphone	x		x	x
roe_net	x		x	x
nonInsurancePrem	x		x	x
_cons	x	x	x	x

Source: Author's construct (2021)