Time Is (Not) Money: The decay of machine learning models using cellphone data to predict wealth levels over time

Oscar Barriga-Cabanillas[1], Mira Korb[2], and Travis Lybbert[3]

[1] *The World Bank*
[2] *University of California, Davis*
[3] *University of California, Davis*

November 10, 2024

Abstract

Cellphone data has proven successful in predicting socioeconomic status. However, data limitations have hindered the study of the capacity of these models to predict welfare over time. By relying on a panel sample with information collected two years apart, we test for decay in the ability of an algorithm to predict wealth levels. We find evidence of model decay, with the predictive capacity of a model trained on the first panel wave being 15-30% lower than a model trained on the second wave survey and contemporary cellphone data. We link the lower performance to re-ranking of households across the wealth distribution, changes in the distribution of cellphone features over time and the rise of internet-based communication apps. Finally, we explain how the COVID-19 pandemic serves as a mechanism through which these effects could occur.

JEL Classification: C8, D0, O1
Keywords: Poverty, mobile phone data, machine learning, cash transfers, targeting, Haiti

1 Introduction

Most social programs in developing countries are articulated around the explicit objective of reducing poverty (Elbers, Gunning and Kinsey, 2007). This objective makes targeting the poor an integral part of program design, with widely studied benefits on improving the program's cost-effectiveness; see for example (Alatas et al., 2012; Coady, Grosh and Hoddinott, 2004; Brown, Ravallion and Van de Walle, 2016). Detailed information on the geographic distribution of poverty and wealth is a key parameter to guide decisions on the allocation of limited resources. However, in developing countries the required information is often not available with the spatial or temporal coverage necessary to inform the roll out of programs.

In recent years, non-traditional sources of data have shown potential to compensate for data gaps. Among these applications, the usage of Cellphone Detail Records (CDR) has shown potential to inform the targeting of humanitarian aid and measure the effects of policy interventions. These applications work by leveraging the digital footprints from an individual's communication transactions, and through the combination of feature engineering and machine learning, accurately predict wealth levels for small geographic regions and cellphone subscribers' socioeconomic characteristics.

We build on this literature by testing the capacity of these models to impute welfare levels over time. Specifically, we investigate whether the performance of a model trained on CDR data to predict wealth levels declines over time, a phenomenon referred to in the machine learning literature as model decay. We match a panel sample of surveyed households in Haiti to the participants' CDR data. Results show that the predictive capacity of the model trained to predict relative wealth in the initial panel wave goes down by 15-30% when used to predict the observed wealth of the household in the second wave, collected two years after.

We investigate several explanations for the loss of predictive capacity, linking each to concepts in the model decay literature:

1. Changes in wealth ranking between survey rounds that cannot be captured by the model
2. Differential attrition among categories where prediction works best
3. Changes in the distribution of predictive features over time

4. Rise in the use of communication apps, where the directions and volumes of calls and texts are masked, excluding an important source of information from the training algorithms

Finally, we discuss the extent to which the COVID-19 pandemic could be driving the observed effects listed above. While the existing literature investigates how the pandemic affected algorithms designed to predict outcomes in the health and financial sectors, specifically in developed country contexts, we make a contribution by discussing the extent to which these limitations are present in algorithms using CDR data to predict wealth levels in a developing country context.

Our results are the first formal assessment of the capacity of machine learning models to properly predict welfare levels at different points in time. Past applications of predicting individual welfare levels using CDR data show success when used at single points in time. Our results point toward the need of constantly updating socioeconomic information even in the era of big data, a situation that is not unusual for commercial applications where similar algorithms are used to estimate credit scoring.

2 Data and Methods

Survey Data. Our primarily analysis relies on two surveys. First, we use an nationally representative in-person survey with information on 4,267 households, collected between May-October 2018. The survey was conducted by Finmark Trust as part of its 'FinScope Consumer,' and is close to a Living Standards Measurement Survey (LSMS), asking questions related to household demographics, asset ownership, food consumption, food security and access to financial services. Our second source of information is a phone follow-up survey conducted in April 2021 on the original survey participants, where we collected a subset of questions focused on asset ownership and welfare. Since participation in the Finscope survey was in person, we only contacted the original 2,869 participants who provided a phone number at the time of the original survey. We ultimately were able to recontact and match 505 individuals to their cellphone records, which we refer to as the panel sample.

Call Detail Records. The call detail records (CDR) data are provided by the largest Mobile

Network Operator (MNO) in Haiti, which accounts for about 75% of the cellphone market in the country. The CDR data contain all transactions made in the network, providing information for each phone call, text message, and mobile money transfer placed on the operator's network. For each transaction, we observe a unique identifier for the parties involved in the event (e.g., the two subscribers on a phone call), a timestamp for the event and the location of the cellular tower routing the transactions. The CDR data is anonymized, so we cannot see the content of calls or text messages.

The underlying CDR data represents week-level statistics, so we must decide how many days of cellphone transaction data to use in the feature generation process. While a wider time window affords a more diverse set of transactions from which to extract information, it is also important to consider how much wealth varies during this time window. We generate features using CDR data from three time windows, each of which covers a four month period. Window 1 takes place just after the in-person survey (12/10/2018 - 3/10/2019), while Window 2 occurs midway between the in-person survey and the follow-up phone survey (12/10/2019 - 3/10/2020) and Window 3 (12/10/2020 - 3/10/2021) occurs just before the phone survey. Our choice of a four month window aligns with existing applications predicting individual wealth levels, which extract features from six months to 1 year of CDR data (Blumenstock, Cadamuro and On (2015); Aiken et al. (2022)).

Informed consent and matching. As part of the phone survey, participants provided 'informed consent' to match their information with their CDR data. For the sample in the in-person survey, we are able to match a total of 1,129 respondents with their CDR data for Window 1. This sample includes both those individuals successfully contacted in the phone-follow up and those whose lines had been disconnected for more than six months after the survey.[1] Of these 1,129 individuals, only 505 individuals could be contacted in the follow-up phone survey and were matched to all three windows of CDR data.

Wealth Index Our main outcome used to test the predictive capacity of CDR data is a wealth

[1] Between January and March 2021 we made four attempts to contact each number. In consultation with our IRB, we used survey responses from respondents who either provided informed consent or whose line had been disconnected for more than six months, making it impossible to re-contact the original survey respondent. Out of 2,869 respondents with a phone, 909 belonged to a different MNO network, 167 refuse to participate in the phone survey and 661 numbers remained active but could not be contacted.

index calculated via principal components analysis (PCA), which includes 14 underlying binary asset variables. We run the PCA separately over the in-person and phone follow-up samples, using assets that were covered by questions present in both surveys. Overall, the relative weights of the variables are similar across surveys and samples; see Table 1.[2]

Asset	Loadings 2018 - Window 1	Loadings 2018 - Panel	Loadings 2020 - Panel	Asset Mean 2018 - Window 1	Asset Mean 2018 - Panel	Asset Mean 2020 - Panel
tv	0.51	0.44	0.42	0.49	0.58	0.54
radio	0.35	0.26	0.31	0.59	0.65	0.58
fridge	0.36	0.41	0.37	0.19	0.29	0.32
fan	0.43	0.44	0.41	0.24	0.34	0.37
bike	0.05	0.06	0.03	0.04	0.06	0.08
motorbike	0.03	0.05	0.10	0.07	0.08	0.18
car	0.11	0.14	0.12	0.05	0.07	0.07
computer	0.14	0.19	0.22	0.07	0.11	0.16
washing machine	0.03	0.05	0.03	0.01	0.02	0.02
wardrobe	0.34	0.38	0.25	0.29	0.36	0.37
ac	0.01	0.01	0.02	0.01	0.02	0.01
electric appliances	0.27	0.29	0.42	0.19	0.25	0.54
gas	0.20	0.25	0.33	0.10	0.15	0.29
bed	0.18	0.15	0.07	0.78	0.82	0.92
N	1129	505	505	1129	505	505

Table 1: Magnitude of First Principal Component for 2018 field survey and 2021 phone survey

Non-random attrition. Since our goal is to test the capacity of CDR data to predict a wealth index, we test for non-random attrition between the samples in the 2018 in-person survey and the phone follow-up (1,129 and 505 participants, respectively.) We estimate a probit model on the in-person survey data using 1,048 of the 1,129 households that could be matched to the CDR data, who also had complete information on head of household characteristics. This allows us to investigate the correlation of different observable characteristics on the probability of observing an individual participating in the phone survey in 2021.

We show in Table 2 that households headed by wealthier and more educated individuals tend to be more likely to remain in the panel. This is in line with past results where better-off individuals are more likely to keep their cellphone numbers for longer periods (Barriga-Cabanillas and Lybbert, 2021).

[2]The wealth index estimated over the panel sample explains 64.9% of variance in asset ownership in the 2018 in-person survey but 81.3 % of the variance in asset ownership in the 2021 follow-up phone survey.

Table 2

	Remains in panel
Intercept	-0.7826***
	(0.2601)
Age 25-34	0.1944
	(0.1863)
Age 35-44	0.4543**
	(0.1887)
Age 45-54	0.3431*
	(0.1947)
Age 55-64	0.2840
	(0.2105)
Age 65+	0.4065*
	(0.2305)
Preschool/primary	0.0922
	(0.1510)
Secondary	0.2691*
	(0.1578)
University or higher education	0.5773***
	(0.2114)
Male	0.1966**
	(0.0824)
Urban	0.0444
	(0.1349)
Wealth Index	0.4239***
	(0.0638)
N	1048

Feature engineering In its raw form, the CDR data provides a detailed record of a user's cellphone activity. We use feature engineering to extract information about each user's behavioral patterns of cellphone use that would correlated with socioeconomic characteristics, in this case a wealth index.[3] We extract features for the three different time windows using information from four month intervals. The panel sample of 505 individual is present for all the three windows but the original 1,129 individuals are only present in the first window due to attrition from the cellphone network. For each phone number, we extract 998 features in each of the three time windows. However, when estimating our predictive model we exclude features relating to the precise location of cellphone towers and international calls. In addition, we adjust the features in Windows 2 and 3 relating to monetary variables, such as mobile money expenditures, for inflation so that all expenditures are in 2018 values.

2.1 Prediction model

We adopt the machine-learning methods used by previous work (Aiken et al. (2022); Aiken et al. (2023); Blumenstock, Cadamuro and On (2015)) to train a model that predicts wealth from CDR features. Our preferred algorithm uses gradient boosting. The model is trained and evaluated with five-fold cross validation, where hyperparameters are tuned independently on each fold, allowing us to generate out-of-sample estimates of accuracy and out-of-sample predictions of wealth for every household in the sample. Once the optimal hyperparameters have been determined, we re-train the model over the entire sample. [4]

The optimal hyperparameters of the model trained on the 2018 panel sample include a learning

[3] Features are generated using Bandicoot, an open-source toolbox for CDR analysis. For a full description of the method see De Montjoye, Rocher and Pentland (2016). Features extracted include information about an individual's overall behavior (average call duration and percent initiated conversations, percent of nocturnal interactions, inter-event time between two phones calls), spatial patterns based on cell tower locations (the number of unique antennas visited and the radius of gyration), social network (the entropy of their contacts and the balance of interactions per contact) and recharge patterns (including the average amount recharged and the time between recharges).

[4] Parameter grid:

- Winsorization of features: either no winsorization or winsorize the top and bottom 0.5% of each feature
- Minimum data in each leaf: 10, 20, 50
- Number of estimators: 20, 50, 100
- Learning rate: 0.05, 0.075, 0.1

rate of 0.1, 20 estimators, 50 data points in each leaf and winsorizing the top and bottom 0.5% of each feature.

3 Results

As a benchmark, we start with the features from Window 1 and proceed to estimate independent models that predict wealth based on the in-person survey in 2018 for the Window 1 and panel sample. We compare the results for predictions generated by estimating the model over the Window 1 sample (N = 1129) with those results generated by estimating the model over the sub-sample of these households that are present in the panel (N = 505).

As Table 3 shows, prediction performance is comparable between the model trained on the 2018 Window 1 sample for predicting 2018 wealth levels (Column 1) and the model trained and predicting for the 2018 panel sample (Column 2), even if the panel sample contains about half the observations and is affected by non-random attrition.

Table 3: Predicting household wealth

Statistic	Train 2018, Test 2018	Train 2018, Test 2018 Panel	Train 2020, Test 2020 Panel	Train 2018, Test 2020 Panel
Pearson	0.405	0.438	0.318	0.366
Spearman	0.372	0.411	0.306	0.356
R2	0.163	0.192	0.101	0.130
N	1129.000	505.000	505.000	505.000

Next, we asses the capacity of the model trained on panel sample using features from Window 1 to predict wealth levels for these same individuals in 2020 using the features from Window 3. This allows us to directly compare how well the 2018 model can predict the 2020 observed wealth levels. From Table 3, we can see that across all measures of model fit, the model trained on the 2018 data performs worse at predicting wealth for panel households in 2020 (Column 4) than 2018 (Column 2). The Pearson correlation coefficient drops from 0.438 to 0.366, while the Spearman correlation coefficient drops from 0.411 to 0.356. R-squared also falls from 0.192 to 0.130. For a detailed discussion about the poor fit of the model, across all samples, we refer to Barriga-Cabanillas et al. (2021). Next, we turn to investigating the drivers of this model decay.

4 Drivers of Model Decay

The term model decay is typically used to describe a phenomenon where the predictive capacity of a model declines over time. The literature describes two main drivers of model decay: data drift and concept drift. Data drift refers to a meaningful change in the input data or the distribution of the variables used in the prediction model over time, while concept drift describes situations where the relationship between the input data (CDR) and the target variable (wealth index) changes over time, while the distribution of the inputs might remain the same. In practice, it is often difficult to distinguish between these two phenomena as both can operate at the same time.

We explore four potential explanations for the deterioration of our model's performance between 2018 and 2021: (1) attrition from the panel and differential model performance across categories, (2) changes in the distribution of predictive features, (3) re-ranking of wealth between survey rounds that cannot be captured by the model and (4) the increasing use of communication apps, such as WhatsApp, that is not captured by the voice and text data transactions. We attempt to link each explanation to the ideas of data and concept drift from the model decay literature.

4.1 Attrition and differential accuracy across categories

In our setting, we may be concerned if the model predicts well on sub-groups of individuals that drop out of the panel between 2018 and 2021. Recalling that less wealthy and less educated households tend to drop out of the panel, if we find evidence that model fit is higher for these groups, the lower predictive capacity could be related to the exiting of these individuals from the sample.

To investigate this, we calculate the R^2 and Spearman correlation coefficient between the true and predicted wealth index from the model trained on features from Window 1 and predicting wealth for all individuals that could be matched to the Window 1 data, by sub-group. This allows us to examine how model fit varies by household characteristics such as gender, age and education level of the head of household, as well as whether the household is located in a rural or urban area.

As we can see in Table 4, the model trained on the Window 1 data to predict the 2018 wealth levels does not equally predict across individual characteristics. However, we do not find consistent evidence that model fit is lower for more educated or wealthier individuals, relative to less educated and less wealthy individuals. In regards to education, while we find that the R-squared for the most educated sub-group is lower relative to the other education levels, we find that the Spearman correlation is higher. When examining model fit across wealth terciles, we see that the fit is best, although still poor, for the wealthiest tercile. Thus, while non-random attrition of households from the panel where predictive performance was highest could be an explanation for worse model performance on the 2021 data, our findings on this are inconclusive and we turn to other potential explanations of model decay.

Table 4: Model Fit by Subgroup

Group	Spearman	R2	N
Female	0.331	0.135	617
Male	0.395	0.169	512
Rural	0.037	-1.111	124
Urban	0.372	0.160	1005
Age 15-24	0.340	-0.093	65
Age 25-34	0.300	0.068	261
Age 35-44	0.419	0.188	276
Age 45-54	0.359	0.174	258
Age 55-64	0.341	0.196	159
Age 65+	0.424	0.223	110
no education	0.384	0.093	102
preschool/primary	0.281	0.043	351
secondary	0.255	0.020	496
university or higher	0.486	-0.095	99
Wealth Tercile 1	0.062	-9.853	448
Wealth Tercile 2	0.105	-2.471	305
Wealth Tercile 3	0.214	-2.615	376
Total	0.372	0.163	1129

4.2 Changes in the distribution of the predictive features

Data drift can be driven by changes in the distribution of the features used in the predictive model. In our setting, this would arise if the distribution of variables in the CDR data displays meaningful changes between Window 1 and Window 3. An advantage of our data is that both survey rounds were collected in similar time periods of the year, which allows us to eliminate the effects of seasonal cellphone usage patterns.

Table 5 shows the 10 most important variables from the model trained on the 2018 data, used to predict wealth levels in 2018, with panel households only. Feature importance is calculated based on the total number of times each feature is used to split nodes across all trees used in the gradient boosting algorithm.

Table 5: Features with highest importance. Model trained on 2018 panel sample

Feature	Importance
number_of_antennas_weekday_day	6
recharges_sum	5
interevent_time_weekend_night_call_median	4
recharges_mean	4
interevent_time_allweek_day_text_std	3
percent_pareto_durations_allweek_allday	3
entropy_of_contacts_weekday_allday_call	3
number_of_antennas_weekday_allday	3
percent_nocturnal_allweek_call	3
interevent_time_weekday_day_call_kurtosis	3

We plot the histograms and provide summary statistics for each of these features across three time windows of CDR data, to illustrate how the distributions of these variables are changing over time. Below each plot, the top table provides the summary statistics for each of the most important features across individuals that are present in all three windows of the CDR data. The second table provides the mean of the feature for only observations in the Finscope data that could be matched to all three windows of CDR data and are part of the panel sample (N = 505), by wealth tercile (as defined by the wealth index that is calculated using the panel sample for asset data in 2018) and window. Taken together, we argue that significant changes in the distributions of the 2018 model's most important predictors over time could be a meaningful driver of model decay.[5]

The COVID-19 pandemic is an important mechanism that could explain the changing distribution of important features from 2018 to 2021. The pandemic likely influenced the extent and the ways that people used their phones. Call and text volume likely increased as people tried to comply with social distancing protocols, while mobile money transactions also likely increased as a strat-

[5]Figures and tables exploring the distributions over time for the remaining 6 variables from the set of most important predictors are included in the Appendix.

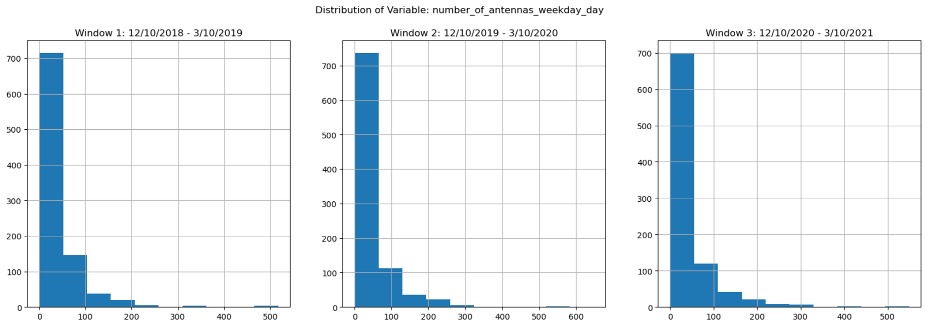


Figure 1: Number of Antennas Weekday

egy to cope with negative income shocks. Similarly, since peoples' daily routines were dramatically changed by the pandemic, the the times and locations where they used their phones might also have been affected. Indeed, we see some evidence in favor of this hypothesis by comparing Table 5 and Table 6: features relating to mobile money transactions are more important for predicting wealth in the model trained on the 2020 features than the model trained on the 2018 features. Furthermore, we also see that features relating to the percent of calls and texts occurring at home become more important.

Table 6: Features with highest importance. Model trained on 2020 panel sample

Feature	Importance
mobiledata_num_transactions	10
entropy_of_antennas_allweek_allday	8
call_duration_allweek_allday_call_kurtosis	5
balance_of_contacts_allweek_day_text_kurtosis	4
entropy_of_antennas_weekday_day	4
call_duration_weekend_day_call_std	3
call_duration_allweek_day_call_mean	2
percent_nocturnal_weekday_call	2
mobilemoney_all_other_amount_mean	2
percent_at_home_allweek_night	2

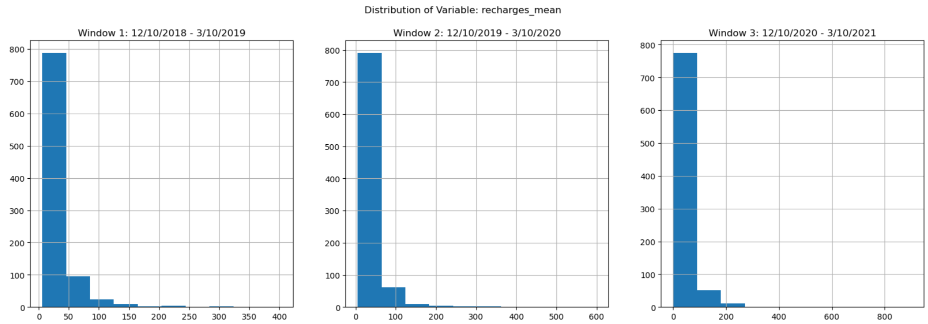


Figure 2: Mean Recharges

Statistic	Window 1	Window 2	Window 3
Mean	40.390	44.068	42.136
Median	24.000	24.000	20.500
SD	51.317	60.342	60.661
Min	0.000	0.000	0.000
Max	517.000	647.000	549.000
N	931.000	931.000	931.000

Panel Wealth Tercile	Window 1	Window 2	Window 3	N
Low	24.80	27.62	31.47	169
Med	47.42	54.85	52.95	168
High	70.02	79.31	72.91	168
Total	47.37	53.98	52.32	505

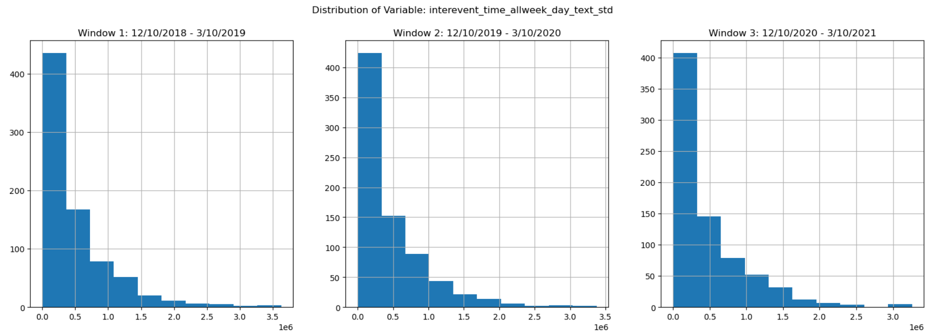


Figure 3: Inter-event Time, All Week Texts

Statistic	Window 1	Window 2	Window 3
Mean	31.851	33.915	41.840
Median	22.728	23.224	26.948
SD	32.935	40.914	57.972
Min	5.655	5.145	0.000
Max	403.151	599.945	903.164
N	931.000	931.000	931.000

Panel Wealth Tercile	Window 1	Window 2	Window 3	N
Low	28.92	29.64	39.34	169
Med	27.78	29.47	35.98	168
High	46.34	42.49	59.23	168
Total	34.31	33.91	44.92	505

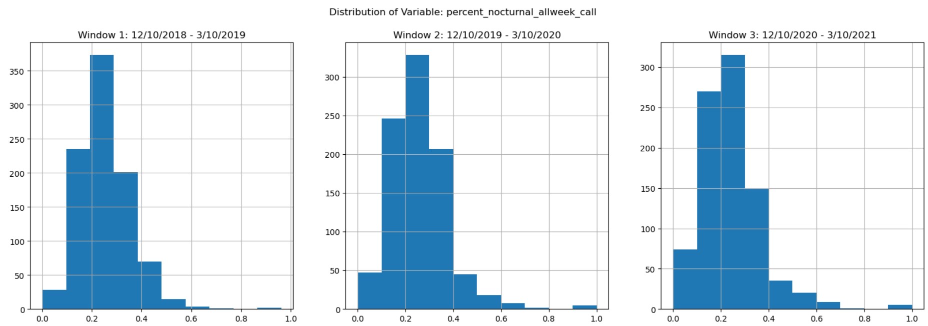


Figure 4: Percent Nocturnal Calls

Statistic	Window 1	Window 2	Window 3
Mean	478069.401	446452.346	464205.653
Median	306511.080	289282.197	271555.113
SD	559036.848	516544.842	536920.665
Min	0.000	0.000	0.000
Max	3624953.000	3371807.816	3260349.396
N	931.000	931.000	931.000

Panel Wealth Tercile	Window 1	Window 2	Window 3	N
Low	491172.15	512514.57	455416.17	169
Med	472844.88	456895.56	496070.77	168
High	483517.43	474335.54	525890.06	168
Total	482327.22	480053.02	493554.74	505

Statistic	Window 1	Window 2	Window 3
Mean	0.252	0.259	0.241
Median	0.243	0.249	0.224
SD	0.103	0.123	0.130
Min	0.000	0.000	0.000
Max	0.962	1.000	1.000
N	931.000	931.000	931.000

Panel Wealth Tercile	Window 1	Window 2	Window 3	N
Low	0.25	0.26	0.25	169
Med	0.26	0.25	0.25	168
High	0.23	0.23	0.20	168
Total	0.24	0.25	0.23	505

4.3 Wealth re-ranking

As discussed, the model trained on the 2018 data (Window 1) fails to predict the wealth levels of the panel sample when using the feature levels in 2021 (Window 3). This could be driven by the re-ranking of individual's wealth over time due to shocks. Linking back to the model decay literature, we are worried about concept drift: the meaning of what we are trying to predict (wealth) may be evolving over time, even if the CDR data stays relatively constant. In this way, the relationship between CDR features and wealth may be changing from 2018 to 2021, worsening the performance of a model trained on 2018 data to predict wealth levels from CDR data in 2021.

To test for this, we compare the wealth distribution for the panel sample between 2018 and 2021. For completeness, we show the re-ranking both when independently estimating the PCA in 2018 and 2021, as well as applying the basis vector from 2018 to the asset data from 2021.

Comparing the fraction of households in each 2018 and 2021 wealth quintile suggests that there is some re-ranking in the wealth distribution over time. When looking at the wealth index generated with the basis vector and asset data from 2021 (Table 7), while many of the poorest households (Quintile 1) in 2018 remain the poorest households in 2021, many of these households are estimated to be relatively wealthier in 2021 as they have moved into the second or third quintiles. Almost all of the richest households in 2018 (Quintile 5), remain in the top two quintiles in 2021. Households falling in Quintiles 2-4 in 2018 tended to be the most likely to move slightly up or down in the wealth rankings in 2021. From Table 8, we also see that when looking at the wealth index generated using the 2018 basis vector but the 2021 asset data, the re-ranking of households is similar.

Wealth Quintile 2020, 2020 loadings Wealth Quintile 2018, 2018 loadings	1	2	3	4	5
1	0.406	0.287	0.149	0.119	0.040
2	0.308	0.250	0.173	0.135	0.135
3	0.206	0.225	0.255	0.157	0.157
4	0.099	0.129	0.188	0.376	0.208
5	0.041	0.113	0.155	0.216	0.474

Table 7: Wealth index in 2021 estimated using basis vector from PCA on 2021 data and asset data from 2021

Wealth Quintile 2020, 2018 loadings Wealth Quintile 2018, 2018 loadings	1	2	3	4	5
1	0.485	0.188	0.158	0.129	0.040
2	0.288	0.279	0.154	0.135	0.144
3	0.196	0.206	0.284	0.186	0.127
4	0.089	0.119	0.248	0.337	0.208
5	0.031	0.103	0.155	0.216	0.495

Table 8: Wealth index in 2021 estimated using basis vector from PCA on 2018 data and asset data from 2021

It is difficult to discern whether this re-ranking is (1) an actual reflection of the household's wealth status changing over time or (2) a reflection of changing definitions of wealth over time. For the former explanation, we must assume the set of assets that accurately captures wealth is available in 2018 and does not change over time. Thus, any observed sorting of individuals in the panel sample into different wealth quintiles occurs from some exogenous shock that changes asset ownership. We try to minimize the extent to which the model might fail to capture changing definitions of wealth over time by only using assets for which ownership has not become radically cheaper over the two years, such as having a landline. In both cases, the underlying concern is not the re-ranking of households in the wealth distribution, but rather that the model cannot accurately capture this re-ranking.

The COVID-19 pandemic is a potential explanation for this re-ranking of households over time. First, fixing the set of assets used to define wealth, if certain households sold assets to smooth consumption, they would appear relatively less wealthy in 2021. However, it is also possible that COVID redefined the set of assets associated with being wealthy. In both cases, we are concerned that our model is not equipped to pick up structural shifts in household wealth and the sources of this wealth over time.

4.4 Rise in Communication Apps

A final potential explanation for the deterioration of the model between 2018 and 2021 is the rise in usage of communication apps, such as WhatsApp, in Haiti. These apps allow for communication via call or text through the Internet-based app. As consumers transitioned towards WhatsApp, we would expect to see a reduction in the number of call and text transactions. However, the higher data usage does not contain the same level of information (or signals) related with a person's wealth. For instance, the direction, frequency and number of unique contacts cannot be extracted from the internet usage data, leaving only information about volume and location.

As a next step, we will show the changes in the total volume of transactions and explore the extent to which using only features derived from data usage reduces our model's predictive capacity.

References

Aiken, Emily L., Guadalupe Bedoya, Joshua E. Blumenstock, and Aidan Coville. 2023. "Program targeting with machine learning and mobile phone data: Evidence from an anti-poverty intervention in Afghanistan." *Journal of Development Economics*, 161: 103016.

Aiken, Emily, Suzanne Bellue, Dean Karlan, Chris Udry, and Joshua E Blumenstock. 2022. "Machine learning and phone data can improve targeting of humanitarian aid." *Nature*, 603(7903): 864–870.

Alatas, Vivi, Abhijit Banerjee, Rema Hanna, Benjamin A Olken, and Julia Tobias. 2012. "Targeting the poor: evidence from a field experiment in Indonesia." *American Economic Review*, 102(4): 1206–40.

Barriga-Cabanillas, Oscar, and Travis Lybbert. 2021. "Liquidity or Convenience? Heterogeneous Impacts of Mobile Airtime Loans on Network Usage and Communication Expenditure." *Working Paper.*

Barriga-Cabanillas, Oscar, Travis Lybbert, Joshua Blumenstock, and Daniel Putman. 2021. "The potential and limitations of big data in development economics: The use of cell phone data for the targeting and impact evaluation of a cash transfer program in Haiti." *Working Paper.*

Blumenstock, Joshua, Gabriel Cadamuro, and Robert On. 2015. "Predicting poverty and wealth from mobile phone metadata." *Science*, 350(6264): 1073–1076.

Brown, Caitlin, Martin Ravallion, and Dominique Van de Walle. 2016. *A poor means test? Econometric targeting in Africa.* The World Bank.

Coady, David, Margaret Grosh, and John Hoddinott. 2004. "Targeting outcomes redux." *The World Bank Research Observer*, 19(1): 61–85.

De Montjoye, Alexandre-Yves, Luc Rocher, and Alex Pentland. 2016. "bandicoot: a Python Toolbox for Mobile Phon eMetadata."

Elbers, Chris, Jan Willem Gunning, and Bill Kinsey. 2007. "Growth and risk: Methodology and micro evidence." *World Bank Economic Review*, 21(1): 1–20.

5 Appendix

6 Constructing the wealth index

I construct the wealth index by taking the first principal component from running PCA over the following set of dummies for the presence of an asset in an household:

1. Television
2. Radio
3. Fridge
4. Fan
5. Bicycle
6. Motorbike
7. Private car
8. Computer
9. Washing machine
10. Wardrobe or dresser
11. Air conditioning
12. Electrical appliances
13. Gas appliances
14. Bed

Note that the 2018 Finscope includes information on the following additional assets: mobile phone access, generator use, antenna, camera, water source and sewage system access. I do not include these variables in the construction of the wealth index because they are not present in the asset questions for the 2020 Finscope survey and I wanted to calculate the wealth index for the two years using a consistent set of variables so that differences in predictive power between the two years would not be driven by the number/set of variables available. I also drop the following assets: VCR, landline and jewelry.

6.1 Machine learning algorithm

I adopt the GBM methodology used by Aiken et al. (2022) to train a model that predicts wealth index from CDR features. Specifically, I train a gradient boosting regressor with Microsoft's light GBM over the Window 1 and the panel samples, separately. Following Aiken et al. (2022), I tune the hyperparameters of the model over three-fold cross validation, with parameters chosen from the following grid:

- Winsorization of features: No winsorization, 0.5% limit
- Minimum data in leaf: 10, 20, 50
- Number of leaves: 5, 10, 20
- Number of estimators: 20, 50, 100
- Learning rate: 0.05, 0.075, 0.1

I train and evaluate the model over five-fold cross validation, with hyperparameters tuned independently on each fold, to obtain out-of-sample estimates of accuracy and out-of-sample predictions of wealth index for each observation in the Window 1 and panel samples. I then re-train the model on all the data (for each of the two samples separately), record feature importances (the total number of times a feature is split on over the entire forest) and use the final model to generate wealth predictions for every subscriber each of the samples.

6.2 Investigating Drivers of Model Decay

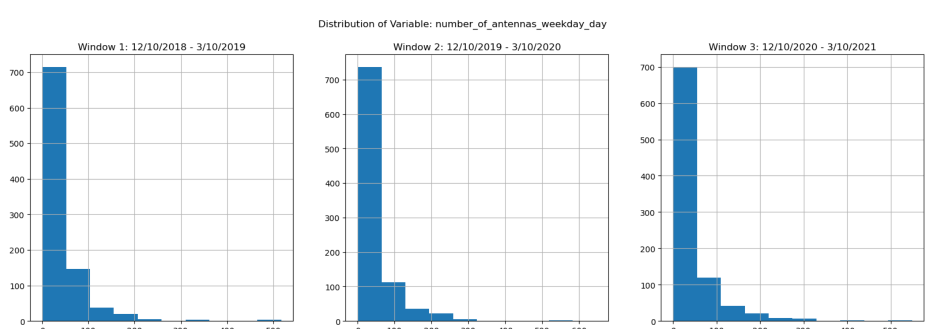
6.2.1 Distribution of most important predictors from 2018 model

The table below identifies the 10 most important predictors from the model trained on the 2018 data.

Feature	Importance
number_of_antennas_weekday_day	6
recharges_sum	5
interevent_time_weekend_night_call_median	4
recharges_mean	4
interevent_time_allweek_day_text_std	3
percent_pareto_durations_allweek_allday	3
entropy_of_contacts_weekday_allday_call	3
number_of_antennas_weekday_allday	3
percent_nocturnal_allweek_call	3
interevent_time_weekday_day_call_kurtosis	3

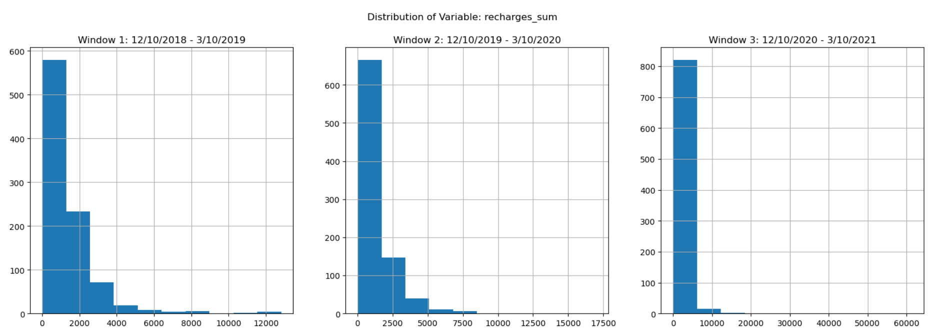
Table 9: Features with highest importance. Model trained on Window 1 Panel sample

I plot the histograms and provide summary statistics for each of these variables across three time windows of CDR data, to illustrate how the distributions of these variables is changing over time. Below each plot, the top table provides the summary statistics of each of the most important features across individuals that are present in all three windows of the CDR data. The second table provides the mean of the feature for only observations in the Finscope data that could be matched to all three windows of CDR data and are part of the panel sample (N = 505), by wealth tercile (as defined by the wealth index that is calculated using the panel sample in for asset data in 2018) and window.



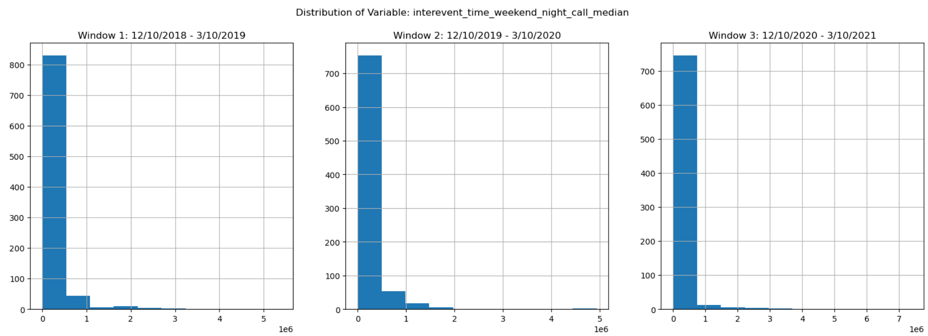
Statistic	Window 1	Window 2	Window 3
Mean	40.390	44.068	42.136
Median	24.000	24.000	20.500
SD	51.317	60.342	60.661
Min	0.000	0.000	0.000
Max	517.000	647.000	549.000
N	931.000	931.000	931.000

Panel Wealth Tercile	Window 1	Window 2	Window 3	N
Low	24.80	27.62	31.47	169
Med	47.42	54.85	52.95	168
High	70.02	79.31	72.91	168
Total	47.37	53.98	52.32	505



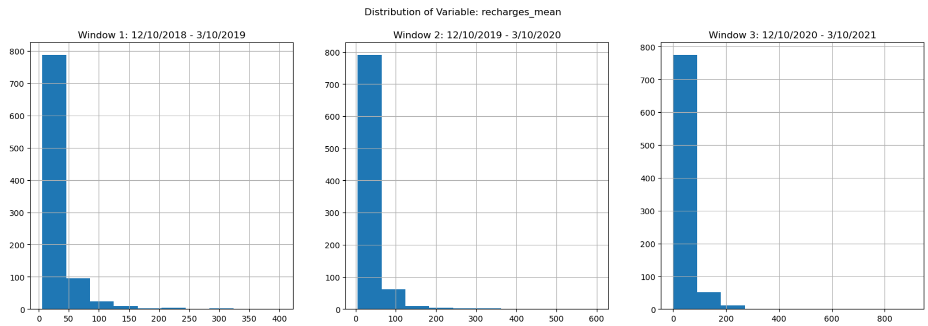
Statistic	Window 1	Window 2	Window 3
Mean	1352.649	1269.169	1439.156
Median	949.620	852.711	840.616
SD	1530.748	1478.748	2670.891
Min	9.090	7.658	0.000
Max	12813.640	17021.109	61415.179
N	931.000	931.000	931.000

Panel Wealth Tercile	Window 1	Window 2	Window 3	N
Low	1215.09	1015.66	1640.27	169
Med	1262.57	1272.78	1331.92	168
High	2172.68	1893.59	2153.77	168
Total	1548.21	1396.77	1711.17	505



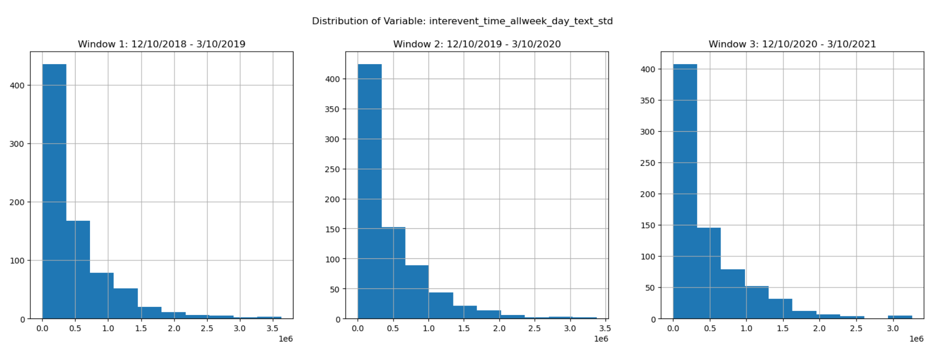
Statistic	Window 1	Window 2	Window 3
Mean	119962.163	115861.351	136751.623
Median	2493.000	2304.000	2656.000
SD	397330.426	415140.066	482936.439
Min	14.000	9.000	0.000
Max	5392421.000	4933226.000	7396019.000
N	931.000	931.000	931.000

Panel Wealth Tercile	Window 1	Window 2	Window 3	N
Low	175211.95	71874.25	110828.40	169
Med	73786.77	115646.98	99190.98	168
High	63736.71	67317.14	82474.11	168
Total	104060.55	84918.34	97387.12	505



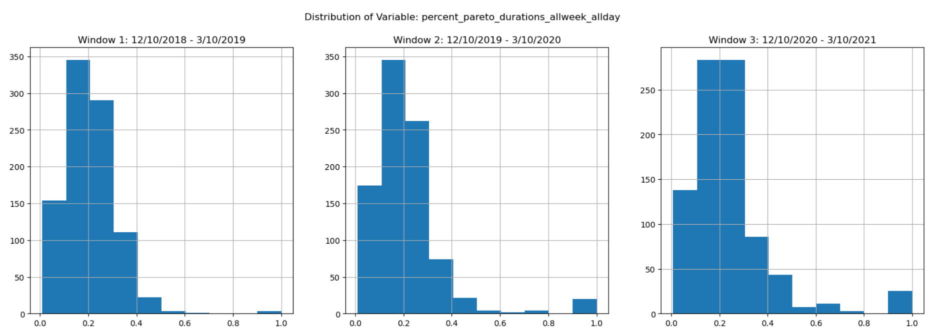
Statistic	Window 1	Window 2	Window 3
Mean	1352.649	1269.169	1439.156
Median	949.620	852.711	840.616
SD	1530.748	1478.748	2670.891
Min	9.090	7.658	0.000
Max	12813.640	17021.109	61415.179
N	931.000	931.000	931.000

Panel Wealth Tercile	Window 1	Window 2	Window 3	N
Low	1215.09	1015.66	1640.27	169
Med	1262.57	1272.78	1331.92	168
High	2172.68	1893.59	2153.77	168
Total	1548.21	1396.77	1711.17	505



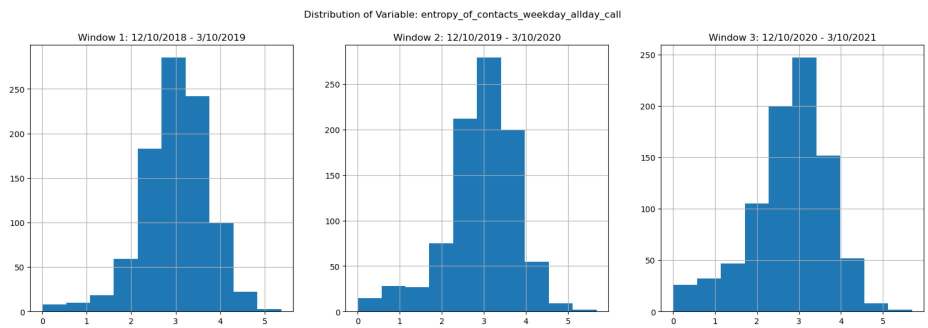
Statistic	Window 1	Window 2	Window 3
Mean	478069.401	446452.346	464205.653
Median	306511.080	289282.197	271555.113
SD	559036.848	516544.842	536920.665
Min	0.000	0.000	0.000
Max	3624953.000	3371807.816	3260349.396
N	931.000	931.000	931.000

Panel Wealth Tercile	Window 1	Window 2	Window 3	N
Low	491172.15	512514.57	455416.17	169
Med	472844.88	456895.56	496070.77	168
High	483517.43	474335.54	525890.06	168
Total	482327.22	480053.02	493554.74	505



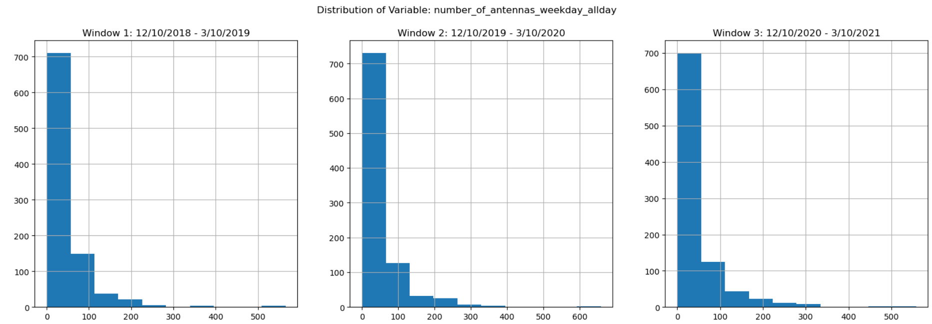
Statistic	Window 1	Window 2	Window 3
Mean	0.208	0.215	0.243
Median	0.200	0.190	0.211
SD	0.108	0.158	0.178
Min	0.009	0.010	0.007
Max	1.000	1.000	1.000
N	931.000	931.000	931.000

Panel Wealth Tercile	Window 1	Window 2	Window 3	N
Low	0.21	0.21	0.23	169
Med	0.20	0.18	0.22	168
High	0.21	0.19	0.21	168
Total	0.21	0.19	0.22	505



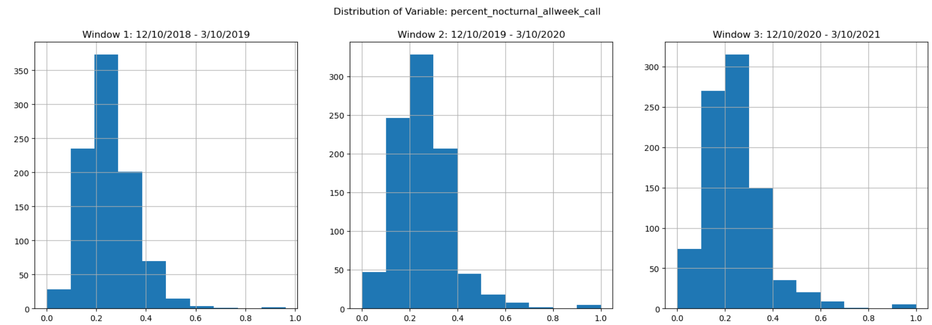
Statistic	Window 1	Window 2	Window 3
Mean	2.998	2.939	2.777
Median	3.060	3.014	2.902
SD	0.742	0.858	0.944
Min	-0.000	-0.000	-0.000
Max	5.369	5.675	5.694
N	931.000	931.000	931.000

Panel Wealth Tercile	Window 1	Window 2	Window 3	N
Low	3.02	2.99	2.83	169
Med	3.09	3.12	2.94	168
High	3.37	3.39	3.25	168
Total	3.16	3.17	3.00	505



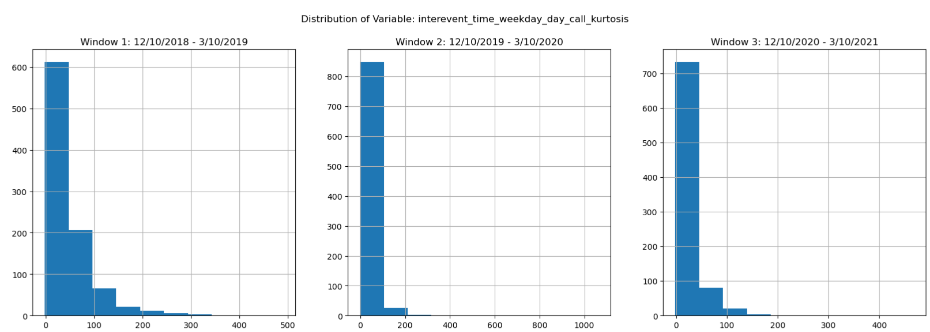
Statistic	Window 1	Window 2	Window 3
Mean	44.383	47.556	44.943
Median	26.000	25.000	22.000
SD	55.868	64.807	64.573
Min	0.000	0.000	0.000
Max	565.000	658.000	557.000
N	931.000	931.000	931.000

Panel Wealth Tercile	Window 1	Window 2	Window 3	N
Low	27.02	30.09	34.02	169
Med	52.36	59.60	56.73	168
High	77.05	85.74	77.08	168
Total	52.10	58.48	55.90	505



Statistic	Window 1	Window 2	Window 3
Mean	0.252	0.259	0.241
Median	0.243	0.249	0.224
SD	0.103	0.123	0.130
Min	0.000	0.000	0.000
Max	0.962	1.000	1.000
N	931.000	931.000	931.000

Panel Wealth Tercile	Window 1	Window 2	Window 3	N
Low	0.25	0.26	0.25	169
Med	0.26	0.25	0.25	168
High	0.23	0.23	0.20	168
Total	0.24	0.25	0.23	505



Statistic	Window 1	Window 2	Window 3
Mean	46.461	27.715	23.127
Median	29.943	17.498	13.675
SD	50.003	46.468	30.952
Min	-2.000	-2.000	-2.000
Max	490.945	1064.270	468.679
N	931.000	931.000	931.000

Panel Wealth Tercile	Window 1	Window 2	Window 3	N
Low	33.37	23.14	23.76	169
Med	47.57	27.30	24.23	168
High	68.59	39.80	31.76	168
Total	49.77	30.12	26.62	505