Impact of Digital Literacy on Financial Outcomes – A Cross-Country Analysis

Schaff Mirza, Jovera Shakeel, Khan Abdullah and Shehzil Munir
26/12/2024

Abstract

The impact of digital literacy on financial outcomes has been well-explored. However, the onset of AI necessitates a pressing need for more granular, cross-country analyses that incorporate local variations in digital infrastructure and socioeconomic conditions. Using data from three sources in 82 countries, we employ a Directed Acyclic Graph (DAG) to examine both the direct and indirect effects of digital literacy on financial well-being. Our study uses two econometrics models, Ordinary Least Squares Regression (OLS) and Structural Equation Model (SEM), along with the machine learning approach of Random Forests. Our results confirm our initial hypothesis that digital literacy has a positive impact on financial well-being through financial inclusion. Through our models, we find that the indirect link through financial inclusion dominates the direct impact of digital literacy on financial well-being, as it accounts for socio-economic, institutional, and individual factors.

1. Introduction

Global financial systems have changed due to the quick development of digital technology, and digital literacy is now seen as a key component that promotes resilience, financial inclusion, and sound decision-making. Even though these themes are receiving more attention, there still needs to be more knowledge regarding how digital literacy affects financial results across nations when viewed using thorough, aggregated data. This research attempts to close that gap by performing a cross-country analysis utilizing data from the World Bank, ITU Data Hub, and Meta Data for Good from 2021. In addition to examining the wide-ranging effects of digital literacy, our research digs into important aspects of inequality.

Our final data models analyse data for 82 countries, and therefore findings are generalizable. However, we focus policy implications on and tailor model controls for third-world countries such Pakistan, where rural areas face particularly low digital literacy rates and lack financial inclusion infrastructure. As of 2021, only 23% of Pakistan's rural population met basic digital competency thresholds, curtailing the adoption of mobile-based financial services and hindering access to formal financial channels. Such disparities underscore a broader global challenge: nations with sizable unbanked populations often exhibit significant digital literacy gaps that aggravate financial inequalities. By considering Pakistan alongside other countries—each with its own socioeconomic complexities and infrastructures—we highlight how tailored digital literacy initiatives can spur more equitable financial systems.

In addition to investigating cross-country patterns, we control for inequality, stability and infrastructural limitations. This multi-layered perspective allows us to examine whether strengthening digital literacy can reduce longstanding barriers to financial inclusion. Ultimately, our findings provide policymakers and stakeholders with actionable insights into how targeted interventions in digital skills training, infrastructure expansion, and regulatory frameworks can collectively foster financial equity around the globe.

2. Literature Review

Over the past decade, rapid advances in digital technology have reshaped global financial systems, prompting growing scholarly interest in the role of digital literacy in financial inclusion and economic well-being. Traditional frameworks of financial literacy—centered on numeracy, budgeting, and understanding basic financial products—are now being augmented by digital competencies that enable individuals to navigate mobile banking, digital wallets, and internet-based financial tools. This section synthesizes the existing literature on digital literacy and its effect on financial outcomes specifically through financial inclusion, emphasizing both cross-country evidence and the specific context of Pakistan.

A recurring theme in the literature is that digital literacy and financial inclusion form a reinforcing cycle. Putri et al. (2022) find in an ASEAN-wide cross-country study that higher financial inclusion—measured by access to banking services and mobile money—positively correlates with gains in financial literacy, especially in countries that initially display lower rates of formal financial access. Their analysis, which compares financial inclusion and literacy scores using World Bank Global Findex data, underscores the value of incremental improvements in digital infrastructure for boosting overall financial well-being in emerging markets. Although the study focuses on ASEAN countries, its lessons generalize to other regions with fragmented or nascent digital ecosystems.

Related work by Rojas-Suarez and Fiorito (2021) illustrates the strategic sequence by which countries can bolster digital financial inclusion. Using a decision tree model across Latin America, Africa, and Asia, they argue that increasing digital literacy should precede the introduction of more complex digital financial services such as mobile banking. Their framework highlights both infrastructural variables (e.g., mobile network coverage) and human capital factors (e.g., internet usage skills), showing that digital literacy stands as a critical prerequisite for sustaining financial inclusion. This logic underpins a wider consensus: without robust digital capabilities, the adoption of digital financial products—even if they are made accessible—will remain shallow.

Several studies specifically highlight the intermediary role of mobile-based financial services in connecting digital literacy to financial well-being. Lyons et al. (2020) examine digital inclusion in South Asia and Sub-Saharan Africa, showing that mobile money usage substantially reduces poverty levels and increases income stability among low-income households. Higher levels of digital literacy emerge as a linchpin, enabling households to exploit basic mobile banking features—such as transfers and savings—that ultimately lead to greater economic security. Similarly, Salman et al. (2024) demonstrate that rural households in Pakistan that adopt digital wallets see improvements in savings, credit access, and overall economic resilience, contingent on their level of financial and digital literacy. Thus, digital literacy not only fosters initial adoption of new technologies but also promotes more impactful and sustained usage.

In the context of financial stability, Pham and Linh (2020) warn of the potential risks of accelerating financial inclusion without parallel growth in digital financial literacy. Their panel data from Asian countries suggest that widening access to credit through digital platforms can lead to over-indebtedness where consumers lack the digital or financial savvy to manage complex products. This interplay between new forms of credit and consumer capabilities underscores the double-edged nature of digital inclusion: while it promotes stability and resilience at a macro level, insufficient literacy raises systemic vulnerabilities. These cautions highlight the importance of policy interventions that focus on both the demand side (e.g., consumer education) and the supply side (e.g., robust regulatory frameworks).

Against this backdrop of cross-country findings, the case of Pakistan stands out for its large rural population, low formal banking penetration, and emerging digital payment landscape. Ameen and Gorman (2009) identify low digital skills as a fundamental development hurdle in Pakistan, with direct repercussions for financial engagement. Their qualitative work—drawing on interviews with educators and policymakers—reveals that poor digital readiness in rural areas restricts individuals' capacity to use e-banking solutions such as mobile wallets. This disparity exacerbates existing inequalities, limiting the scope of economic opportunities for lower-income and female populations. Confirming this challenge, more recent data indicate that only 23% of Pakistan's rural population is digitally literate, leaving a significant portion of potential users unable to engage with Easypaisa, JazzCash, or similar mobile-based financial platforms (Ameen and Gorman, 2009; Choung et al., 2023).

Further underscoring Pakistan's scenario, Dura (2022) and Ariana et al. (2023) demonstrate from Indonesian evidence that digital literacy is a key driver of financial performance, resilience, and capability. Dura's (2022) quantitative study of SMEs in East Java shows that 56% of digitally competent enterprises reported marked improvements in financial performance once they adopted mobile banking and digital payment systems. Ariana et al. (2023) similarly use structural equation modeling (SEM) to confirm that digital literacy—alongside more general financial literacy—enhances financial resilience through intermediary channels such as financial inclusion and financial decision-making. Translating these results to Pakistan, one would expect that boosting digital literacy levels, especially in rural communities, would elevate the adoption of digital tools, expand financial decision-making capabilities, and ultimately enhance economic resilience.

Empirical examinations of specific skill sets shed additional light on why certain digital competencies may matter more than others. Studies by Sjam (2024) and Widyastuti et al. (2024) in Indonesia highlight that the ability to use digital tools for banking or mobile payments is especially instrumental to fostering positive financial outcomes, more so than broader or purely technical abilities like basic programming. In Pakistan, where 70% of rural households do not have steady digital access (Salman et al., 2024), targeted training for everyday financial applications, rather than generalized computing instruction, appears to hold the greatest promise. This resonates with findings from Rojas-Suarez and Fiorito (2021), who argue that efficient sequencing of capacity building—starting with practical digital financial services—yields a greater payoff in emerging markets.

Finally, a cross-cutting theme in the literature is the interdependence between digital literacy, financial literacy, and broader enabling infrastructure. Both Putri et al. (2022) and Lyons et al. (2020) emphasize how enhancements in internet coverage, affordability of mobile devices, and targeted public policies accelerate the positive effects of digital literacy. Choung et al. (2023) corroborate these structural elements in the context of South Korea, suggesting that even in technologically advanced economies, digital literacy must coincide with inclusive systems, regulatory safeguards, and consumer protection frameworks to maximize financial well-being. In Pakistan, this implies that bridging the digital divide must go hand in hand with fostering trust in digital payment channels, lowering the costs of mobile data, and ensuring adequate consumer safeguards.

In summary, existing research provides robust evidence that digital literacy drives meaningful gains in financial inclusion and economic outcomes, but also highlights the risks of expanding digital services without adequate consumer capacity and infrastructural support. While cross-country studies underscore the positive relationship between digital literacy and indicators such as income stability, savings, and poverty reduction, context-specific challenges—particularly in Pakistan—demonstrate that infrastructure gaps, gender disparities, and rural-urban divides can neutralize potential gains. Taken together, these insights underscore a pressing need for more granular, cross-country analyses that incorporate local variations in digital infrastructure and socioeconomic conditions. By investigating the impact of digital literacy on financial outcomes across 82 countries, this study aims to extend the literature on the relationship between targeted digital interventions and greater financial equity and resilience.

3. Data Methodology

3.1. Data Sources

To investigate the connection between digital literacy and financial inclusion in different nations, we used three primary datasets. Although the final dataset concentrates on 2021 for accuracy and consistency, the data sources cover the years 2001 to 2023.

1. World Bank Data (2011 onwards): The first source was the World Bank, where we used variables about digital literacy and financial inclusion exactly as they were. This dataset was used to extract intermediate financial outcomes i.e. financial inclusion and digital use of financial services, and final financial outcomes i.e. perceived financial wellbeing.

2. Meta's Data for Good (2017 onwards): Meta's Data for Good platform served as the second data source. It was cleaned to ensure the country names and years were correctly aligned with the rest of the dataset. This procedure allowed us to prepare the data for cross-country comparisons, and for extracting our control variables.

3. International Telecommunication Union (ITU) Data (2001 onwards): The Internet Inclusivity Index from the ITU served as our last data source. This data needed substantial adjustment, in contrast to the other datasets. We changed the variables' original row-based structure to a column-based one to accommodate the structure required for our analysis. The variables were renamed to better reflect the scholarly setting of our investigation on how digital literacy affects financial inclusion. This was used to extract independent variables such as digital skills, and some control variables.

3.2. Data Cleaning and Specifications

3.2.1. Time Frame

Starting in 2010, the datasets covered several years. However, as many of our target variables (financial outcomes) only had complete data for 2021, we limited our research to that year. To create a final dataset with only data from 2021, any rows that did not relate to that year had to be eliminated.

3.2.2. Managing Missing Data

Two imputation techniques were used to treat missing data:

First Method: Imputation of Recent Values
We used the most recent value available after 2017 to fill in the missing data for variables that had missing values. We were able to use the most recent data for missing values in 2021 as a result.

Second Method: k-Nearest Neighbours (KNN) Imputation
To handle the remaining missing values after imputing with recent data, we applied the K-Nearest Neighbors (KNN) imputation technique. Here's how it works: for each missing value x in the dataset (which were typically those of supply-side an country variables), we use the known values of x as the target (y) for training a KNN model. The model is trained using the corresponding control variables as input features for those observations where x is known. Once trained, the model predicts the missing values of x based on a weighted average of the nearest neighbors, identified using the control variables. This systematic approach enabled us to address all missing values in the dataset for the year 2021.

3.2.3. Variable Selection

1. Outcome Variables (Y): We concentrated on variables that showed perceived financial well-being as a proxy for financial well-being, especially those that showed how concerned people were about financial stability. We divided these into four groups: "Very worried," "somewhat worried," "a little worried," and "not worried at all."

2. Multicollinearity Check: We employed the Variable Inflation Factor (VIF) to assess multicollinearity, enhancing the reliability of our model by eliminating strongly correlated variables that might distort the findings. This enabled us to examine the r-squared value, and we made certain that we selected values under 10, with some exceptions for intermediate financial results.

3.2.4. Principal Component Analysis (PCA)

We performed a Principal Component Analysis (PCA) to develop a weighted index that acted as the main measure of financial outcomes in our research. The PCA enabled us to identify the inherent structure of the data and condense the indicators into a smaller number of components, enhancing the efficiency of the analysis. 4 indices were made based on "Very worried," "somewhat worried," "worried," and "not worried at all" into which these were divided. For further discussion on this, see section 3.2.6.

3.2.5. Analytical Approach

Our analysis focuses on correlations rather than causal inference, exploring the associations between digital literacy and financial inclusion across 82 countries in 2021. The results will highlight the strength and direction of these relationships, shedding light on how digital literacy impacts financial outcomes using a cross-country approach.

3.2.6. Variable Descriptions

The Worried Indexes consist of four variables created using Principal Component Analysis (PCA) from worry-related data. The original data had these variables bifurcated into four categories: (i) worried about not being able to pay for medical costs in case of a serious illness, (ii) worried about not being able to pay for school fees, (iii) worried about not having enough money for monthly expenses, and (iv) worried about not having enough money for old age. Each category was initially measured on a scale of not worried, somewhat worried, and very worried, which formed the basis for the following indexes:

- not_worried_index: Represents individuals who are "not worried."
- somewhat_worried_index: Captures individuals who are "somewhat worried."
- very_worried_index: Represents individuals who are "very worried."
- worried_index: Aggregates overall worry intensity across all categories.

Since the descriptions of the remaining variables are straightforward, they have been included in the appendix (refer to A1).

3.3. Exploratory Data Analysis

We can divide the preliminary data analysis into two parts. The first part deals with variations across different income levels for some interesting variables and the second part deals with variations across regions for those (interesting) variables.

1. *Across Income Levels:*
    a. Internet-Related Metrics: Analysis across high and low-income levels provides some key insights as to how access to digital literacy varies between these two key sub-groups. For mobile subscriptions, high-income countries tend to have nearly universal rates. In contrast, low-income countries show a lower peak around 50%, indicating much fewer mobile subscriptions per capita. High-income countries tend to have nearly universal 3G coverage as well, with low very low variance. On the other hand, the contrast is much sharper here between the two income groups, with low income countries having a much higher variance in coverage with the mean being much lower.

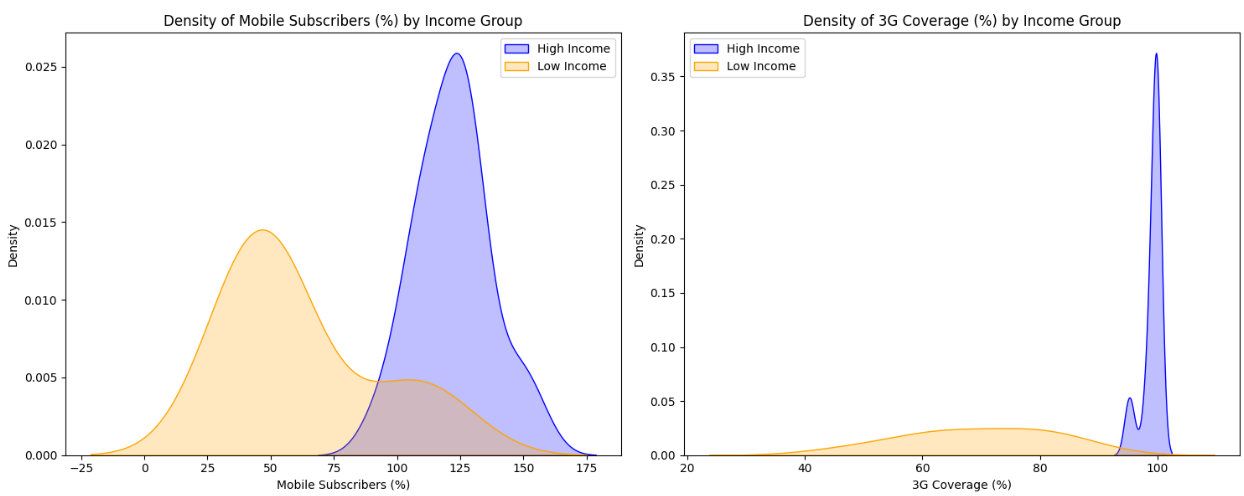

Figure 1

    b. Internet and Educational Attainment: The relationship between the internet and educational attainment tends to be linear and positively correlated, as shown in the scatter plot - higher education levels correlate with higher internet usage percentages. High-income countries (purple) cluster at the top-right have both high educational attainment and high percentage of internet users, while low-income countries (orange) are concentrated at the bottom-left, showing lower values for both metrics.

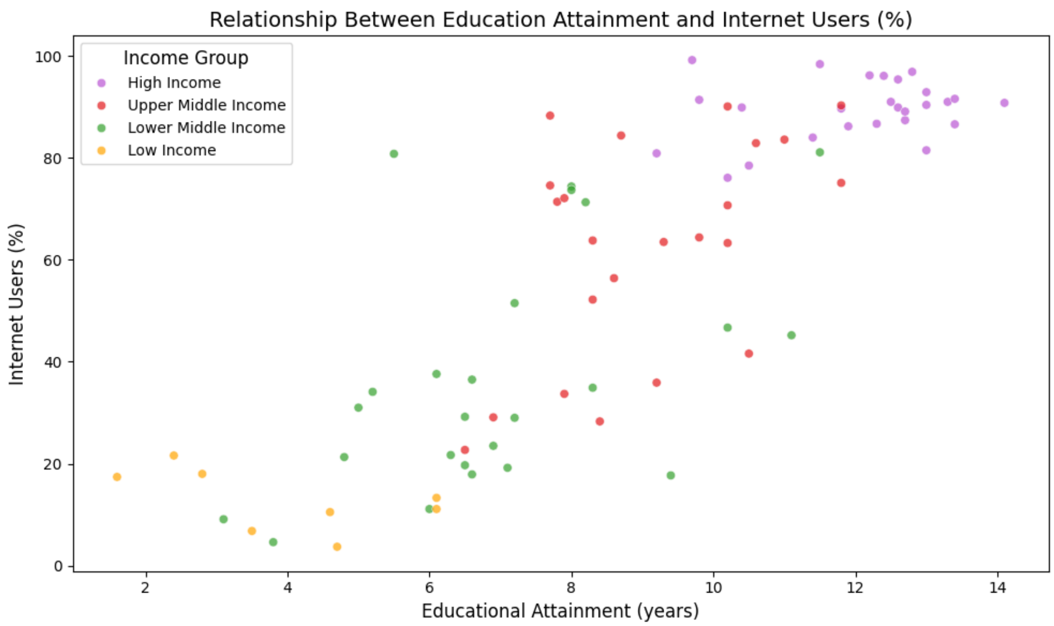

Figure 2

c. Variation in Intermediate Financial Outcomes: All Intermediate Financial Outcomes have similar mean amongst themselves across high-income and low-income countries, except for borrowed_for_health_pct (% of people who borrowed money for health purposes). The opposite relationship of borrowed_for_health_pct as compared to others makes logical sense: lower-income countries tend to borrow more money for healthcare because they tend to have lower salaries and wealth.

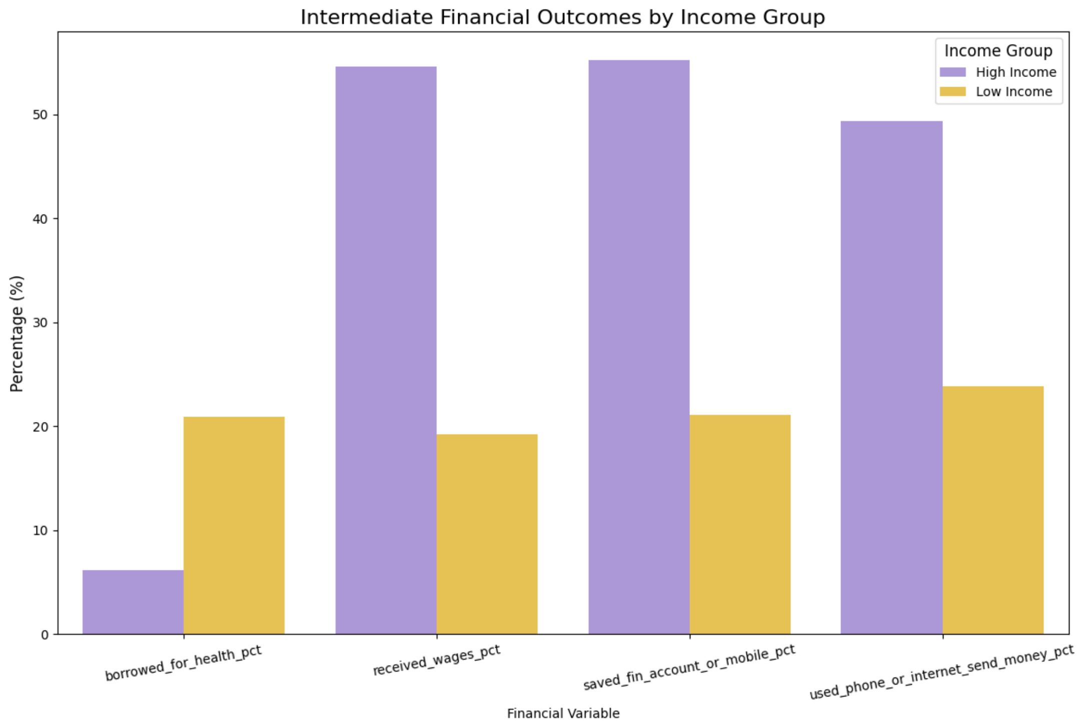

Figure 3

d. Programming Skills: For High-Income countries, the mean percentage of people having programming skills is more than thrice as high as that of Low-Income countries. The variance is also much larger for High-Income countries, with the percentage of people having programming skills going up to 30% for certain countries; on the other hand, the maximum percentage for Low-Income countries is around 3.5%.

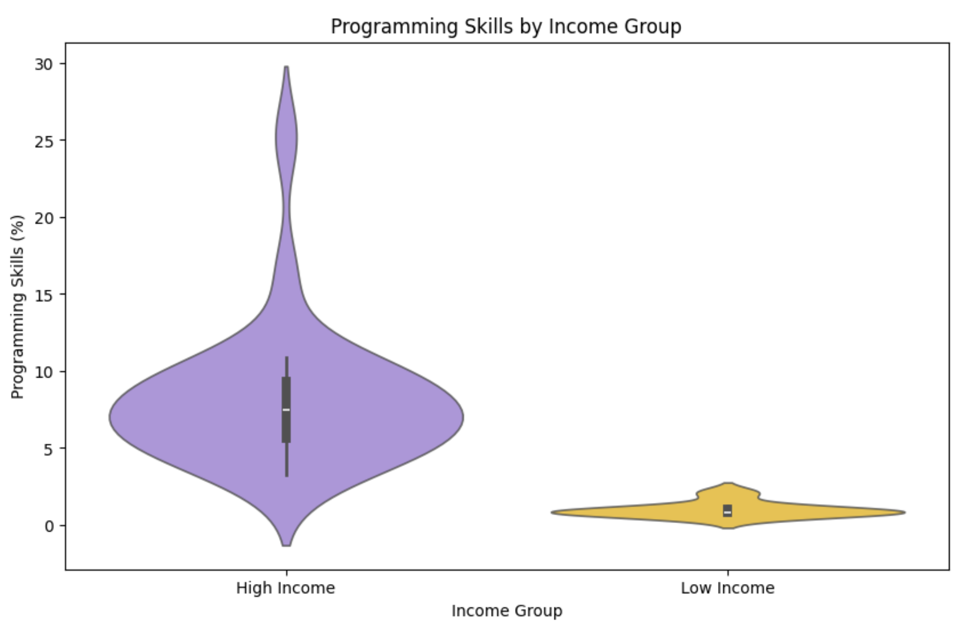

Figure 4

2. Across Regions:

a. Internet-Related Metrics: The graph shows the distribution across regions of internet users (in %). Africa and Europe have the most skewed distributions but in opposite directions. Distribution for Africa indicates that most African countries tend to have a small percentage of internet users. On the other hand, most European countries tend to have a high percentage of internet users. It is easy to notice that, once again, regions that are developed and therefore have high income tend to have higher percentage of internet users than the regions which are not as development.

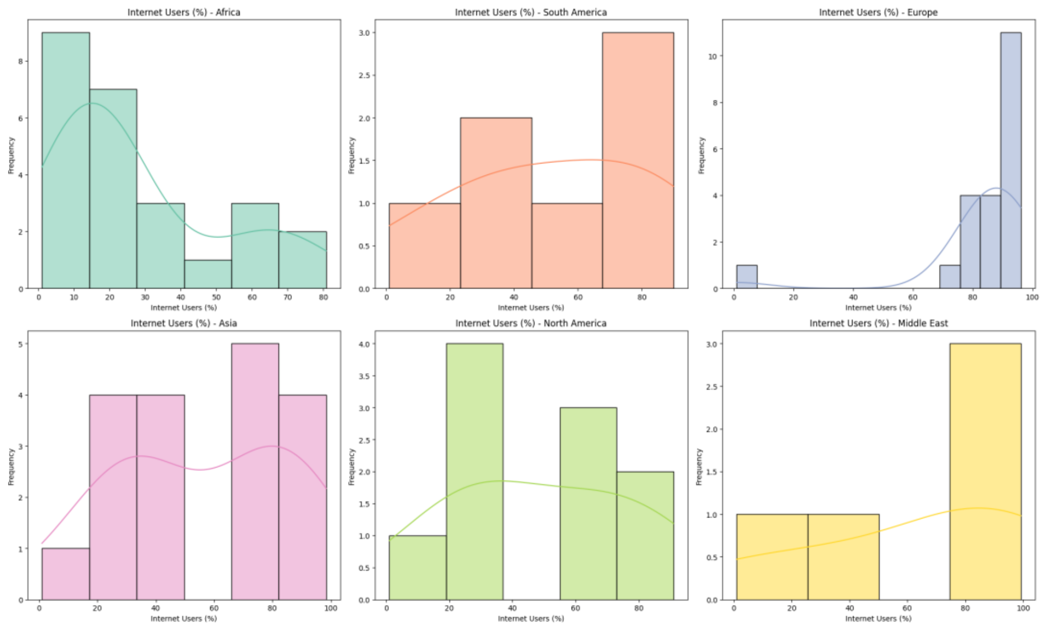

Figure 5

b. Educational Attainment: Mean educational attainment (in years) varies across regions, with Europe having the highest mean and Africa the lowest. The distribution also varies. For Asia, there is a very high variation, indicating certain countries have much better education systems than others. As for Europe, the variation is much lower, indicating sustained high level of education.

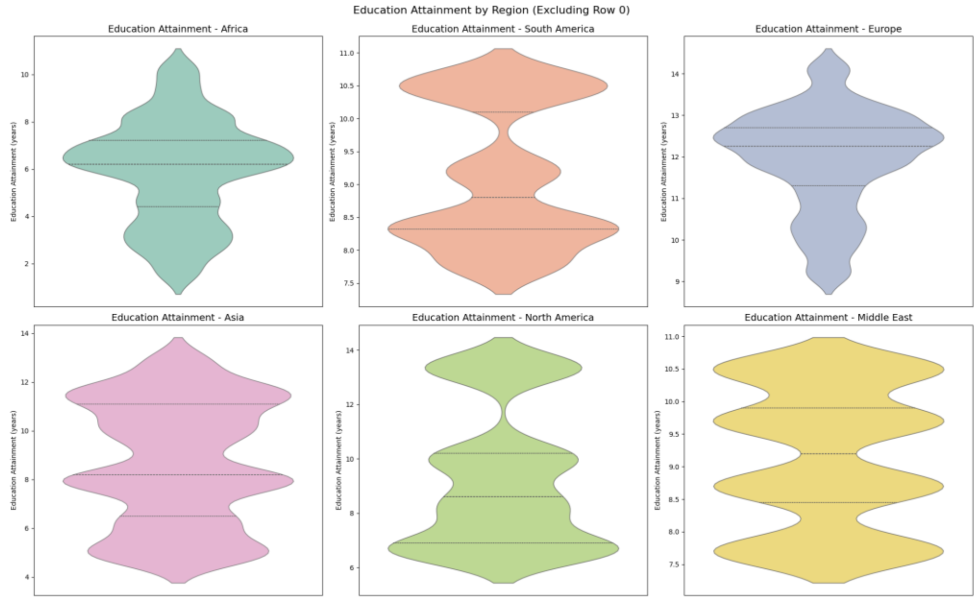

Figure 6

c. Skill_Internet_Banking, Skill_Uploading_Content, and Skill_Arithmetic_Formula_Spreadsheet: Very similar patterns across these variables for all regions in terms of their mean relative to each other. However, the variation in region values changes a between these variables, indicating that one skill doesn't necessarily imply the other skills and that there is less than perfect correlation between these skills.

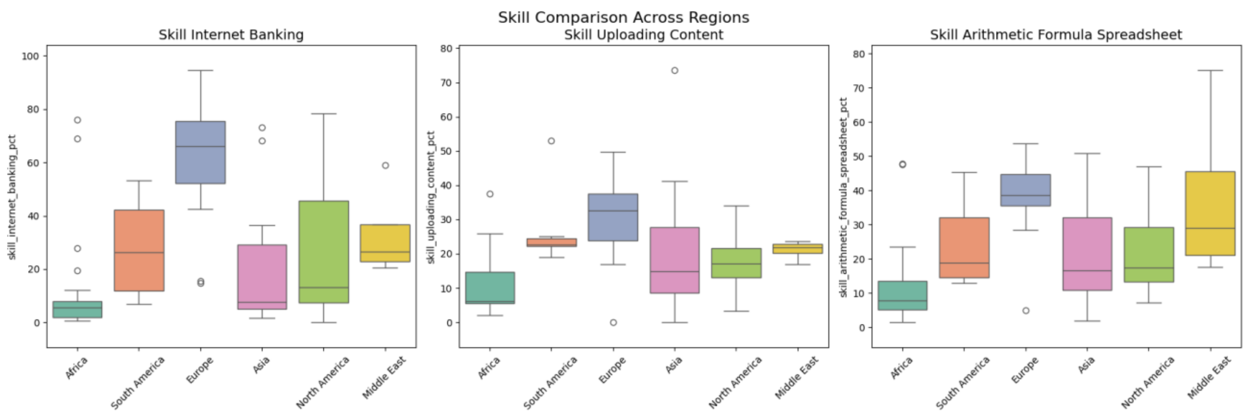

Figure 7

d. Used_phone_or_internet_send_money: Again, similar pattern across regions, with Europe having the highest mean and Africa one of the lowest. One thing to notice: variation in this variable is much higher than that of other variables. Therefore, what region a country falls in has a slightly lower influence on the whether or the residents of that country use phone or internet to send money. Further analytics might make the picture clearer and show what exactly determines the value of this variable.

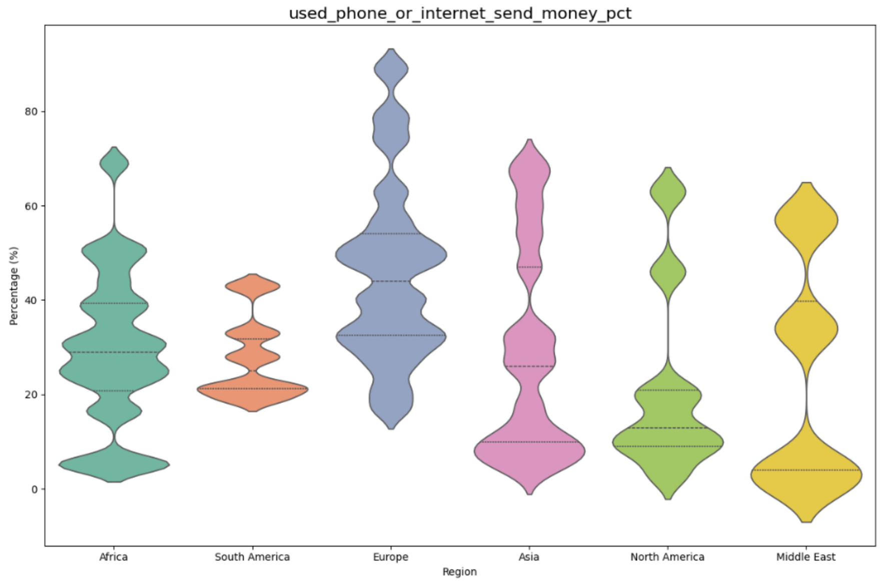

Figure 8

4. Model Methodology

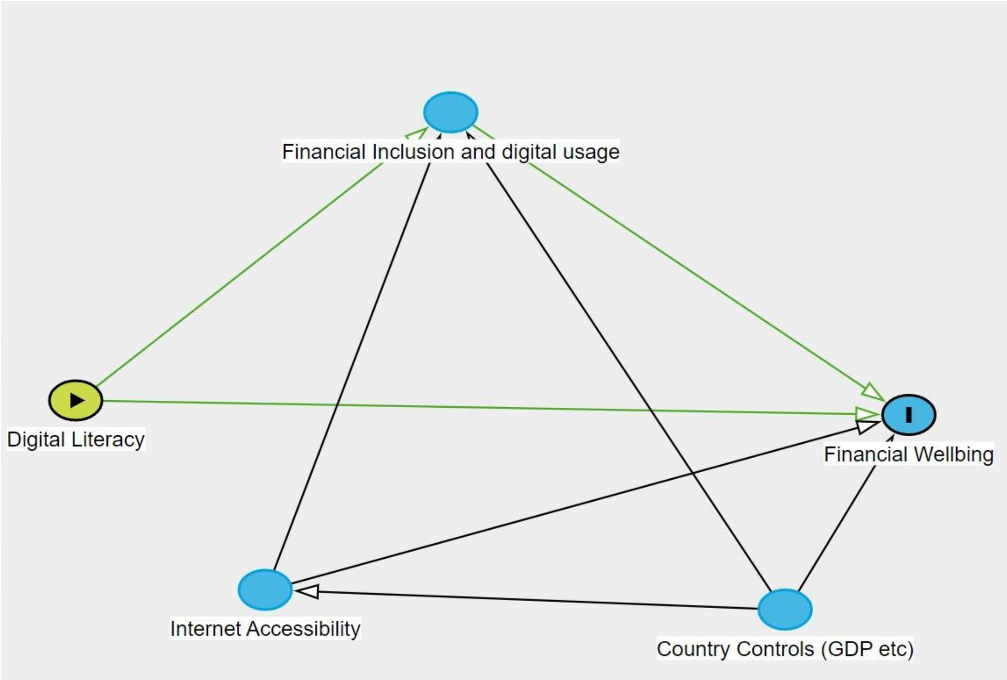

Figure 9: Casual DAG showcasing paths that form basis for all 3 methods

Based on the literature review, a Casual Directed Acyclic Graph (DAG) was also constructed to show the different pathways through which digital literacy can impact financial outcomes. It is key to note that there can be both a direct effect of digital literacy on financial outcomes, as well as an indirect effect through financial inclusion. This DAG also enables us to identify that country controls and internet accessibility controls need to be added to minimize bias in the estimations. By applying the local Markov's assumption, factors outside this DAG, such as other factors influencing financial inclusion, can be ignored to be having a causal impact on financial outcomes and can therefore be excluded from this analysis.

Based on this DAG, below, we describe three data models that we used to shape and model our data. The models are Ordinary Least Squares (OLS) Regression, Structural Equation Modelling (SEM) and Random Forest. All these models have certain details and nuances associated with them as they differ in their assumptions and algorithm implementation

4.1. OLS and SEM Assumptions

Before moving on, let us discuss several assumptions that must be acknowledged to ensure the validity and reliability of the analysis, and which affect both OLS and SEM:

i) Linearity within the Model

OLS and SEM assume that the relationships between variables can be adequately captured through linear equations. This implies that the effects of predictors on outcomes are proportional and additive. While this assumption simplifies estimation, it may not fully capture complex non-linear relationships present in real-world data. Nonetheless, given the nature of the constructs being studied, the linearity assumption is deemed reasonable for this analysis.

ii) Sufficient Sample Size

A key requirement for both methodologies is a sufficiently large sample size to ensure stable numerical computation, convergence, and accurate estimation of parameters. As a general rule of thumb, a sample size of at least 200 for SEM and 100 for OLS is considered necessary for meaningful and reliable solutions. This threshold helps avoid convergence failures and improper solutions with negative variance estimates. While this study falls short of the ideal threshold, with a sample size of 82, considerations of statistical power, significance levels, and effect sizes are used to determine whether this sample size is adequate to avoid both Type I errors and Type II errors.

iii) Multicollinearity Among Variables

Multicollinearity, or high correlations among explanatory variables, can distort the estimation of relationships in both methods. This issue inflates standard errors and undermines the reliability of path coefficients. To address this, as mentioned earlier in the paper, the variance inflation factor (VIF) was calculated for all explanatory variables in the model. Variables with high VIF values, indicating significant multicollinearity, were excluded during model specification to ensure a parsimonious and reliable econometric specification.

iv) Multivariate Normality

Both methods assume multivariate normality, meaning that all variables and their linear combinations should follow a normal distribution. This assumption underpins the Maximum Likelihood Estimation (MLE) method used for parameter estimation. While perfect adherence to multivariate normality is difficult to confirm, the assumption is accepted for this analysis as it aligns with standard econometric practices.

v) Completeness of Data

Both methods require a complete dataset to ensure reliable parameter estimation. Missing data can introduce bias and reduce statistical power. In this study, missing data were addressed during the data cleaning process using k-Nearest Neighbours (KNN) imputation. This method predicts missing values based on the values of similar observations, ensuring that the dataset used for analysis is complete and unbiased.

vi) Independence of Observations

Both also assume that the data collected from one observation (e.g., an individual) does not influence the data collected from another. This assumption of independence is critical, as any interdependencies within the dataset could bias parameter estimates and undermine the validity of the results. In this study, the data was sourced from the World Bank, Meta, and ITU, all of which adhere to rigorous international standards for data collection. These standards ensure that the independence of observations is maintained, thereby meeting this assumption for the analysis.

4.2. Ordinary Least Squares (OLS)

In the context of our study, OLS can identify any linear relationships that exist between digital literacy and financial outcomes in the data.

As described before, our data is arranged in four categories of variables. We want to find out the relationship between X variables (digital literacy) and Y variables (financial outcomes). We have one category of intermediate variables which measure financial inclusion. And lastly, we have a category of controls variables that account for the confounding variation in digital literacy, financial inclusion and financial outcomes. To estimate the effect of digital literacy on financial outcomes, we simply regress y on x. However, with this method, there is a likely possibility of our estimates being biased due to omission of variables that potentially have an effect on financial outcomes. For instance, the income level of a country affects both digital literacy and financial outcomes. Omitting it means that we have a confounding variable that drives both y and x. To account for that, we introduce income level as a control variable. Therefore, in our analysis, we carry out and report results from two models: i) a model that estimates results of regressing y on x without accounting for confounding variables, and ii) a model that estimates results of regressing y on x while accounting for confounding variables by including them as controls in the regression. These two models make up our *Direct Link Regression.*

$$\begin{aligned} Y = \ & \beta_0 + \beta_1 internet\_banking + \beta_2 uploading\_content \\ & + \beta_3 arithmetic\_formula\_spreadsheet + \beta_4 software\_run\_over\_internet \\ & + \beta_5 skill\_programming \ \ \ldots \ \ \beta_k controls + \epsilon \end{aligned}$$

Y = {Worried Index, Very Worried Index, Somewhat Worried Index, Not Worried Index}

We also have another set of variables that fall in the financial inclusion set. As explained in the Directed Acyclic Graph (DAG), one causal pathway in our data is from digital literacy (x) to financial outcomes (y) through financial inclusion (intermediate). While OLS cannot exactly model this relationship, we can model the direct impact of digital literacy on financial inclusion, a process very similar to that of impact of digital literacy on financial outcomes. We call this the *Indirect Link Regression,* and once again, divide it into two separate models: i) Regression of financial inclusion on digital literacy without the control variables, and ii) Regression of financial inclusion on digital literacy with the control variables.

$$Y = \beta_0 + \beta_1 internet\_banking + \beta_2 uploading\_content + \beta_3 arithmetic\_formula\_spreadsheet + \beta_4 software\_run\_over\_internet + \beta_5 skill\_programming \;\ldots\; \beta_k controls + \epsilon$$

Y = {Borrowing from formal sources, Sending money via internet, Saving through financial accounts}

The motivations behind using OLS to model our data are many. Firstly, OLS provides an easy-to-understand model. That is, it makes clear how much influence a particular x variable has on y variable (for example, the effect of digital literacy with regard to the ability to upload content on worried_index). Thus, it is simple enough to comprehend and implement while still being robust (if the conditions are met). In fact, if the conditions are met and our assumptions are true, OLS is the best estimator. Furthermore, OLS allows for control over certain aspects of the data, resulting in more accurate coefficients by accounting for biases that may exist. Lastly, OLS allows us to model relationships in a way as to quantify the magnitude and direction of those relationships, and therefore, enables us to predict possible scenarios with accuracy.

4.3. Structural Equation Modelling

As highlighted in the literature, one critical pathway through which digital literacy may influence a person's financial well-being is by enhancing financial inclusion—a relationship that is further illustrated in the DAG. Addressing this relationship using simple OLS regressions or even machine learning techniques, such as random forests, presents significant challenges. These methods, while effective in identifying patterns and correlations, often fail to capture the intricate web of direct and indirect effects that operate across multiple pathways. This limitation arises because they cannot decompose these pathways to isolate causal mechanisms. Consequently, they fall short in moving us closer to causal identification.

To overcome these challenges, this paper also employs Structural Equation Modelling (SEM). This econometric technique is uniquely suited to handle such complex relationships by allowing for the simultaneous analysis of direct and indirect pathways. SEM facilitates the decomposition of these pathways, enabling the estimation of both direct and indirect effects. This makes it particularly valuable for disentangling the layered influences that digital literacy exerts on financial well-being through financial inclusion.

Another significant advantage of SEM is its ability to incorporate latent variables—unobserved constructs that cannot be directly measured but are instead inferred from observable indicators.

Realistically, the variables and indices used in this study do not directly represent the core concepts of interest: Digital Literacy, Financial Inclusion, and Financial Well-Being. Instead, they act as indicators for these broader constructs.

For example, while skills such as the ability to upload content on the internet, operate spreadsheets, or engage in programming are undeniably components of digital literacy, they fail to fully encapsulate the multidimensional nature of this concept. Digital literacy also involves aspects such as the ability to leverage digital tools for personal budgeting and expense tracking and adapt to new technologies that enhance productivity and economic mobility. These aspects are challenging to measure directly, making digital literacy a prime candidate for representation as a latent variable.

Similarly, indicators like whether an individual has borrowed from formal institutions, saved in a financial account or mobile wallet, or used digital platforms to transfer money signify financial inclusion. However, these indicators do not comprehensively capture the extent of a person's integration into the financial system. Factors such as the regularity of financial activity, access to credit under fair terms, and the ability to leverage financial tools for economic security also play pivotal roles in defining financial inclusion, none of which are entirely reflected in the observed indicators.

Lastly, indices measuring levels of concern regarding paying bills, education fees, and meeting everyday expenses serve as indicators for financial well-being. However, financial well-being encompasses far more, including the stability of income sources, future financial security, and the freedom to make choices without undue financial stress. These broader dimensions are not fully captured by the observable variables, highlighting the necessity of conceptualizing financial well-being as a latent variable.

In essence, these explanations underscore a fundamental idea: the variables used in this analysis serve merely as indicators for the underlying constructs we aim to study. SEM enables the utilization of latent variables, which represent these unobservable constructs. By constructing latent variables from their relevant indicators, SEM provides a more accurate and holistic representation of Digital Literacy, Financial Inclusion, and Financial Well-Being. This capability is essential for advancing the rigor and depth of the analysis in this paper.

The selection of variables to serve as relevant indicators for the latent constructs in this study adheres to established methodologies, ensuring consistency with prior research. The variables retained for analysis were carefully chosen based on their Variance Inflation Factor (VIF) as the paper discussed earlier. Additionally, rather than using raw values, the z-scores of these variables were calculated using the Standard Scalar package.

This transformation was undertaken for several reasons. Firstly, SEM assumes that all variables are on comparable scales to facilitate stable numerical computation. When variables exist on vastly different scales, the optimization of the Maximum Likelihood function—the core estimation mechanism in SEM—can face significant challenges. Secondly, unscaled data can lead to issues in estimating standard errors via the Fisher Information Matrix, potentially resulting in an ill-conditioned matrix that is not positive definite. Such a matrix would undermine the reliability of the statistical estimates. Lastly, scaling enhances the interpretability of the factor loadings and the relationships between latent variables. By converting raw values to z-scores, these relationships can be interpreted in terms of standard deviations, allowing for clearer insights into the underlying constructs.

After scaling, the variables used to construct the latent variables are categorized as follows:

i) Digital Literacy

a. Percentage of people with the ability to program.
b. Percentage of people with the ability to use arithmetic formulae on spreadsheets.
c. Percentage of people with the ability to upload content on the internet.
d. Percentage of people with the ability to run software over the internet.

ii) Financial Inclusion

a. Percentage of people who have borrowed from formal institutions.
b. Percentage of people who have received wages.
c. Percentage of people who have saved in a financial or mobile account.
d. Percentage of people who have used a phone or the internet to send money.

iii) Financial Well-Being

a. Not Worried Index.
b. Somewhat Worried Index.
c. Worried Index.
d. Very Worried Index.
e. Percentage of people who can arrange emergency funds without difficulty.

$$X_{\text{Digital Literacy},i} = \lambda_{i1}\eta_1 + \epsilon_{i1}, \quad i = 1, \dots, 4$$
$$X_{\text{Financial Inclusion},j} = \lambda_{j2}\eta_2 + \epsilon_{j2}, \quad j = 1, \dots, 4$$
$$X_{\text{Financial Wellbeing},k} = \lambda_{k3}\eta_3 + \epsilon_{k3}, \quad k = 1, \dots, 5$$

Where,

λ_{i1}: The factor loading, representing the strength of the relationship between the indicator and the Digital Literacy latent variable (η_1).
λ_{j2}: The factor loading, representing the strength of the relationship between the indicator and the Financial Inclusion latent variable(η_2).

λ_{k3}: The factor loading, representing the strength of the relationship between the indicator and the Financial Wellbeing latent variable(η_3).
ϵ_{i1}: The measurement error term, accounting for variance in the indicator unrelated to the Digital Literacy latent variable.
ϵ_{j2}: The measurement error term, accounting for variance in the indicator unrelated to the Financial inclusion latent variable.
ϵ_{k3}: The measurement error term, accounting for variance in the indicator unrelated to the Financial Wellbeing latent variable.

Structural equation

$$\eta_2 = \beta_{21}\eta_1 + \delta_2 C + \zeta_2$$
$$\eta_3 = \beta_{31}\eta_1 + \beta_{32}\eta_2 + \delta C + \zeta_3$$

Where,

β_{21}: Path coefficient measuring the direct effect of Digital Literacy(η_1) on Financial Inclusion (η_2)
β_{31}: Path coefficient measuring the direct effect of Digital Literacy(η_1) on Financial Wellbeing(η_3)
β_{32}: Path coefficient measuring the direct effect of Financial Inclusion(η_2) on Financial Wellbeing(η_3)
δ: Coefficient measuring the effect of the control variables (C) on Financial Wellbeing(η_3)
ζ_2: Error term capturing unexplained variation in Financial Inclusion
ζ_3: Error term capturing unexplained variation in Financial Wellbeing

Three distinct models will be estimated in this analysis. The first model excludes controls entirely, serving as a baseline for comparison. The second model incorporates a reduced set of controls deemed statistically insignificant or partially captured by other variables. Lastly, the third model includes the full set of controls identified earlier in the paper, ensuring a comprehensive examination of the relationships under consideration. Such controls removed included 2g and 4G coverage (whilst 3G coverage was still included), dummies of market structures, and urban population amongst others. This approach aims to mitigate the risk of overfitting, which can arise when an excessive number of controls are included. All three models are overidentified as they have positive degrees of freedom.

4.4. Random Forest

Random Forest is a machine learning method that builds a collection of decision trees, with each tree trained on a randomly selected subset of the data. Each tree independently models the relationship between the input variables and the target variable. The model's final prediction is derived by averaging the results of the individual trees (for regression) or taking a majority vote (for classification). By combining many weak learners (decision trees), Random Forest reduces sensitivity to outliers and overfitting, improving accuracy and generalizability. This ensemble approach makes it a robust and reliable method for a wide range of tasks.

In our study, the way Random Forest works is that it creates multiple trees - each from a bootstrap sample (random sample with replacement). It creates these trees by identifying the best learner (which splits data the best) through Mean Squared Error (MSE) for those samples. And at the end, it aggregates the results of those trees using the following equation:

$$\hat{Y} = \frac{1}{T}\sum_{t=1}^{T} h_t(X)$$

We apply similar methodology here as we have with both OLS and SEM with respect to controls. Estimating results without controls means that the results are biased. To account for that, we first apply Random Forest on only x and y variables. Then, we apply Random Forest on x and y as well as control variables to get better estimations. This gives us two model results that we can compare to identify the influence of both x and control variables.

Our data is exceptionally high dimensional with many features distributed across multiple categories. In such a dataset, Random Forest performs particularly well since it can manage complex data and multiple data types together. Random Forest provides an insight into feature importance, that is, which features contribute most to the target variable. This is especially important in the context of our research where the key goal is to identify variables that contribute most to the financial outcomes and therefore should be most targeted by the policies designed around digital literacy. For instance, if the value of the percentage of people with internet banking skills contributes most to the not worried index, it tells the policy that the increased education campaigns to improve these skills has a higher chance of improving financial outcomes in the country. And lastly, Random Forest is free from most of the assumptions of SEM and OLS (as listed above). The only assumptions that Random Forest adheres to are sufficient sample size (and therefore sufficient trees), independence and no strong multicollinearity. And most important of all, Random Forest does not assume a linear relationship (unlike SEM and OLS). This means that if the data is non-linearly distributed, which is often the case in the real world, Random Forest does a good job of modelling the data.

5. Findings and Results

5.1. Findings from Ordinary Least Squares (OLS)

Using data from 82 nations in 2021, we apply Ordinary Least Squares (OLS) regression models both with and without controls to explore the impact of digital literacy on financial inclusion, financial well-being, and associated economic results. This method emphasizes the intricate relationships between digital ability and financial consequences, illuminating the mediating influence of socioeconomic and infrastructural aspects. This analysis is divided into direct and indirect links, as explained earlier in our Directed Acyclic Graph (DAG), allowing for a thorough grasp of the causal sequence. For the indirect link, the regression analysis examines the connection between digital literacy skills and intermediate financial outcomes. These intermediate results offer an understanding of how digital literacy promotes wider financial inclusion and decision-making. Models were analyzed with and without controls, allowing for an exploration of the relationship between personal digital skills and broader factors such as regional infrastructure, economic disparity, and institutional backing. We follow the general rule that t-values greater than 2.015 indicate that the variable is significant at the 5% level.

5.1.1. Direct Link

Equation:

$$\begin{aligned} Y = \ & \beta_0 + \beta_1 internet\_banking + \beta_2 uploading\_content \\ & + \beta_3 arithmetic\_formula\_spreadsheet + \beta_4 software\_run\_over\_internet \\ & + \beta_5 skill\_programming \ \ \ldots \ \ \beta_k controls + \epsilon \end{aligned}$$

Y = {Worried Index, Very Worried Index, Somewhat Worried Index, Not Worried Index}

1. *Worried Index*

	OLS Regression Results for worried_index			
Metric/Feature	Without Controls Coefficient	Without Controls t-stat	With Controls Coefficient	With Controls t-stat
const	6.025	38.098	8.293	2.217
skill_internet_banking_pct	-0.025	-4.412	-0.028	-3.158
skill_uploading_content_pct	0.001	0.179	0.001	0.087
skill_arithmetic_formula_spreadsheet_pct	0.006	0.572	-0.004	-0.224
skill_software_run_over_the_internet_pct	0.003	0.349	0.014	1.335
skill_programming_pct	-0.008	-0.282	0.026	0.589

Figure 10: Regression Coefficients for the Worried Index

The regression findings for the worried index indicate that Internet banking proficiency is an essential element in alleviating financial stress. With controls, a significant negative coefficient

($\beta = -0.0277$, $p<0.05$) emphasizes the importance of digital literacy in enhancing financial results. This suggests that Internet banking skills directly alleviate financial worries, regardless of other influences.

The inclusion of controls, like regional traits and income brackets, enhances the explanatory power of the model from 0.431 to 0.720, highlighting that structural disparities intensify financial concerns. Notably, other digital abilities, like coding or spreadsheet skills, do not demonstrate considerable effects on financial anxiety. For example, a 1 percentage point change in programming is associated with only a 0.026 change in the worried index, ceterus paribus. This indicates that various types of digital literacy do not all hold equal importance for financial health. The results highlight the significance of practical digital skills specifically designed for financial situations, such as online banking, and over broader or technical abilities.

The notable regional variations seen in the model further highlight the systemic aspects of financial anxiety. For example, areas with stronger digital infrastructure, like North America, indicate much lower levels of financial anxiety. These regional impacts evidence structural inequality in explaining financial results, where the availability of infrastructure and institutional assistance shield individuals from monetary stress.

2. *Very Worried Index*

	OLS Regression Results for very_worried_index			
Metric/Feature	**Without Controls Coefficient**	**Without Controls t-stat**	**With Controls Coefficient**	**With Controls t-stat**
const	14.913	21.576	21.618	1.807
skill_internet_banking_pct	-0.066	-2.653	0.001	0.050
skill_uploading_content_pct	-0.038	-1.177	-0.003	-0.082
skill_arithmetic_formula_spreadsheet_pct	-0.087	-1.797	0.022	0.433
skill_software_run_over_the_internet_pct	-0.053	-1.485	-0.031	-0.935
skill_programming_pct	0.034	0.281	-0.283	-2.038

Figure 11: Regression Coefficients for the Very Worried Index

For this index, which indicates severe financial stress, online banking proficiency stands out as an essential factor in the absence of controls ($\beta=-0.0663$, $p=0.01$). This outcome emphasizes the increased significance of digital tools in easing serious financial issues. Once controlled, the model exhibits strong explanatory capability ($R2=0.910$), indicating that digital literacy and related structural (both individual and infrastructural) factors account for a significant portion of the differences in severe financial anxiety.

Contrary to the worried index, the programming skill has a significant negative impact on this index. A significant feature of this index is the pronounced impact of regional differences. Areas characterized by low internet access or inadequate financial infrastructures, like Africa, display increased instances of severe financial anxiety. Africa displays a positive coefficient of 2.1986, however, it is insignificant. Regardless, these results posit that although personal digital literacy is crucial, its advantages depend on favorable structural conditions. Additionally, the decreasing marginal returns of digital literacy are apparent in this context: in well-developed areas, enhancements in digital abilities have a diminished impact on alleviating financial hardship since the initial levels of financial inclusion are already elevated. This corresponds with the concept of diminishing returns, where extra inputs generate increasingly smaller outputs in saturated systems.

3. *Somewhat Worried Index*

	OLS Regression Results for somewhat_worried_index			
Metric/Feature	**Without Controls Coefficient**	**Without Controls t-stat**	**With Controls Coefficient**	**With Controls t-stat**
const	7.707	30.724	8.655	1.957
skill_internet_banking_pct	-0.035	-3.882	-0.031	-2.975
skill_uploading_content_pct	0.006	0.480	-0.020	-1.635
skill_arithmetic_formula_spreadsheet_pct	0.022	1.226	-0.012	-0.674
skill_software_run_over_the_internet_pct	-0.003	-0.197	0.004	0.307
skill_programming_pct	0.009	0.210	0.070	1.358

Figure 12: Regression Coefficients for the Somewhat Worried Index

Skills in Internet banking show a negative correlation with moderate financial concerns ($\beta=-0.0308, p=0.01$), which is significant at the 5% level. However, since this is not an index that demonstrates 'extremes,' the impact is not as prominent as with the other indices. The model's reduced explanatory capability ($R2=0.292$) indicates that moderate concerns are affected by elements beyond the variables considered, like individual behavioral characteristics or psychological resilience.

In contrast to the very worried index, structural factors like income or local infrastructure have less impact on moderate financial concerns. For example, a 1 percentage point change in 4G coverage is associated with a 0.0174 decrease in the somewhat worried index, ceterus paribus. As expected, the cost of acquiring a smartphone has a significant positive effect on the somewhat worried index. This may suggest the threshold nature of these issues, which do not escalate into serious distress unless worsened by systemic factors. The findings suggest that while digital literacy plays a role, it is not crucial nor sufficient to tackle moderate financial issues, emphasizing a complex interplay of individual and contextual elements.

4. *Not Worried Index*

	OLS Regression Results for not_worried_index			
Metric/Feature	**Without Controls Coefficient**	**Without Controls t-stat**	**With Controls Coefficient**	**With Controls t-stat**
const	3.481	6.303	-9.833	-0.853
skill_internet_banking_pct	0.085	4.265	0.054	1.984
skill_uploading_content_pct	0.013	0.488	0.035	1.099
skill_arithmetic_formula_spreadsheet_pct	0.041	1.062	0.040	0.833
skill_software_run_over_the_internet_pct	0.062	2.143	0.049	1.541
skill_programming_pct	-0.049	-0.509	0.061	0.459

Figure 13: Regression Coefficients for the Not Worried Index

The not worried index shows a significant positive relationship with online banking skills (β=0.0852, p<0.01). This underscores the empowering qualities of digital financial tools, allowing individuals to address and manage their financial requirements more effectively. The model explains a significant portion of variation (R2=0.893), emphasizing the important influence of digital literacy on improving financial confidence.

Interestingly, unlike the worried index or the very worried index, this measure demonstrates reduced sensitivity to regional variations. While structural elements such as broadband availability and income levels are crucial, their influence is less critical than that of individual digital skills. This reinforces the idea that financial confidence is more strongly associated with personal skills and perceived control than with external constraints.

The results suggest that initiatives aimed at enhancing digital literacy, particularly in online banking competencies, can significantly elevate financial confidence. This has important policy implications because building personal skills can offer substantial benefits, even in cases where structural improvements take more time to appear.

5.1.2. Indirect Link

Equation:

$$\begin{aligned} Y = \ \beta_0 + \beta_1 internet\_banking &+ \beta_2 uploading\_content \\ &+ \beta_3 arithmetic\_formula\_spreadsheet + \beta_4 software\_run\_over\_internet \\ &+ \beta_5 skill\_programming \ \ \ldots \ \ \beta_k controls + \epsilon \end{aligned}$$

Y = {Borrowing from formal sources, Sending money via internet, Saving through financial accounts}

5. *Borrowing from Formal Sources*

	OLS Regression Results for borrowed_from_formal_pct			
Metric/Feature	**Without Controls Coefficient**	**Without Controls t-stat**	**With Controls Coefficient**	**With Controls t-stat**
const	4.846	1.996	4.079	0.074
skill_internet_banking_pct	0.144	1.634	0.012	0.096
skill_uploading_content_pct	0.124	1.080	-0.031	-0.206
skill_arithmetic_formula_spreadsheet_pct	0.400	2.341	0.224	0.973
skill_software_run_over_the_internet_pct	0.446	3.530	0.328	2.143
skill_programming_pct	-0.955	-2.239	-1.120	-1.756

Figure 14: Regression Coefficients for Borrowing from Formal Sources

The findings for this dependent variable highlight the complex influence of digital literacy on formal borrowing habits. Without controls, abilities like utilizing spreadsheets (β=0.4000, p=0.022) and the ability to run software (β=0.4458,p=0.001) show notable positive correlations with borrowing from formal sources, whereas programming abilities have a negative impact (β=−0.9553, p=0.028). This discovery indicates that practical and applied digital abilities improve access to credit, whereas technical skills might relate to situations where formal borrowing is less common, like in informal or entrepreneurial areas.

The addition of controls significantly enhances the model's explanatory power (R2=0.859). The consideration of factors such as income level, regional variations, and infrastructure emphasizes the mediating influence of structural elements. For example, gni_per_capita appears as a significant positive factor in borrowing, indicating the vital importance of economic stability. A 1 dollar change in GNI per capita is associated with a 0.06% change in borrowing behavior, all else being equal. Simultaneously, the negative influence of personal digital skills in the controlled model indicates that borrowing behavior is influenced more by institutional accessibility and cost than by individual digital abilities.

6. Sending Money via the Internet

	OLS Regression Results for used_phone_or_internet_send_mon			
Metric/Feature	**Without Controls Coefficient**	**Without Controls t-stat**	**With Controls Coefficient**	**With Controls t-stat**
const	20.179	5.591	-113.121	5.591
skill_internet_banking_pct	0.498	3.816	0.817	3.816
skill_uploading_content_pct	-0.020	-0.116	-0.136	-0.116
skill_arithmetic_formula_spreadsheet_pct	0.154	0.607	0.388	0.607
skill_software_run_over_the_internet_pct	-0.091	-0.487	-0.031	-0.487
skill_programming_pct	-0.662	-1.043	-1.600	-1.043

Figure 15: Regression Coefficients for Sending Money via the Internet

The dependent variable using the phone or internet to send money represents the use of digital means for monetary transactions. Abilities in online banking (β=0.4984,p<0.01) are significantly linked to a heightened usage of phones or the internet for transferring money in the uncontrolled model, highlighting the importance of this particular skill in digital financial inclusion. Nonetheless, other digital abilities like uploading content or operating software do not show notable impacts, indicating that money transfer behaviors are mainly influenced by targeted, task-related skills.

By incorporating controls, the model's ability to explain increases significantly (R2=0.785), and the importance of Internet banking abilities remains (β=0.8165, p<0.01). Regional disparities are especially noticeable: areas like Africa exhibit greater digital tool adoption when systemic obstacles, such as costs and infrastructure, are alleviated. These results emphasize the dual influence of personal abilities and systemic elements, whereby the latter enhances the effect of digital literacy.

Notably, socioeconomic factors like income bracket and internet availability at schools have a reduced impact in this model than in others, suggesting that digital money transfer is more affected by individual choices and less reliant on overarching structural circumstances. This indicates a significant level of scalability for digital financial instruments in settings with limited resources.

7. *Savings through Financial Accounts*

	OLS Regression Results for saved_fin_account_or_mobile_pct			
Metric/Feature	**Without Controls Coefficient**	**Without Controls t-stat**	**With Controls Coefficient**	**With Controls t-stat**
const	13.216	4.858	4.454	0.099
skill_internet_banking_pct	0.423	4.293	0.403	3.816
skill_uploading_content_pct	0.154	1.199	0.118	0.947
skill_arithmetic_formula_spreadsheet_pct	0.002	0.009	0.197	1.045
skill_software_run_over_the_internet_pct	0.344	2.433	0.290	2.310
skill_programming_pct	-0.704	-1.473	-0.375	-0.717

Figure 16: Regression Coefficients for Savings through Financial Accounts

The findings highlight the crucial importance of digital literacy in fostering savings behavior via formal avenues. Abilities in online banking (β=0.8165,p=0.00) show significant positive impacts, indicating the effectiveness of these skills in managing and utilizing digital savings platforms. Nevertheless, abilities such as programming continue to remain insignificant, underscoring the minimal importance of technical knowledge in shaping fundamental financial habits.

Incorporating controls improves the model's ability to explain savings behavior (R2=0.919), uncovering essential mediators like gni_per_capita and regional inequalities. For instance, affluent areas with strong digital infrastructure show elevated savings rates, whereas areas with restricted broadband access encounter considerable obstacles. The diminished importance of personal skills in the controlled model indicates that systemic elements, such as institutional trust and economic stability, play a comparable role in influencing savings behavior. A point of further discussion is the significant negative impact of the percentage of internet users saving through financial accounts. This likely arises from expanded consumption opportunities via the Internet, a transition to informal or alternative savings methods such as mobile wallets, and enhanced access to investment choices that divert funds from conventional savings. Behavioral factors, including exposure to online advertisements and trends, along with greater access to credit, further diminish the tendency to save. Moreover, internet users might be part of demographics that have reduced savings rates because of factors such as high urban living expenses, or gig economy jobs.

5.2. Findings from Random Forest

Leveraging the ability of random forests to detect non-linear and interactive relationships, we investigate various target variables, as specified above. Our findings highlight the intricate elements of financial inclusion and well-being, illustrating the crucial role of digital skills, literacy, and infrastructure alongside socioeconomic disparities.

Utilizing models with and without controls offers a greater understanding of the fundamental dynamics at play. This dual approach not only provides an understanding of the direct impacts of predictors but also examines broader contextual factors that affect financial outcomes. Models without controls emphasize the natural predictive power of features, demonstrating their direct influence on outcomes. Conversely, models that incorporate controls address structural and regional variations, thereby elucidating the overall effects of predictors. Elements such as income brackets, geographic markers, and infrastructure aspects (such as internet access) give context to the analysis by depicting the essential socio-economic framework. Incorporating these variables frequently alters the comparative significance of predictors. This emphasizes that although Internet banking abilities are crucial, their effect is dependent on wider socioeconomic circumstances, albeit to varying degrees in different models.

1. *Deposit with Formal Accounts (deposit_has_fin_account_pct)*

Metric/Feature	Without Controls	With Controls
skill_arithmetic_formula_spreadsheet_pct	0.1128	0.0206
skill_internet_banking_pct	0.7059	0.3604
skill_software_run_over_the_internet_pct	0.1016	0.021
skill_uploading_content_pct	0.0462	0.0043
skill_programming_pct	0.0335	0.0056

Figure 17: Feature Importances for Depositing in a Financial Account

The results indicate that digital literacy, especially skills related to Internet banking, is a key factor in determining financial inclusion. In the model lacking controls, skills in Internet banking represent 70.59% of the explained variance. Nevertheless, once controls are implemented, their significance decreases to 36.04%, emphasizing the mediating influence of socioeconomic and infrastructural elements. GNI per capita (11.1%) and fixed broadband expenses (10.25%) surface as crucial indicators with controls, highlighting the structural obstacles to financial inclusion in economically disadvantaged areas. These results indicate that enhancing digital literacy by itself may not be enough; tackling infrastructure expenses and income inequalities is also vital. Such findings become crucial for us as we explain further policy in our paper. It is also interesting to control for income groups for policy-making. Control for the high-income group accounts for 4% of the explained variance. Particularly noteworthy is the impact of educational attainment (3.14%) and internet access in schools (1.86%) in explaining the variance in depositing using a formal account. This gives credence to the impact of developing early digital literacy among children for improved financial outcomes later.

2. *Very Worried Index (very_worried_index)*

Metric/Feature	Without Controls	With Controls
skill_arithmetic_formula_spreadsheet_pct	0.3628	0.0453
skill_internet_banking_pct	0.2744	0.0316
skill_software_run_over_the_internet_pct	0.1991	0.0055
skill_uploading_content_pct	0.109	0.0024
skill_programming_pct	0.0546	0.0038

Figure 18: Feature Importances for the Very Worried Index

This index shows strong predictive capabilities, achieving an R^2 of 0.7660 in the controlled model. As expected in our hypothesis, access to the Internet in schools appears as a significant factor (11.33%), in addition to literacy (10.51%) and skills in Internet banking (3.16%). These results indicate that significant financial worries are greatly shaped by basic education and initial digital engagement. In contrast to moderate concerns, a higher degree of worry is more connected to infrastructural and digital elements, such as the number of Internet users (5.8%), demonstrating their stronger link to systemic obstacles. More importantly, skills like running software over the Internet have an approximately negligible role (0.55%) in explaining variance once controls are introduced.

3. *Worried Index (worried_index)*

Metric/Feature	Without Controls	With Controls
skill_arithmetic_formula_spreadsheet_pct	0.1901	0.0566
skill_internet_banking_pct	0.511	0.1372
skill_software_run_over_the_internet_pct	0.1044	0.0103
skill_uploading_content_pct	0.087	0.0162
skill_programming_pct	0.1075	0.0256

Figure 19: Feature Importances for the Worried Index

For the aggregate worried index, adding controls slightly enhances the model's R^2 from 0.4259 to 0.4605, indicating the diverse impacts of individual and contextual elements. However, the small increase might be because this is not an index of extreme financial worry, therefore, there might be other controls that would explain the variation better. Digital abilities, such as online banking (13.72%) and spreadsheet proficiency (5.66%), are important, along with literacy (12.28%). Socioeconomic and infrastructural factors, including school internet access (5.65%), the cost of smartphones (6.56%), and broadband expenses (7.01%), emphasize the critical role of digital connectivity in addressing financial worries. This variable reflects a combination of personal and structural elements, indicating that alleviating financial concerns necessitates both individual empowerment and systemic enhancements.

4. *Somewhat Worried Index (somewhat_worried_index)*

Metric/Feature	Without Controls	With Controls
skill_arithmetic_formula_spreadsheet_pct	0.188	0.0139
skill_internet_banking_pct	0.4322	0.0886
skill_software_run_over_the_internet_pct	0.101	0.0181
skill_uploading_content_pct	0.1237	0.019
skill_programming_pct	0.155	0.0281

Figure 20: Feature Importances for the Somewhat Worried

This model exhibits reduced effectiveness, showing an R^2 of 0.2681 in the controlled model. The relatively limited explanatory power indicates that moderate financial concerns might be influenced by other unaccounted factors, like mental health or social safety nets. However, literacy (22.14%) and skills in Internet banking (8.86%) continue to be important predictors. The significance of infrastructure is less pronounced here, like school internet access (1.12%), suggesting that moderate concerns might be shaped less by access and more by individual and psychological elements.

5. *Not Worried Index (not_worried_index)*

Metric/Feature	Without Controls	With Controls
skill_arithmetic_formula_spreadsheet_pct	0.1358	0.0085
skill_internet_banking_pct	0.3837	0.0512
skill_software_run_over_the_internet_pct	0.3229	0.063
skill_uploading_content_pct	0.0857	0.0065
skill_programming_pct	0.0719	0.0113

Figure 21: Feature Importances for the Not Worried Index

The not_worried_index shows the highest performance, achieving an R^2 of 0.7926 in the controlled model. This highlights the model's capability to identify the factors influencing financial confidence when posed with a financial concern. Literacy stands out as the primary predictor (19.36%), with digital abilities like internet banking (5.12%) and software application (6.30%) following. Notably, internet access in schools represents 8.79% of the explained variance, emphasizing the importance of early digital exposure in building financial confidence. In the absence of controls, internet banking skills (38.37%) and software usage (32.29%) prevail, indicating that these abilities are essential for financial security but are shaped by various contextual factors.

6. *Borrowed Any Percentage of Money (borrowed_any_pct)*

Metric/Feature	Without Controls	With Controls
skill_arithmetic_formula_spreadsheet_pct	0.2492	0.0147
skill_internet_banking_pct	0.1341	0.0141
skill_software_run_over_the_internet_pct	0.2496	0.0137
skill_uploading_content_pct	0.1693	0.0148
skill_programming_pct	0.1978	0.0156

Figure 22: Feature Importances for Borrowing from a Formal Channel

> The models analyzing borrowing behavior exhibit inadequate performance, showing negative R^2 values for both cases, with or without controls. This suggests that the factors considered in the analysis do not effectively identify the main influences on formal borrowing. Nevertheless, GNI per capita (22.54%) prevails in the model with controls, underscoring the systemic disparities in borrowing access. The low significance of digital skills, like programming (1.56%), indicates that borrowing habits are shaped more by access and affordability than by digital literacy. This conclusion should be highly indicative of policymaking in Pakistan, which continues to largely focus on education rather than ease of access or removal of borrowing constraints.

The random forest analysis emphasizes both the significance of specific predictors and their interactions. For instance, the relationship between literacy and Internet banking abilities enhances their joint effect on financial confidence and inclusion. In a similar vein, the interaction between infrastructure (such as broadband expenses) and digital literacy highlights the multiplied difficulties encountered by underserved communities. These interactions imply that interventions need to be multi-faceted, tackling digital skills, education, and infrastructure at the same time. Additionally, the differences in feature significance among target variables suggest that the factors influencing financial results are specific to the context. For example, although online banking is essential for deposits and financial assurance, it has a lesser impact on addressing significant financial issues. This highlights the necessity for focused measures designed for particular financial results.

5.3. Findings from Structural Equation Modelling (SEM)

Our findings from SEM are important regardless of other methodologies used since they allow us to identify the magnitude and significance of the direct and indirect effect of Digital Literacy on Financial Outcomes. Simultaneously, they quantify the importance of the proposed channel Financial Inclusion, through which we hypothesize Digital Literacy mainly impacts Financial Outcomes.

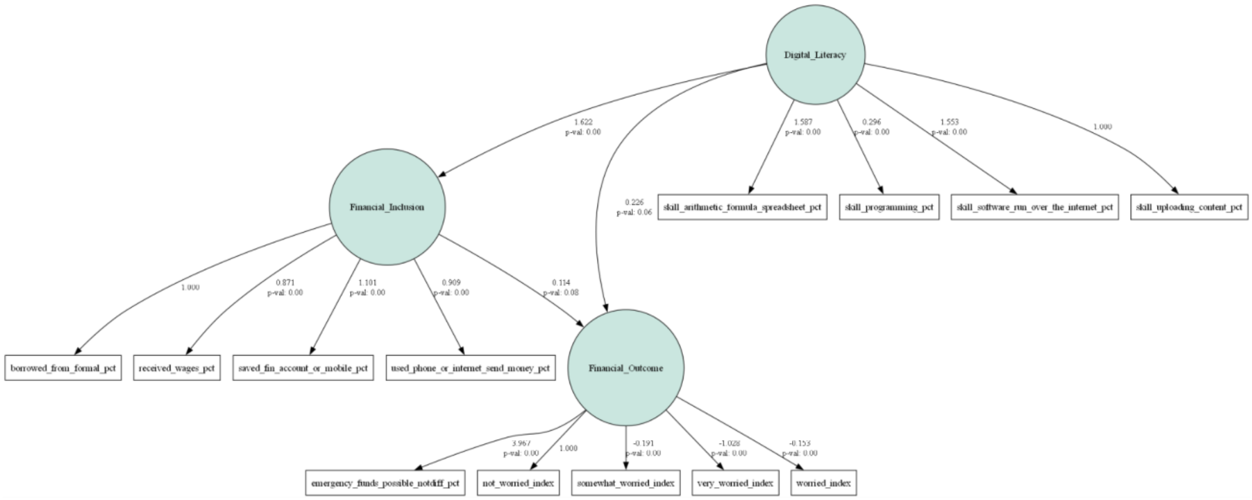

Figure 23: Path Diagram with no controls

Our first model without any controls just serves as a measure to demonstrate how robust the magnitude and significance of our effects is to controls. We see here that all the factor loadings (which represent the relationships between latent variables and their observed indicators) are significant at the 1% level. These findings show that are measured variables or indicators do a substantial job in forming our latent variables. A 1 standard deviation change in the latent variable is associated with factor loading* standard deviation change in the value of the measured indicator. For example, a 1 standard deviation change in Digital Literacy is associated with a 0.296 standard deviation (of the variable programming skills percentage) increase in programming skills percentage.

We can also note from Figure 23 that our path coefficients are all significant at the 10% level. Without any controls, a standard deviation increase in Digital Literacy is associated with a direct increase in Financial Outcomes by 0.226 standard deviations, and an indirect increase in Financial Outcomes (through Financial Inclusion) by 0.185 standard deviations (1.622*0.114). This results in a total effect of 0.41 standard deviations.

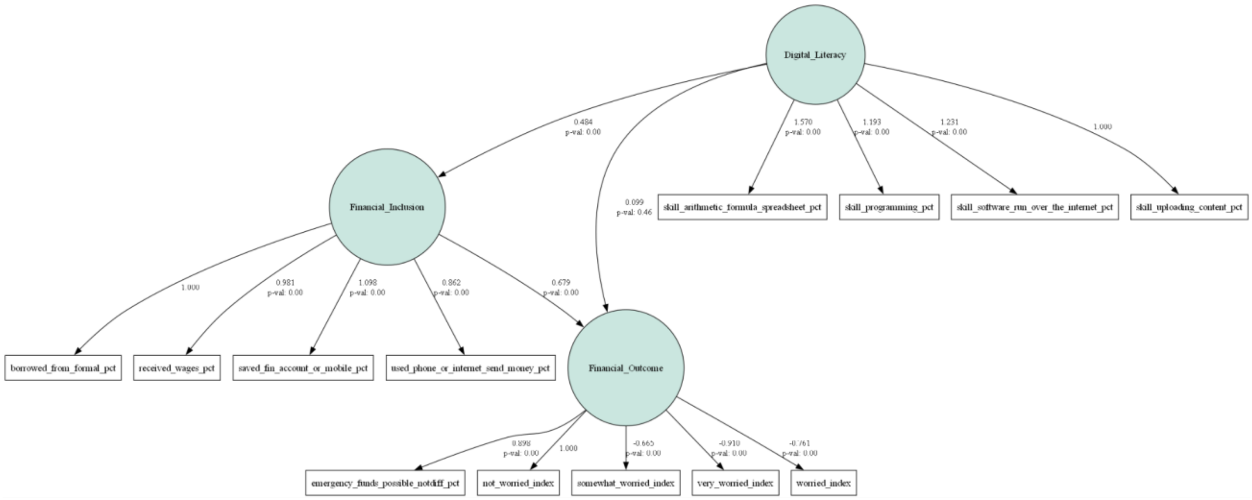

Figure 24: Path Diagram with selected controls

Our second model includes some select controls from the forty shortlisted through VIF. Inclusion does not impact the ability of our observed variables to predict our latent variables (factor loadings remain significant at the 1% level). However, the factor loadings do fluctuate (some increase, while some decrease) once controls are added.

We observe that as even some controls are added, our model moves toward perfect mediation as the direct effect (path coefficient between Digital Literacy and Financial Outcomes) becomes insignificant. The indirect effect remains significant so that an increase in Digital Literacy by 1 standard deviation is associated with an indirect increase in Financial Outcomes by 0.329 standard deviations (0.484*0.679). It is important to note here that the magnitude of the indirect effect increases mainly because of the increase in the path coefficient between Financial Inclusion and Financial Outcomes.

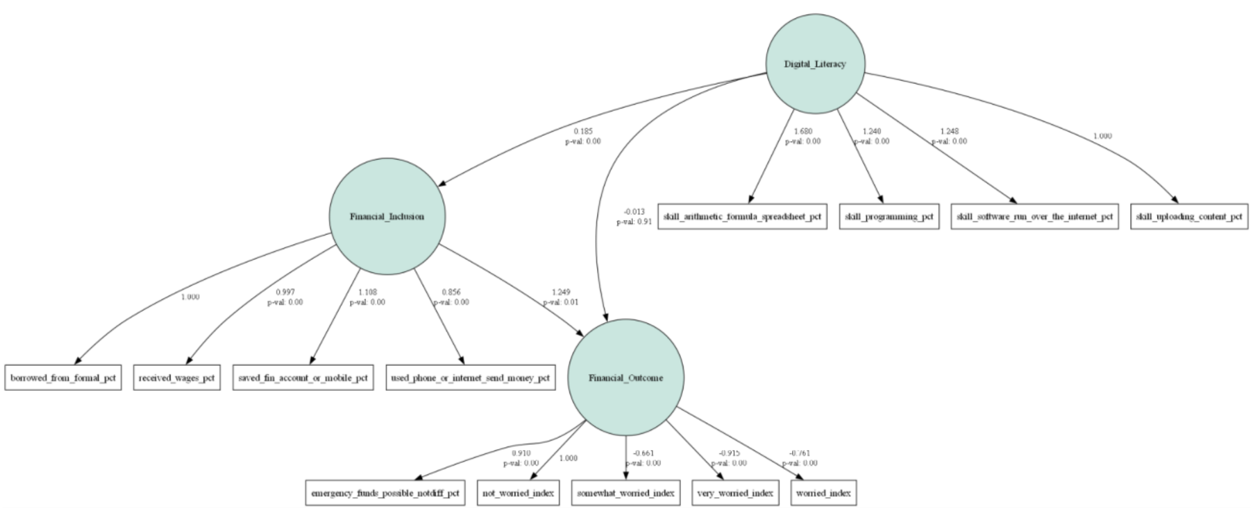

Figure 25: Path Diagram with all 40 controls

The inclusion of all 40 controls in our model does not change much in terms of our factor loadings or perfect mediation being achieved. The direct effect remains insignificant, while the indirect effect weakens to an associated increase of 0.231 standard deviations (0.185*1.249) in Financial Outcomes through the Digital Literacy → Financial Inclusion pathway. Regardless, the indirect effect is significant and therefore robust to controls, affirming that Digital Literacy has a strong correlation with Financial Outcomes, and this correlation is manifested through Digital Literacy's impact on Financial Inclusion and Financial Inclusion's in turn impact on Financial Outcomes.

Metric	Model with All Controls	Model with Fewer Controls	Interpretation
Chi² p-value	0.0102 (poor fit)	0 (poor fit)	Both models fail, but the all-controls model is marginally better.
TLI	0.971 (excellent fit)	0.820 (mediocre fit)	The all-controls model has a significantly better TLI.
BIC	414.91	274.50	The fewer-controls model is more parsimonious (better BIC).
RMSEA	0.0352 (excellent fit)	0.102 (poor fit)	The all-controls model is significantly better in terms of RMSEA.

Figure 26: Fit and performance metrics for SEM models

Figure 26 above highlights and compares the results of some fit and performance metrics that were most commonly used in the literature for such analysis for our 2 SEM models. We leave out the model without controls because we see significant changes in mediation once controls are added and endogeneity bias is lowered.

The Chi² - test is failed by both models (p-values are below 5%) which may indicate that there are significant differences between the observed covariance matrix (actual data) and the model-implied covariance matrix (what the model predicts the data should look like). However, this test may be misleading because we run the models on a relatively small sample size (below 200) and the Chi² distribution is extremely sensitive to sample size (Saris et al., 2008). This can also be further proved by the Tucker-Lewis Index (TLI), which also comments on how well the model fits the data compared to a null model (a baseline with no relationships) while penalizing model complexity. The TLI for the control model is within the range of an excellent fit (Kline, 2015; Schermelleh-Engel et al., 2003). The Root Mean Square Error of Approximation (RMSEA) has a similar purpose, while focusing on the error in approximating the population data and also determines the all controls model to be an excellent fit (Kline, 2015; Schermelleh-Engel et al., 2003).

The above discussion prompts us to compare both the models, and while the control model is deemed to be better by the three chosen fit metrics, the Bayesian Information Criterion (BIC) is lower, and significantly so, for the select controls model. Therefore, the model with all controls captures data structure more accurately, while the model with fewer controls sacrifices some fit quality and gains in parsimony, avoiding potential overfitting. Since our main purpose is to make theoretical claims with robust evidence rather than prediction, we favor the all-control model.

Overall, our SEM findings indicate that our observed variables are suitable indicators of our latent variables (the variables of interest in our study) as p-values are all significant at the 1% level. We achieve perfect mediation as we add controls, indicating that Digital Literacy does not have a direct association with Financial Outcomes. However, it does significantly associate indirectly with Financial Outcomes through our hypothesized pathway: Financial Inclusion.

6. Conclusion and Policy Implications

This study utilized three distinct methodologies to analyze the relationship between digital literacy and financial outcomes, focusing on the pathways through which this association operates. Structural Equation Modeling (SEM) revealed full mediation via the pathway Digital Literacy → Financial Inclusion → Financial Outcomes, as the direct effect of Digital Literacy → Financial Outcomes was found to be statistically insignificant. This result was robust to the number of controls included in the model, underscoring the critical role of financial inclusion in mediating this relationship.

The analysis also highlighted that specific digital skills disproportionately influence both digital literacy and, consequently, financial outcomes. Notably, the ability to engage in internet banking emerged as a pivotal skill, aligning with the findings of the SEM analysis. In contrast, for indices capturing extreme financial distress, basic digital skills—such as operating spreadsheets—were observed to have greater relevance than advanced skills like programming.

From a policy perspective, it is important to recognize the diminishing marginal returns to digital literacy, particularly in developed regions. Consequently, emphasizing digital literacy initiatives is more impactful in developing economies like Pakistan, where the baseline digital literacy is low. However, the results also emphasize that digital literacy alone is insufficient; the development of financial infrastructure is equally crucial, as it serves as the primary pathway through which improvements in digital literacy translate into better financial outcomes.

While the cross-sectional nature of the data limits definitive causal claims, the combined use of OLS, random forest, and SEM methodologies provides evidence to inform policy decisions:

1. Promoting digital literacy without concurrent improvements in financial infrastructure risks creating a demand-supply mismatch, where individuals become digitally literate but face limited opportunities for financial inclusion. Policies must prioritize simultaneous development to ensure that the benefits of digital literacy are realized.

2. The ongoing significance of internet banking abilities highlights the need for hands-on, transaction-based digital education. In regions with limited banking access, training initiatives should focus on mobile banking applications and strategies to prevent digital fraud, empowering individuals to engage securely in financial activities.
3. Policies must address gaps in digital infrastructure, such as broadband and mobile network accessibility. The importance of internet access in schools, particularly for individuals in the "Very Worried" financial distress index, underscores the long-term advantages of fostering early digital engagement.
4. Digital literacy initiatives should be tailored to meet the specific needs of marginalized and economically disadvantaged communities, ensuring that resources are allocated effectively to maximize impact.

In summary, the findings emphasize the intertwined nature of digital literacy and financial infrastructure, particularly in developing economies. Efforts to enhance digital literacy must be complemented by robust investments in financial systems and infrastructure to achieve meaningful and sustainable improvements in financial outcomes.

7. Limitations of the study

Firstly, although our models offer valuable insights, their interpretability is constrained by the inherent risk of overfitting, particularly with smaller datasets like the ones we have employed. The sample size may restrict the statistical power, particularly in subgroup or multi-group SEM analyses, as smaller respondent subsets could produce less dependable estimates. Also, our models are limited by their ability to derive correlations rather than cause-and-effect relationships. Therefore, exploring longitudinal datasets or quasi-experimental approaches (for instance, utilizing regional differences in the deployment of digital infrastructure) is required to determine causality. Moreover, endogeneity concerns, like reverse causation (e.g., financially secure people enhancing their digital abilities) or missing variables (e.g., local infrastructure), could have affected the outcomes of our analysis. Aspects like macroeconomic stability, policy measures, or historical disparities were not distinctly accounted for, potentially complicating the outcomes. While strong controls reduce these effects, using instrumental variable techniques or longitudinal data would enhance causal assertions. Furthermore, there are measurement constraints in SEM, whereby its analysis reveals connections, but it does not completely tackle endogeneity or omitted variable bias. For instance, unseen elements such as individual choices or past obstacles to obtaining digital literacy may affect both digital literacy and financial results.

Secondly, our study uses aggregated data (e.g., employing country-level statistics) which may obscure differences within a country, particularly between urban and rural communities. Upcoming studies ought to take into account disaggregated datasets to reflect these subtleties. Stratified analyses would improve representativeness. Moreover, in our random forest analysis, the issue of regional inequalities could still arise. The rankings of feature importance demonstrate overall trends, yet the significance of predictors can differ across regions. Hence, analyses at the subnational level may offer more focused insights. Also, the dependence on cross-sectional data limits the capacity to understand how digital literacy and financial inclusion evolve and interact over time.

Thirdly, while random forests work for non-linear relationships and capture implicit interactions between variables, they do not naturally offer an understanding of interactions between features. Future research may employ techniques like SHAP (SHapley Additive exPlanations) values to measure these interactions. While SEM defines and tests interactions between variables and can model causal pathways, SHAP can complement by providing key interactions between variables, and showing how feature combinations affect the predictors.

Furthermore, there is an assumption of homogeneity in our analysis whereby the research presumes consistent impacts of factors such as digital literacy among all participants. In reality, these impacts are probably diverse based on socio-economic, cultural, and infrastructural elements. Similarly, the analysis emphasizes formal financial inclusion (borrowed_from_formal_pct, saved_fin_account_or_mobile_pct) while overlooking the importance of informal financial services, which are significant in rural or underdeveloped regions. Indices such as worried_index rely on subjective evaluations, which might be biased or influenced by intermediate factors (e.g., recent economic occurrences).

8. Future Research

This study also provides a guide for future research in this domain and in relation to the research question. One particular avenue could be delving into heterogeneity, specifically examining why digital literacy skills (e.g., skill_internet_banking_pct) have different effects across various regions (e.g., North America vs. Africa). This could involve subgroup analysis or interaction terms using region-level indicators to account for variations in culture, infrastructure, and policy.

Additionally, exploring panel data analysis could help control for time fixed effects and potentially employ causal techniques such as difference-in-differences to yield more unbiased results. Other techniques, such as instrumental variables or natural experiments, can also be investigated to establish causal inferences.

Future research could also focus on identifying more accurate indicators for all latent variables considered in this study, particularly financial well-being, as this study had to rely on self-reported data. This would help mitigate bias in the results presented.

Lastly, future research can explore analyses using micro-level data rather than aggregated data, which would help incorporate behavioral factors and examine the effects of gender and other socio-demographic factors.

Works Cited

Ameen, K., & Gorman, G. E. (2009). Information and digital literacy: a stumbling block to development? *Library Management*, *30*(1/2), 99–112. https://doi.org/10.1108/01435120910927565

Ariana, I. M., et al. (2023). The effects of financial literacy and digital literacy on financial resilience: Serial mediation roles of financial inclusion and financial decisions. *Uncertain Supply Chain Management*, *12*(1), 999–1014. https://www.researchgate.net/publication/377584380_The_effects_of_financial_literacy_and_digital_literacy_on_financial_resilience_Serial_mediation_roles_of_financial_inclusion_and_financial_decisions

Choung, Y., Chatterjee, S., & Pak, T.-Y. (2023). Digital financial literacy and financial well-being. *Finance Research Letters.* https://www.econstor.eu/bitstream/10419/276210/3/Choung-Chatterjee-Pak-FRL.pdf

Dura, J. (2022). Determinants of financial literacy and digital literacy on financial performance in driving post-pandemic economic recovery. *Journal of Contemporary Eastern Asia*, *21*(2), 47–69. https://www.researchgate.net/publication/355969790_Digital_and_Financial_Literacy_as_Determinants_of_Digital_Payments_and_Personal_Finance

Kline, R. B. (2015). *Principles and practice of structural equation modeling* (4th ed.). Guilford Publications. https://psycnet.apa.org/record/2015-56948-000

Lyons, A., Kass-Hanna, J., & Greenlee, A. (2020). Impacts of financial and digital inclusion on poverty in South Asia and Sub-Saharan Africa. *SSRN Electronic Journal.* https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3684265

Pham, M. H., & Linh, T. P. (2020). The impact of financial inclusion on financial stability in Asian countries. https://www.researchgate.net/publication/342565890_The_Impact_of_Financial_Inclusion_on_Financial_Stability_in_Asian_Countries

Putri, V. Y., & Nurhayati, I. (2022). Does financial inclusion improve financial literacy? An ASEAN cross-country analysis. https://www.atlantis-press.com/proceedings/icveast-23/125993589

Rojas-Suarez, L., & Fiorito, A. (2021). Building digital financial inclusion: A decision tree approach.
https://www.jstor.org/stable/resrep58233

Salman, M., Rauf, N., & Murtaza, Z. (2024). Unlocking financial inclusion and economic empowerment in rural Pakistan: The interplay of financial literacy and infrastructure development in the impact of digital wallets.
https://journals.smarcons.com/index.php/jems/article/view/308

Saris, W. E., et al. (2008). Testing structural equation models or detection of misspecifications? ESADE, Universitat Pompeu Fabra, & Radboud University.
https://www.researchgate.net/publication/233293890_Testing_Structural_Equation_Models_or_Detection_of_Misspecifications

Schermelleh-Engel, K., et al. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online*, *8*(2), 23–74.
https://psycnet.apa.org/record/2003-08119-003

Sjam, A. A. (2024). The dual lens: Unveiling the dynamics of financial and digital literacy in financial decision making. *AKSARA: Jurnal Ilmu Pendidikan Nonformal*, *10*(1), 99–112.
https://doi.org/10.37905/aksara.10.1.99-112.2024
https://ejurnal.pps.ung.ac.id/index.php/Aksara/article/download/2338/1623

Widyastuti, U., et al. (2024). The nexus of digital financial inclusion, digital financial literacy, and demographic factors: Lessons from Indonesia. *Cogent Business & Management*, *11*(1), 1–23.
https://www.tandfonline.com/doi/full/10.1080/23311975.2024.2322778

Appendix

A.1. Variable Descriptions

Variable Type	Variable Name	Description
Control Variables	Global_peace_index	Variable created by META to measure peace in a country. Score 1-5, 1 = best.
	Adult_pop	Adult population by country and year.
	mkt_structure_fibre	Measures the level of competition in the fiber market of a country for a specific year, categorized as monopoly, partial competition, or full competition.
	mkt_structure_fixed_wireless_broadband	Measures the level of competition in the fixed wireless broadband market of a country for a specific year, categorized as monopoly, partial competition, or full competition.
	region	Maps countries to their regions. Categories: North America, Africa, South America, Asia, Europe, Middle East, and Oceania.
	income_group	Divides countries into income levels: high income, upper middle income, lower middle income, and low income.
	schools_internet_access_pct	Measures the % of schools with internet access.
	population_under_poverty_line_pct	Measures the % of the population under the poverty line.
	gni_per_capita	Measures Gross National Income per capita.
	internet_users_pct	Measures internet users as a % of the total population.
	mobile_subscribers_pct	Measures mobile subscribers as a % of the total population.
	mobile_upload_speed	Measures average mobile upload speed in Mbps.
	2G_coverage_pct	Measures the % of the total population with 2G coverage.
	gini_coefficient	Score (0-100): 0 = perfect income equality; 100 = perfect income inequality.

	3G_coverage_pct	Measures the % of the total population with 3G coverage.
	4G_coverage_pct	Measures the % of the total population with 4G coverage.
	smartphone_cost	Score (0-100): 100 = most affordable.
	mobile_cost_postpaid	Measures cost of postpaid network as a % of monthly GNI per capita.
	fixed_broadband_cost	Measures cost of fixed-line broadband as a % of monthly GNI per capita.
	urban_population	Measures urban population as a % of the total population.
	efinance_content	Qualitative rating (0-2): 2 = best.
	support_digital_literacy	Qualitative rating (0-3): 3 = best.
	edu_attainment_pct	Measures years of schooling.
	literacy_pct	Measures literacy rate.
	mobile_cost_prepaid	Measures cost of prepaid network as a % of monthly GNI per capita.
X Variables	skill_internet_banking_pct	Proportion of individuals with internet banking skills.
	skill_uploading_content_pct	Proportion of individuals with the skill to upload content.
	skill_arithmetic_formula_spreadsheet_pct	Proportion of individuals able to apply basic and arithmetic formulas in a spreadsheet.
	skill_software_run_over_the_internet_pct	Proportion of individuals using internet-based software for editing text, spreadsheets, or presentations.
	skill_programming_pct	Proportion of individuals capable of writing a computer program using a programming language.
Intermediate Financial Variables	borrowed_from_formal_pct	% of the population (age 15+) that borrowed from a formal financial institution.
	used_phone_or_internet_send_money_pct	% of the population (age 15+) that used mobile phones or the internet to send money.
	saved_fin_account_or_mobile_pct	% of the population (age 15+) that saved at a financial institution or via a mobile money account.
	received_wages_pct	% of the population (age 15+) that received wages.
	borrowed_for_health_pct	% of the population (age 15+) that borrowed for health or medical purposes.

Y Variables	emergency_funds_possible_somewhat_diff_pct	% of the population (age 15+) reporting emergency funds in 7 days were possible but somewhat difficult.
	emergency_funds_possible_notdiff_pct	% of the population (age 15+) reporting emergency funds in 7 days were possible and not difficult.
	emergency_funds_possible_pct	% of the population (age 15+) reporting emergency funds in 7 days were possible.
	deposit_has_fin_account_pct	% of the population (age 15+) with a financial institution account who made a deposit.

Our cleaned dataset and result files can be found here.

Credit Statement

All four authors contributed to all sections of the paper. Overall contribution of each author is 25% each in this research. The following credit statement underlines contribution of each member to sections in order of importance.
Schaff Mirza: Software, Methodology, Analysis, Data Curation, Visualization, Writing, Conceptualization. Khan Abdullah Bin Azim: Visualization, Software, Analysis, Data Curation, Methodology, Writing, Conceptualization. Jovera Shakeel: Analysis, Methodology, Resources, Conceptualization, Software, Writing, Visualization. Shehzil Munir: Conceptualization, Methodology, Resources, Analysis, Software, Writing, Visualization.