Where are the unbanked in **Belize?**

Using Machine Learning Small Area Estimation to improve Financial Inclusion geographic targeting

Jonathan Hersh
Chapman University Argyros School of Business

Lucia Martin Rivero
Inter-American Development Bank

Janelle Leslie
Inter-American Development Bank

IDB
Inter-American Development Bank

Cataloging-in-Publication data provided by the
Inter-American Development Bank
Felipe Herrera Library

Hersh, Jonathan.
Where are the unbanked in Belize? Using machine learning small area estimation to improve financial inclusion geographic targeting / Jonathan Hersh, Lucia Martin Rivero, Janelle Leslie.

p. cm. – (IDB Monograph; 939)
Includes bibliographic references.

1. Financial services industry-Social aspects-Belize. 2. Machine learning-Belize. 3. Finance-Social aspects-Belize. 4. Poor-Latin America-Finance, Personal. I. Martin Rivero, Lucia. II. Leslie, Janelle. III. Banco Interamericano de Desarrollo. Country Department Central America, Haiti, Mexico, Panama and the Dominican Republic. IV. Title. V. Series.

IDB-MG-939

JEL classification: D14 Household Saving Personal Finance, G21 Banks Depository Institutions Micro Finance Institutions Mortgages, G51 Household Saving, Borrowing, Debt, and Wealth, G53 Financial Literacy, O31 Innovation and Invention: Processes and Incentives, O35 Social Innovation.

Keywords: Belize, Central America, financial inclusion, big data, machine learning, small area estimation, social innovation, geographic targeting, geographic information systems

July 2021

Jonathan Hersh
Chapman University Argyros School of Business

Lucia Martin Rivero
Inter-American Development Bank

Janelle Leslie
Inter-American Development Bank

Where are the unbanked in **Belize?**

Using Machine Learning Small Area Estimation to improve Financial Inclusion geographic targeting

The Inter-American Development Bank (IADB) is grateful to Sheree Smiling Craig and Emory Ford from the Central Bank of Belize (CBB) for providing us with access to the financial inclusion module included in the April 2019 national Labor Force Survey (LFS). The LFS is a national survey that is conducted by the Statistical Institute of Belize (SIB). The IADB also acknowledges the instrumental comments and suggestions received from the staff of the CBB and the SIB, as well as Maria Deza (IADB), during preliminary presentations of this exercise.

Executive Summary

Access and use of suitable financial services are crucial components of economic and social development. In recognition of the role that financial inclusion can play in financial stability and, ultimately, economic prosperity, Belizean financial authorities launched in 2019 an ambitious National Financial Inclusion Strategy (NFIS) program, which outlines the policy approach being taken to improve financial inclusion in the country. The NFIS includes four Thematic Financial Inclusion Task Forces charged with increasing the share of adults who possess a bank account from 66% in 2019 to 80% by 2022.

The NFIS draws on data collected during the 2019 Labor Force Survey, which reports on the nature of Belize's unbanked population at the level of the country's administrative divisions, of which Belize has six districts. According to the document, having insufficient funds was identified as the most common reason for not utilizing financial services, with those in rural areas and those possessing lower levels of education found to be least likely to own a bank account. However, from the data gathered, only aggregate conclusions and recommendations can be obtained. Identifying the distribution of unbanked individuals within the landscape of Belize's districts and designing policies tailored to these different realities, requires data at smaller geographical units.

The objective of this paper is twofold. First, we employ Machine Learning Based Small Area Estimation (ML-SAE) in order to develop more granular estimates of Financial Inclusion at a smaller geographical level known as Enumeration Districts. While Belize is divided into six districts, there are more than 700 Enumeration Districts.

Identifying exactly where the unbanked are located in each of the Enumeration Districts in the country will support the geographical targeting of the Financial Inclusion Strategy. Focusing the interventions on some regions rather than others, could be more effective as opposed to a broader policy approach in closing geographic disparities. If the unbanked population is relatively evenly dispersed, a broad campaign to increase financial access across the country may suffice. However, if we are able to identify large clusters (high concentrations of data points in one area relative to surrounding areas) of unbanked individuals, it is likely that specific policies targeted at geographic areas where the unbanked populations are clustered may be more effective than a broad strategy.

To have a deep understanding of the population's financial characteristics at the Enumeration district level, we built five measures of access to banking services: 1) whether a household has a bank account; 2) whether a household member cites

the bank being too far away as the reason for not having a bank account; 3) whether the reason they do not have a bank account is due to not having enough income; 4) whether they have used online banking services, and 5) whether they have used any formal lending services.[1] ML-SAE was then used to identify the degree of concentration of unbanked individuals in each Enumeration Districts, given the selected Financial Inclusion metrics. Furthermore, the clusters of unbanked persons were analyzed to reveal patterns specific to each Financial Inclusion metric tracked.

We find significant clustering in Financial Inclusion metrics that are not apparent in the district level averages. For instance, we are able to identify that households with bank accounts are concentrated in urban areas, such as Belize City and Belmopan City, as well as tourism hubs, such as Placencia Village and San Pedro Town. Meanwhile, households without bank accounts were found to be of highest concentration in rural areas in the south and north of Toledo, as well agricultural areas in the Orange Walk and Cayo Districts. Households that stated they do not use banks because they are located too far away, are also clustered in the South and West of the country. Interestingly, the specific pattern of clustering is distinct for each of the five Financial Inclusion metrics tracked, though low population densities, as well as cultural differences impact the use of financial services[2] and are significant factors to be considered in the interpretation of the data. The results suggest that there is significant benefit from incorporating geographic targeting into programs aimed at promoting Financial Inclusion.

The second objective of this paper is to analyze the factors that influence the use of financial services and instruments in order to propose appropriate adjustments in the strategies implemented by authorities in each geographical area. The analysis of factors correlated with higher levels of Financial Inclusion at the Enumeration District level indicates that targeting should focus on non-Belizean born individuals, as well as areas with more indigenous populations, which were identified as the populations most vulnerable to financial exclusion. Poverty was also found to be strongly correlated with low levels of Financial Inclusion. Consequently, a subsidy for opening bank accounts targeted towards the poorest households could be considered by financial authorities.

The spatial distribution of Financial Inclusion indicators and the factors influencing the adoption of financial services shed light on specific recommendations for each of the four Thematic Financial Inclusion Task Forces included in the plan of the NFIS.

We recommend that the ICT and Financial Infrastructures team considers increasing financial access points for Enumeration Districts where there is low banking usage, paying particular focus on the rural and inhabited areas, particularly in southern Toledo. To the Tailored

1 The data used to estimate access to banking services was gathered using a Financial Inclusion module that was appended to the 2019 April Labor Force Survey (LFS).
2 Some EDs demonstrating low Financial Inclusion according to the metrics used in this study, contain Mennonite communities, which utilize their own credit and loan systems separate from banks.

Financial Products and Innovation team, we recommend considering financial products that are at low cost or subsidized, tailored to individuals who currently do not have bank accounts. To the Financial Consumer Production and Financial Literacy teams, we recommend creating introductory banking finance literature in Q'eqchi' (Kekchi) Mayan, Mopan, Yucatec Mayan, and German. Finally, to the Data Collection, Analysis and Reporting team, we commend reporting data disaggregated by gender, age, rural/urban, and by district in order to better understand the demand for financial services.

Ultimately, fostering economic growth that promotes inclusivity requires placing heightened effort and attention to those regions which are socioeconomically weakest, as identified in this study. Placing the necessary emphasis on wealth disparities evident in geographic regions, through geographically targeted policies, should enhance policy effectiveness through a more efficient distribution of resources and reduce the risk of widening the gaps between better off regions and those which are more in need. Orienting policies targeted towards addressing the specific socioeconomic factors and deficiencies identified in this paper by enumeration district, provides an avenue to accelerate the attainment of improved financial inclusivity and stability to both consumers and financial service providers that is needed to stimulate sustainable economic growth in Belize.

1. Introduction

Financial Inclusion is a crucial component of a development strategy. A bank account is a vital savings tool that allows households to save for future investments, or to weather the storm when they face unintended negative shocks. When a household has access to a bank account, it is easier for them to acquire needed lending for business and personal investments. Bank access makes it easier for poorer households to receive remittances as well as government transfers. Financial Inclusion has been shown to have macroeconomic affects as well. Sethi and Acharya (2018) use data on a panel of 31 countries from 2004-2010 and find Financial Inclusion has a causal link to economic growth.

It is not without noticing these issues that Belize embarked on an ambitious National Financial Inclusion Strategy (NFIS) program in 2019. The aim of the program is to grow the share of adults who possess a bank account from the current rate of 66% in 2019 to 80% of adults by 2022.

The NFIS provides a comprehensive overview of the current situation of Financial Inclusion. In Belize, the banking sector shows significant levels of concentration among the three largest banks. The country's credit penetration rate (73.7% in 2019) has remained below Latin America and the Caribbean (LAC) levels (83%). Similarly, depositors at commercial banks per 1,000 adults (656.5) stood below regional levels in 2018. Some of the challenges to financial infrastructure is that Belize does not yet have a credit bureau or credit registry in place.

Belize is among the countries in the LAC region with the poorest territorial geographic outreach indicators. Branches and ATMs are in all districts, but the financial infrastructure remains concentrated in urban areas, such as Belize City and Belmopan. Mobile wallets have recently been introduced but have yet to be fully adopted[3], while there is moderate interoperability between financial institutions as three banks utilize the Visa network, while one other financial institution uses a proprietary debit card and operates a small network for credit unions to access their ATMs and Point of Sale (POS) system.

Using a Financial Inclusion module that was appended to the 2019 April Labor Force Survey (LFS), the NFIS reports that 66% of adults in Belize have an account at a bank or credit union - an indicator typically used to measure Financial Inclusion. The account ownership level is along the 2017 levels reported for the LAC region (53%) and of middle-income countries (65%), yet it is below levels observed in upper middle-income countries (73%).

Across districts, adults from the Belize and Corozal Districts report the highest levels of account ownership,

3 Belize Bank is piloting the E-kyash Mobile Wallet, and National Bank is developing a mobile eWallet called NBB Pay. https://www.e-kyash.com/about-us/

while the Toledo District stands at the bottom. While 61% of adults living in rural areas report owning a formal account, a gap of 8 percentage points still favors the urban adult population. A gender gaps exists, being more significant in Cayo and Toledo districts. Some of the most cited reasons Belizeans listed for not to have an account included not having enough money (69%), followed by considering they have no need for financial services (31%)[4].

Even though the 2019 Labor Force Survey provides key insights of Financial Inclusion, its design only yields data for the unbanked population at the district level. That is, we only know which districts hold more unbanked individuals. However, we do not have information on which of the more than 700 enumeration districts (ED) hold more unbanked persons.[5]

To close this information gap, this paper employs Machine Learning Based Small Area Estimation (ML-SAE) to develop spatially disaggregated estimates of measures of Financial Inclusion at the ED level. These machine learning methods reveal which EDs hold more unbanked individuals and discuss patterns in EDs that have more or fewer unbanked persons.

This information could serve to improve the implementation of the NFIS. While there are many methods of encouraging the unbanked populations to open bank accounts, a crucial component of this strategy is knowing where unbanked are located. With the current data, which shows unbanked individuals at the district level, it is not possible to reveal these unbanked clusters (concentration of populations in certain geographic areas). Thus, we have much to gain by using methods which can reveal clusters of unbanked individuals at a lower geographic level.

In any marketing effort, geographic targeting - that is, focusing the intervention on some regions versus others - will be more effective if the response of the target population to Financial Inclusion strategies varies across the different regions of a country. Geographic targeting has been shown to be an effective means of increasing access to the banking sector for the poorest households (Kochar, 2018). By isolating areas with the largest gap in financial inclusivity, specific policy can be crafted with the goal of targeting unbanked in these areas.

Specifically, we will use machine learning methods that can generate ED level measures of Financial Inclusion metrics, rather than at the current district level of spatial aggregation. Small area estimation techniques are commonly used in development economics, typically for the purpose of estimating poverty in a smaller geographic level (Elbers et al., 2003). We employ a machine learning prediction approach, specifically a random forest model, to estimate the prediction model between the survey's household characteristics and the Financial Inclusion metrics.

Financial Inclusion is defined as equal access and opportunity to a variety of

4 The survey allowed respondents to cite more than one reason. Hence, response percentages exceed 100%.
5 Belize is divided into six districts, which in turn are further divided into enumeration districts (ED). The ED are the smallest geographical unit, and they are considered as a census/statistical "building block".

financial services such as individual bank accounts, business and personal credit, and insurance products. There are several ways to view Financial Inclusion, and if we want to create quantitative measures of Financial Inclusion, we must use precise definitions. Fortunately, we can use the Financial Inclusion module that was appended to the 2019 April Labor Force Survey (LFS) which asks several questions about Financial Inclusion at the District level.

We focus on five measures of Financial Inclusion: 1) Whether a household has a bank account; 2) Whether the reason they do not have a bank account is due to the bank being too far; 3) Whether the household says they do not have enough income to open a bank account, 4) Whether they have used online banking services, and 5) Whether they have used any formal lending services. These questions were chosen based on being the best proxies for what are viewed as international standards for Financial Inclusion, which focus on access, usage, and barriers to usage of core financial products such as savings, credit, and insurance (Cámara and Tuesta, 2014; Demirguc-Kunt and Klapper, 2012).

Our modeling approach generates a novel picture of Financial Inclusion at the spatially refined ED level - vs the District level in the LFS -, showing estimates of access, usage, and barriers to use. As we will see, we find that all five Financial Inclusion Metrics show significant spatial clustering. That is, households responding affirmative to each of the metrics are concentrated in different areas of the country, which highlights the potential advantages of incorporating geographic targeting into programs addressed at alleviating Financial Inclusion.

We also examine factors that are correlated with higher levels of Financial Inclusion for EDs. We find that ED poverty levels (Hersh et al., 2020) are strongly correlated with lower Financial Inclusion. Urban EDs, and EDs with larger populations also tend to have higher levels of Financial Inclusion. EDs with a larger share of non-Belizean born individuals, or with more ethnically indigenous individuals tend to be less financially included.

The results suggest that there are efficiency gains to targeting financial access towards these underrepresented groups and geographies. First, the clusters of unbanked EDs should be targeted, and more outreach efforts should be placed in these locations. Second, targeting should focus on non-Belizean born individuals, as well as areas with more indigenous populations. Finally, because poverty is so strongly correlated with low levels of Financial Inclusion, a subsidy for opening bank accounts targeted towards the poorest households could be considered.

2. Data

Survey Data

We rely upon two critical surveys for the analysis: the April 2019 Labor Force Survey (LFS), which includes a financial module, asking relevant questions on households' access to various financial products at the District level; and the 2010 Belizean Census. Typically, with survey-to-survey imputation, it is recommended to align temporally as much as possible the training (LFS) to the target (Census) survey (Elbers, et al., 2003). Belize was scheduled to conduct a national Census in 2020. However, due to the Covid-19 pandemic, the 2020 Census was postponed and thus not available at the time of the writing of this document.

We derived characteristics of households and generated a picture of their financial information from the April 2019 Labor Force Survey. The survey samples 8,155 individuals in 2,216 households. Because it is easier to model household-level characteristics, we transformed all individual information into household level aggregates. As we discussed in the introduction, Financial Inclusion is defined broadly, but we focus on questions in the LFS that proxy for measures of access, use, and barriers to use for savings and credit. From the LFS FI module we defined five binary measures meant to proxy for Financial Inclusion more broadly. The measures of Financial Inclusion we derive are[6]:

1. Does anyone in the household have an account at a credit union or a bank? (FI module question 1)
2. Is the reason you do not have an account at a bank or credit union because financial institutions are too far away? (FI module question 3-A)
3. Is the reason you do not have an account at a bank or credit union because you do not have enough money to use them? (FI module question 3-F)
4. In the past 12 months have you used Internet / online banking? (FI module question 4-E)
5. In the past 12 months, have you borrowed any money from a bank, credit union or another type of formal institution? (FI module question 7)

In Figures 1 A-E we present bar charts at the District level, summarizing the fraction of households responding in the affirmative for the given question. The first measure tracks any usage of formal banking activity and should be considered a baseline interaction with the financial system. Figure 1A shows Toledo and Orange Walk have the greatest percentage of households without banking access, with roughly 40-50% of the households without a bank account. Belize and Corozal appear to have the greatest penetration of banking access, although roughly one fourth of households in these regions do not have a bank account.

6 For reasons of brevity and clarity we have paraphrased from the exact wording in the survey. The exact survey wording is provided in the Appendix.

The second and third measures track specific impediments to personal finance demand by the households. The former tracks whether the lack of bank use happens because they are physically too far away. Only 3% of households overall cite this as a reason, but many more cite this in Toledo (7.9%) or Stann Creek (5.4%) than in Belize City (0.8%). The third measure tracks whether the main reason individuals may not be able to use the financial system is because they do not have enough money to part with any amount needed to open a bank account. Figure 1B shows Toledo and Cayo as the districts with the greatest number of households who cite this as a primary reason: roughly 30% of households say they possess insufficient money to open a bank account.

We also look at household engagement with the financial market for two less commonly used financial products. The fourth measure tracks whether households have used any online banking in the past 12 months. Figure 1C shows online banking penetration is low overall. Belize City shows the highest online banking penetration, at around roughly 18%. The last measure captures whether households, when facing credit constraints, are able to meet these constraints by accessing formal banking institutions. Belize City and Corozal show the greatest number of households accessing formal borrowing, with roughly 25% doing so. Residents of Toledo, however, access formal banking roughly 10% of the time.

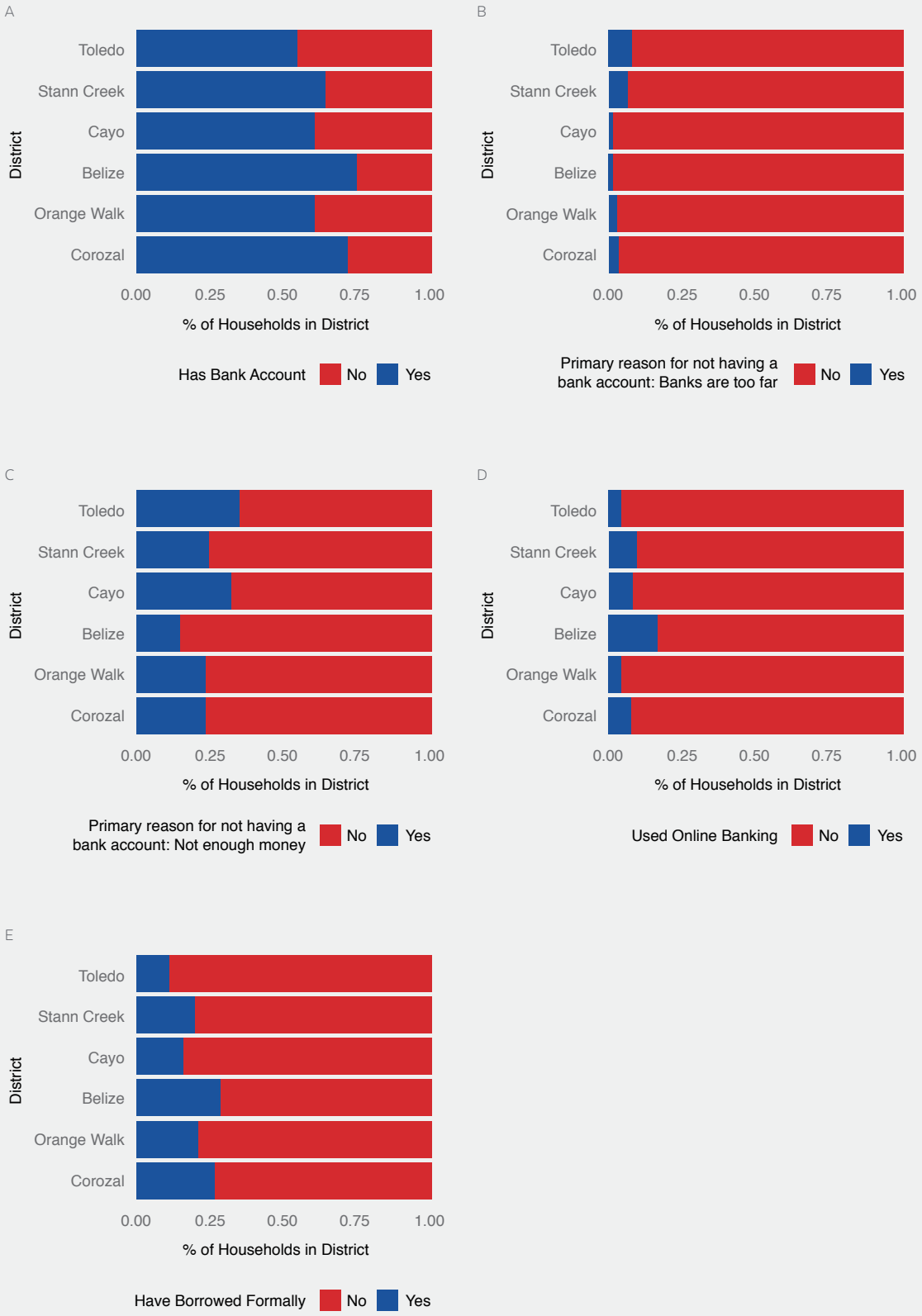
While these District-level analyses show broad usage at the District level, to build estimates of ED levels of Financial Inclusion, we will apply the Small Area Estimation technique. This takes the household level model estimated against the LFS and applies that model to matched household variables in the Census[7]. We will then build a statistical model that finds patterns relating these household characteristics and the measures of Financial Inclusion shown in Figure 1. That model will be applied to every household in the Census, giving us a predicted measure of financial usage for every household in the Census, which we can then aggregate to the ED level.

Household-Level Covariates in the LFS

To build the prediction model, we must first create household level features or variables that can predict household heterogeneity in Financial Inclusion. Table 1 shows the household-level variables that we compiled, along with a household level summary of the financial inclusion metrics shown in Figures 1 A-D. The household level predictors include several household assets (computers, mobile phone, air-condition, etc.) along with demographic characteristics of the household (dependency ratio, number of children, etc.). We also include some measure of public services such as whether the home is connected to the electrical grid, or whether it has access to a poor-quality toilet or not. The 2019 NFIS found that there persist a 7.1 percentage point gender gap in access to bank accounts, with males having a higher rate of bank account ownership. While gender remains an important consideration for individual targeting, our methodological strategy targets household level characteristics

7 For more detail we refer the reader to Hersh et al., (2020) in particular the methods section.

and not individual characteristics such as gender. Further investigation may be necessary to uncover the difference in Financial Inclusion between female headed households and male headed ones.



Figures 1 A-E: Description statistics from the Financial Inclusion module from the April 2019 Labor Force Survey. Responses are weighted according to survey weights

Variable Name	Mean	Std. Dev
Household Financial Inclusion Metrics		
Household does not have a bank account	33%	(0.47)
Primary reason for household not having a bank account because closest bank is too far away	3%	(0.16)
Primary reason for household not having a bank account because not enough money	23%	(0.42)
Anyone in household has ever used online banking	10%	(0.30)
Anyone in household has borrowed formally	22%	(0.41)
Household-Level Characteristics		
Household is in an urban area	48%	(0.50)
Household owns home	61%	(0.49)
Household rents home	27%	(0.44)
House out walls made of poor material	31%	(0.46)
House floors made of poor material	37%	(0.48)
Toilet not in septic or sewer	23%	(0.42)
House is on electric grid	93%	(0.25)
Number of bedrooms in house	2.31	(1.03)
Household has an air conditioner	10%	(0.31)
Household has a refrigerator	80%	(0.40)
Household has a microwave	44%	(0.50)
Household has a washing machine	80%	(0.40)
Household has a stereo	60%	(0.49)
Household has a DVD player	28%	(0.45)
Household has a TV	80%	(0.40)
Household has a cellphone	94%	(0.23)
Household has a computer	37%	(0.48)
Household has a vehicle	41%	(0.49)
Household has cable	47%	(0.50)
Household has internet	59%	(0.49)
Number of household members	3.62	(2.06)
Number of children in household	1.29	(1.49)
Number of dependents in household	1.68	(1.47)
Number of adults in household	2.34	(1.19)
Total household size (persons)	10.51	(10.88)
Years of education head of household	8.22	(5.30)
Total Households (N)		**2171**

Table 2: Description statistics from the Financial Inclusion module from the April 2019 Labor Force Survey. Responses are weighted according to household weights.

The limiting factor for building complex models of the relationship between household characteristics and Financial Inclusion is typically not the LFS survey but the Census. According to the Small Area Estimation (SAE)[8] methodology, for each variable identified in the LFS survey, there must be a similar variable that measures the same characteristic captured in the Census. The Census measures the head of household's education status, and not more detailed labor and industry information as available in the LFS. We generated 26 household-characteristics variables which we used to predict the five Financial Inclusion metrics. We now turn to the estimation and modeling description.

8 Small Area Estimation, or SAE abbreviated here, refers to the statistical method of estimating statistics for sub-populations of interest. Here we apply SAE to the Labor Force Survey to estimate statistics for the enumeration district sub-populations.

3. Methods

Using Machine Learning to Estimate Enumeration District Financial Inclusion Metrics

If the Financial Inclusion module sampled every household, our task would be easy. However, because of the cost limitations of surveys, it is often not feasible to survey every area in which we would like to learn a particular statistic. The SAE methodology allows us to extend a survey -in this case the Financial Inclusion module - into areas that were never sampled, using clever statistical prediction techniques. These techniques are commonly used across a variety of disciplines, including agriculture (Rao, 2004), biostatistics (Yu et al., 2007), and political science (Buil et al., 2019). Examples in development abound as there are often significant spatial data coverage gaps that make SAE particularly attractive (Serajuddin, 2015).

We follow the standard methodology but deviate in that we use machine learning methods rather than parametric approaches, such as those used in Elbers et al., (2003). We employ machine learning methods as they produce more consistently accurate models relating the household characteristics to the outcome of interest. This remains important even when we do not have a tremendous number of household-level variables to use for prediction. (Afzal et al., 2015).

As is standard with any machine learning modeling exercise, we first split our estimation data into a training and test set. The training set will be used to develop our model, and the test set we will only use for final model evaluation. The logic of this approach is that machine learning methods can be made as flexible and complicated as we wish, and past a certain point of complexity they may learn characteristics that describe inherent noise rather than actual relationships between the predictors and the outcome variable (James et al., 2013). We use a training sample of 75% (N_train = 1,629) of our original LFS data, and reserve 25% (N_test = 542) of our LFS data for testing or evaluating our model.

We estimate five random forest models, one for each of the household level Financial Inclusion metrics we want to calculate at the ED level. Our random forest model is standard. Each employ 500 classification trees, where we cross-validate to select the optimal number of variables sampled at every decision node in building the classification trees.[9]

9 For more detail on random forests, we refer the curious reader to James et al. (2013) or Hersh and Harding (2018).

4. Model Results and Diagnostics

Each model predicting Financial Inclusion is a binary model, thus we employ binary classification diagnostics to this problem. Two important diagnostics are the Receiver Operator Curve (ROC) and the AUC (area under the curve) metric (Parker, 2011). Recall from our household level model we recover a probability that a household uses a particular financial instrument. The receiver operator curve plots these predicted probabilities as a function of their true positive fraction and false positive fraction. The logic here is that a very good model has high true positives with low false positives. The ideal model is one that looks like a capital gamma in Greek (Γ). This model would have all of the graph area "under" the curve. Thus, a perfect model has an AUC of 1, and a model with no predictive power would have an AUC of 0.

The ROC plots are shown in the Appendix in figures A1 A-E, and the AUC metrics (both applied to the test or validation set) are shown below in table 2. We see that the best performing model measures whether online banking is used. The AUCs vary between 0.584 and 0.0775.

AUC (Test Set)	Outcome Variable
0.698	Household does not have a bank account
0.695	Primary reason for household not having a bank account because bank is too far
0.665	Primary reason for household not having a bank account because not enough money
0.775	Anyone in household has ever used online banking
0.584	Anyone in household has borrowed formally

Table 2 AUC metrics, estimated against the test set, for the five machine learning models estimating the given measure of Financial Inclusion.

5. Enumeration District Estimation of Financial Inclusion Metrics

Satisfied with the predictive performance of the machine learning models, we apply the prediction model to the Census observations. This generates, for each household in the Census, an estimate of the five different metrics of Financial Inclusion.

The results are plotted in the maps shown in Figures 2 and A2-A5 in the Appendix. We see significant heterogeneity in the distribution of all five of the maps. First looking at the map in Figure 2 estimating the fraction of households in EDs without bank accounts, we see the majority of the households without banks are estimated to be in the north of Toledo, and to a lesser extent the south of Stann Creek, however these EDs largely comprise sparsely populated farmlands. More notably, EDs in the south of Toledo containing more inhabitants, also demonstrated low bank account ownership. Looking at map A2 which plots the fraction of households that state banks are too far, we see that most households who do not use banks due to this fact are primarily located in the South and West. In particular, rural areas of Toledo show a high number of households who cite banks as being too far. The Cayo district also has a large number of households mentioning that their distance from banks is too great. Primarily we see that residents in rural districts identify geographic distance from banks as a barrier, which makes sense given that the market supply of banks may be focused in cities rather than rural areas.

Turning to map A3 which shows households that are too poor to have bank accounts, we see two significant clusters: one in southern Toledo and the other in the north in Corozal District, though EDs for the latter are widely comprised of Mennonite communities, which due to cultural practices, may not utilize the national financial system in the same ways as the wider population. Online banking usage appears quite restricted to the urban areas such as Belize City, San Ignacio, Orange Walk, and Corozal (map A4). Finally bank borrowing seems significantly concentrated in the northern half of the country, with Belize City, Orange Walk and Corozal over-represented relative to its counterparts in the south (map A5)[10].

10 Tables A1 - A4 in the Appendix list the top 20 EDs with the strongest estimated metrics for Financial Inclusion

1. Factors Which Affect Enumeration District Level Adoption of Financial Inclusion

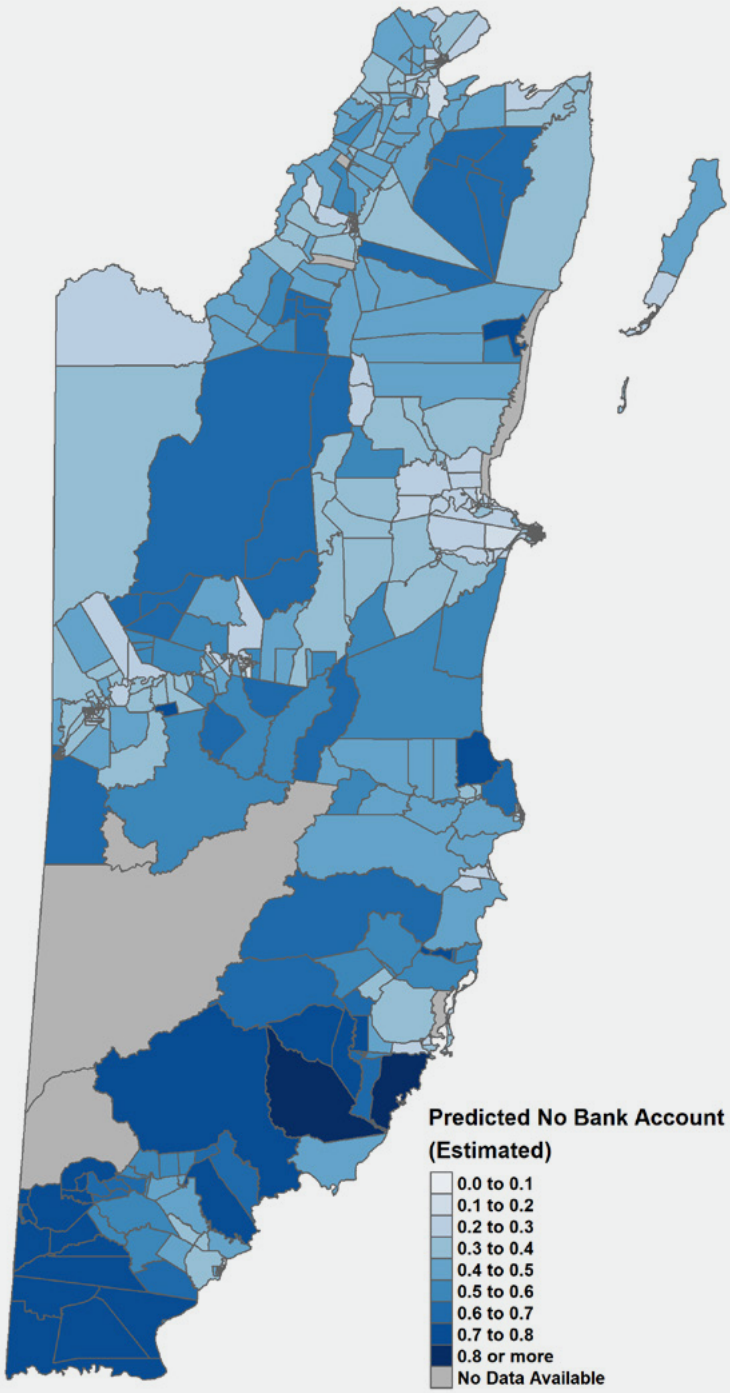
Once we have our ED level estimates of Financial Inclusion, we can then ask what are factors which influence the adoption of Financial Inclusion. We now estimate a regression model, where the dependent variable is the ED level Financial Inclusion metric, and the predictor or independent variables are ED level characteristics, such as average poverty rate of an ED or whether the ED is urban or rural. We focus on only two ED level outcomes: 1) the fraction of households in an ED that have bank accounts, and 2) the fraction of households in an ED that do not have bank accounts due to not having enough money.

Table A6 presents the regression results. We focus on the coefficients in the table, which describe factors that are associated with the degree of Financial Inclusion of the EDs. One of the strongest predictors of Financial Inclusion is the poverty rate of an ED. Increasing the poverty rate by ten percentage points is associated with a 2.45 percentage points decrease in fraction of households that have bank accounts, and a 1.89 percentage points increase in households who state they are too poor to open a bank account. We can further see this relationship between poverty and Financial Inclusion by plotting ED-level household average Financial Inclusion metrics against ED poverty rates. Figure 3 panel A plots each ED by the fraction of household unbanked against the ED poverty rate. We see a strong positive relationship between these two factors. The least poor EDs have a low number of unbanked households on average. We estimate that EDs in the top decile of poverty rates have 58% unbanked households on average, compared with 25% in the lowest decile of poverty. In panel B of Figure 3 we plot the ED poverty rate against the fraction of households that declare they are too poor to open a bank account. EDs that have poverty rates in the lowest decile have 21% of households who claim they are too poor to open a bank account, compared to 40% of households in the highest decile of poverty.

Language and ethnicity appear to be strongly correlated with ED level Financial Inclusion. We see that EDs which contain more indigenous or Carib language speakers - Q'eqchi' (Kekchi) Mayan, Mopan, Garifuna, or Yucatec Mayan - have lower rates of Financial Inclusion. The coefficients suggest that a ten percentage points increase in the rate of speakers of these languages increases the unbanked fraction of households by 4.6 percentage points. EDs with a higher rate of German speakers, who generally reside in Mennonite communities, also demonstrate this pattern with nearly the same strength in magnitude. Though, the circumstances determining the participation of Mennonite communities in the formal financial sector may be heavily influenced by cultural and religious practices, and thus may require enhanced data for improved analysis.

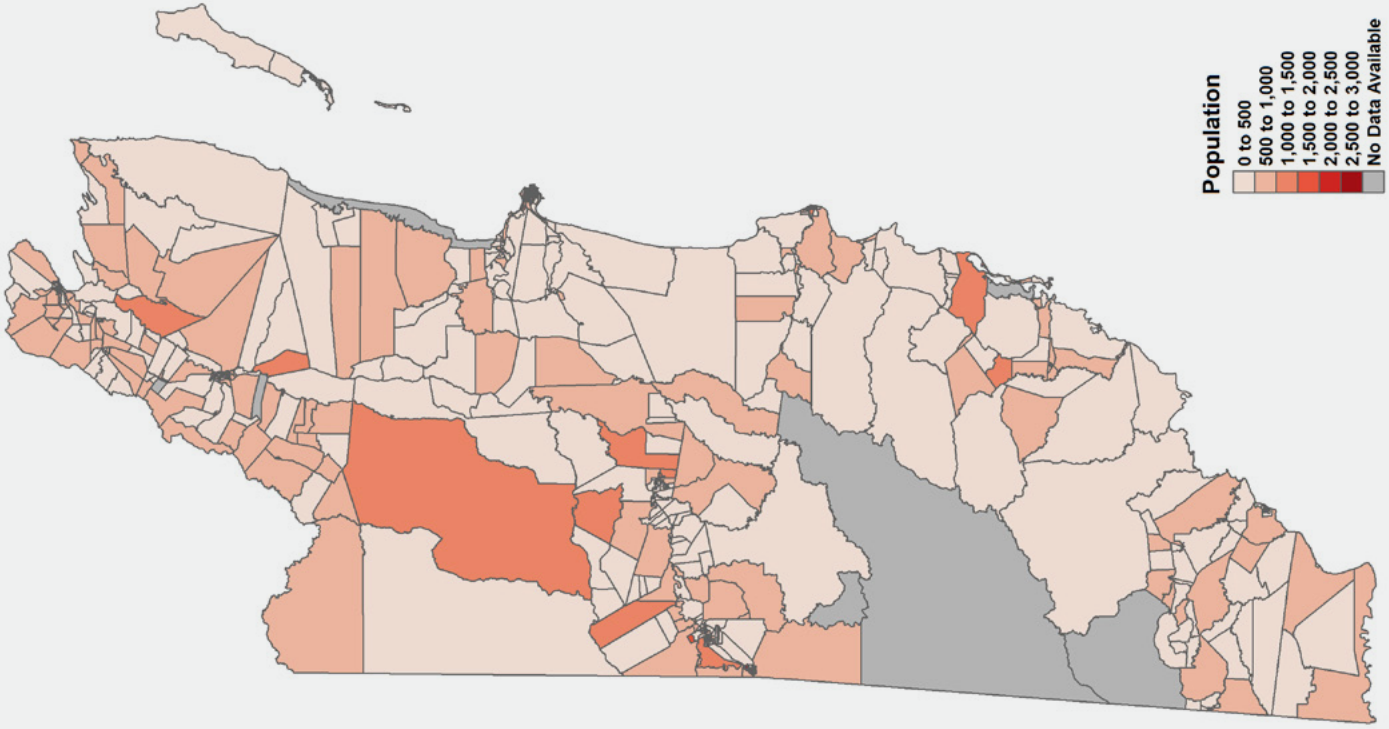
The regression table reveals other ED level characteristics that are associated with the degree of

Financial Inclusion. Urban areas are associated with higher number of banked households. The fraction of an ED that is natively born is also strongly related with bank use and access. Finally, infrastructure is strongly correlated with Financial Inclusion. EDs with a higher number of households with fixed phone access is associated with a higher rate of bank access.



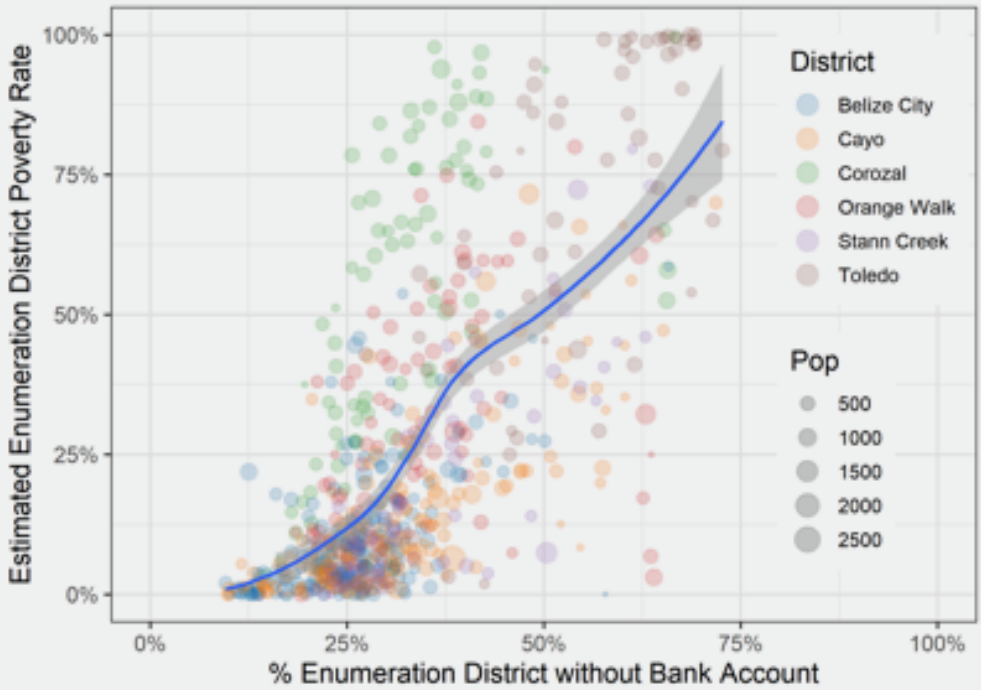
A

Figures 2 A-B Panel A: estimated enumeration district average share of households without bank accounts. Panel B (Next Page): enumeration district population (Census 2010).

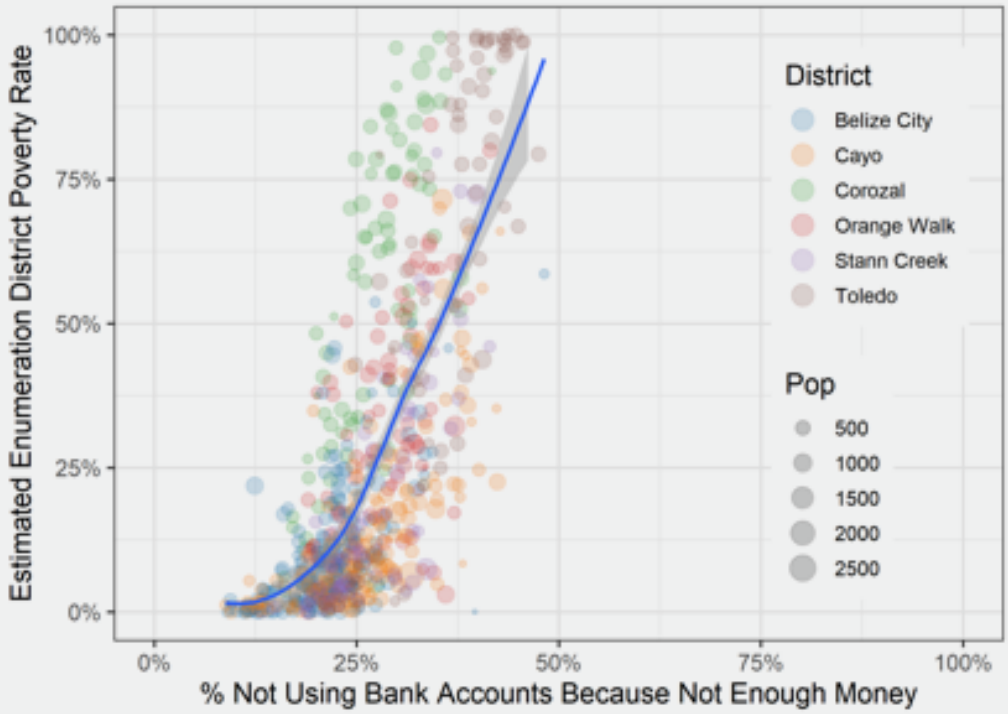

Population
0 to 500
500 to 1,000
1,000 to 1,500
1,500 to 2,000
2,000 to 2,500
2,500 to 3,000
No Data Available

B

A



B



Figures 3 A-B: Panel A: ED Poverty rate against estimated fraction of households ED without Bank account. Panel B: ED Poverty rate against estimated fraction of households who do have a bank account due to not having enough money.

2. Policy Recommendations to Increase Financial Inclusion

As part of the National Financial Inclusion Strategy (NFIS) that began in 2019, the plan includes a series of Thematic Financial Inclusion Task Forces (FITFs) to coordinate and consult on the proposed actions of the NFIS. Four FITFs were established including 1) Enabling ICT and Financial Infrastructures; 2) Tailored Financial Products and Innovation; 3) Financial Consumer Protection and Financial Literacy and 4) Data Collection. We commend the Central Bank of Belize and the government for establishing these tasks forces. Nevertheless, given the spatial and demographic clustering of Financial Inclusion, we make the following recommendations to the specific teams.

We recommend that ICT and Financial Infrastructures team consider increasing financial access points in EDs where there is low banking usage, paying particular focus on the inhabited rural areas of Toledo and Cayo. In the South, there appears to be several EDs with many unbanked persons. Not all individuals in these regions cite lack of available funds as the reason for not having a bank account, indicating that accessibility may pose impediments to Financial Inclusion. Roughly 15% of the households in rural districts of Cayo state banks are too far as an impediment to usage, indicating subsidies for providing banking services may increase bank access for these households.

To the Tailored Financial Products and Innovation team, we recommend considering financial products that are subsidized or low cost and tailored to individuals who currently do not have bank accounts. Poverty is the strongest predictor of being unbanked, and individuals often cite not having available funds as a reason for not having a bank account. It also may be beneficial to consider combining these with specific governmental transfer programs.

To the Financial Consumer Production and Financial Literacy teams, we recommend creating introductory banking finance literature in Q'eqchi' (Kekchi) Mayan, Mopan, Yucatec Mayan, and German. High use of these languages in EDs is predictive of low levels of Financial Inclusion. Thus, creating targeted literature in these languages may encourage participation and foster trust.

Finally, to the Data Collection, Analysis and Reporting team we commend their reporting disaggregation by gender, age, rural/urbaneness, and district. We encourage this team to complete an exercise similar to this study at the end of 2022 to capture the effects of the programs at ED Financial Inclusion metrics.


$5
Belize
$2
Bank of Belize
Bank of Belize

6. Conclusion

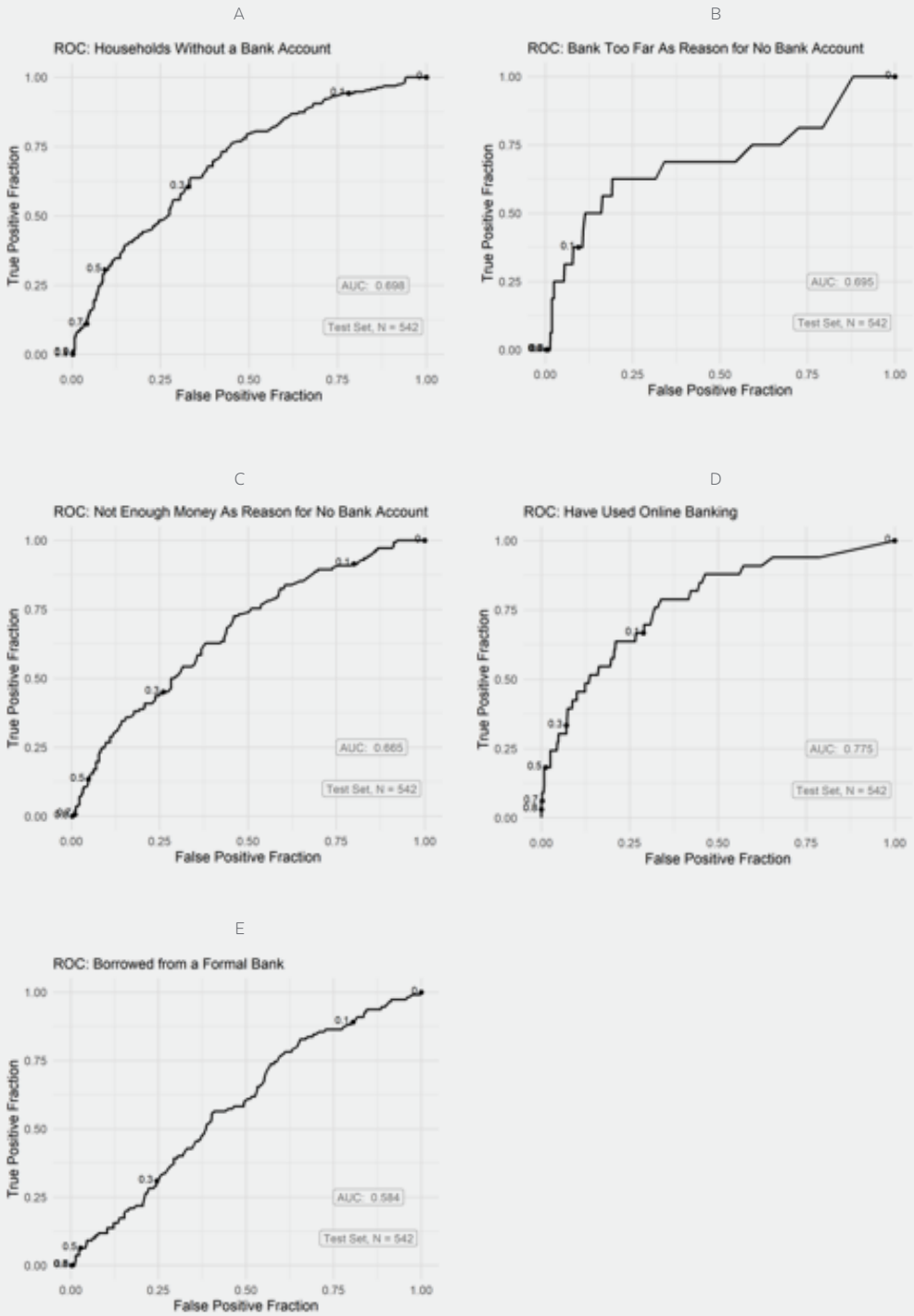
Financial Inclusion continues to be an important tool for developing countries. From this analysis we have found that the simple district averages for lack of access to various financial instruments obscures a richer picture of Financial Inclusion at the ED or Sub-District level.

We find significant clustering of the five Financial Inclusion metrics analyzed in this paper, with the pattern being distinct for each type of financial product. For instance, households without bank accounts are concentrated geographically in rural Toledo, households citing geographic distance as a barrier are clustered in the South and West of Belize, with households that are too poor to have bank accounts also located in in the southern Toledo and in Corozal.

These results suggest that the National Financial Inclusion Strategy would benefit from including geographically targeted policies in its National Financial Inclusion Strategy. Without focusing on the geographic disparities in Financial Inclusion, the gains in banking access would likely accrue in the areas that are already ahead in terms of financial access.

Examining the results more closely, we find the poverty rate of an Enumeration District is the strongest predictor of the fraction of unbanked households, with poorer district being associated with many fewer unbanked households. This suggests gains to be made through the adoption of specific savings vehicles intended to target the poor and unbanked. Finally, we see that language and ethnicity are associated with lower levels of Financial Inclusion, suggesting specific outreach is required for these underrepresented groups. Only with concerted efforts will the gains in financial access be spread broadly throughout the country.

7. Appendix



Figures A1 A-E: Receiver Operator Curve (ROC) plots for the respective machine learning models. ROC plot and AUC reflect calculations against the validation/test set which was not used to estimate the model, giving an approximate measure of out-of-sample fit.

Maps

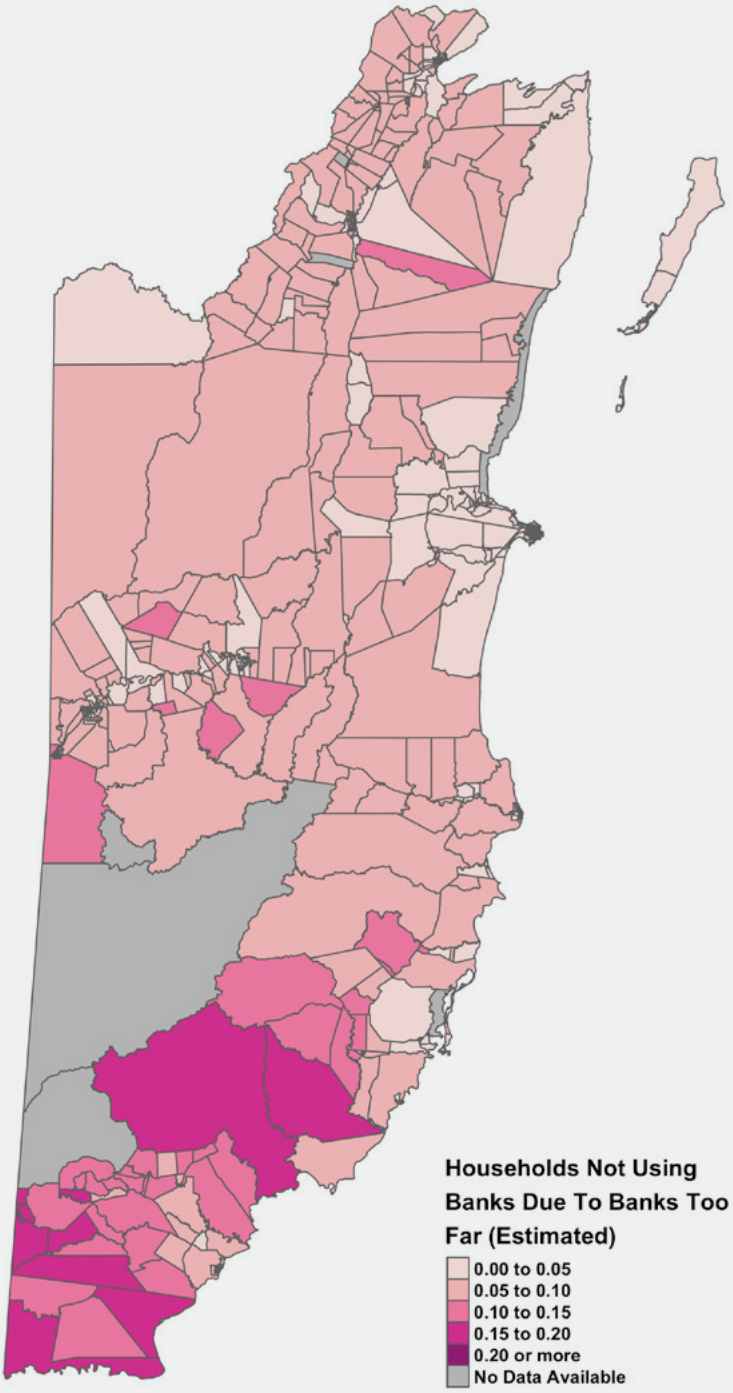


Figure A2 Estimated Enumeration District Average Share of Households Not Using Bank Accounts Due to Bank Being Too Far

Maps

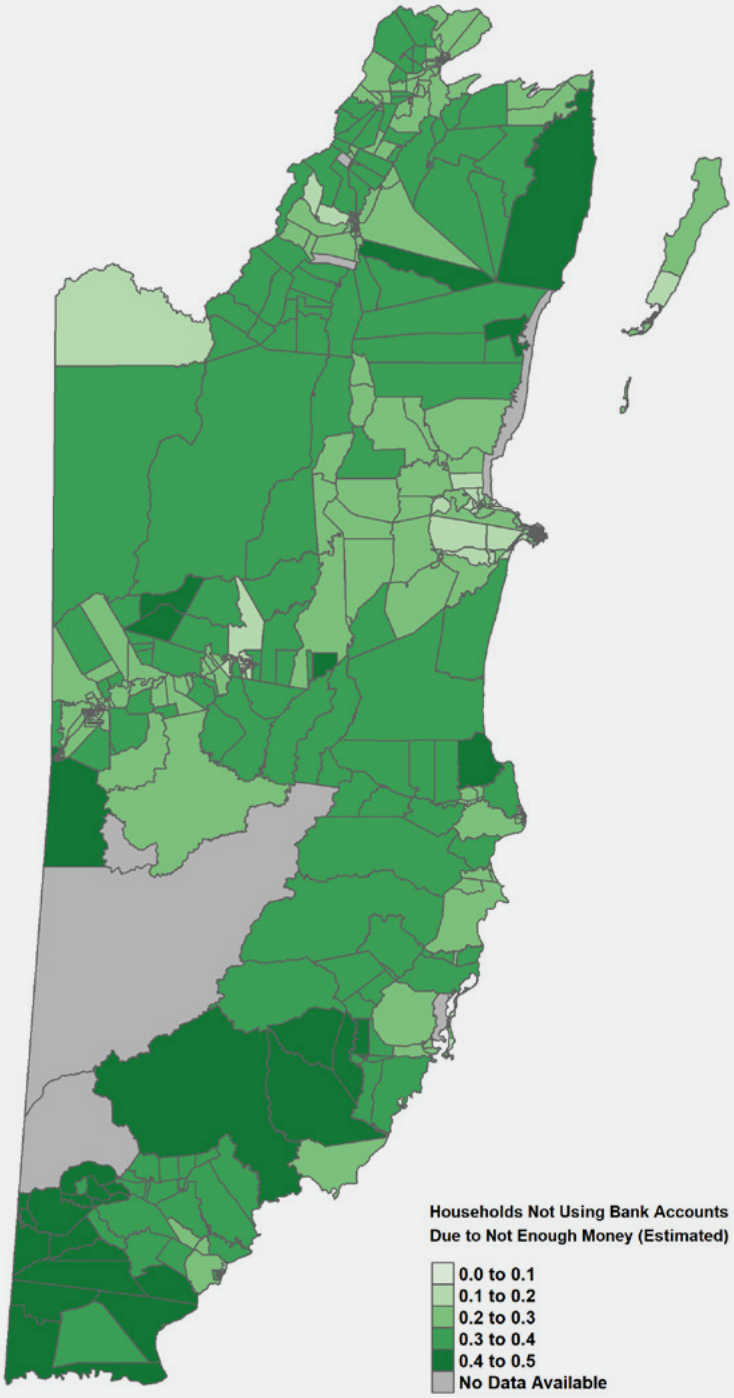


Figure A3 Estimated Enumeration District Average Share of Households Not Using Bank Accounts Due to Not Having Enough Money

Maps

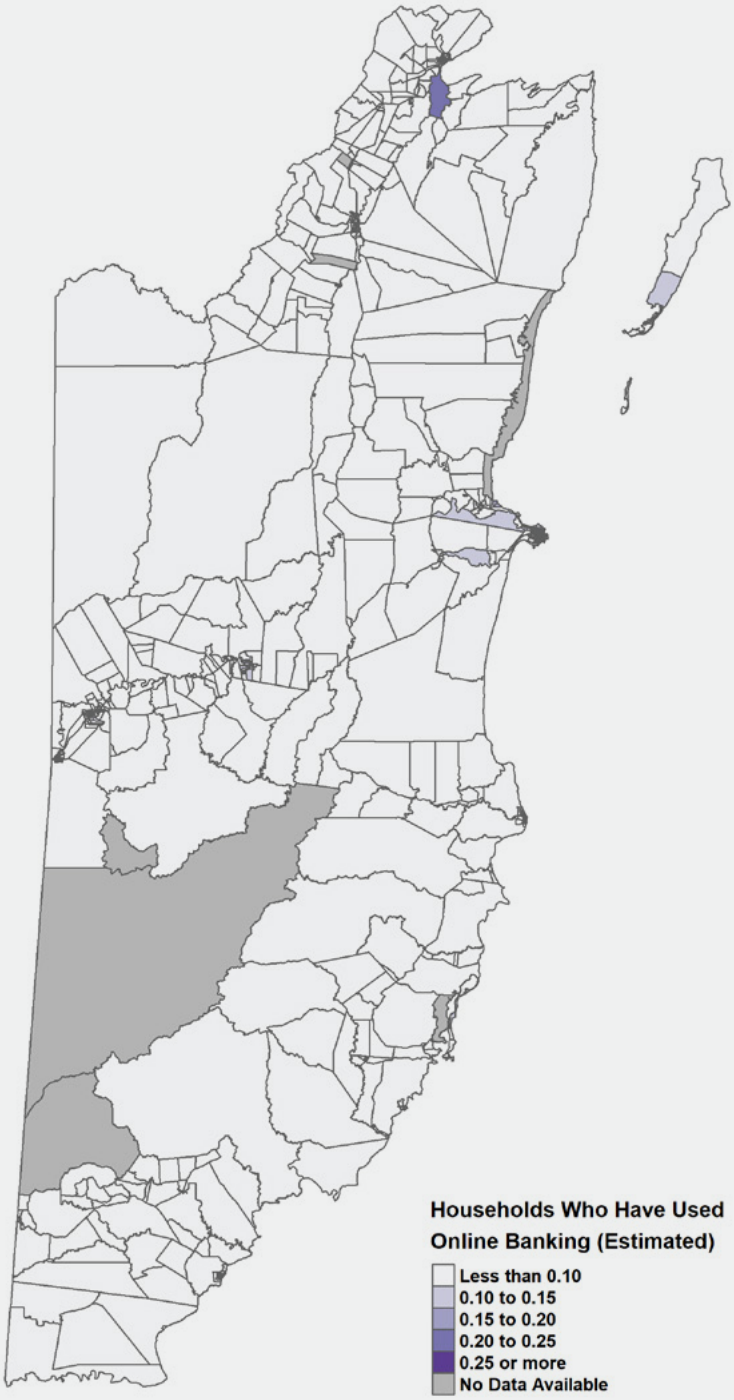


Figure A4 Estimated Enumeration District Average Share of Households Who Have Used Online Banking in Previous 12 Months

Maps

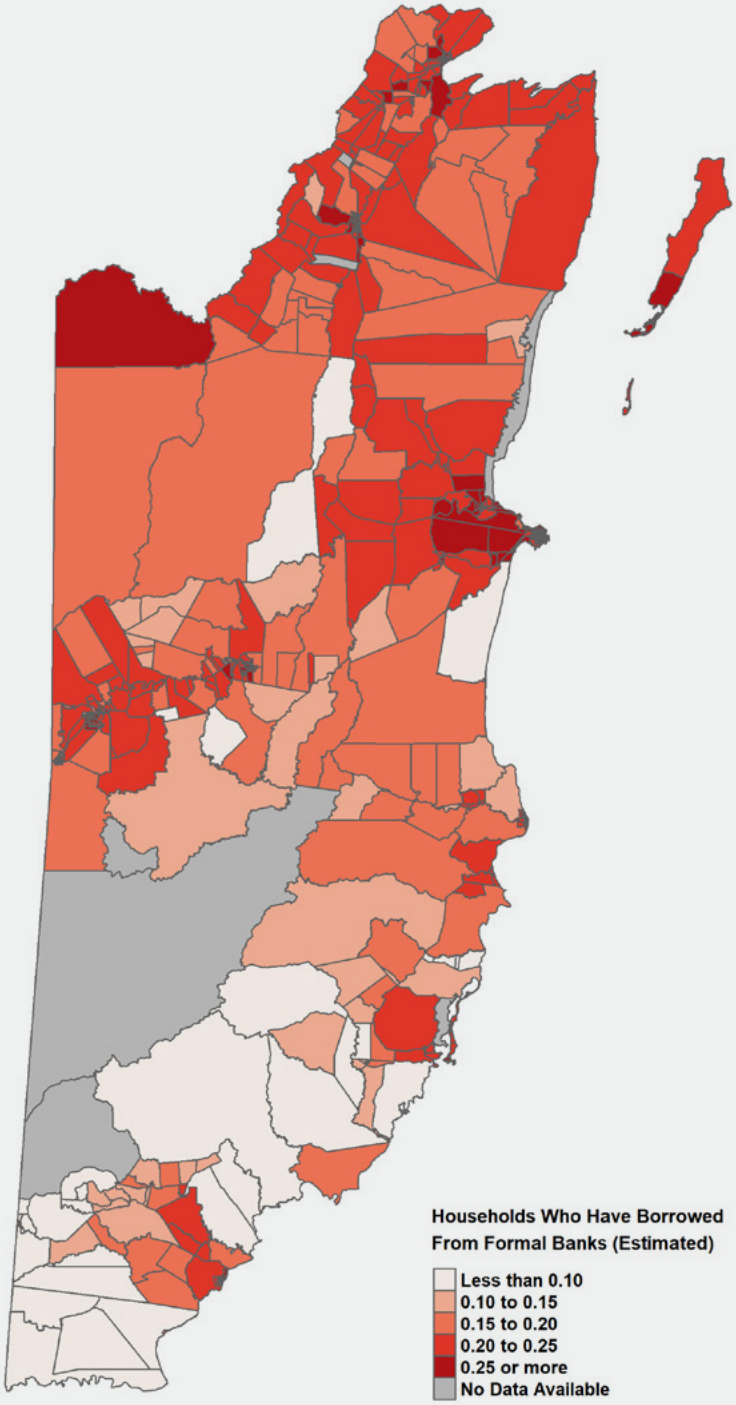


Figure A5 Estimated Enumeration District Average Share of Households Who Have Borrowed Formally from Banks

TABLES

District	Locality	ED	Estimated % of Households Without Bank Accounts
6	12	215	0.727
4	12	12816	0.718
6	12	2202	0.715
6	12	3207	0.692
6	12	2211	0.692
6	12	1706	0.69
6	12	4202	0.688
6	12	1203	0.687
6	12	1214	0.685
6	12	1207	0.68
5	12	3710	0.677
6	12	2706	0.675
6	12	5211	0.668
6	12	4211	0.668
1	12	3209	0.666
3	12	102	0.658
6	12	2214	0.658
6	12	214	0.658
1	12	4209	0.657
1	12	2209	0.656

Table A1 Top 20 EDs with Largest Estimated Share of Households Without Bank Accounts

District	Locality	ED	Estimated % of Households Who Cite Bank Too Far as Reason for Not Using Banks
6	12	3207	0.197
6	12	216	0.186
6	12	215	0.173
6	12	1207	0.167
6	12	2202	0.166
6	12	4211	0.162
6	12	2214	0.161
6	12	211	0.159
6	12	5211	0.154
6	12	1706	0.154
6	12	2207	0.152
6	12	6211	0.146
6	12	2216	0.145
4	12	816	0.142
6	12	1214	0.137
6	12	214	0.135
6	12	3202	0.134
4	12	12816	0.13
6	12	1211	0.129
6	12	2706	0.127

Table A2 Top 20 EDs with Largest Estimated Share of Households Who Cite Bank Too Far as Primary Reason for Not Having a Bank Account

District	Locality	ED	Estimated % of Households Who Cite Not Having Enough Money as Reason for Not Using Banks
3	12	102	0.482
6	12	215	0.475
6	12	211	0.457
6	12	3207	0.454
6	12	2202	0.45
6	12	1214	0.447
6	12	2211	0.439
6	12	4211	0.436
6	12	1207	0.434
6	12	1706	0.433
6	12	4202	0.433
6	12	2214	0.432
6	12	5211	0.43
5	12	3710	0.429
4	7	206	0.427
4	12	308	0.424
6	12	256	0.423
4	12	7816	0.423
1	12	213	0.418
6	12	216	0.417

Table A3 Top 20 EDs with Largest Estimated Share of Households Who Cite Not Having Enough Money as Primary Reason for Not Having a Bank Account

District	Locality	ED	Estimated % of Households Using Online Banking
2	12	212	0
4	8	3111	0
2	12	4201	0.002
3	12	117	0.002
4	12	9816	0.004
3	12	102	0.005
4	12	10816	0.005
1	12	213	0.009
4	12	12816	0.009
2	12	2213	0.009
2	12	4213	0.009
6	12	215	0.01
6	12	4202	0.01
6	12	3216	0.01
6	12	5211	0.011
2	12	1213	0.011
1	12	4209	0.011
6	12	1706	0.011
6	12	2211	0.012
6	12	207	0.012

Table A4 20 EDs with Lowest Estimated Online Banking Penetration

District	Locality	ED	Estimated % of Households Using Formal Borrowing
3	12	117	0.038
4	12	12816	0.049
5	12	1207	0.06
6	12	1203	0.061
6	12	214	0.069
6	12	1207	0.069
5	12	2207	0.07
6	12	215	0.071
6	12	2207	0.071
6	12	1216	0.077
6	12	2211	0.077
5	12	207	0.078
4	8	5108	0.08
6	12	2216	0.08
6	12	2202	0.08
6	12	1706	0.081
6	12	5211	0.082
6	12	202	0.082
6	12	4211	0.082
4	7	206	0.083

Table A5 20 EDs with Lowest Estimated Usage of Formal Borrowing

Dependent Variable:	% of households in Enumeration District that have bank accounts	% of households in Enumeration District who do not have a bank account due to not having enough money
Urban = 1	0.027*** p = 0.003	-0.037*** p = 0.00002
Religion = Pentecostal	-0.041 p = 0.150	0.047* p = 0.079
Religion = Roman Catholic	0.099*** p = 0.0003	-0.094*** p = 0.0003
Religion = Seventh Day Adventist	0.102*** p = 0.010	-0.064* p=0.082
Fixed Phone = 1	0.291*** p = 0.000	-0.269*** p = 0.000
Log of ED Population	0.04 p = 0.443	0.218*** p = 0.00001
Language = English	0.165*** p = 0.000	-0.163*** p = 0.000
Language = Spanish	0.007 p = 0.883	-0.007 p = 0.868
Language = Garifuna/ Yucatec/Kekchi/Mopan	-0.461*** p = 0.000	0.371*** p = 0.000
Language = German	-0.459*** p = 0.000	0.175*** p = 0.00003
Ethnicity = Creole	-0.243*** p = 0.000	0.192*** p = 0.00000
Ethnicity = Kekchi Maya	0.063* p = 0.065	-0.047 p = 0.142
Ethnicity = Mestizo	-0.181*** p = 0.0001	0.211*** p = 0.00000
Born in Belize	0.284*** p = 0.000	-0.121*** p = 0.0001
Poverty Rate (20%)	-0.245*** p = 0.000	0.189*** p = 0.000
Constant	0.415*** p = 0.000	0.355*** p = 0.000
Observations	684	684
R2	0.813	0.805
Adjusted R2	0.809	0.8
Residual Std. Error (df = 668)	0.092	0.085
F Statistic (df = 15; 668)	193.518***	183.501***
Note: *p<0.1; **p<0.05; **		

Table A6 Enumeration District Level Characteristics Associated with Estimated Enumeration District Level Financial Inclusion Measures

APRIL 2019 LABOR FORCE SURVEY FINANCIAL INCLUSION MODEL SELECTED QUESTIONS

FI module question 1

An account can be used to save money, to make or receive payments, or to receive wages or financial help. Do you, either by yourself or together with someone else, currently have an account at a bank (for example, Atlantic or Belize Bank or any other bank), or a credit union (for example Holy Redeemer or any other credit union)?

- Yes, No, DK/NS, Refused

FI module question 3-A

Please tell me whether each of the following is A REASON why you, personally, DO NOT have an account at a bank or a credit union. Is it _____ ?

a) Because financial institutions are too far away?

- Yes, No, DK/NS, Refused

FI module question 3-F

Please tell me whether each of the following is A REASON why you, personally, DO NOT have an account at a bank or a credit union. Is it _____ ?

f) Because you don't have enough money to use financial institutions

- Yes, No, DK/NS, Refused

FI module question 4-E

Financial transactions include for example making or receiving payments, sending money to others, and depositing or taking out money from an account. In the PAST 12 MONTHS, have you ever made a transaction using any of the following, whether with your own account or another person's account?

e) Internet / Online banking website

- Yes, No, DK/NS, Refused

FI module question 7

In the PAST 12 MONTHS, have you, by yourself or together with someone else, borrowed any money from any of the following sources?

From a bank, credit union or another type of formal financial institution

- Yes, No, DK/NS, Refused

References

Breiman, L. (2001). Random forests. Machine learning, 45(1), 5-32.

Buil-Gil, D., Moretti, A., Shlomo, N., & Medina, J. (2019). Worry about crime in Europe: A model-based small area estimation from the European Social Survey. European Journal of Criminology, 1477370819845752.

Cámara, N., & Tuesta, D. (2014). Measuring financial inclusion: A muldimensional index. BBVA Research Paper, (14/26).

Central Bank of Belize. (2019). National Financial Inclusion Strategy 2019-2022. Retrieved from https://www.centralbank.org.bz/publications-research/nfis

Elbers, C., Lanjouw, J. O., & Lanjouw, P. (2003). Micro-level estimation of poverty and inequality. Econometrica, 71(1), 355-364.

Demirguc-Kunt, A., & Klapper, L. (2012). Measuring financial inclusion: The global findex database. The World Bank.

Goldstein, A., Kapelner, A., Bleich, J., & Pitkin, E. (2015). Peeking inside the black box: Visualizing statistical learning with plots of individual conditional expectation. Journal of Computational and Graphical Statistics, 24(1), 44-65.

Hersh, J., Engstrom, R., & Mann, M. (2020). Open data for algorithms: mapping poverty in Belize using open satellite derived features and machine learning. Information Technology for Development, 1-30.

Hersh, J., & Harding, M. (2018). Big Data in economics. IZA World of Labor, (451).

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An introduction to statistical learning (Vol. 112, p. 18). New York: springer.

Kochar, A. (2018). Branchless banking: Evaluating the doorstep delivery of financial services in rural India. Journal of Development Economics, 135, 160-175.

Nicodemus, K. K., Malley, J. D., Strobl, C., & Ziegler, A. (2010). The behaviour of random forest permutation-based variable importance measures under predictor correlation. BMC bioinformatics, 11(1), 1-13.

Park, D. K., Gelman, A., & Bafumi, J. (2004). Bayesian multilevel estimation with poststratification: State-level estimates from national polls. Political Analysis, 375-385.

Parker, C. (2011, December). An analysis of performance measures for binary classifiers. In 2011 IEEE 11th International Conference on Data Mining (pp. 517-526). IEEE.

Rao, J. N. K. (2004). Small area estimation with applications to agriculture. Agricultural survey methods, 9, 139-147.

Sethi, D., & Acharya, D. (2018). Financial inclusion and economic growth linkage: Some cross country evidence. Journal of Financial Economic Policy.

Serajuddin, U., Uematsu, H., Wieser, C., Yoshida, N., & Dabalen, A. (2015). Data deprivation: another deprivation to end. The World Bank.

Yu, H., Meng, Y. Y., Mendez-Luck, C. A., Jhawar, M., & Wallace, S. P. (2007). Small-area estimation of health insurance coverage for California legislative districts. American journal of public health, 97(4), 731-737.

Where are the unbanked in **Belize?**

Using Machine Learning Small Area Estimation to improve Financial Inclusion geographic targeting

IDB
Inter-American Development Bank