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**


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 use 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
• Assess 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
13. Outstanding housing loan
14. Borrowed for business venture
15. Saving for old age

Negative Group
13. Insufficient funds
14. Banking services too 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

Four Types of Analytics

- Descriptive Analytics
- Predictive Analytics
- Diagnostic Analytics
- Prescriptive Analytics
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
-13.110  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

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

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

---

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: \( \text{check\_bal} \sim \text{own(all subsets)} + \text{access} + \text{bill\_pay} \)

**Analysis of Variance Table**

<table>
<thead>
<tr>
<th></th>
<th>Df</th>
<th>Sum Sq</th>
<th>Mean Sq</th>
<th>F value</th>
<th>Pr(&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>check_bal</td>
<td>1</td>
<td>57335</td>
<td>57335</td>
<td>1903.4105</td>
<td>&lt; 2.2e-16 ***</td>
</tr>
<tr>
<td>acc_male</td>
<td>1</td>
<td>60</td>
<td>60</td>
<td>1.9957</td>
<td>0.1602113</td>
</tr>
<tr>
<td>acc_out_labor</td>
<td>1</td>
<td>3084</td>
<td>3084</td>
<td>102.3900</td>
<td>&lt; 2.2e-16 ***</td>
</tr>
<tr>
<td>acc_fem</td>
<td>1</td>
<td>439</td>
<td>439</td>
<td>11.4527</td>
<td>0.0012991 ***</td>
</tr>
<tr>
<td>acc_young_adult</td>
<td>1</td>
<td>482</td>
<td>482</td>
<td>15.9894</td>
<td>0.0001078 ***</td>
</tr>
<tr>
<td>acc_older_adult</td>
<td>1</td>
<td>11</td>
<td>11</td>
<td>0.3677</td>
<td>0.5435140</td>
</tr>
<tr>
<td>acc_inc_poorest_40</td>
<td>1</td>
<td>100</td>
<td>100</td>
<td>3.3049</td>
<td>0.0734472</td>
</tr>
<tr>
<td>acc_inc_below_hs</td>
<td>1</td>
<td>184</td>
<td>184</td>
<td>6.1225</td>
<td>0.0146770 *</td>
</tr>
<tr>
<td>acc_inc_poorest_60</td>
<td>1</td>
<td>1104</td>
<td>1104</td>
<td>36.6470</td>
<td>1.5086-08 ***</td>
</tr>
<tr>
<td>acc_inc_richest_60</td>
<td>1</td>
<td>105</td>
<td>105</td>
<td>3.4860</td>
<td>0.0624176</td>
</tr>
<tr>
<td>acc_rural</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0.0003</td>
<td>0.9880065</td>
</tr>
<tr>
<td>access</td>
<td>1</td>
<td>7955</td>
<td>7955</td>
<td>264.0493</td>
<td>&lt; 2.2e-16 ***</td>
</tr>
<tr>
<td>bill_pay</td>
<td>1</td>
<td>5124</td>
<td>5124</td>
<td>170.1044</td>
<td>&lt; 2.2e-16 ***</td>
</tr>
<tr>
<td>Residuals</td>
<td>126</td>
<td>3795</td>
<td>30</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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

---

**Modeling Comparison**

- \( \text{check\_bal} \sim 20 \text{ variables} \)
  
  Adjusted R-squared = 0.9578

- \( \text{check\_bal} \sim \text{own} + \text{access} + \text{bill\_pay} \)
  
  Adjusted R-squared = 0.9493

- \( \text{check\_bal} \sim \text{own(all subsets)} + \text{access} + \text{bill\_pay} \)
  
  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.

<table>
<thead>
<tr>
<th>Coefficients</th>
<th>Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>24.6746</td>
</tr>
<tr>
<td>acc_male</td>
<td>-9.8008</td>
</tr>
<tr>
<td>acc_in_labor</td>
<td>-2.3556</td>
</tr>
<tr>
<td>acc_out_labor</td>
<td>-1.0930</td>
</tr>
<tr>
<td>acc_fem</td>
<td>-8.5474</td>
</tr>
<tr>
<td>acc_young_adult</td>
<td>3.4710</td>
</tr>
<tr>
<td>acc_older_adult</td>
<td>18.2119</td>
</tr>
<tr>
<td>acc_ed_below_hs</td>
<td>0.5553</td>
</tr>
<tr>
<td>acc_ed_hs_or_above</td>
<td>-0.5808</td>
</tr>
<tr>
<td>acc_inc_poorest_40</td>
<td>5.2585</td>
</tr>
<tr>
<td>acc_inc_richest_60</td>
<td>4.9514</td>
</tr>
<tr>
<td>acc_rural</td>
<td>-4.3566</td>
</tr>
<tr>
<td>access</td>
<td>2.1219</td>
</tr>
<tr>
<td>bill_pay</td>
<td>16.7333</td>
</tr>
</tbody>
</table>

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

Track: Analytics, Technology, The Internet of Things

ID#: 1423