The effect of literacy and bank penetration on Financial Inclusion in India: A statistical analysis

by

Santanu Dutta  and Pinky Dutta

Abstract : In this paper quantile regression model is used to test the significance of the impact of literacy percentage and number of bank offices (per 10^5 people) on the conditional percentiles of the number of accounts per hundred adults, in 35 states and union territories in India. The latter is a measure of financial inclusion in a region.

We find that bank penetration, measured by number of bank offices (per 10^5 people), has more significant impact on the level of financial inclusion, than literacy percentage. However literacy has significant positive impact on the higher (80 or above) percentiles of the conditional distribution of the number of accounts per hundred adults in a region, with a given level of literacy. If a state/union territory is among the top 20 percent states/union territories in terms of financial inclusion (with a given literacy percentage), then increase in literacy percentage can further improve the financial inclusion in that region. However literacy alone cannot improve the level of financial inclusion significantly, especially if the state/union territory is not among the top 30 percent states/union territories in terms of financial inclusion with similar literacy percentage. In those regions, increase in bank penetration is likely to improve the level of financial inclusion significantly rather than improving literacy percentage.

Key words: Financial inclusion in India, literacy, bank penetration, quantile regression.

Authors: Santanu Dutta, Assistant Professor, Dept. of Mathematical Sciences, Tezpur University, India.
Email: tezpur1976@gmail.com
Pinky Dutta, Research Scholar, Dept. of Business Administration, Tezpur University, India.
Email: dutta.pinky1@gmail.com

Editor: Subrata Chakraborty, subrata_arya@yahoo.co.in

1. Introduction

Financial inclusion can be defined as easy access to formal financial services or systems and their usage by all members of the economy. The committee on financial inclusion, of government of India, has defined financial inclusion as the process of ensuring timely access to financial services and adequate credit where needed by vulnerable groups such as the weaker sections and low income groups at an affordable cost (Rangarajan Committee, 2008). The process of financial inclusion consists of ensuring bank accounts to each household and offering
their inclusion in the banking system (Reddy, 2007). Access to financial services promotes social inclusion, and builds self-confidence and empowerment. In an address Dr. K. C. Chakrabarty, Deputy Governor, Reserve Bank of India at the National Finance Conclave 2010, has mentioned that financial inclusion is no longer a policy choice but it is a policy compulsion today. And banking is a key driver for inclusive growth.

There are various socio-cultural, economic issues that hinder the process of financial inclusion. For instance on demand side, it includes lack of awareness and illiteracy (see Throat, 2007). From supply side, lack of avenues for investment (such as poor bank penetration), unwillingness of banks to do financial inclusion or high cost involved in financial inclusion seem to be some likely reasons for financial exclusion. However deputy governor of RBI has recently clarified that the latter two reasons are myths, i.e. the cost involved in financial inclusion is not unbearable by the banks and that it is not true that the banks are unwilling to do financial inclusion (see Chakrabarty (2010)).

So it seems that literacy (on demand side) and bank penetration (on the supply side) are two important factors that effect the level of financial inclusion in a region. A common measure of financial inclusion is the number of accounts per 100 adults (above 19 years) in a region (see Throat, 2007). This due to the fact that holding a bank account (savings/credit/fixed deposit or recurring) itself confers a sense of identity, status and empowerment and provides access to the national payment system (RBI, 2008). A measure of bank penetration is branch density, which is the number of bank offices per 1,00,000 people (see Chakrabarty (2010)).

1.1 Objective: In this paper we study the effect on literacy percentage and branch density on financial inclusion in 35 states and union territories in India. We use quantile regression to evaluate the significance of the impact of these two factors. The motivation and the advantages of quantile regression are discussed later.

1.2 Data: The data on adult population (above 19 years) and literacy percentage for each state/union territory is obtained from 2001 census report. We consider adult population as only an adult can have a bank account. Data on the number of bank offices and total number of deposit (savings/current/term deposit) accounts for each state/union territory are obtained from the Reserve Bank of India’s BSR report 2009.
1.3 **Literature review:** Sangwan (2008) carried out an extensive study to estimate the relationship between the percentage of adults having saving and credit accounts with the branch density, the literacy percentage and some other factors, for 35 states/union territories in India. He observed that there is a positive significant impact of branch density, and negative but insignificant impact of literacy on financial inclusion. The latter observation is somewhat surprising, especially because Sarma and Pais (2010) have reported that literacy is positively and significantly associated with financial inclusion.

Sangwan (2008) used a simple linear OLS regression model to study the effects of literacy and bank density on financial inclusion. OLS regression models the effect of covariates on the average value of the conditional distribution of the dependent variable. A particular covariate, such as literacy percentage, may not have significant impact on the average value of the conditional distribution of the dependent variable (viz. percentage of adults having bank account), but it may have significant impact on extreme quantiles (such as the 10th or 90th percentiles) of the conditional distribution. Quantile regression, introduced by Koenker and Bassett (1978), models the effect of covariates on various quantiles (such as percentiles) of the conditional distribution of the dependent variable. In this paper we use quantile regression model to investigate the extent to which literacy and bank penetration effect ten different (from 5th to 95th) percentiles of the conditional distribution of the number of accounts per 100 adults in a state/union territory, given a fixed level of literacy and/or bank penetration.

1.4 **Summary of the paper:** In the next section we provide an overview of quantile regression. Finally we present data analysis. We consider three models, namely

- The model I explaining the marginal effect of literacy percentage on various conditional percentiles of number of accounts per 100 adults.
- The model II for marginal effect of branch density on the conditional percentiles accounts per 100 adults.
- The model III for joint effect of the two covariates on the conditional percentiles.

For each model the coefficients are computed and tested using R software package. From the 1st model we find that literacy percentage has significant positive impact on the higher conditional percentiles of the number of accounts per 100 adults, i.e. if there exists a healthy
level of financial inclusion in a state, then it can be improved further by improving the literacy percentage. From 2\textsuperscript{nd} and 3\textsuperscript{rd} model we see that branch density has significant impact on financial inclusion, and that literacy percentage cannot alone improve the level of financial inclusion in a state without significantly improving branch density.

2. Methodology & Data analysis

2.1 Methodology:  Ordinary least-squares regression models the relationship between one or more covariates \(X\) and the conditional mean of the response variable \(Y\), given \(X = x\). Quantile regression extends the regression model to conditional quantiles of the response variable, such as the 90th percentile. Quantile regression is especially useful in applications where extremes are important. It provides a more complete picture of the conditional distribution of \(Y\), given \(X = x\) (see Chen (2005)). The \(\theta\)th quantile of \(Y\), given \(X = x\), is equal to \(F^{-1}(\theta)\), where \(F\) is the distribution function of the conditional distribution of \(Y\), given \(X = x\).

In the \(\theta\)th (linear) quantile regression model, \(F^{-1}(\theta)\) is expressed as a linear combination of the values of the covariates (see, Koenker and Bassett (1978), Chen (2005), Kuan (2007) and references there in). In our application \(Y\) represents the number of account holders per 100 adults, in a state.

The higher percentiles of \(Y\) represent the states where this number is high, and the lower percentiles represent the states where number of accounts holders per 100 adults is low. We study the impact of literacy percentage \(X_1\) and branch density \(X_2\) on different percentiles of \(Y\). We consider three classes of models. They are described below.

Model I: \[ Y = \alpha_1 + \beta_1 X_1 + \epsilon_1 \] (Effect of literacy on financial inclusion)

The \(\theta\) the quantile of the conditional distribution of \(Y\), given \(X_1 = x_1\), equals \(\alpha_1 + \beta_1 x_1\) and \(\epsilon_1\) is the error component. We test, \(H_0: \beta_1 = 0\), against a two sided alternative hypothesis.

Model II: \[ Y = \alpha_2 + \beta_2 X_2 + \epsilon_2 \] (Effect of branch density on financial inclusion)

The \(\theta\) the quantile of the conditional distribution of \(Y\), given \(X_2 = x_2\), equals \(\alpha_2 + \beta_2 x_2\) and \(\epsilon_2\) is the error component. We test, \(H_0: \beta_2 = 0\), against a two sided alternative hypothesis.
Model III: \( Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \epsilon \) (Joint effect of literacy and branch density on financial inclusion)

The \( \theta \) the quantile of the conditional distribution of \( Y \), given \( X_1 = x_1 \) and \( X_2 = x_2 \), equals \( \alpha + \beta_1 x_1 + \beta_2 x_2 \) and \( \epsilon \) is the error component. We test, \( H_0 : \beta_i \neq 0 \) against \( H_1 : \beta_i = 0 \), \( i=1,2 \).

Note: The details of the model assumptions, the methodology for estimation of \( \beta_1 \), \( \beta_2 \) and testing of hypotheses (on these parameters) are available in Kuan (2007) and references there in. We skip the technical details, and emphasize only on the computational aspects using the free software R for statistical computing. In this paper \( \theta \) is expressed in percentage.

If \( \beta_i \) equals zero (or significantly close to zero) then \( X_i \) has no impact (or no significant impact) on the \( \theta \)th percentile of the condition distribution of \( Y \), given the covariates, for \( i=1,2 \). So in each one of the above models, we compute the values of \( \beta_i \) and the p-value of the test \( H_0 : \beta_i \neq 0 \) against \( H_1 : \beta_i = 0 \), \( i=1,2 \). These values have been computed using the \( rq \) function in \( quantreg \) package in R (version 2.10.1). p-value larger than level of significance implies accepting \( H_0 \).

2.2 Data Analysis: Our data consists of the values of literacy percentage \( X_1 \), branch density \( X_2 \) and \( Y \) which denotes the number of accounts per 100 adults, for 35 states and union territories in India. These figures are obtained RBI report on `basic statistical returns of scheduled commercial banks in India”, volume 38, 2009, and the 2001 census report.

In Table 1, 2, 3 we report the estimated values of \( \beta_i \) and p-values of the test

\( H_0 : \beta_i \neq 0 \) against \( H_1 : \beta_i = 0 \), \( i=1,2 \), for Model I, II and III respectively, Corresponding to \( \theta = 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9 \) and 0.95.

Table 1: Values of \( \beta_i \) and the p-values, for different percentiles, in Model I

<table>
<thead>
<tr>
<th>( \theta )</th>
<th>0.05</th>
<th>0.10</th>
<th>0.20</th>
<th>0.30</th>
<th>0.40</th>
<th>0.50</th>
<th>0.60</th>
<th>0.70</th>
<th>0.80</th>
<th>0.90</th>
<th>0.95</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \beta_1 )</td>
<td>0.54</td>
<td>0.09</td>
<td>1.07</td>
<td>1.03</td>
<td>1.21</td>
<td>1.18</td>
<td>1.9</td>
<td>2.57</td>
<td>3.6</td>
<td>6.12</td>
<td>6.05</td>
</tr>
<tr>
<td>p-value</td>
<td>0.62</td>
<td>0.82</td>
<td>0.15</td>
<td>0.10</td>
<td>0.06</td>
<td>0.20</td>
<td>0.05</td>
<td>0.06</td>
<td>0.02</td>
<td>0.002</td>
<td>0.0001</td>
</tr>
</tbody>
</table>

Table 2: Values of \( \beta_2 \) and the p-values, for different percentiles, in Model II

<table>
<thead>
<tr>
<th>( \theta )</th>
<th>0.05</th>
<th>0.1</th>
<th>0.2</th>
<th>0.3</th>
<th>0.4</th>
<th>0.5</th>
<th>0.6</th>
<th>0.7</th>
<th>0.8</th>
<th>0.9</th>
<th>0.95</th>
</tr>
</thead>
<tbody>
<tr>
<td>p-value</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 3: Values of $\beta_i$, $i=1, 2$ and the p-values for, for different percentiles, in Model III

<table>
<thead>
<tr>
<th>$\theta$</th>
<th>0.05</th>
<th>0.10</th>
<th>0.20</th>
<th>0.30</th>
<th>0.40</th>
<th>0.50</th>
<th>0.60</th>
<th>0.07</th>
<th>0.08</th>
<th>0.09</th>
<th>0.95</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_1$</td>
<td>-0.36</td>
<td>-0.54</td>
<td>-0.52</td>
<td>-0.25</td>
<td>-0.11</td>
<td>-0.12</td>
<td>0.19</td>
<td>0.48</td>
<td>0.22</td>
<td>0.31</td>
<td>1.14</td>
</tr>
<tr>
<td>p-value</td>
<td>0.59</td>
<td>0.34</td>
<td>0.40</td>
<td>0.63</td>
<td>0.83</td>
<td>0.79</td>
<td>0.71</td>
<td>0.43</td>
<td>0.77</td>
<td>0.78</td>
<td>0.29</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>0.37</td>
<td>0.44</td>
<td>0.61</td>
<td>0.76</td>
<td>0.75</td>
<td>0.78</td>
<td>0.78</td>
<td>0.74</td>
<td>0.75</td>
<td>1.07</td>
<td>0.89</td>
</tr>
<tr>
<td>p-value</td>
<td>0.07</td>
<td>0.03</td>
<td>0.004</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00009</td>
<td>0.00007</td>
</tr>
</tbody>
</table>

Observations and their interpretation:

1. From Table 1, we see that the values of $\beta_1$ (in Model I) are positive and increase rapidly as $\theta$ is increased. The p-values decrease drastically as $\theta$ is increased. In fact, the p-values are clearly less than 0.05 for $\theta$ exceeding 70 percent. So there is significant positive effect (at 5 percent level of significance) of literacy percentage $X_1$ on the 80th, 90th and 95th percentiles of the conditional distribution of number of accounts (per 100 adults) $Y$, given $X_1$. However, the effect of $X_1$ on the lower percentiles is insignificant.

   This observation has a simple interpretation. Suppose we rank the states/union territories, with similar literacy percentage, in increasing order of the number of accounts per 100 adults. Then if a state/union territory is among the top 20 percent in that list, then increase in literacy percentage can help to further improve the financial inclusion in that state/union territory. But increase in literacy percentage does not necessarily imply significant improvement in financial inclusion in every state/union territory, especially if it is not among the top 20 percent states/union territories in terms of financial inclusion (with similar literacy percentage).

   To investigate whether increase in literacy percentage will have any impact on financial inclusion in Assam, we rank Assam and six other states (viz. Nagaland,
Meghalaya, Madhya Pradesh, Orissa, Rajasthan and Andhra Pradesh), with similar literacy percentage (60 to 66 percent), in increasing order of the number of accounts per 100 adults in these states. Assam is not among the top 20 percent in this list, so increase in literacy percentage will not have significant impact on financial inclusion in Assam. Similar analysis can be performed for other states as well.

2. From Table 2, we see that the values of $\beta_2$ (in Model II) are positive and the p-values less than 0.05, for $\theta$ exceeding 5 percent. So clearly the branch density $X_2$ has significant impact on almost all the percentiles of the conditional distribution of $Y$, given $X_2$.

This observation implies that increase in branch density helps to significantly improve financial inclusion in a state. Therefore, branch density is a key factor in improving financial inclusion in a state.

3. In Model III we study the joint effect of literacy percentage and branch density on financial inclusion. From Table 3 we see that the p-values, of the test $H_0: \beta_1=0$, exceed 0.10 for all values of $\theta$. In contrast, the p-values of the test $H_0: \beta_2=0$ are less than 0.05 for all $\theta$, exceeding 0.05. So $X_2$ has significant impact (at 5 percent level of significance) on almost all the percentiles of the conditional distribution of $Y$, given $X_1$ and $X_2$. In contrast $X_1$ has no significant impact (at even 10 percent level of significance) on any quantile of the conditional distribution of $Y$, given $X_1$ and $X_2$.

The above observations imply that among the two covariates, the branch density is the most significant factor which effects the level of financial inclusion in a state. This supports the findings in Sangwan (2008).

**Conclusion:** Branch density in a state measures the opportunity for financial inclusion in that state. Literacy is a prerequisite for creating investment awareness, and hence intuitively it seems to be a key tool for financial inclusion. But the above observations imply that literacy alone cannot guarantee high level financial inclusion in a state. Branch density has significant impact on financial inclusion. It is not possible to achieve financial inclusion only by creating investment awareness, without significantly improving the investment opportunities in a state. However if a state/union territory is among the top 20 percent in terms of financial
inclusion, with a given level of literacy. Then increase in literacy percentage can further improve the level of financial inclusion in that region. Our data analysis provides support to these claims.

References:
