Final report RWA-24217 October 2025

Harnessing mobile money data to modernise economic measurement in Rwanda

Emily Aiken
Didace Irafasha
Christian Lippitsch
José Vicente Rodríguez Mora
Dominique Ntirushwa

IGC

DIRECTED BY LSE UNIVERSITY OF OXFORD

FUNDED BY UKaid from the British people

Executive Summary

How can governments in low-income countries generate faster, more granular, and more accurate economic statistics? Modern payment systems, such as mobile money (MoMo) and bank transfers, naturally generate large volumes of transaction data that offer an unprecedented opportunity to measure economic activity in detail and can generate faster, more granular measures of economic activity in a low-income setting. In a collaboration with the Central Bank of Rwanda (BNR) and the National Statistics Agency (NISR) we have started exploring the use of transaction-level data to build high-frequency indicators of income and consumption that complement survey-based national accounts. In this report, we describe the data environment in Rwanda and initial explorative approaches to categorize MoMo users and financial transactions for the goal of generating of a full panel of economic activity and towards to following core objectives of our project:

- **Structure and merge transaction data** from all mobile money providers and banks to create a vast and representative consumption and income dataset.
- **Develop new economic indicators** by comparing this transaction-based data with traditional survey-based data to improve the accuracy of national accounts and measure evolving economic trends.
- **Collaborate with policymakers and researchers** to answer key questions about income inequality, consumption patterns, and policy responses to economic shocks.
- **Harness mobile payment data to capture the informal sector**, which is often missed by traditional data collection methods.

We secured the full universe of MoMo transactions for one month and started constructing an anonymized monthly cross-section linking transaction between users. A typical month has several million of transactions and millions of unique users. In a first step, we began categorizing users according to their role as consumers or entrepreneurs. We use an algorithmic approach to categorize MoMo subscribers, that are not identified as *formal* merchants, as mere subscribers or *informal* merchants by comparing transaction patterns of formal merchants with the remaining subscribers. This allowed us to assign a probability of being an informal merchant to any given subscriber. In the future, we will use additional data sources, including official consumption as well as our own surveys, to verify our results and map estimates probabilities to actual binary categorizations.

Based on this, we will in a next step categorize financial transaction as consumption or business-related transactions. We further describe remaining challenges and suggestive approaches to address these. In particular, we describe approaches to account for remaining use of cash transaction and approaches to account for non-representativeness of MoMo users to eventually obtain representative aggregates of economic activity in Rwanda.

1. Introduction

How can governments in low-income countries generate faster, more granular, and more accurate economic statistics? Modern payment systems, such as mobile money (MoMo) and bank transfers, naturally generate large volumes of transaction data that offer an unprecedented opportunity to measure economic activity in detail. Each time a payment is made, a record is created that identifies who paid, who received the payment, and how much money was transferred. If this information is not directly available, one can further infer the purpose of the transaction by incorporating additional information about the parties involved. Over time, this data enables the tracing of consumption and income patterns, effectively generating a massive, ongoing consumption survey with vast granularity across socioeconomic, geographic, and temporal dimensions.

Despite the promise of this data, two major challenges remain: (1) transaction data is often limited to a single financial institution, making it non-representative of the broader economy, and (2) most existing studies focus on OECD countries, while low-income and developing countries stand to benefit the most from the insights this data can provide given often worse data availability. Many low-income countries lack comprehensive and timely economic measurements, with a significant number not producing quarterly national accounts or household budget surveys.

Our project seeks to overcome these limitations by leveraging the entire universe of electronically mediated transactions in Rwanda. Over the past year, in a collaboration between the International Growth Centre (IGC), the National Bank of Rwanda (BNR), and the National Institute of Statistics of Rwanda (NISR) we have begun exploring whether anonymized MoMo and (eventually) bank transaction data could be used to construct high-frequency measures of income and consumption. Thanks to this collaboration, we secured access to the full universe of MoMo transactions and *monthly* bank statements across Rwanda across several months. Note, that this is an ongoing project and all discussed results present intermediary steps towards the building of this novel data set.

The novelty of this project lies in its application to a non-OECD country where electronic payment coverage is exceedingly high, even compared to many developed economies. We have been exploring whether Rwanda's robust mobile payment infrastructure allows us to generate real-time, highly detailed economic measurements that can be scaled to other developing countries. By enhancing traditional economic measurements and creating innovative statistical frameworks, our project aims to address global development challenges regarding robust data availability while providing a model for similar studies in other IGC countries.

The core objectives of our project are:

- **Structure and merge transaction data** from all mobile money providers and banks to create a vast and representative consumption and income dataset.
- **Develop new economic indicators** by comparing this transaction-based data with traditional survey-based data to improve the accuracy of national accounts and measure evolving economic trends.

- **Collaborate with policymakers and researchers** to answer key questions about income inequality, consumption patterns, and policy responses to economic shocks. By working closely with Rwandan institutions, we will ensure that the dataset will serve as an important input to policy makers and can inform economic planning.
- **Harness mobile payment data to capture the informal sector**, which is often missed by traditional data collection methods. With 86% of the population using mobile money, Rwanda is uniquely positioned to provide a comprehensive view of its economy, including informal transactions.

The project builds on a methodology first developed using banking data in Spain, where researchers successfully produced near real-time national accounts of private consumption (Buda et al, 2023). In Rwanda, we have started testing whether a similar approach could be adapted to a very different context. Working closely with BNR and NISR, we reviewed data architecture, and we are currently developing classification systems of different types of financial *transactors*. Next steps will include the categorization of different transaction *types* (including consumption) and apply different methods to adjust for the role of cash which remains central to many households' economic lives.

In this report, we describe i) potential use cases of the above-described panel data set, ii) current challenges, and iii) the data and our empirical approach including initial findings. We conclude with an outlook on next steps.

2. Potential use cases

Our transaction-based database has the potential to function as a real-time "microscope" on Rwanda's economy, unlocking use cases that extend far beyond conventional surveys. First, it can complement **national accounts** adding granularity, high frequency and potentially a larger degree of reliability compared to national accounts based on consumption which rely on surveys. These surveys, however, have been found to be biased along several dimensions, understating certain consumption categories across the income distribution and for specific income groups (Buda et al., 2023). Daily aggregates of private consumption (overall and by category), labour income, taxes and transfers can be computed directly from MoMo flows, then re-weighted for cash prevalence to match official totals. This transforms static quarterly estimates into dynamic indicators that policymakers can monitor, and respond to, within days of fiscal or monetary interventions, natural disasters or market shocks.

Second, by linking anonymized mobile-money and bank panels with census covariates, we can construct granular **input–output matrices**, tracing supplier–buyer relationships across formal and informal sectors. This makes it possible to map, for instance, how a subsidy to smallholder coffee farmers ripples through transporters, agro-dealers, processors and consumer markets. In economic terms, the data will allow us to measure Keynesian multipliers.

Beyond macro-aggregation, our data enable **micro-analyses** of, for example, taxation, social protection and firm behaviour. Unlike surveys that rely on self-reports, transaction data reveal actual spending and receipt patterns; we can therefore evaluate how a tax or subsidy alters consumption baskets and saving rates across household types, or detect discrepancies between reported and observed revenues for VAT-registered merchants to improve audit targeting. Embedding randomized policy experiments becomes far more efficient as treatment and control groups can be monitored passively through transaction panels, allowing direct measurement of both first-round impacts on recipients and broader spillovers to their trading partners.

In the domain of **urban studies**, transaction-anchored geolocation data yield powerful insights into mobility and settlement dynamics. Every MoMo payment carries a cell-tower stamp, enabling us to identify daily commuting flows, market-day migrations and the emergence of new urban corridors complementing smartphone "ping" studies by adding volumes of economic activity. This continuous information of economic activity can guide infrastructure planning, crisis-response logistics and service-delivery optimization in rapidly evolving urban areas.

For **social protection**, transaction panels provide a precise lens on vulnerability and inclusion. By observing how cash transfers change spending on essential goods, peer-to-peer support networks and local business turnover, we gain a holistic view of program efficacy and community spillovers. In tourism, similar methods can quantify the direct and indirect gains from new resorts or cultural events as we observe not just ticket or accommodation payments but downstream expenditures by visitors in transport, food markets and artisan outlets across the country.

These use cases build on and potentially significantly advance the existing literature on digital metadata in development economics, as it adds the critical element of economic transactions patterns to mere information on existing mobile phone data which often study movement and phone call patterns (recently discussed at VoxDev (Hanney, 2025)). Studies such as Aiken et al. (2022) use mobile-phone metadata to study the targeting of COVID-19 transfers in Togo and combine this data with information on consumption from a survey. Similar to the above-discussed bias in consumption surveys, our approach overcomes this bias by observing actual transaction flows. Work by Dube, Blumenstock and Callen (2025) exploited call-volume "dips" to proxy religiosity in Afghanistan yet could not access the material consequences of beliefs. Transaction logs would allow us to track behavioral shifts in spending around religious observances or climate shocks. And while Gollin et al. (2021) used high-frequency mobility mapping with app pings in East Africa, our geotagged payment streams offer a more representative and economically relevant measure of movement. In short, by combining the full-universe of transaction data with network analytics and machine learning, our project delivers a rich, economy-wide sensor that both complements and transcends prior phone-metadata research.

3. Main challenges

During our exploratory phase we have mapped Rwanda's mobile-money landscape and, in partnership with BNR and NISR, identified several methodological and data-related challenges. In the next phase, we will collaborate closely with government agencies and integrate complementary survey, census, and administrative datasets to address these issues towards the above-stated goals of the project. Below we elaborate on these challenges and describe ways to address them.

Self-employed and the informal economy

Our administrative records of MoMo transactions include transactions by all actors on the mobile money network, including subscribers (MoMo users who transact using standard wallets registered via their cell phone numbers, incurring transaction fees for peer-to-peer payments), businesses/merchants (MoMo users associated who have purchased MoMo business accounts, called "MoMoPay" or "MoMo Codes", which assign transaction fees to businesses rather than clients), and agents (who help MoMo users cash in and cash out money). A challenge in the Rwandan context is that a limited number of businesses have purchased MoMoPay accounts, and a large number of "informal merchants" conduct business transactions using standard subscriber accounts. We therefore cannot assume that all standard subscriber accounts are used only for personal consumption.

Our first technical challenge, therefore, has been to try to identify which MoMo subscribers use their accounts for business transactions. We leverage our transaction data to attempt to estimate, or predict, which subscribers are personal accounts and which are acting as informal merchants. For example, wallets with many small incoming payments from diverse senders and concentrated outflows to supplier-type accounts could exhibit the behavioral signature of micro-enterprises rather than households. We have used the rich list of network statistics (e.g. transaction counts, value dispersion and supplier concentration) and fed them into a machine-learning classifier trained on *formal*, identified merchants to predict the likelihood of a mere subscriber to be an *informal* merchant based on her transaction characteristics. Iterating this process allowed us to refined both role assignment (firm, self-employed or consumer) and sector labels. As even self-employed users often use their MoMo account for business and private purpose, we aim to conduct user surveys to ultimately partition each wallet's activity into business versus personal expenditures and produce the first dynamic measure of Rwanda's informal economy. We will describe the approach to identify likely self-employed in more detail below.

Representativeness and cash adjustment

Because a sizable share of income and spending in Rwanda occurs in cash, relying solely on digital transactions would understate economic activity. Also, the use of cash will not be random across a list of socioeconomic characteristics but correlated with them (e.g. rural vs. urban, older vs. younger consumers) which would not just understate our measures of consumption but lead to systematic bias.

To correct for this, we aim to incorporate several secondary data sources including consumption and financial access surveys to learn more about the use of cash and digital payments. Eventually, we aim to stratify the population along various dimensions including district, settlement type, gender and education, and estimate cell-level cash-to-digital ratios. These ratios could then be used to reweight digital aggregates so that each demographic-geographic cell aligns with survey benchmarks.

Bank transaction-level data

We are currently discussing the access to bank data at the individual transaction level with our government partners. Omitting bank transaction-level data leaves a significant gap in our consumption-income panels and obscures how economic shocks propagate through Rwanda's formal sector. Relying on currently available monthly bank aggregates hides the rich detail of individual salary payments, merchant receipts, interbank transfers and automatic debits that drive large-value flows. Without counterparty identifiers and timestamps, we cannot reliably distinguish consumption outlays from wage credits or peer transfers, nor can we link these events to MoMo activity to build a truly unified, high-frequency household and firm panel. As a result, we observe only the immediate, "first-round" effects of a policy subsidy or weather event on mobile-money users but remain blind to the crucial second- and third-degree multiplier effects mediated by formal-sector salaries, supplier payments and government disbursements. Incorporating bank transaction-level records would expand our framework enabling us to classify every payment, merge digital and bank data seamlessly, and trace the full chain of economic reactions in near real time.

In addition, our preliminary analysis shows that a sizeable share of MoMo transactions involves bank accounts in one or both directions. Yet we currently lack any visibility into the identity or purpose of the transaction identity on the bank's side for a given MoMo transaction. This ambiguity means that many entries could represent non-consumption transfers, salary disbursements, self-transfers or genuine consumption purchases, and we cannot untangle these without bank details. While MoMo logs alone allow for sophisticated mapping of transaction patterns, they inevitably yield biased or incomplete insights when large-value, bank-mediated flows enter and exit the system. Access to granular bank data would fill this, providing the missing context needed to ensure our estimates of consumption, income and shock multipliers are both accurate and truly reflective of Rwanda's entire economic landscape.

4. Data structure

The dataset made available by BNR comprises the complete universe of mobile-money transactions currently available for a single month—over 258 million individual transfers—captured at the transaction level and complemented with anonymized sender and recipient metadata. We have identified 5.06 million unique senders and 5.23 million unique recipients, along with attributes such as agent type and customer profile, but currently lacking transfers' geo-location. Preliminary cross-tabulations that show transaction density between transaction profiles and types reveal, for example, that person-to-person "P2P" transfers account for roughly 25 % of transaction count but just under 20 % of volume, while merchant inflows exhibit high frequency but lower ticket sizes (see figures 1 and 2).

The outward-flow heatmaps show that "P2P" transfers by subscribers comprise about 31 % of transaction counts but only 14 % of transaction volume, reflecting many small peer-to-peer payments. In contrast, deposits by services providers account for 15% of transaction volume but below 1% of transaction accounts indicating few but large transactions. Similar patterns hold for inward transactions.

Figure 3 shows a heatmap of transaction patterns between sender and receiver types. It confirms the pattern of significant discrepancies between the number and volume of transaction. 41% of all transaction occur between subscribers, yet this only amounts to 16% of volume.

Note, that several profile elements require further scrutiny to understand their composition. This holds naturally for the *Blank* profile. We are working together with BNR and the telecommunication providers to clarify open questions.

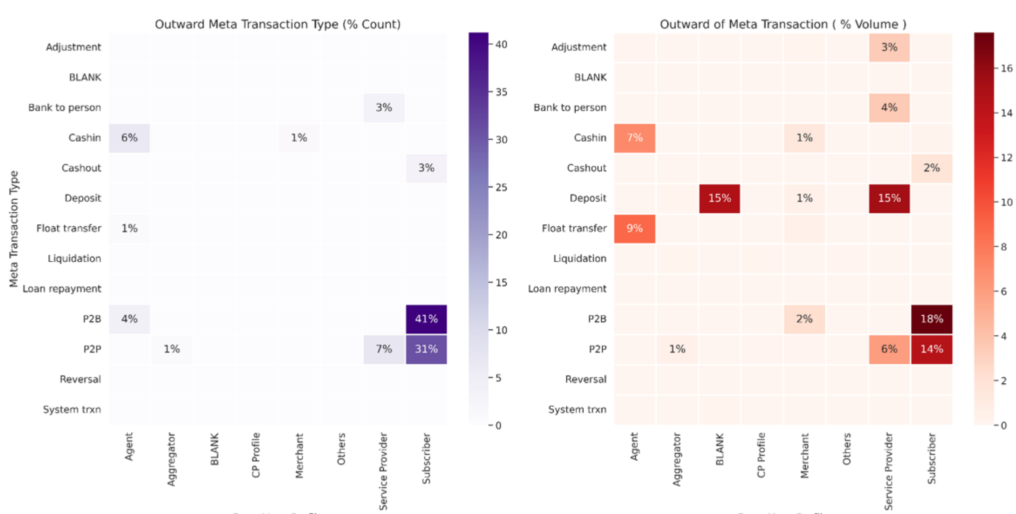

Figure 1: Outward transaction by transaction profile and type (% count and volume) Data: BNR.

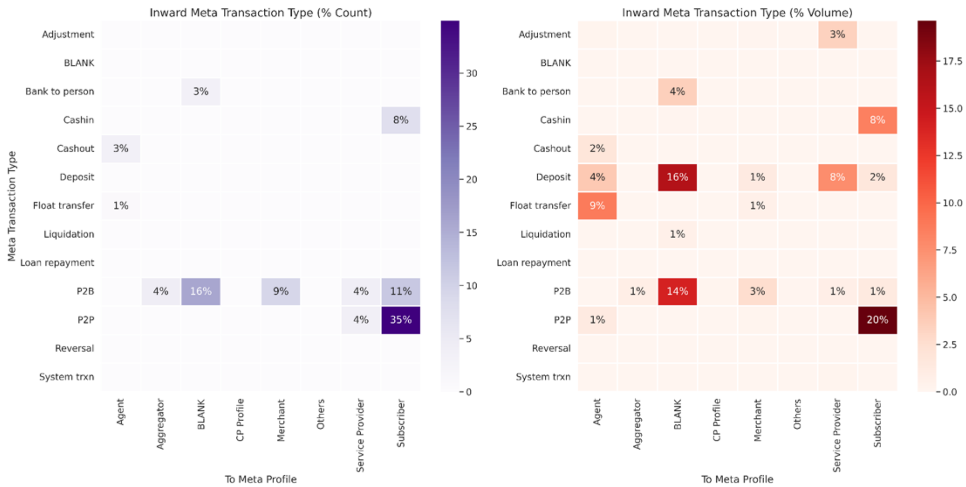

Figure 2: Inward transaction by transaction profile and type (% count and volume). Data: BNR.

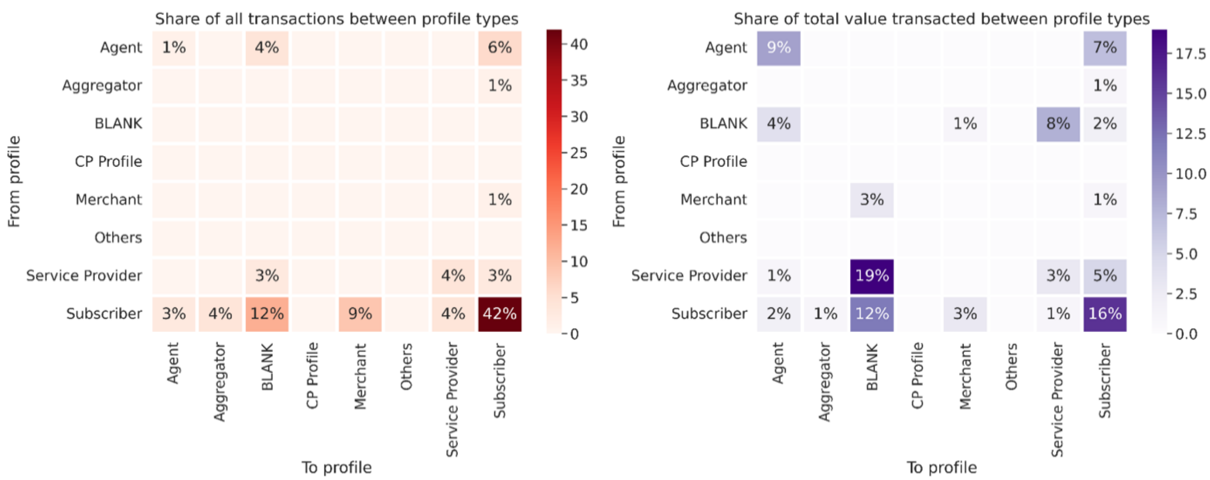

Figure 3: Transaction between different profile types (% count and volume) Data: BNR

5. Empirical strategy

The broad steps map the route from the raw transaction-level data set towards a consumption-income (and potentially investment) panel:

1. Assign individuals to different *type* categories.

These include formal firms, formal merchants, informal merchants, and mere subscribers. However, among the relevant profile types, the metadata provides us with only identifiers for the former two groups and *subscribers* as a whole. Informal merchants and wage workers fall into this category. Hence, we will need to use patterns in the data to make informed predictions about who is an informal merchant amongst the set of subscribers.

2. Assign individual transactions into spending categories.

We are currently working on the first step which we will describe in more detail below. We will then describe our envisioned approach for step 2.

Assigning individuals to different type categories.

Our approach in predicting informal merchants among subscribers relies fundamentally on the assumption that *informal* merchants show transaction and network patterns very similar to formal *merchants*, which are identified in the metadata. Figure 4 shows the distribution of five different transaction pattern metrics. In simple terms, our empirical, data-driven strategy exploits the overlap of the box plot diagrams to predict the likelihood of a subscriber being an informal merchant.

Our approach follows a similar approach to standard propensity score matching (PSM). In the same spirit of PSM, where one models the probability of a binary treatment indicator from covariates, we model the probability that a MoMo user is a known formal merchant using a labelled training set. The key challenge in this initial stage, before extending the model to estimate the likelihood that ordinary subscribers are informal merchants, is selecting the specification with the strongest predictive performance. We do this by exploiting the fact that for formal merchants we observe the true status and can directly compare predicted probabilities to actual labels.

This entails optimizing two components simultaneously: (a) the choice of statistical model, and (b) the set of features—i.e., transaction and network-derived statistics—that feed into it. To identify the most informative combinations, we compiled an extensive list of candidate statistics and evaluated various models, comparing those that integrate multiple statistics against versions that use each statistic in isolation.

Below is a (non-exhaustive) list of the transaction and network categories and specific statistics we considered.

Transaction counts and volumes

- Count of transactions (received and sent)
- Total volume transacted (received and sent)
- Ratios of receiving to sending (count and amount)
- Return transactions

Network features

- Unique contacts (received and sent)
- "Alter network" connectedness (number of edges, count, amount)

Temporal features

- Active days
- Average and standard deviation in daily transaction counts

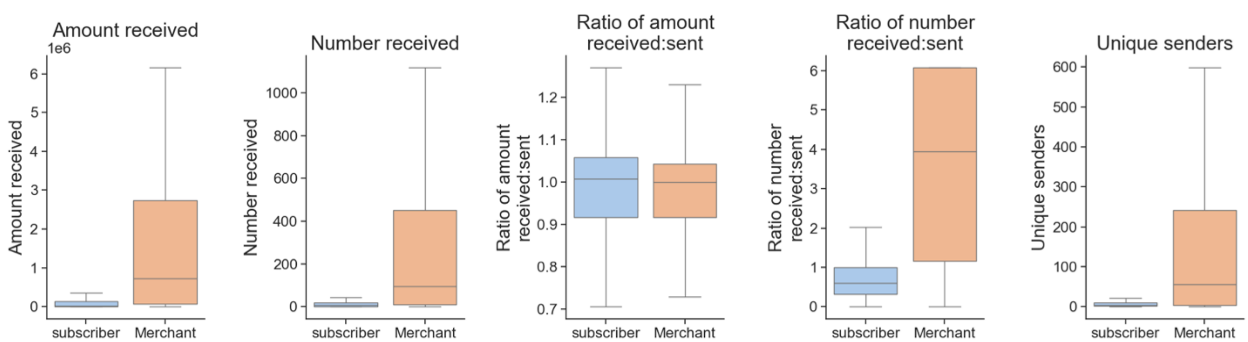

Figure SEQ Figure * ARABIC 4: Box plot distribution of subscribers and formal merchants among different transaction metrics. Data: BNR.

- Transaction volatility (standard deviation divided by average)
- Outlier days
- End of day transactions

Transaction types and amounts

- Number of received transactions with the most-commonly transacted value
- Number, total value, and percentage of cash in and cash out transactions
- Transaction size distribution

Transaction types and amounts

- Number of received transactions with the most-commonly transacted value
- Number, total value, and percentage of cash in and cash out transactions

To choose the best specification, we jointly optimize over two dimensions: the statistical learning algorithm and the set of input features (the transaction and network-derived statistics). We first sample a dataset of 1,000 accounts (500 registered subscribers and 500 registered formal merchants) to use for model training and evaluation, to keep runtime reasonable. We divide these into a balanced 75% training set and 25% test set. We evaluate all methods – both approaches using a single statistics and approaches that combine statistics using machine learning (which are described in more detail below) – only on the held-out test set.

For model selection, we follow standard practice from classification and rely on the *Receiver Operating Characteristic* (ROC) curve and its summary metric, the *Area Under the Curve* (AUC). The ROC curve plots the true positive rate against the false positive rate across all possible decision thresholds, showing how well a model separates known formal merchants from others without fixing a cutoff. The AUC is the integral under that curve and quantifies overall discriminative ability: a value of 1 represents perfect separation, 0.5 is no better than random.

We systematically compare models that use a single statistic in isolation to versions that combine multiple statistics, in order to assess whether and which combinations yield complementary gains. As an illustrative example, Figure 5 plot ROC curves and compute AUCs for several temporal volatility features when each is used alone in a given classifier; this example graph makes it clear which individual volatility metrics carry the most signal and provides a baseline against which multi-feature models can be judged. The learning algorithms we evaluate include Random Forest Classifier, Logistic Regression, Support Vector Machine, and Gradient Boosting Classifier. Figure 6 shows that among these that Gradient Boosting performs the best based on a sample of 1,000 transactions. We further increase the sample size which, however, shows only marginal improvements in the performance indicator.

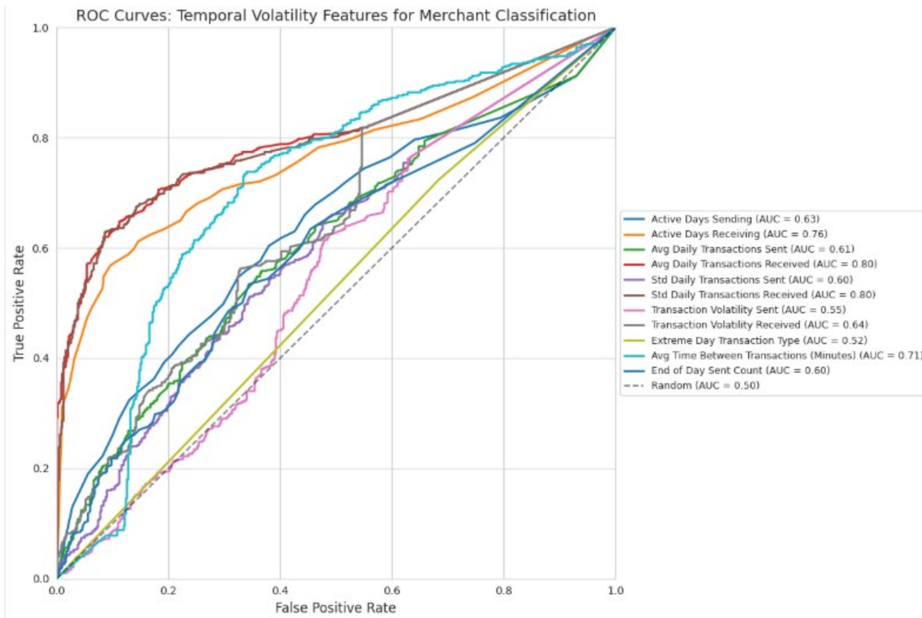

Figure 5: ROC and AUC for single temporal volatility features

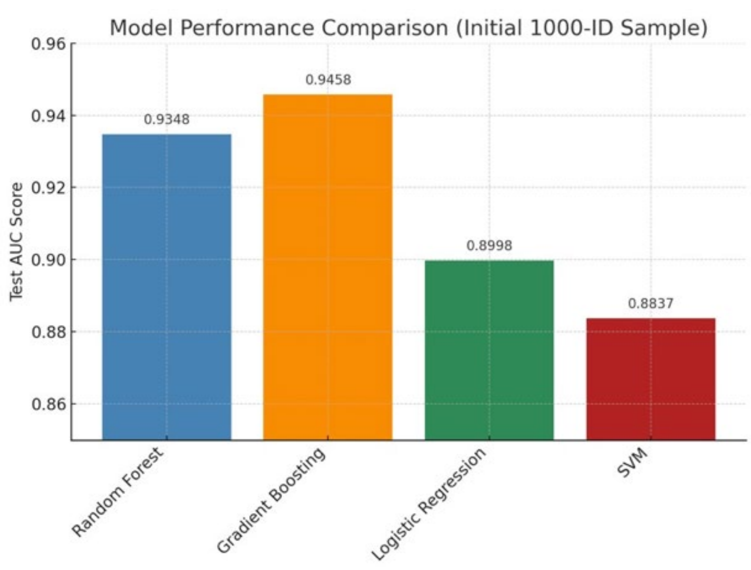

Figure 6: Performance comparison between different statistical models

We also evaluate the performance of individual inputs statistics across statistical models depicted in Figure 7. The results show that except for bank features, all inputs statistics generally perform very well and across all models. We further tested the marginal impact on performance statistics by testing all and dropping one individually to evaluate its additional test performance. Based on this, we decided to add a wide range of statistical measures as each individual comes with non-sizeable computational costs.

With that in hand, the next step is to apply the selected model to estimate the probability that ordinary subscribers are informal merchants. Because the model outputs a continuous score, we must still make a binary classification—i.e., decide a probability threshold above which a subscriber is flagged as a merchant. In an ideal world, the joint distribution of subscribers over the chosen statistics would be bimodal, revealing two distinct latent groups and making the threshold choice straightforward in practice. Figure 8 shows the outcomes of our predictive model as the distribution of predicted likelihoods of being an informal merchant showing that such a clean separation does not exist. To address this, we aim to test different options.

First, we incorporate additional information including secondary data on the overall prevalence of informal merchants among subscribers can be used to "back out" a threshold that targets that known share. Rwanda's FinScope survey offers rich information on how people conduct financial transactions and for what purposes plus additional information such as socioeconomic characteristics and district locations. It also includes the share of MoMo users that conduct business via MoMo. While we hope to get access to more granular data location data from mobile money operators in future, we usedBNR's anonymized know-your-customer data. Figure 9 shows the district-level distribution of SIM-card registrations which we believe serves as a good proxy for the share of MoMo activity. This allows us to disaggregate our MoMo data geographically and then use the share of MoMo business uses and our share of identified formal MoMo business users to approximate the location of probability decision threshold (in Figure 8 arbitrarily set at 50%) that is consistent with the residual share of MoMo business users.

A complementary and more robust strategy is to collect ground-truth labels directly from surveying a sample of subscribers asking whether they are informal merchants or wage workers and merge those responses with our existing features. That labelled subset can then be used to calibrate the threshold and validate the probabilistic scores.

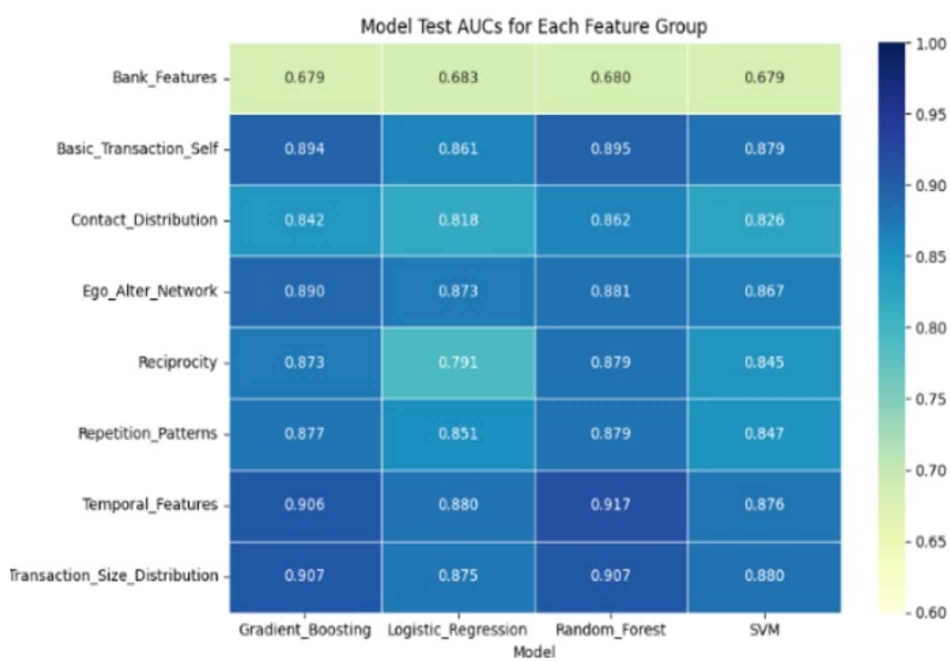

Figure 7: Performance of different combinations of statistical model and individual input statistic

Another approach, at a more aggregated level, is to compare levels of aggregate consumption, and related changes, from Rwanda recently conducted representative consumption survey, and compare the related aggregates of consumption from the MoMo data. For this, we would test different thresholds, in an iterative process, and measure consumption only of those subscribers below this threshold. We believe that while almost certainly this will not produce *levels* of consumption that align with the official statistics, changes, i.e. growth rates, could be replicated if the systematic bias of consumption through MoMo is relatively modest. This would be the case if, as the economy grows, consumption through cash and MoMo growth in proportional (while we can further account for the growth of MoMo accounts).

Assign individual transactions into spending categories.

Our next step is to assign individual transactions into spending categories. As described above, having eventually classifying each counterparty ID as a firm or an individual, and, among firms, distinguishing known formal merchants from others to the extent possible. We then exclude flows that do not represent consumption or income: (i) firm-to-firm transactions are dropped because they reflect intermediate or investment activity, (ii) transfers between private individuals are excluded as non-economic consumption, (iii) tax payments other than VAT are removed, (iv) government transfers and subsidies to individuals (except regular wages) are filtered out, and (v) interactions involving the government are retained only when they are fees paid to or wages disbursed by the government. Note, we will also rely on official data about household structures to make informed inferences about intra-household consumption, which likely differs from intra-household spending patterns (e.g., one person receives income while another person pays for consumption)

Consumption is defined as payments from individuals to firms (formal or informal merchants), with the merchant's sector or inferred activity determining the consumption category. For this categorization, we will, again, leverage patterns we detect in the data to infer spending categories, which we will validate against secondary data. For instance, wage income could be defined as regular receipts from firms to individuals at a specific re-occurring time of the month. Additionally, we aim to use above-described

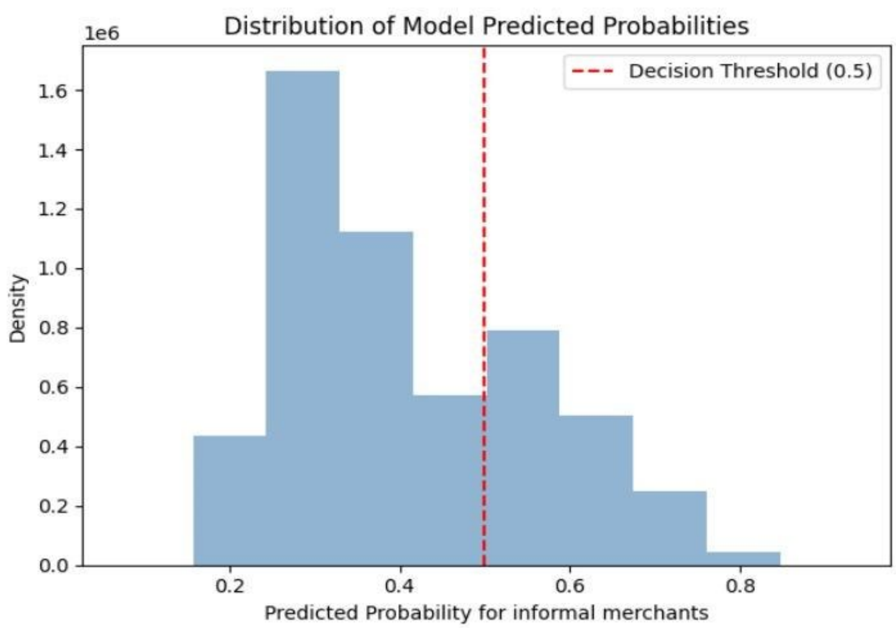

Figure 8: Outcomes of model predictions of being an informal merchant

surveys to further obtain information about spending patterns and categories to verify data from other sources and from patterns we detect from the data.

Aggregation: representativeness and cash adjustment

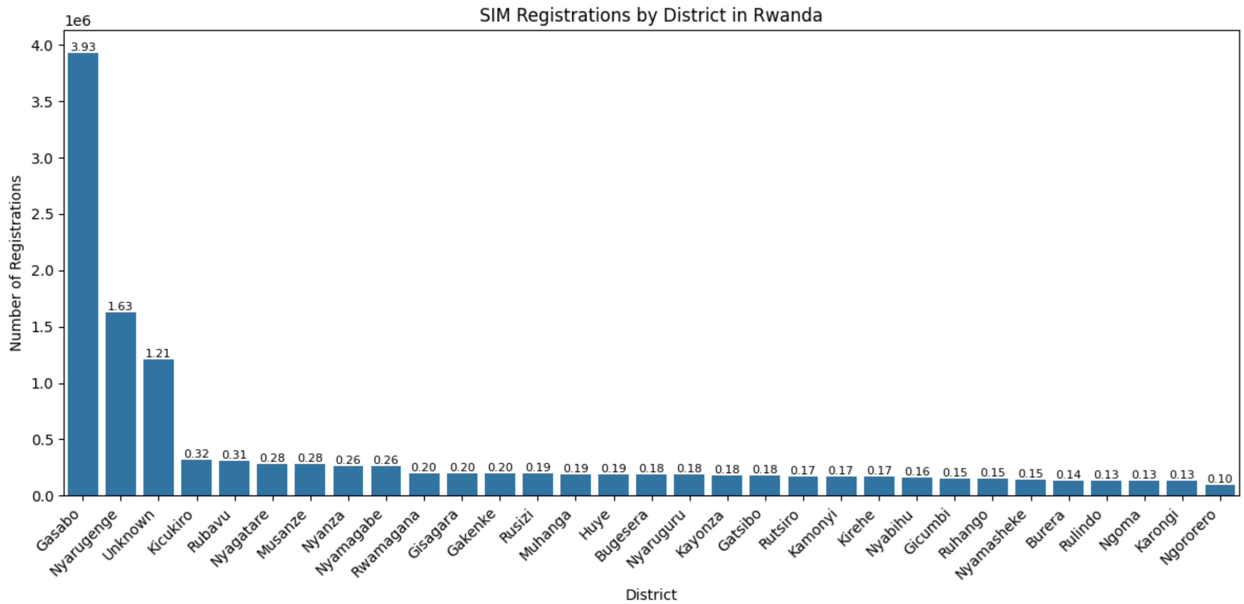

Figure 9: BNR data on SIM-card registration

As noted earlier, a key concern is that digital transaction data omit a non-random share of activity conducted in cash. As described above, we aim to use secondary survey and benchmark data to estimate cash-to-digital ratios across stratified cells (e.g., by district, settlement type, gender, and education) and reweight the raw digital aggregates so each cell aligns with its expected economic mass. These adjusted series serve as the foundation for more representative consumption and income measures, with ongoing validation against independent benchmarks and internal consistency checks to refine the correction over time.

6. Way forward

The report described the initial exploratory phase of this project during which developed a deeper understanding of the structure of the different data sources and implemented the first stages of a categorization of different MoMo users. With the current intermediate results in hand, we will finalize the informal merchant classification by refining the machine learning models and calibrating the binary threshold using both auxiliary prevalence information and newly collected survey data to distinguish wage workers from informal merchants.

Based on this, we will identify the transaction taxonomy by operationalizing the distinction between income, consumption, and investment flows by combining categorical counterparty labels with behavioral heuristics and filtering out non-relevant transactions. Eventually, we will use additional data on individuals' use of financial transactions together with socioeconomic characteristics to create a reweighted aggregate panel of Rwanda's economy. This can be continuously updated and validated, anchoring the panel in reality and correcting for coverage biases.

References

Aiken, E., Bellue, S., Karlan, D., Udry, C., Blumenstock, J. E. (2022) "*Machine learning and phone data can improve targeting of humanitarian aid",* Nature 603, 864-870.

Buda, G., Carvalho, V. M., Hansen, S., Ortiz, Á., Rodrgio, T., Rodríguez Mora, J. V. (2023), "*National Accounts in a World of Naturally Occurring Data: A Proof of Concept for Consumption*", CEPR Discussion Paper No. DP17519.

Dube, O., Blumenstock, J. E., Callen, M. (2025) "*Measuring religion from behavior: violence, economic shocks and religious adherence in Afghanistan",* unpublished manuscript, http://odube.net/wp-content/uploads/2025/03/Measuring_religion_from_behavior.pdf.

Gollin, D., Blanchard, P., Kirchberger, M. (2021) "*Perpetual motion: human mobility and spatial frictions in three African countries",* CEPR discussion paper No. DP16661.

Hanney, O. (2025), "*How economists are using data sources in development economics*", VoxDev, blogpost, https://voxdev.org/topic/how-economists-are-using-new-data-sources-development-economics.

theigc.org