KENYA BANKERS
ASSOCIATION
One Industry. Transforming Kenya.

WPS/02/18

# Credit Risk Analysis for Low Income Earners

Davis Bundi Ntwiga

KBA Centre for Research on Financial Markets and Policy®
Working Paper Series

24

KENYA BANKERS
ASSOCIATION
One Industry. Transforming Kenya.

# Working Paper Series

## Centre for Research on Financial Markets and Policy

The Centre for Research on Financial Markets and Policy® was established by the Kenya Bankers Association in 2012 to offer an array of research, commentary, and dialogue regarding critical policy matters that impact on financial markets in Kenya. The Centre sponsors original research, provides thoughtful commentary, and hosts dialogues and conferences involving scholars and practitioners on key financial market issues. Through these activities, the Centre acts as a platform for intellectual engagement and dialogue between financial market experts, the banking sector and the policy makers in Kenya. It therefore contributes to an informed discussion that influences critical financial market debates and policies.

The Kenya Bankers Association (KBA) *Working Papers Series* disseminates research findings of studies conducted by the KBA Centre for Research on Financial Markets and Policy. *The Working Papers* constitute "work in progress" and are published to stimulate discussion and contribute to the advancement of the banking industry's knowledge of matters of markets, economic outcomes and policy. Constructive feedback on the *Working Papers* is welcome. *The Working Papers* are published in the names of the author(s). Therefore their views do not necessarily represent those of the KBA.

# Credit Risk Analysis for Low Income Earners

By Davis Bundi Ntwiga*

August 2018

## Abstract

*The low income earners have been excluded from financial services due to their limited ability to access credit as they lack good financial options. Their income is volatile, fluctuating daily and they lack reliable ways to harness the power of their low incomes. This challenge can be met through use of mobile technology to collect data on the socioeconomic activities of the low income earners at minimum costs. The success of the mobile financial services in Kenya cannot be understated coupled with increase in mobile penetration. A mobile based technology for micro-credit already exists through M-Shwari started in the year 2012 on the M-pesa platform to further increase financial inclusion. This paper pro- poses a decision support system that is mobile based for credit scoring, classification and peer group lending of the low income earners in Kenya. This facility is referred to as Mobile Micro Credit System. We hypothesize that, first, mobile micro-credit lending for low income peer groups is similar to that of the conventional individual lending. Second, credit scores and credit quality levels among low income men and women is the same. Third, a mobile based micro credit can further enhance financial inclusion among the low income earners. A comparison is made between the peer groups and individual customers in terms of their credit scores and credit quality levels. The data for the study was extracted from the financial diaries dataset by Financial Sector Deepening. Customers were clustered in peer groups and as individual households based on gender groupings of men only, women only, or men and women. The credit scoring factors were estimated and resulting data trained the hidden Markov model, the classification technique used. The hidden Markov model emit- ted the credit scores and the credit quality levels of the individual households and the peer groups in which the groups had stronger credit scores and credit quality levels compared to those of the individual households. Peer groups for women only, and those of men and women had superior credit scores when compared to men only peer groups. The* *optimal peer group size for lending is between four and eight members. The current mobile financial services offer a baseline to implement a mobile micro credit service for the low income earners. This is an incentive for financial services providers to consider providing mobile based micro credit loans to low income earners.*

***Keywords:*** *Peer groups, credit score, hidden Markov model, mobile micro credit system, men, women, credit quality and low income earners*

* *Davis Bundi is affiliated with School of Mathematics, University of Nairobi*
*dbundi@uonbi.ac.ke*

# 1.0 Introduction

**In November 2012, Safaricom Kenya and the Commercial Bank of Africa (a registered commercial bank in Kenya) started a mobile based Micro-credit facility dubbed M-shwari ("Shwari" means calm in Swahili) based on the M-Pesa platform. M-Pesa has become one of the most successful mobile phone based financial services in the developing world (Jack and Suri, 2011; Mbiti and Weil, 2011).**

The product offers a combination of savings and access to micro-credit loan and the target market was to capture both the banked and unbanked in an effort to increase financial inclusion (Cook and Mckay, 2015). The features of M-Shwari are: allows customers to save and access credit; the cost of moving money between M-Pesa and M-Shwari is free; account maintenance is free; no transfer of funds between M-Shwari and a bank account, only between M-Shwari and M-pesa, then M-pesa to a bank account; and credit scoring algorithm used for accessing a loan from M-Shwari is based on a set of telecommunication variables from Safaricom data related to airtime, M-pesa transactions, airtime credit and length of time as a customer (Cook and Mckay, 2015). Financial inclusion is typically defined as the proportion of individuals that use financial services. There is a heightened interest on the importance of financial inclusion for the economic and social development.

The features highlighted by Cook and Mckay (2015) limit the ability of the low income earners to access the micro-credit facility due to limited resources. The low income market segment that is excluded from financial inclusion lack good financial options. Their income is volatile, fluctuating daily and they lack reliable ways to harness the power of their low income (FSD, 2014a). Most likely, they are also excluded from M-Shwari products. The low and volatile income can offer opportunities to find innovative ways to lend to these groups. Peer groups can be formed that are mobile based for micro-credit lending. A study by FSD (2014b) observed that compared to men, women's financial aspirations and behaviors are heavily colored by social and cultural norms. Serving women requires serving individuals with particularly low and inconsistent incomes when compared to men. The low income refers to a large section of households in Kenya with extremely low level of income, and therefore un-served by financial institutions.

The World Bank Findex data reveal that in the year 2015, around 55 percent of the global unbanked was women. In the developing countries, 59 percent of men have an account compared with 50 percent of women (FSD, 2014b). This gap can be closed with imaginative services design and delivery that account for the economic and social realities that are common among both men and women but with special emphasis to women, who are more marginalized in formal financial services (FSD, 2013; FSD, 2014b; FSD, 2016). The need to capture the low income earners is because poor people are normally more highly motivated in managing their limited resources. They save, invest and even re-invest as a means to lower the fluctuations in their income and try to maximize on the power of the low incomes. Women tend to operate in horizontal social networks that increases their social collateral and this lowers moral hazard in the peer group lending setups they form (Simtowe, Zeller and Phiri, 2006; FSD, 2014b). The value of information of each other in a peer group improves the group risk mitigation. Peer groupings in credit lending acts as social collateral (Holmstrom, 1979).

Low income earners are more inclined to financial service providers with whom they can develop long term relationships. A provider can attract this group of customers with products that offers interdependence, uplifts and values the support system common among the low income people (FSD, 2014d). Loan products are designed in a manner that the lender and borrower are willing to adapt to the unpredictable circumstances in terms of repayment due to the shocks low income earners experience on daily basis. Thus, products for low income earners must resonate with the local systems for resource exchange and upliftment (FSD, 2014d).

In this paper, we propose a decision support system for a mobile micro credit system (MMCS). The system is based on hidden Markov model (HMM) technique to estimate the credit risk profiles of peer groups formed by the low income earners to save and borrow from the MMCS. The credit scores and the credit quality levels form the credit risk profile. The low income earners are grouped based on gender as men and women, men only and women only.

The Kenya financial diaries (KFD) project data of the year 2012-2013 formed the data for this study. We selected 14 variables of the respondents data on age, gender, marital status, working age, household size, education, income, expenditure, savings, income variability, social network strength, resources given and received, mobile money flows and mobile money transactions are used for HMM learning and training. The hidden Markov model is a classification technique that can be utilized to build a classification model to classify the customers into different risk groups (Benyacoub, Bernoussi and Zoglat, 2014).

The study is significant for financial institutions as they can loop in new customer segments, expand market share, better price risk and improve the credit risk management processes. The financial market regulators can enhance and increase financial inclusion in Kenya. In the academic sector, there is improvement in credit risk models with an increase in research output in consumer credit risk.

The purpose of the decision support system is to encourage financial services providers to introduce peer group saving and lending. This is to offer low income earners through peer groups, a micro credit facility that is mobile based with an aim of reducing financial exclusion in Kenya from the current rate of 17.4 percent adults (FSD, 2016). The success of the mobile financial services, low cost of collecting data using mobile technology, a need to further increase financial inclusion, availability of advanced data mining techniques and a wide array of data on consumers, and to offer an alternative financial services to the low income earners in Kenya are the key motivators for this study. The hypotheses that capture the objectives of the study are:

$H_{1a}$: Peer groups and individuals credit scores are the same among the low income earners.

$H_{1b}$: Credit scores and credit quality levels among low income men and women is the same.

$H_{1c}$: A mobile based micro-credit system can further enhance financial inclusion among the low income earners.

## 1.1 Assumptions

No mathematical system or model is perfect as these models only depict those characteristics of direct interest to the researcher. First, the past financial transactions of the low income earners reflect their future activities as the past can represent the future. The true default is elusive as actual default estimation depends on the information that is available.

Second, the hidden Markov model has the ability to dynamically track changes in the credit scores of the low income earners. Modeling default events needs a dynamic process like HMM, a stochastic process to estimate when the default occur. Third, the macro economic factors are not captured in the model.

We note that economic conditions are not the only cause of change in credit risk as massive increase in defaults and bankruptcy are observable even in good economic conditions. Fourth, most credit risk analysis are done with the five C's of credit analysis (capital, collateral, conditions, character and capacity) but not the case in this study. Fifth, the model can use time dependent data to overcome the challenge of model deterioration due to use of historical data with the variables in the study assumed to be ideal for analysis and credit scoring of the low income earners.

# 2.0 Literature Review

**The ability to access micro credit loan promotes individual outcomes (FSD, 2010). As lending is at the heart of an economy financial architecture, M-shwari innovation offers mobile micro- credit to middle income earners (Deesai, 2012).**

The perceptions about mobile banking and technology determine the rate of adoption as income alone is not a sufficient indicator (Ivatury and Pickens, 2006). A study by FSD (2016) observed that 17.4 percent of the adult population in Kenya does not have access to any form of financial services. An individual is more likely to use a mobile phone followed by an informal group (known as *chamas* in Swahili), a bank, Savings and credit organizations and then a Micro-Finance institution for financial services. Thus, the use of mobile phone is more prominent in Kenya for access of financial services.

A need exists on how the low income segment of the market can access financial services as this enhances financial inclusion. Studies on micro-credit facilities observe that a positive change in income is observable when the poor use these facilities (Ahlin and Jiang, 2008; Chavan and Ramakumar, 2002). A key pointer is the poor to save and access credit as this brings an economy from stagnation to full development. The ability to accumulate savings is critical to assist the low income earners benefit from micro-credit program (Ahlin and Jiang, 2008). Women have a higher rate of using informal financial services at 51.4 percent compared to men at 30.9 percent while in using mobile financial services, the difference between men (75.5 percent) and women (67.5 percent) is 8 percent (FSD, 2016). This shows that women are higher consumers of informal financial services as their networks tends to be horizontal when compared to men whose social networks are vertical (FSD, 2014b). The gap of lending to low income earners can be reduced if the social collateral possessed by women can help to reduce the moral hazard. Peer groups that are mobile based can be formed as moral hazard is lower in groups that are formed endogenously through peer selection (Simtowe *et al*., 2006).

The FinAccess Household Survey noted the main reasons people join a group are: access to a lump sum for emergencies, daily needs, social reasons, keep money safe, acquire a lump sum for investment and commitment to save (FSD, 2016). This compares well to the financial diaries study in which it indicates the implications of the findings to the financial services providers as: they should provide products that cater to small and inconsistent incomes; offer better tools for managing day to day transactions and risks; assist women better leverage their social networks; accessibility; and services that endure and support women to face major life transitions (FSD, 2014b).

HMM is the classification method in this study and the technique is increasingly becoming popular due to its strong mathematical structure and theoretical basis for wide applications. Rabiner (1989) offers an introduction to HMM as a generalization of a mixture model with the hidden variables and observed emissions. Hassan and Nath (2005) outlines some advantages of HMM from its strong statistical foundation, ability to handle new data robustly, predict similar patterns efficiently, computationally efficient to develop and evaluate due to availability of training algorithms. Bilmes (2006) observes that HMM have no theoretical limitations given enough hidden and observation distributions and sufficient training data. The applications of HMM for solving different problems are available. Netzer and Srinivasan (2008) developed a customer relationship dynamics model to estimate the effects of encounters between the customer and the firm, that is, the customer firm relationship and customer's choice behaviour. A model for the default rates in a bond portfolio using HMM is undertaken (Crowder, Davis and Giampieri, 2005). A model for credit card transaction processing sequence as a stochastic process using HMM shows a drastic reduction in the number of false positive transactions (Srivastava, Kundu, Sural and Majumdar, 2008). The HMM is dynamic in observing a sudden downgrading of a customer credit worthiness. Quirini and Vannucci (2014) note that HMM and related tools are essential to assess the credit risk in order to gain profitability in the complex and fluctuating credit market.

A model based on HMM by Hassan and Nath (2005) for the stock market forecasting showed a 100 percent accuracy rate in the prediction. Ntwiga (2016) used social network (media) data to credit score obligors and estimate default rates in a loan portfolio using HMM. The model had an accuracy rate of between 53 percent and 73 percent and offers promising prospects of using social data to improve on credit risk modeling. The work of Daniel and Grissen (2015) did not use HMM but the mobile phone usage data to predict loan repayment in a developing country. The behavioural signatures in mobile phone data are able to predict default with accuracy than the approach of credit scoring using financial histories. If the data is used, the bank can reduce defaults by 41 percent while still accepting 75 percent of the borrowers.

Researchers have considered the influence of socio-demographic characteristics on credit risk of consumers. The male customers have a higher probability of default compared to the female clients, but the recovery rates were found to be similar between the male and female customers (Marrez and Schmit, 2009). For the age factor, the age group of 20 - 25 year olds are more likely to default compared to the group of 61 - 70 year olds. The married defaults in 24 percent of the cases while the single customers tend to default in 36 percent of the cases when marital status was considered.

The work of Marrez and Schmit (2009) further considers the age, gender and marital status among other variables. The young cardholders have higher chance of default compared to the older cardholders. Gender and marital status are statistically significant in affecting default rates for credit card holders. Formal financial inclusion increases significantly with the level of education with 73 percent for those with primary education and 98 percent for those with tertiary education (FSD, 2016).

The socio-demographics are important variables in this study but we also wish to compare the credit scores of the peer group borrowers and individual borrowers for optimality. Bhole and Ogden (2010) observe that group lending yields a higher safeguard to the financial institution compared to individual lending. This was compared with changes in social sanctions or cross reporting among the group members. Gomez and Santor (2003) noted that group lending outperforms conventional individual lending. The duo mentioned that for a peer lending group to be effective, the size of the loan, level of trust and enforcement of social norms with the group or surrounding neighbourhood are important. Ahlin and Jiang (2008) considers the group size and noted that as group size increases, efficiency reaches a moderate threshold and recommends a group of four to ten members in group borrowing.

# 3.0 Methodology

**This section highlights the hidden Markov model classification method, credit scoring analysis, the peer group dynamics, individual household data and the conceptual framework to conceptualize the key variables of the study and how these are interlinked to achieve the objectives stated.**

The data used in this study was extracted from the Kenya Financial Diaries (KFD) project of the year 2012 - 2013 by the Financial Sector Deepening Kenya (FSDK). The KFD project was designed to improve the understanding of how low income earners in Kenya earn and spend money, which offered new and deeper insights on their financial lives (FSD, 2014c). The study covered 300 households from 14 communities in 5 areas: Eldoret, Makueni, Mombasa, Nairobi and Vihiga. It recorded in great detail how they earn and spend their money, share resources and utilize financial devices available in bad and good times.

Sampling during KFD study was purposive and thus limited in making a generalization about the Kenya as a whole but offers deep insights on the financial lives of the low income earners in Kenya. The sample mix was 60 - 40 rural-urban. The data assists to generate hypotheses difficult to generate from larger quantitative surveys. The diaries questionnaires and interviews were designed to capture information in nine broad domains: well being, household flows, households own production and consumption of food, cash on hand, flows of money, outflows of money, flows of money associated with financial devices, and major events in life. See the Financial diaries study data set, general guidelines and related literature (FSD, 2014b; FSD, 2014c).

**Table 1** highlights the variables selected from the financial diaries study (FSD, 2014b) to form the data of the study. The variables indicated with frequency of once are those that are captured once from the financial diaries data set while those with frequency as monthly are obtained through the monthly average value. We selected 280 households for further analysis from the 300 households in the KFD dataset as some data entries were missing or incomplete in the set of the 20 households excluded.

The data from the variables in **Table 1** was then scaled on (0, 1] to convert it into probabilities to ease compatibility with HMM. The 280 households were analyzed as individuals, as peer groups and on gender groupings. The credit scores and credit quality levels were compared for these different groups.

The statistical software, Ms Excel, STATA version 14 were used to analyze the data combined with Matlab version 7.1 for developing the HMM analysis toolkit.

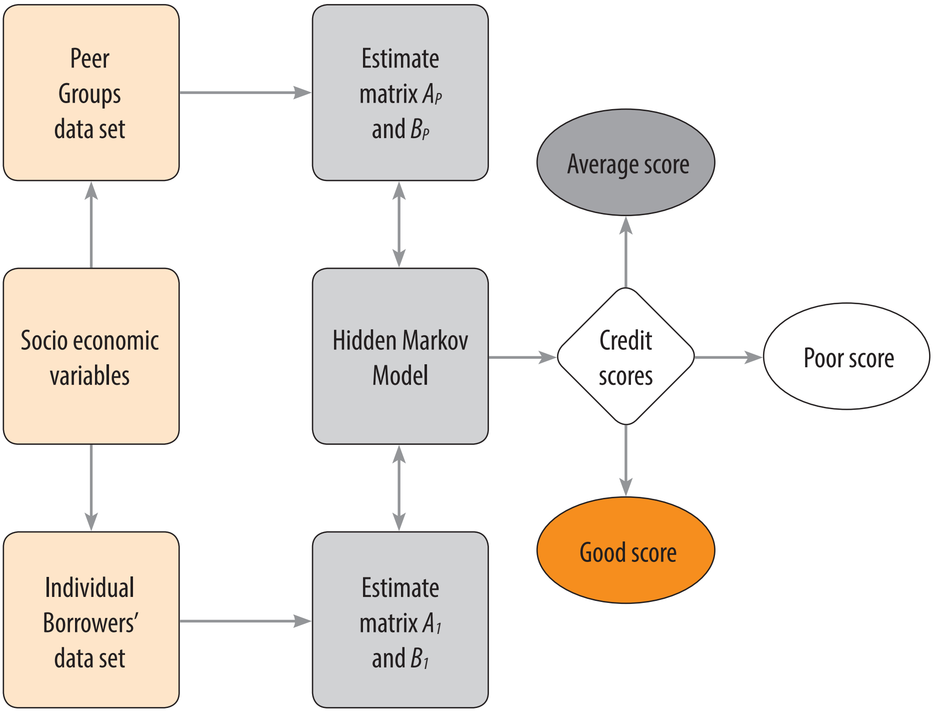
**Table 1: Selected Variables from the KFD project study used in this research**

| Variables | Frequency of use | Description |
|---|---|---|
| Age | Once | Age in years and if of working age |
| Marital status | Once | Single, Married, Live together, Widowed or Separated |
| Gender | Once | Female or Male |
| Household size | Once | Number of dependants or house members |
| Education | Once | Years spent in school |
| Mobile money flows | Once | Amount received or sent via phone |
| Mobile money transactions | Once | Number of transactions via phone |
| Income variability | Once | Changes in income in the time period |
| Social network | Once | Weak, medium, high |
| Resources given | Once | Out flows |
| Resources received | Once | In flows |
| Income | Monthly | Low, medium, high |
| Expenditure | Monthly | Low, medium, high |
| Savings | Monthly | Low, medium, high |

### 3.1 Conceptual Framework

We conceptualize the key ideas of this research that involves the use of social and economic variables from the financial diaries study of the year 2014. The credit scores of the individual and peer groups for lending are estimated using HMM from the data set in FSD (2014b). The socio-economic variables for each of the 280 households are treated as data for individuals or for peer groups. We then estimate the transition matrix and observation matrix for each individual household and each peer group formed, respectively. The HMM is trained based on the matrices and the HMM output are the credit scores and credit quality levels. **Figure 1** highlights the steps in achieving the objectives of the study. The households' financial activities data to the HMM output, the credit risk analysis and the classification into different credit profiles.

**Figure 1: The credit scoring and classification process framework**



### 3.2 Hidden Markov Model

We classify the MMCS customers based on their credit scores and credit quality levels captured by use of their socio-economic factors in **Table 1**. We refer to these factors as the credit scoring factors (CSFs) which are used for HMM learning and training. A transition matrix has high and low scores for the customers and the observation matrix has the three credit quality levels of poor, average and good. The two matrices are derived from the financial diaries data (FSD, 2014b) of the customers using the CSFs.

A HMM can be characterized by the following: number of states in the model, state transition probabilities, observation probability distribution that characterizes each state, initial state distribution, and the observation symbols (Rabiner, 1989)

- The number of states in the model with the set of states denoted as

$$S=\{S_1,S_2\} = \{Low, High\}$$

- The state transition probability distribution

$A = \{a_{ij}\}$ ,where

$$a_{ij} = P[q_{t+1} = S_j \mid q_t = S_i] \quad t = 1,2,\ldots,T$$

- The number of distinct observation symbols per state. We denote the set of observation symbols corresponding to the physical output of the system being modeled as

$$V=\{v_1, v_2, v_3\} = \{Poor, Average, Good\} = \{P, A, G\}$$

- The observation symbol probability matrix

$$B=\{b_j, (k)\}, \text{ where } b_{j(k)}= P[v_k / S_j], \quad 1 \leq j \leq 2$$

- The initial state probability vector

$$\pi = \{ \pi_j \}, \text{ where } \pi = P[q_1 = S_j], \quad 1 \leq j \leq 2$$

We use the notation as the set of parameters of the model in the study for both the individual members and the peer groups formed.

**Figure 2: A two state Markov model for the customer's transitions**

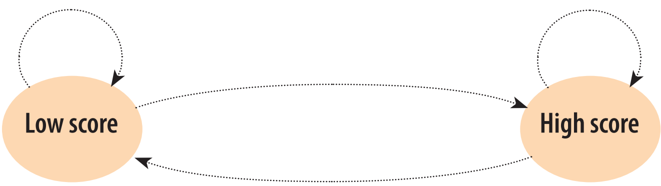


**Figure 2** is a Markov process showing the ergodic transitions between low score and high score for the state transition of the households.

### 3.3 Credit Scoring

The probabilistic relationship between the households financial activities, the credit scores and the credit quality level are estimated by the HMM. The classifications offer insights on who among the low earners, that is, the peer borrowers and individual borrowers have stronger credit scores to qualify for a micro-credit loan. The data from the 280 households using a set of 14 variables (see **Table 1**) is accumulated together to have enough data for HMM training and learning. The output is the credit scores and credit quality levels of these households.

Let $0 < \beta_i \leq 1$
be the credit score of customer .

$i \qquad (i = 1, 2, ..., N)$

We express

$$\beta_i = P(O_1, O_2, ...O_N \mid \lambda)$$

as the credit score from the observations of the *N=280* households.

Let $\alpha_i (i = 1, 2, ..., n)$ be the number of customers in a given credit quality score, where . $\alpha = \{P, A, G\}$ The estimation of the credit quality level is based on the credit score which is emitted directly by the HMM. The model emits the credit score and credit quality level at the same time. Credit scores and credit quality levels are dynamic and changes according to the prevailing conditions. The increase in information about a customer over time can be incorporated in the system. This includes repayment history, saving and withdrawal patterns as well as other data from the mobile phone usage.

### 3.4 Group Borrowers

Let $n_k$ be the group size of group $k$ with $k \geq 2$ The number of customers is represented as $N = n_1 + n_2 + ... + n_k$. Let $0 < \gamma_k \leq 1$ be the credit score of group $k$. The credit score is estimated as $\gamma_k = P(O_k^1, O_k^2, ..., O_k^n \mid \lambda_k)$ , where $\lambda_k = (A_k, B_k, \pi_k)$ estimated from the $n$ members forming group $k$. The observations are the emissions from the HMM during and after the initial period. Let $w_{nk}$ be the number of peer group members in a given credit quality level, where $w = \{P, A, G\}$. The estimated credit quality level is based on the credit score, the HMM emissions.

The selection of the peer groups was based on both systematic sampling and random sampling using the computer software Matlab. The gender variable is used to separate the sample and offer more insight on the credit scoring for the two groups of households. Descriptive statistics using tables, graphs, frequencies, percentages, standard deviation, coefficient of variation and mean provided more insights on the data. Inferential statistics applied lillietest to test for normality, Friedman test which is non-parametric test to test for difference between groups, correlation coefficient for the linear relationship and the Chi-square test to ascertain if distribution of the categories in the poor, average and credit scores differs significantly when gender is considered, individual households and the peer groups. Inferential statistics tested two of the hypotheses in the study.

# 4.0 Results and Discussions

**A Matlab computer program generated the results tabulated or discussed in this section. We compare the individual households and peer groups formed based on gender groupings as men and women, men only and women only. The sample comprises of 57.14 percent men and 42.86 percent women who were the households head. The data was ordered such that the first 160 households were men and the 120 were women, a total of 280 households. Credit scores and credit quality levels output from the HMM form the basis for the discussions.**

**Table 2: Credit quality levels compared against the gender**

| Credit Quality | Number (men/ women) | Percentage (men & women) | Number (Men) | Number (Women) | Percentage (Men) | Percentage (Women) |
|---|---|---|---|---|---|---|
| Poor | 120 | 42.86 | 92 | 28 | 57.50 | 9.56 |
| Average | 37 | 13.21 | 27 | 10 | 16.87 | 59.24 |
| Good | 123 | 43.93 | 41 | 82 | 25.63 | 31.20 |
| **Total** | **280** | **100** | **160** | **120** | **100** | **100** |

A comparison between the male and female household heads is made to understand the individual behavior. **Table 2** shows the comparison between men and women against the credit quality levels. The percentage of individual women in the average and good credit quality levels is 91.44 percent as compared to 42.5 percent for the men. The grouping of men and women together has 57.14 percent in the average and good credit quality levels as individuals. The comparison was made based on the proportion of each gender in the sample of study.

**Table 3: Systematic sampling of peer group sizes and the percentage of the groups in the respective credit quality levels**

| | Peer Group Size | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Credit Quality | 2 | 4 | 5 | 7 | 8 | 10 | 14 | 20 |
| **Poor** | 58.57 | 61.43 | 65.00 | 67.44 | 68.57 | 75.00 | 75.00 | 71.42 |
| **Average** | 23.57 | 17.14 | 13.30 | 11.62 | 11.43 | 10.71 | 10.00 | 14.29 |
| **Good** | 17.86 | 21.43 | 21.70 | 20.94 | 20.00 | 14.29 | 15.00 | 14.29 |

In **Table 3**, the group sizes of the peer groups are varied between 2 to 20 members for both men and women groupings, and the credit quality levels are noted in percentages. The percentage of the households in each peer group of the poor credit quality level increases from a group size of 2 members to 14 members and a reduction is observed in the average credit quality level. Mixed results are observable in the good credit quality level as the group size increases. In **Table 2**, 30 percent of the households have average or good credit quality when they are treated as individuals. These changes with the peer groups in **Table 3** where a peer group of 2 members have 41.43 percent of the members with average or good credit quality. This credit quality of average and good decreases as group size increases.

**Table 4: Peer group sizes and percentage of credit quality levels based on gender**

| | Peer Group Size | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | 2 | | | 4 | | | 8 | | |
| Gender | P | A | G | P | A | G | P | A | G |
| Men | 100.00 | 0.00 | 0.00 | 83.33 | 0.00 | 16.67 | 56.25 | 43.75 | 0.00 |
| Women | 0.00 | 50.00 | 50.00 | 12.50 | 56.25 | 31.25 | 8.33 | 66.67 | 25.00 |

**Table 4** highlights the dynamics observable in terms of gender when grouped in peer groups. For a group size with 2 members, 100 percent of the women peer groups had average or good credit level as compared to none for the men peer groups. For four members, 16.67 percent of men peer groups and 87.50 percent of the women peer groups had average or good credit quality levels.  Eight members in a peer group show that 43.75 percent of men groups and 91.67 percent of women groups had the average or good credit quality levels. Peer groups that comprised of women only had superior credit quality levels compared to groups with men only.

**Table 5** highlights peer groups of size 2 members to 16 members and the number of times each group was randomly selected. Three gender groupings were considered, men and women, men only and women only. The credit quality levels were computed for each grouping, to estimate the percentage of poor, average and good credit quality. For the category of men and women as members of the peer group, the average and good credit quality was observed in the peer group with 12 members with 84.62 percent, 3 members with 90.00 percent and 6 members with 100.00 percent. For the peer groups comprising of men only, the average and good credit quality was observed in the groups with 11 members with 64.29 percent, 7 members with 63.64 percent and 14 members with 60.00 percent. The peer group made up of women only had an average and good credit quality in the groups of 8 members with 91.67 and 3 members with 86.67. On average, the peer groups comprising of men and women had 76.22 percent of members in the average and good credit quality levels, the groupings of men only with 34.37 percent and women only with 91.37 percent. Evidently, the women only peer groups had superior performance, followed by men and women groups, then men only groups exhibited the lowest credit quality levels.

**Table 5: Randomly selected peer groups of varying sizes comprising men and women, men only and women only**

| | Men & Women | | | | Men only | | | | Women only | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Grp Size | Credit Quality | | | | Credit Quality | | | | Credit Quality | | | |
| | f | P | A | G | f | P | A | G | f | P | A | G |
| 2 | 6 | 66.67 | 33.33 | 0.00 | 2 | 100.00 | 0.00 | 0.00 | 6 | 0.00 | 50.00 | 50.00 |
| 3 | 10 | 10.00 | 50.00 | 40.00 | 13 | 53.85 | 46.15 | 0.00 | 15 | 13.33 | 40.00 | 46.67 |
| 4 | 10 | 40.00 | 50.00 | 10.00 | 6 | 83.33 | 0.00 | 16.67 | 16 | 12.50 | 56.25 | 31.25 |
| 5 | 13 | 23.08 | 46.15 | 30.77 | 11 | 63.64 | 36.36 | 0.00 | 12 | 8.33 | 58.33 | 33.33 |
| 6 | 7 | 0.00 | 71.43 | 28.57 | 11 | 72.73 | 27.27 | 0.00 | 8 | 12.50 | 50.00 | 37.50 |
| 7 | 13 | 30.77 | 53.85 | 15.38 | 11 | 36.36 | 63.64 | 0.00 | 7 | 0.00 | 42.86 | 57.14 |
| 8 | 11 | 27.27 | 54.55 | 18.18 | 16 | 56.25 | 43.75 | 0.00 | 12 | 8.33 | 66.67 | 25.00 |
| 9 | 7 | 14.29 | 85.71 | 0.00 | 8 | 37.50 | 62.50 | 0.00 | 7 | 14.29 | 57.14 | 28.57 |
| 10 | 14 | 7.14 | 78.57 | 14.29 | 14 | 64.29 | 35.71 | 0.00 | 6 | 16.67 | 50.00 | 33.33 |
| 11 | 12 | 25.00 | 75.00 | 0.00 | 14 | 35.71 | 64.29 | 0.00 | 15 | 26.67 | 46.67 | 26.67 |
| 12 | 13 | 15.38 | 69.23 | 15.38 | 10 | 80.00 | 20.00 | 0.00 | 8 | 0.00 | 62.50 | 37.50 |

| | Men & Women | | | | Men only | | | | Women only | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Grp Size | Credit Quality | | | | Credit Quality | | | | Credit Quality | | | |
| 13 | 14 | 21.43 | 57.14 | 21.43 | 16 | 75.00 | 25.00 | 0.00 | 13 | 7.69 | 53.85 | 38.46 |
| 14 | 8 | 37.50 | 62.50 | 0.00 | 5 | 40.00 | 60.00 | 0.00 | 11 | 9.09 | 90.91 | 0.00 |
| 15 | 9 | 11.11 | 88.89 | 0.00 | 14 | 85.71 | 14.29 | 0.00 | 17 | 0.00 | 82.35 | 17.65 |
| 16 | 7 | 28.57 | 71.43 | 0.00 | 2 | 100.00 | 0.00 | 0.00 | 4 | 0.00 | 75.00 | 25.00 |
| Total | 154 | | | | 153 | | | | 157 | | | |
| Average | | 23.88 | 63.19 | 12.93 | | 65.62 | 33.26 | 1.11 | | 8.63 | 58.83 | 32.54 |

*f - frequency

## 4.1 Peer groups vs individual households based on gender

The data for the individual households and peer groups were not normally distributed when tested with lillietest in Matlab. A non-parametric test, Friedman test was used to test for differences between different peer group sizes when the sample of 280 households was analyzed.

**Table 6: Friedman test for peer groups with three gender groupings**

| | Men and Women | | | | | Men only | | | | | Women only | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Source | SS | df | MS | $x^2$ | $P>x^2$ | SS | df | MS | $x^2$ | $P>x^2$ | SS | df | MS | $x^2$ | $P>x^2$ |
| Columns | 0.7 | 2 | 0.35 | 0.79 | 0.67 | 8.1 | 2 | 4.05 | 9.53 | 0.01 | 17.1 | 2 | 8.55 | 19.36 | 0.00 |
| Error | 25.8 | 28 | 0.92 | | | 17.4 | 28 | 0.62 | | | 9.4 | 28 | 0.33 | | |
| Total | 26.5 | 44 | | | | 25.5 | 44 | | | | 26.5 | 44 | | | |

In **Table 6**, we are interested in testing if the group size affects the credit quality level. For the peer groups with both men and women, $\chi^2 = 0.79 \quad (p = 0.67)$, which is not statistically significant, indicating that the peer group size did not influence the credit quality levels of the households. For the peer groups with men only, the results are statistically significant with $\chi^2 = 9.53 \quad (p = 0.01)$ while for the groups with women only, $\chi^2 = 19.36 \quad (p < 0.001)$ which is statistically significant. Peer groups with men only or women only credit quality levels are affected by the size of the peer groups. This is not the case for peer groups with both men and women. A possibility is that men and women offer different attributes to the group which reduces variability and thus increasing the group positive attributes.

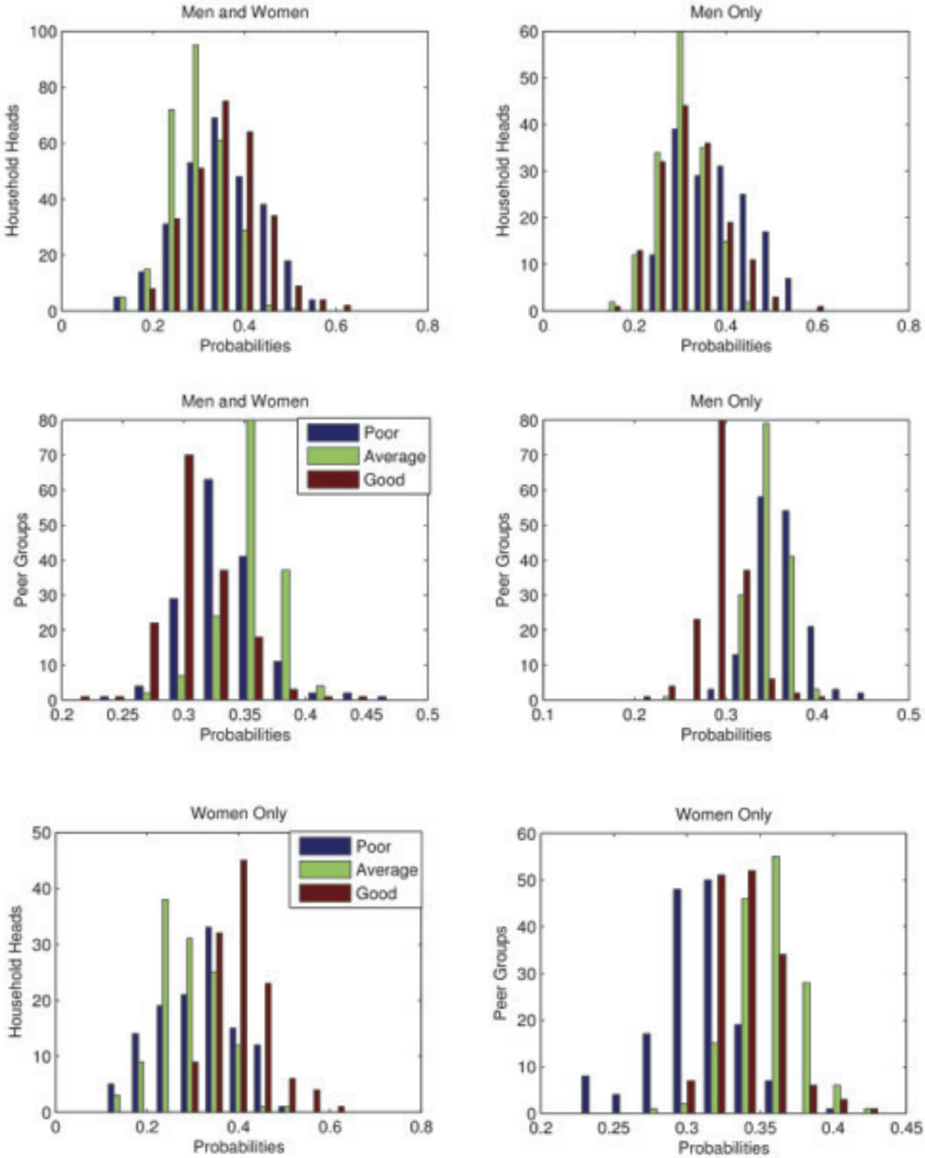

Men and Women
Household Heads
Probabilities
Men Only
Household Heads
Probabilities
Men and Women
Peer Groups
Probabilities
Poor
Average
Good
Men Only
Peer Groups
Probabilities
Women Only
Household Heads
Probabilities
Poor
Average
Good
Women Only
Peer Groups
Probabilities

**Figure 3**: Credit scores of the peer groups and individual households with gender groupings. The credit scores based on the credit quality levels (poor, average and good) when the individual households and the peer groups are analyzed as men and women, men only and women only.

In **Figure 3**, the credit quality levels of women show a marked difference with the men only and men and women peer groups. Most of the low credit scores for the women only peer groups are poor in credit quality when compared to the good credit quality with low credit scores for the men only and also the men and women peer groups. A major difference is observed between the peer groups and the individual household heads credit quality levels of poor, average and good.

**Table 7: Descriptive statistics for the credit scores and credit quality levels of the peer groups and individuals**

| Individual Households | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Men and Women | | | Men only | | | Women only | | |
| Statistic | P | A | G | P | A | G | P | A | G |
| Mean | 0.350 | 0.295 | 0.355 | 0.378 | 0.301 | 0.321 | 0.312 | 0.287 | 0.401 |
| Stdev | 0.088 | 0.061 | 0.081 | 0.078 | 0.056 | 0.076 | 0.087 | 0.066 | 0.062 |
| CV (%) | 25.10 | 20.56 | 22.68 | 20.49 | 18.47 | 27.78 | 123.08 | 15.51 | 18.700 |
| Skewness | -0.025 | 0.084 | 0.283 | 0.324 | -0.040 | 0.626 | -0.073 | 0.288 | 0.764 |
| **Peer Groups** | | | | | | | | | |
| | Men and Women | | | Men only | | | Women only | | |
| Statistic | P | A | G | P | A | G | P | A | G |
| Mean | 0.334 | 0.355 | 0.311 | 0.360 | 0.346 | 0.294 | 0.307 | 0.355 | 0.338 |
| Stdev | 0.033 | 0.025 | 0.030 | 0.027 | 0.020 | 0.024 | 0.028 | 0.022 | 0.022 |
| CV (%) | 9.74 | 7.060 | 9.77 | 7.42 | 5.91 | 8.20 | 9.26 | 6.22 | 6.61 |
| Skewness | 0.842 | -0.553 | 0.520 | 0.622 | -0.925 | 0.297 | -0.431 | -0.051 | 0.937 |

*Stdev - Standard deviation

*CV - Coefficient of Variation

*P – Poor, A-Average and G-Good

In **Table 7**, the individual households have high variability in the credit scores with the coefficient of variation of between 15.51 percent and 27.78 percent as compared to the peer groups with a coefficient of variation ranging between 5.91 percent and 9.97 percent. Peer groups have more stable credit scores in comparison to the individual households. For the three credit qual- ity levels, the average and good credit quality probabilities for the individuals range between 62.2 percent and 68.8 percent while the peer groups rating ranges between 64.00 percent and 69.3 percent. Peer groups tend to have lower credit risk or higher credit scores as compared to the individual households.

**Table 8: Friedman test for peer groups and individual households**

| | Men and Women | | | | | Men only | | | | | Women only | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Source** | **SS** | **df** | **MS** | $x^2$ | $P>x^2$ | **SS** | **df** | **MS** | $x^2$ | $P>x^2$ | **SS** | **df** | **MS** | $x^2$ | $P>x^2$ |
| Columns | 45.83 | 5 | 9.17 | 13.11 | 0.023 | 0.67 | 2 | 0.33 | 0.67 | 0.7165 | 4.67 | 2 | 2.33 | 4.67 | 0.097 |
| Error | 6.67 | 10 | 0.67 | | | 5.33 | 4 | 1.33 | | | 1.33 | 4 | 0.33 | | |
| Total | 52.50 | 17 | | | | 6.0 | 8 | | | | 6.0 | 8 | | | |

In **Table 8**, there exists a difference between the individual households and the peer groups. Peer groups and individual household's coefficient of variation differenceshad $\chi^2 = 13.1$ $(p = 0.023)$ which is statistically significant. For the individual households and peer groups, the respective $\chi^2 = 0.67$ $(p = 0.7165)$ a n d $\chi^2 = 4.67$ $(p = 0.097)$ are not statistically significant. We observe that variability between individual household's changes in credit scores are the same, and that applies to the peer groups.

When the peer group size is compared to the credit scores and the credit quality levels, the correlation coefficient differs. Men only with women only peer groups $(r = 0.07895)$, men only with men and women peer groups $(r = 0.5171)$, and women only with men and women peer groups $(r = 0.5137)$. The peer groupings had higher influence on the linear relationship between the men only and women only peer groups.

The peer group sizes of between four and eight members are the optimal sizes for lending. This observation was made based on the steady state of the transition matrices of the peer group analysis. Male and female peer groupings did not alter the optimal group sizes. The estimation of the transition equilibrium factors in the long run effect of the changes in the credit dynamics of the peer groups.

**Table 9: Summary of opportunities available for mobile micro credit system in Kenya**

**Low income earners characteristics**

| | |
|---|---|
| 1 | Social collateral (Horizontal social networks for women and vertical social networks networks for men |
| 2 | Lack of good financial options due to small and inconsistent incomes |
| 3 | Develop long term relationships via mobile technology |
| 4 | Peer selection that can reduce moral hazard and information assymemetry in the group |
| 5 | Members can save as a group or individuals in the same mobile wallet |

**Mobile micro credit system strengths**

| | |
|---|---|
| 1 | Peer groups to diversify risk and utilize the power of group saving |
| 2 | Credit scoring using telecommunication variables |
| 3 | Availability of data mining tools and low cost of data collection |
| 4 | Existence of baseline mobile financial |
| 5 | High mobile penetration rate in Kenya standing at $\approx$ 90 percent |

**Table 9** shows the analysis of the low income earners characteristics and the mobile micro credit system strengths. This indicates an opportunity exists for the financial services stake- holders to consider providing financial services to the low income earners to reduce financial exclusion. We consider and summarize the findings for the three hypotheses in the study.

H1a: **Credit scores for low income peer groups and the individual's credit scores among low income earners are the same.**

Peer groups comprising of women only have the highest credit profiles, followed by groups with men and women, while men only groups had the lowest credit score profiles. The credit scores for the low income peer groups outperform the conventional individual credit scores among low income earners as noted from the analysis. It would be ideal for financial services providers to consider offering mobile micro credit to low income earners through peer groups instead of offering the same to individual members.

### H1b: **Credit scores and credit quality levels among low income men and women is the same**

The credit scores and credit quality levels among the low income earners is higher in women than in men. When both men and women are analyzed as one, the overall scores are higher than those of men only peer groups but lower than those of the women. Financial products where both men and women are involved or women only peer groups would be more stable and can offer the providers lower credit risk levels. Thus, credit scores and credit quality levels among the low income men and women is not the same.

### H1c: **A mobile based micro credit system can further enhance financial inclusion among the low income earners.**

A peer group that is mobile based can enhance the peer groups' credit score. The availability of data for credit risk analysis such as the telecommunication variables, savings, and withdrawal and loan repayment patterns of the group members can be incorporated in the risk analysis process. As the members share one account, the social collateral can lower moral hazard since the group members are peer selected. The success story of mobile financial services in Kenya forms a baseline to cater for the low income earners, with small and inconsistent incomes.

# 5.0 Conclusions

**The Hidden Markov model is an ideal statistical technique to classify customers into different credit risk profiles. The HMM output being the credit scores and credit quality levels for the peer groups and the individual households. Peer groups outperform conventional individuals as the former has lower credit risk as compared to the latter.**

The diversification of risk in the peer groups inherent in individuals motivates a need to develop a mobile micro credit facility to cater for peer groups of the low income earners in Kenya. This can further increase financial inclusion, ride on the success of the mobile financial services in the country and offer the low income earners another financial option.

The peer groups comprising of women only have higher credit quality levels as compared to those of men only. Women had lower variability in the credit scores in each of the credit risk profile when compared to men only or men and women peer groups. Thus, lending to women offers lower credit risk as compared to men even for low income earners. Even though men have higher financial muscle than women due to the societal set up in our country, women offers a more stable lending environment due to their strong horizontal social networks. This can be explained by the fact that women borrow more from informal financial services.

The women peer groups are more stable, less variability in exhibiting credit scores as com- pared to men. The clustering pattern observable in women peer groups exhibits the high levels of social collateral they posses and the informal lending patterns and borrowing that characterizes them. Therefore, incorporating soft data already in use by KCB Mpesa and M-Shwari to offer mobile based loans would be a step in the right direction.

Further research can consider analyzing the credit risk profile of each peer group member and develop a mechanism to reward more consistent savers in the envisioned mobile micro credit system. On the same note, the system can consider how to estimate the moral hazard of the peer group members and cater for it, and allow for peer selection of the peer group membership.

# Reference


1. Ahlin, C., and Jiang, N. (2008). *Can micro-credit bring development? Journal of Development Economics,* 86 (1): 1-21

2. Benyacoub, B., Bernoussi, S., and Zoglat, A. (2014). *Building classification models for customer credit scoring.* Logistics and Operations Management International Conference, 5 (7): 107-111

3. Bilmes, J.A. (2006). *What hidden Markov models can do. IEICE Transactions, Information and Systems* (Special issue on IEICE Transactions), E89-D (3)

4. Bhole, B., and Ogden, S. (2010). *Group lending and individual lending with strategic default.* Journal of Development Economics, 91: 348-363

5. Chavan, P., and Ramakumar, R. (2002). *Micro-credit and rural poverty: An analysis of empirical evidence*. Economics and Political Weekly, 37 (10): 955-965

6. Cook, T., and Mckay, C. (2015). *How M-Shwari works: The story so far.* Access to Finance Forum No. 10 (Reports by CGAP and FSD)

7. Crowder, M., Davis, M., and Giampieri, G. (2005). *A hidden Markov model of default interaction*. Quantitative Finance, 5 (2): 27-34

8. Daniel, B., and Grissen, D. (2015). *Behavior revealed in mobile phone usage predict loan repayment*. Social Science Research Network (SSRN), 2611775: 1 - 10

9. Deesai, S. (2012). *Mobile money for the unbanked: State of the industry,* GSMA, Dubai

10. Financial Sector Deepening (FSD). (2016). *2016 FinAccess household survey. C*ollabo- ration between Central Bank of Kenya, Financial Sector Deepening and Kenya National Bureau of Statistics, Nairobi, Kenya

11. Financial Sector Deepening (FSD). (2014a). S*hilingi kwa shilingi: The financial lives of the poor. Kenya Financial Diaries,* Nairobi, Kenya

12. Financial Sector Deepening (FSD). (2014b). *A buck short: What financial diaries tell us about building financial services that matter to low income women.* Report by Bankable Frontier Associates, Nairobi, Kenya

13. Financial Sector Deepening (FSD). (2014c). *Kenya financial diaries datasets user guide. A guide to the data produced by the 2012 – 2013* Kenya financial diaries, Nairobi, Kenya

14. Financial Sector Deepening (FSD). (2014c). *What do low income people know about money?* FSD insights, Issue 7, Nairobi, Kenya

15. Financial Sector Deepening (FSD). (2013). *FinAccess national survey 2013: Profiling developments in financial access and usage in Kenya,* Nairobi, Kenya

16. Financial Sector Deepening (FSD). (2010). F*inancial capability and the poor: Are we missing the mark? Bankable Frontier Associates, Issue 2*, Nairobi, Kenya

17. Financial Sector Deepening (FSD). (2009). *Costs of collateral in Kenya: Opportunities for reform.* Growth Fin, Nairobi, Kenya

18. Gomez, R., and Santor, E. (2003). *Do peer group members outperform individual borrowers?* A test of peer group lending using Canadian micro credit data. Bank of Canada working paper 2003-33

19. Hassan, R., and Nath, B. (2005). *Stock market forecasting using hidden Markov model: A new approach.* Proceedings of the 2005 5th International Conference on Intelligent Systems Design and Applications (ISDA)

20. Holmstrom, B. (1979). *Moral hazard and observability.* The Bell Journal of Economics, 10 (1): 79-91

21. Ivatury, G., and Pickens, M. (2006). *Mobile phone banking and low income customers: Evidence from South Africa.* Consultative group to assist the poor. World Bank, Washington

22. Jack, W., and Suri, T. (2011). *Mobile money: The economics of M-Pesa.* NBER Working Paper 16721

23. Marrez, H., and Schmit, M. (2009). C*redit risk analysis in micro credit: How does gender matter?* CEB Working paper No. 09/053, Universite Libre Bruxelles

24. Mbiti, I., and Weil, D.N. (2011). *Mobile banking: The impact of M-Pesa in Kenya.* Brown University, Department of Economics, No. 2011-13

25. Netzer, O., and Srinivasan, J.M.L. (2008). *A hidden Markov model of customer relationship dynamics.* Marketing Science, 27 (2): 185-204

26. Ntwiga, D.B. (2016). *Social network analysis for credit risk modeling.* Unpublished PhD thesis, School of Mathematics, University of Nairobi, Kenya

27. Ntwiga, D.B., Weke, P., and Kirumbu, M.K. (2016). T*rust model for social network using singular value decomposition.* Interdisciplinary Description of Complex Systems, 14 (3): 296-302

28. Quirini, L., and Vannucci, L. (2014). *Creditworthiness dynamics and hidden Markov models.* Journal of the Operational Research Society, 65: 323-330

29. Rabiner, L.R. (1989). *A tutorial on hidden Markov model and selected applications in speech recognition.* Proceedings of the IEEE, 77(2)

30. Simtowe, F., Zeller, M., and Phiri, A. (2006). *Determinants of moral hazard in Microfinance: Empirical evidence from joint liability lending programs in Malawi.* African Re- view of Money Finance and Banking, 5-38

31. Srivastava, A., Kundu, A., Sural, S., and Majumdar, A.K. (2008). *Credit card fraud detection using hidden Markov model.* IEEE Transactions on Dependable and Secure Computing, 5 (1)

**Kenya Bankers Association**
13th Floor, International House, Mama Ngina Street
P.O. Box 73100– 00200 NAIROBI
Telephone: 254 20 2221704/2217757/2224014/5
Cell: 0733 812770/0711 562910
Fax: 254 20 2221792
Email: research@kba.co.ke
Website: www.kba.co.ke

KENYA BANKERS
ASSOCIATION

One Industry. Transforming Kenya.