Financial Technology Usage 2017 Predictive Analytics Study

Alan D. Smith, smitha@rmu.edu

ABSTRACT

In 2017, a major global study (EY FinTech Adoption Index 2017) was undertaken that included 20 markets and over 22,000 online interviews. The primary goal was to provide a global perspective on financial usage (FinTech). The basic finding was that on average 1 in 3 digitally active consumers use 2 or more FinTech services. That is significant enough to suggest that FinTech has reached early mass adoption. A common assumption is that FinTech firms struggle to translate innovation and great customer experience into meaningful numbers. The initial findings reflect considerable consumer appetite for new and innovative financial service products that take advantage of new consumer technologies, such as mobile and cloud. This trend is especially true in the historically underserved emerging markets, with China and India leading FinTech adoption across the study. The purpose of this extension is to perform an analytical study, via R-language, a detailed study answering the following questions:

Question 1: ***Choose three useful metrics that you would use to assess Financial Technology Usage. Why did you choose them?***

1. Used a mobile phone or the internet to check account balance in the past year (***mobile balance***).
2. Used a mobile phone or the internet to access an account in the past year. (***mobile account***)
3. Used the internet to buy something online in the past year (***B2C***).

Question 2. ***Are there some interesting correlations between Financial Technology Usage metrics and other factors? What do they mean for policy and practice?***

During our attempt to identify the three financial technology usage metrics, we identified several potential candidates and other metrics we expected to indicate economic strength and conversely the need for economic development. It was by comparing all of these correlations that we settled upon the three above metrics. These correlations are based on data from all three years included in the dataset. The metric indicating having borrowed for a business or farm was weakly correlated with other positive non-technology usage metrics. Therefore, we decided to look at correlations for each year separately. This resulted in no usable data for 2011 and approximately half the amount of correlations produced for 2014 compared to our aggregate correlation results. 2017 yielded data very similar to the aggregate. Interestingly enough, 2014 showed that having borrowed for entrepreneurial reasons was negatively correlated with some of the other positive metrics and positively correlated with the strongest negative metric, having borrowed for medical expenses. However, 2017 data shows the exact opposite correlation pattern of being strongly correlated with the other positive metrics and strongly negatively correlated with other negative metrics. Additionally, the positive metrics of mobile/internet access to an account, credit card usage and ability produce emergency funds all exhibit a weaker correlation with the barrier metrics when compared to the other positive metrics indicating a lesser sensitivity to these barriers. Lastly, the strongest indicator of the need for economic improvement is percentage having borrowed for medical expenses which exhibits the strongest negative correlation to the positive metrics and a positive but less strong correlation to the other negative metrics.

Question 3. ***Are there differences in Financial Technology Usage across countries or regions?***

We decided to divide countries into 4 income levels (i.e., high, upper middle, lower middle, and low) and selected representative countries for each level. In general, as the number of accounts established, the 3 metrics (i.e., mobile balance, mobile account, B2C) also increased. The higher the countries' wealth, so does account established. There were a few exceptions, such as Kenya and Zimbabwe. For example, Kenya's world-leading mobile-money system (M-PESA which was established in 2007) it is now used by over 17m Kenyans. This is equivalent to more than 67% of the adult population and about 25% of the GDP flows through it.

High Income

- Estonia
- Israel
- United States

Upper middle income

- Argentina
- Croatia
- Gabon

Lower middle income

- Kenya
- Honduras
- Moldova

Low income

- Rwanda
- Zimbabwe
- Haiti

Question 4: ***What are some of the factors that you think may drive Financial Technology Usage around the globe? How would you assess them? Explain your choice of technique and be explain 5 interesting results from your Practical Analytical use?***

Adoption of Mobile Phones:

- High adoption of mobile phones in country gives rise to the scenario of an individual to make use of online/mobile services provided financial institutions or mobile money providers
- We can **assess** mobile phone adoption by looking at the internet usage in the country or phone call/SMS activity.

Presence of Non-Traditional Financial Institutions:

- In countries where accessibility to local financial institutions is not great or cost for opening an account in a traditional financial institution is too costly, presence of other options such as mobile money providers increase the use of financial activity through individuals mobile phone
- We can **assess** this looking at the population of mobile money accounts owned by a country.

Employment Rate:

- Higher employment rate leads to individuals having a source of income that allow them to participate in activities such as savings, purchasing, etc. which all require the use of a account from a financial institution or mobile money provider.
- Low employment rates in countries tend to lead to individuals not having an account, which is especially true for developing countries.

- We can **assess** employment rate by using surveys conducted by Census Bureau that collect nationwide information.

Account Accessibility/Convenience:

- Individuals with an account but don't use it often do because services associated with their account are expensive or not convenient enough for them to user in their daily lives.
- If financial institutions improve the capabilities of the services they provide with their accounts, which would convince individuals to use their accounts more to perform various financial activities.
- We can **assess** this by conducting surveys and collecting feedback from account owners about how they use the account and for what activities they would like to use the account for.

Desire to Access Global Marketplace:

- Items not available locally
- Cheaper prices from online sellers
- Asses via percentage reporting making a purchase over the internet in the last year

Question 5. ***Suggest and Present 5 Recommendations.***

1. Ensure access to health insurance in all countries to reduce the percentage reporting having borrowed for medical expenses.

2. Implement cost controls to ensure fair *and consistent* pricing for medical services/drugs regardless of health insurance coverage to reduce the percentage reporting having borrowed for medical expenses.

3. Require all governments to issue required identification documentation so that citizens may open an account.

4. Incentivize financial institutions to offer micro loans in low income countries for very specific short term needs which help in dealing with the most significant barrier to owning an account.

Model Interpretations & Assumptions

- Only 2017 data were selected with regional aggregations removed.
- Dependent variable were multiplied by 100 (ranges from 0-100).
- Independent variables were normalized and imputed.
- Coefficients for normalized independent variables discussion, magnitudes may be compared across variables because they are normalized.
- Please note, 1-unit change in the independent variable corresponding to the coefficient's effect on the dependent variable is no longer 1%.
- Instead, 1-normalized % is centered on mean & scaled on standard deviation.

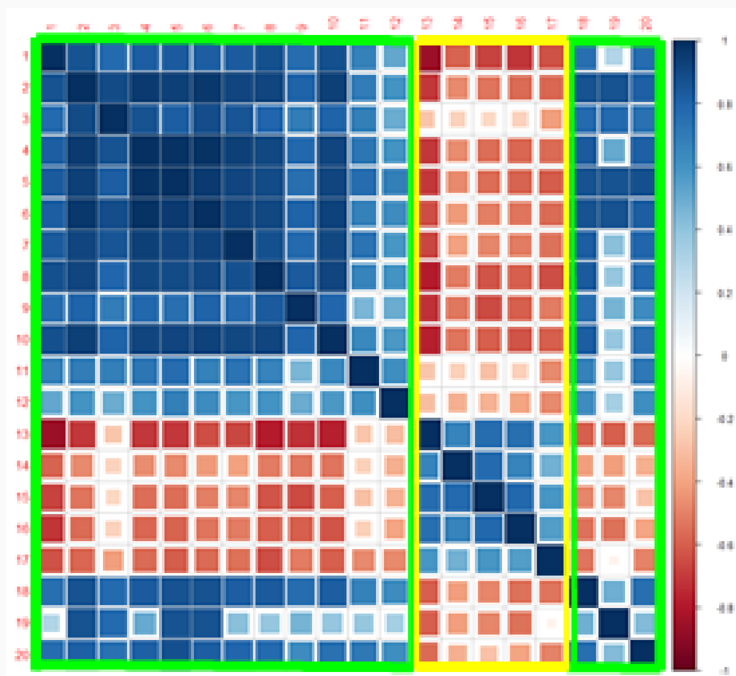
Review metrics

Positive Group

1. Account ownership
2. Mobile or internet: check account balance.
3. Mobile or internet: access an account
4. Internet: Pay bills or buy an item online
5. Internet: Buy an item online
6. Internet: Pay bills
7. Paid utility bills using an account
8. Received wages into an account
9. Received government payments into account
10. Used debit card to make purchase
11. Used credit card
12. Ability to produce emergency funds
18. Outstanding housing loan
19. Borrowed for business venture
20. Saving for old age

Negative Group

13. Insufficient funds
14. Banking services to expensive
15. Institutions are too far away
16. Lack of necessary documentation
17. Borrowed for medical expenses

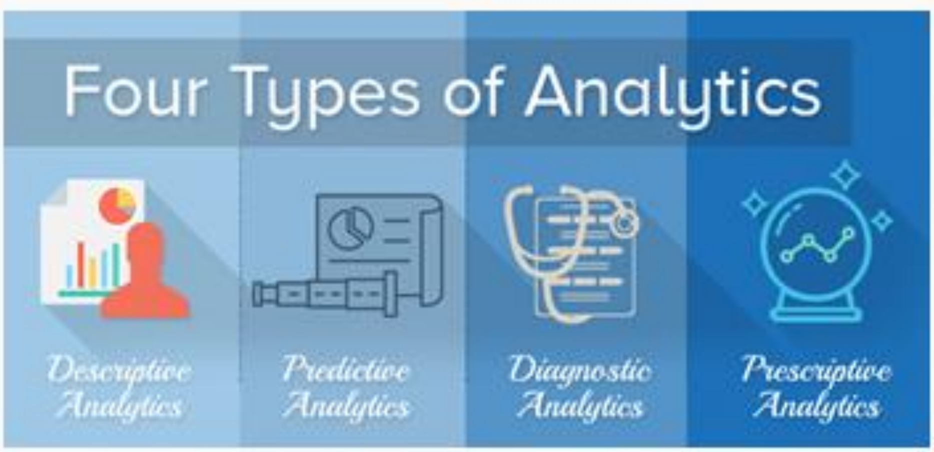


Model Development

Steps to develop model predicting Fin Tech Usage Metric:

- Test models candidate dependent variables.
- Compare adjusted R-squared values of candidate models to determine the maximum variability explained for a specific dependent variable by the selected independent variables.
- Select most important independent variables and create a new model using only those selected variables and observe the decrease in variability.
- If this decrease is acceptable, create a new model using these selected independent variables, exploring the demographic granularities.

Metrics and Data Analytics



Modeling Results: check_bal ~ .

```
Call:
lm(formula = check_bal ~ ., data = gf2017)

Residuals:
     Min       1Q   Median       3Q      Max
-12.6259  -3.2360   0.0399   2.1071  17.3492

Coefficients: (1 not defined because of singularities)
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)           25.5242     0.7119  35.853  < 2e-16 ***
own                    2.9037     1.8033   1.610 0.109955
access                 4.4742     1.3333   3.356 0.001058 **
int_purch_or_bill_pay -5.2388     4.7948  -1.093 0.276741
int_purch              6.0257     2.8053   2.148 0.033711 *
bill_pay              12.0204     3.2092   3.746 0.000277 ***
acc_util              -2.5725     1.4814  -1.737 0.085020 .
acc_wages              2.1612     1.4386   1.502 0.135620
acc_gov_pay            0.1195     0.9118   0.131 0.895979
debit_purch            3.9580     1.5540   2.547 0.012120 *
cc                     1.0852     0.9295   1.168 0.245282
prod_funds            -0.6947     0.6492  -1.070 0.286750
nsf                   -1.7968     1.3472  -1.334 0.184813
too_exp               -0.6044     0.8070  -0.749 0.455360
dist                   1.9867     1.0710   1.855 0.066028 .
lack_doc               0.1852     0.8786   0.211 0.833414
medical_debt          -0.9576     0.6487  -1.476 0.142507
home_loan              2.7584     0.9620   2.867 0.004885 **
business_debt              NA         NA      NA       NA
save_ret              -3.2182     0.9865  -3.262 0.001437 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.92 on 121 degrees of freedom
  (4 observations deleted due to missingness)
Multiple R-squared:  0.9633,    Adjusted R-squared:  0.9578
F-statistic: 176.4 on 18 and 121 DF,  p-value: < 2.2e-16
```

```
Analysis of Variance Table

Response: check_bal
                       Df Sum Sq Mean Sq   F value    Pr(>F)
own                     1  59788   59788 2469.5544 < 2.2e-16 ***
access                  1   9193    9193  379.7084 < 2.2e-16 ***
int_purch_or_bill_pay   1   5873    5873  242.5897 < 2.2e-16 ***
int_purch               1      2       2    0.0831  0.773623
bill_pay                1   1132    1132   46.7690 3.482e-10 ***
acc_util                1     18      18    0.7361  0.392619
acc_wages               1    169     169    6.9815  0.009324 **
acc_gov_pay             1      1       1    0.0320  0.858252
debit_purch             1    174     174    7.1878  0.008364 **
cc                      1      7       7    0.2858  0.593938
prod_funds              1     25      25    1.0530  0.306870
nsf                     1     25      25    1.0421  0.309367
too_exp                 1      1       1    0.0234  0.878767
dist                    1     14      14    0.5896  0.444075
lack_doc                1      0       0    0.0016  0.967783
medical_debt            1     10      10    0.4132  0.521582
home_loan               1    160     160    6.6250  0.011262 *
save_ret                1    258     258   10.6412  0.001437 **
Residuals             121   2929      24
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Modeling Results: access ~ .

```
Call:
lm(formula = access ~ ., data = gf2017a)

Residuals:
     Min       1Q   Median       3Q      Max
-16.6139  -4.5793  -0.9327   4.5501  16.8060

Coefficients: (1 not defined because of singularities)
                      Estimate Std. Error t value Pr(>|t|)
(Intercept)            27.5724     0.9542  28.895  < 2e-16 ***
own                    14.9340     2.0082   7.436 1.44e-11 ***
check_bal               9.6599     2.8024   3.447 0.000773 ***
int_purch_or_bill_pay -12.1362     6.6592  -1.822 0.070772 .
int_purch              -0.5825     3.9976  -0.146 0.884381
bill_pay               20.6170     4.3024   4.792 4.60e-06 ***
acc_util                6.2963     2.0094   3.133 0.002153 **
acc_wages               2.1440     2.0148   1.064 0.289322
acc_gov_pay            -2.2595     1.2408  -1.821 0.070998 .
debit_purch            -8.2158     2.0957  -3.920 0.000145 ***
cc                     -1.2229     1.2981  -0.942 0.347990
prod_funds             -0.5844     0.8919  -0.655 0.513546
nsf                    11.3583     1.5662   7.252 3.77e-11 ***
too_exp                -2.0448     1.1129  -1.837 0.068529 .
dist                   -0.2684     1.4785  -0.182 0.856251
lack_doc                1.7418     1.2006   1.451 0.149336
medical_debt            0.5792     0.8921   0.649 0.517376
home_loan              -1.2044     1.3738  -0.877 0.382328
business_debt               NA         NA      NA       NA
save_ret                2.0861     1.4103   1.479 0.141611
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.889 on 125 degrees of freedom
Multiple R-squared:  0.9118,    Adjusted R-squared:  0.899
F-statistic: 71.75 on 18 and 125 DF,  p-value: < 2.2e-16
```

```
Analysis of Variance Table

Response: access
                       Df Sum Sq Mean Sq  F value    Pr(>F)
own                     1  40053   40053 843.9251 < 2.2e-16 ***
check_bal               1  12586   12586 265.1910 < 2.2e-16 ***
int_purch_or_bill_pay   1     19      19   0.4041 0.5261590
int_purch               1    362     362   7.6207 0.0066394 **
bill_pay                1   1667    1667  35.1310 2.804e-08 ***
acc_util                1    790     790  16.6517 7.949e-05 ***
acc_wages               1    460     460   9.7023 0.0022827 **
acc_gov_pay             1    720     720  15.1808 0.0001583 ***
debit_purch             1    669     669  14.0967 0.0002648 ***
cc                      1     22      22   0.4641 0.4969604
prod_funds              1    125     125   2.6409 0.1066650
nsf                     1   3361    3361  70.8191 7.648e-14 ***
too_exp                 1    142     142   3.0010 0.0856778 .
dist                    1     61      61   1.2878 0.2586268
lack_doc                1    124     124   2.6070 0.1089147
medical_debt            1      5       5   0.1125 0.7378711
home_loan               1     21      21   0.4424 0.5072060
save_ret                1    104     104   2.1879 0.1416107
Residuals             125   5933      47
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Modeling Results: check_bal ~ own + access + bill_pay

```
Call:
lm(formula = check_bal ~ own + access + bill_pay, data = gf2017)

Residuals:
     Min       1Q   Median       3Q      Max
-13.1110  -3.3508  -0.0363   2.7587  19.0788

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  24.7148     0.4568  54.107  < 2e-16 ***
own           5.4423     0.8313   6.547 1.11e-09 ***
access        2.4574     0.9820   2.503   0.0135 *
bill_pay     16.5486     1.0796  15.329  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.395 on 136 degrees of freedom
  (4 observations deleted due to missingness)
Multiple R-squared:  0.9504,    Adjusted R-squared:  0.9493
F-statistic: 868.2 on 3 and 136 DF,  p-value: < 2.2e-16
```

```
Analysis of Variance Table

Response: check_bal
           Df Sum Sq Mean Sq F value    Pr(>F)
own         1  59788   59788 2053.85 < 2.2e-16 ***
access      1   9193    9193  315.79 < 2.2e-16 ***
bill_pay    1   6840    6840  234.97 < 2.2e-16 ***
Residuals 136   3959      29
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Modeling Comparison

- Check_bal ~ .   Adjusted R-squared = 0.9578
- Access ~ .   Adjusted R-squared = 0.8990
- Check_bal ~ own + access + bill_pay   Adjusted R-squared = 0.9493
- Elimination of 16 variables results in less than 1% decrease in adjusted R-squared.
- Eliminate 16 variables as before, except use all demographic subsets for account ownership instead of the aggregate variable.

Modeling Results: check_bal ~ own(all subsets) + access + bill_pay

```
Call:
lm(formula = check_bal ~ ., data = gf2017b)

Residuals:
     Min       1Q   Median       3Q      Max
-12.6844  -3.3251  -0.2087   2.5715  18.6038

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)         24.6746     0.4668  52.863   <2e-16 ***
acc_male            -9.8008    15.9965  -0.613   0.5412
acc_in_labor        -2.3556     4.1713  -0.565   0.5733
acc_out_labor       -1.0930     3.6471  -0.300   0.7649
acc_fem             -8.5474    16.9640  -0.504   0.6152
acc_young_adult      3.4710     3.0258   1.147   0.2535
acc_older_adult     18.2119    12.0019   1.517   0.1317
acc_ed_below_hs      0.5553     1.9534   0.284   0.7767
acc_ed_hs_or_above  -0.5808     2.1905  -0.265   0.7913
acc_inc_poorest_40   5.2585    14.0030   0.376   0.7079
acc_inc_richest_60   4.9514    18.4369   0.269   0.7887
acc_rural           -4.3566     4.7174  -0.924   0.3575
access               2.1219     1.2361   1.717   0.0885 .
bill_pay            16.7333     1.2830  13.042   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.488 on 126 degrees of freedom
  (4 observations deleted due to missingness)
Multiple R-squared:  0.9524,    Adjusted R-squared:  0.9475
F-statistic:   194 on 13 and 126 DF,  p-value: < 2.2e-16
```

```
Analysis of Variance Table

Response: check_bal
                    Df Sum Sq Mean Sq   F value    Pr(>F)
acc_male             1  57335   57335 1903.4105 < 2.2e-16 ***
acc_in_labor         1     60      60    1.9957 0.1602113
acc_out_labor        1   3084    3084  102.3980 < 2.2e-16 ***
acc_fem              1    439     439   14.5857 0.0002091 ***
acc_young_adult      1    482     482   15.9894 0.0001078 ***
acc_older_adult      1     11      11    0.3677 0.5453410
acc_ed_below_hs      1    100     100    3.3049 0.0714472 .
acc_ed_hs_or_above   1    184     184    6.1225 0.0146770 *
acc_inc_poorest_40   1   1104    1104   36.6470 1.508e-08 ***
acc_inc_richest_60   1    105     105    3.4860 0.0642147 .
acc_rural            1      0       0    0.0003 0.9869065
access               1   7955    7955  264.0943 < 2.2e-16 ***
bill_pay             1   5124    5124  170.1044 < 2.2e-16 ***
Residuals          126   3795      30
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Modeling Comparison

- Check_bal ~ 20 variables
  Adjusted R-squared = 0.9578

- Check_bal ~ own + access + bill_pay Adjusted R-squared = 0.9493

- Check_bal ~ own(all subsets) + access + bill_pay Adjusted R-squared = 0.9475

- Introduction of all demographic subsets of account ownership results in a 0.0018 reduction to adjusted R-squared

- However, p-values for almost all of the coefficients for the demographic subsets of account ownership imply the coefficient should be zero. Because we are taking one significant variable and decomposing it into 10 variables, this outcome is plausible. If the number of samples were increased, it is likely that a portion of these variables would become significant.

Model Interpretations: check_bal ~ own(all subsets) + access + bill_pay

- Percentage reporting using mobile/internet to check an account balance is marginally more sensitive to male account ownership. Could be result of fewer female account holders in global data

- Similar results for in and out of labor force and richest 60 poorest 40

- Percentage reporting using mobile/internet to check an account balance is significantly more sensitive to account ownership of adults 25+. Could be the result of barriers facing young adult fin. tech. access

- The sensitivity to education level is almost exactly the same but affects balance checking in opposite directions

- Rural account owners result in a lesser percentage reporting using mobile/internet to check an account balance

Coefficients:

	Estimate
(Intercept)	24.6746
acc_male	-9.8008
acc_in_labor	-2.3556
acc_out_labor	-1.0930
acc_fem	-8.5474
acc_young_adult	3.4710
acc_older_adult	18.2119
acc_ed_below_hs	0.5553
acc_ed_hs_or_above	-0.5808
acc_inc_poorest_40	5.2585
acc_inc_richest_60	4.9514
acc_rural	-4.3566
access	2.1219
bill_pay	16.7333

Model Interpretations - check balances

- Positive relationships include access (HS), Internet purchases(HS), bill pay (HS),debit purchases (S), home loans (HS)

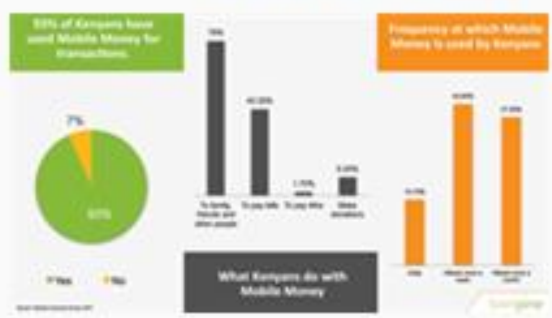
- Negative relationships include saving for retirement (HS).

Model Interpretations - check balances

- Positive relationships include own (HS), check balances (HS), bill pay (HS), account utilities (HS), debit purchases (HS), NSF (HS).
- Negative relationships include debt purchases (HS), too expensive (MS).

Business/Policy Recommendations via Check Balances

- Emphasis on routine/short-term purchases and NSF protection.
- Less emphasis on long-term planning and retirement (if an initiative, much re-education to do).



REFERENCE

EY FinTech Adoption Index 2017: The rapid emergence of FinTech (2017). {Online]. Retrieved https://www.ey.com/Publication/vwLUAssets/ey-fintech-adoption-index-2017/$FILE/ey-fintech-adoption-index-2017.pdf

Track: Analytics, Technology, The Internet of Things

ID#: 1423