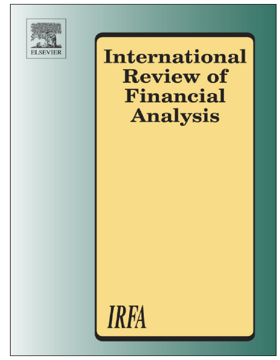
Accepted Manuscript



Using multiple correspondence analysis for finance: A tool for assessing financial inclusion

Mardi Dungey, Firmin Doko Tchatoka, María B. Yanotti

PII:	S1057-5219(18)30397-1
DOI:	doi:10.1016/j.irfa.2018.08.007
Reference:	FINANA 1243
To appear in:	*International Review of Financial Analysis*
Received date:	30 April 2018
Revised date:	20 July 2018
Accepted date:	8 August 2018

Please cite this article as: Mardi Dungey, Firmin Doko Tchatoka, María B. Yanotti , Using multiple correspondence analysis for finance: A tool for assessing financial inclusion. Finana (2018), doi:10.1016/j.irfa.2018.08.007

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

Using Multiple Correspondence Analysis for finance: A tool for assessing financial inclusion *

Mardi Dungey[a,c]
Firmin Doko Tchatoka[b]
María B. Yanotti[a†]

[a] University of Tasmania
Private Bag 84, Hobart, TAS, 7001, Australia
[b] University of Adelaide
10 Pulteney Street, The University of Adelaide, SA, 5005, Australia
[c] CAMA, Australian National University
J.G. Crawford Building 132, Lennox Crossing, ANU, Canberra, ACT, 0200, Australia

July 20, 2018

Abstract

This paper introduces multiple correspondence analysis (MCA) to the literature on financial product choice. MCA is a useful way of assessing the typology of actual or potential consumers, which can then be used to assess the extent to which existing products cover consumer needs. Given the importance of the financial inclusion agenda, this provides a useful means of detecting areas of financial under-servicing. An illustration using bank mortgage data shows how some groups are well-serviced, but others suffer from mismatch between the characteristics of the available and desirable products.

JEL classification: G21, R20 , D81,
Keywords: Multiple Correspondence Analysis, Cluster analysis, Mortgage choice

*We acknowledge funding support from Australian Research Council grant DP120100842.

†Corresponding author contacts: Maria.Yanotti@utas.edu.au; Locked Bag 1317, Tasmanian School of Business and Economics, UTAS, Launceston, 7250, Tasmania, Australia; P: +61 3 6324 3659; F: +61 3 6324 3369. We acknowledge funding support from Australian Research Council grant DP120100842. We are grateful for comments by anonymous reviewers, and participants at AHRC 2015.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

1 Introduction

Global financial inclusion goals aim at 'promoting affordable, timely and adequate access to a wide range of regulated financial products and services and broadening their use by all segments of society', Atkinson and Messy (2013). While financial institutions design their product offerings for profit maximisation, social welfare concerns often identify under-serviced consumer factions; highlighting access restrictions which may be based around characteristics such as ethnicity, gender, employment status or age. Examples include access to banking, savings, mortgages, small loans, use of financial advisors, insurance policies. Rather than directly detectable discriminatory practices, such as on ethnicity, these customers typically form a majority of a customer base with less desirable characteristics, such as being correlated with lower education status, non-nuclear family structures or mobility. Identifying the dominant characteristics of under-serviced groups and how these match with current product offerings provides a means of rethinking how issues of inclusion might be resolved.

This paper introduces the use of multiple correspondence analysis (MCA) to the finance literature as a tool for recognizing mismatches between consumer needs and product offerings. MCA is a dimensionality reduction factor technique in the same family as the more popular principal component analysis (PCA), but specifically designed to deal with nominal or ordinal categorical data, and well suited to the types of data we face when describing consumer characteristics. Although PCA is a very useful tool when constructing multi-dimensional models with numerical variables, it does not work very well when dealing with discrete categorical data; see Gower (1966), Kolenikov and Angeles (2009). MCA deals specifically with

ACCEPTED MANUSCRIPT

categorical data, assigning scale values to the categories of the discrete variables and maximizing the variance of those scores to find: (1) the associations between the variables, and (2) the proximity between individuals.

To our knowledge MCA has not been introduced to the finance literature, although it is a tool used in many fields of research. In the social sciences MCA has been applied to health economics, medicine, marketing, management and sociology; see for example Asselin and Anh (2008), Sourial et al. (2010), Hoffman and Franke (1986), Spearman (1904), Hirschfeld (1935), Guttman (1944), Manté et al. (2013), and Mendes et al. (2012). A good reference comparing MCA with similar and perhaps more familiar methodologies such as PCA, may be found in Tenenhaus and Young (1985); see also Hayashi (1950), Hill (1974), Greenacre (1988), and Le Roux and Rouanet (2004).

In general consumers of financial services select from a pre-determined menu of products offered by financial institutions. For example, Piskorski and Tchistyi (2010) shows that it is cost-effective for a bank to price mortgage products specifically targeted to groups of borrowers with common characteristics.[1] The degree of market completeness for financial products may reduce the welfare of some consumers. MCA, coupled with cluster analysis, provides a means of coming to grips with the extent to which this binds for sample populations.

After introducing the MCA techniques, we present a sample application in financial product choice over mortgage products using data on borrower typology for over half-a-million Australian mortgage borrowers. Mortgage providers design

[1]Gabaix and Laibson (2006) discuss price discrimination between naive and sophisticated financial market consumers and the effect of consumer education. Gary-Bobo and Larribeau (2004) argue, for example, that French lenders practice first-degree price discrimination, rather than self-selection or second-degree discrimination.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

a menu of mortgage products to cater for borrowers' preferences when financing housing through a home loan; see Dunn and Spatt (1988) and Stanton and Wallace (1998). The range of products they offer will also reflect their preferences for profit maximisation and risk minimisation. Customers self-select into different products; however, there may be gaps between the characteristics of their desired products and those on offer. Using MCA we show that the mortgage provider gives a substantial number of options to one set of customers, but relatively few to other (less desirable) groups. The most desirable customers may be able to match their own preferences quite closely, while the less included may have to make substantially more trade-offs to obtain a product relative to their actual preferences.

In our sample application, the MCA results find two important dimensions distinguishing household mortgage decisions, income and wealth. When combined with other borrower characteristics six distinct borrower profiles emerge: (1) young income- and wealth-constrained households with a large proportion of female main borrowers; (2) risk averse, income and wealth constrained, young families (3) senior borrowers (4) young, mobile first-time home buyers (5) settled families (6) and low risk households.

These 6 borrower categories choose between 4 main products; Variable rate mortgages (VRMs) are associated with borrowers with high income and wealth levels, and who are less risk-averse and more financially experienced. Home Equity loans (HE) are well matched to borrowers in the mid-point of their life cycle, with a sound financial position who look for consumption smoothing. Short-term fixed rate mortgages (SFRM) are for the risk averse and income/wealth constrained, while Honeymoon (HM) mortgages typically target those with income constraints.

ACCEPTED MANUSCRIPT

In general, income and wealth constrained borrowers appear to have fewer choices of products which fit their needs. We observe some potentially under-served groups of borrowers whose access to financial markets could be improved by product choices which more closely match their characteristics. It is worth noting that our results are likely to be an under-estimate of under-serviced groups, as our data cover only those who applied for the products on offer – other possible mortgage-seekers whose characteristics are further away from the existing products may choose not to apply at all or choose less regulated financial services.

The remainder of the paper is organized as follows. We begin in Section 2 with a discussion of current areas of concern around financial inclusion and proceed in Section 3 to describe the MCA methodology and how it is combined with cluster analysis to provide an analytical tool suitable for identifying the characteristics of the consumers and the products on offer. Section 4 provides an example based on mortgage product choice. Section 5 concludes.

2 Contemporary issues in financial inclusion

The United Nations Capital Development Fund identifies adequate access to finance and financial inclusion as essential enabling platforms for the majority of the 2030 Millenium goals. With financial inclusion comes the opportunity to save, both for future capital needs and emergencies, plan, and expand opportunities. These characteristics apply not only to the poor, but also to those excluded in more developed nations. Identifying groups which are either currently unserviced by financial markets, or under-serviced by the range of products currently available is key to improving financial participation.

ACCEPTED MANUSCRIPT

The finance literature is replete with papers which investigate the revealed preference of financial market participants for different products. One of the most researched areas is mortgage choice, with a focus on whether households choose fixed rate or variable (adjustable) rate mortgage products. While theory is clear on the determining characteristics – mortgage costs, income, wealth, mobility – the empirical evidence is less clear on which of these are statistically significant. This is likely, at least in part, due to the wide variety of institutional structures in place around the globe; such as the extent of securitization, the role of intermediaries such as Fannie Mae and Freddie Mac in the U.S., mortgage insurance arrangements, and societal preferences.[2]

Global financial inclusion goals focus on the development of services for the unbanked and severely underserviced in many instances; see Atkinson and Messy (2013). Women, the uneducated, younger people, those with broken or no employment, indigeneous, disabled, elderly, immigrant and rural populations have all been identified as at higher risk of being excluded from the formal financial system. While some of the issues are technological, due to geography or institutional structures, many of these attributes are individual consumer or community characteristics. Most countries education programs to promote financial literacy, and have gone to considerable effort to reduce the number of unbanked. However, this is not enough. India, for example, dramatically decreased its unbanked population, but in some areas Morawczynski et al. (2010) report that 70-80% of the poor leave these accounts dormant. The products must not only exist but also meet the needs of the consumer. Existing work for developed markets finds that higher

[2]For example Green and Wachter (2005) discusses the historical context leading to the dominance of 30-year fixed rate mortgages in the U.S., while Warnock and Warnock (2008) compare arrangements across the globe.

ACCEPTED MANUSCRIPT

income, financially mature households are more sensitive to interest rate, pricing and product features when choosing a financial institution, while lower income households are more likely to choose a financial institution based on location or word of mouth; see Devlin (2002), Mylonakis (2007) and Boyd et al. (1994).[3]

Consumers also make choices between what can be a bewildering array of insurance contracts to defray the cost of disaster events. Recently Koijen et al. (2016) estimated that U.S. consumers pay some 3.2% of total wealth as the cost of the incomplete markets for their insurance (health and life) portfolios. Key to reducing these costs is meeting consumer and household characteristics identified in the literature such as age, bequest, access to home ownership and health (including chronic conditions). In one example, Finkelstein et al. (2012) demonstrates the complexity of the trade-off between insurance and savings behaviors when confronted with uncertain health shocks.

More recently fintech has made inroads on increasing financial inclusion. Kenya is a prime example of an economy which has leapfrogged many of the traditional bricks-and-mortar stages of banking development, Demirguc-Kunt and Klapper (2013). Jagtiani and Lemieux (2017) document improvements in credit access for previously excluded consumers in the U.S., particularly due to the use of non-traditional information in assessing risk (harnessing big data and algorithmic advances). However, there are also impediments to fintech solutions based on customer preferences, such as trust. Designing the products to meet the consumers needs is additional to both the problems the provision of capital infrastructure or the idea of technological acceptance on the part of the consumer.[4] These can be

[3] For market survey studies on banking product selection and borrower characteristics see also Black et al. (2003), Stafford (1996), and Talaga and Buch (1998).

[4] The technological acceptance literature is prevalent where the focus is on why uptake of

ACCEPTED MANUSCRIPT

as simple as finding community acceptable and affordable forms of identity – the 2018 World Bank Global ID4D Dataset for 2018 estimates 1.1 billion people have no official form of identity.

All of the issues identified in this Section speak to cases of genuine concern in ascertaining that the products offered by the financial sector meet the complex and varying needs of the population. Incomplete financial markets create the potential for misalignment between product and consumer, and may result in disproportionate disadvantage to those least able to afford it. As we will show in Section 4, MCA can identify both how customers cluster by characteristics, and locate the products offered to serve these needs within the same space. In this way it can help to determine opportunities to offer products which fill important unmet needs from a consumer perspective. Given the increasing amount of categorical information available, via surveys and large financial databases, MCA can provide a useful tool for progressing the financial inclusion agenda.

3 Methodology

A combination of multiple correspondence analysis (MCA) and cluster analysis allow us to locate (potential) consumers and products in the same characteristic space, which provides a means of assessing how well the products meet the needs of differing consumer groups. Since the seminal work of Tenenhaus and Young (1985), MCA method has been extensively used in other disciplines but has remained unfamiliar in the finance literature. In the following sections we review MCA and how it couples with cluster analysis to help build customer typologies.

products has not occurred after the provision of infrastructure; see for example Luarn and Lin (2005).

ACCEPTED MANUSCRIPT

3.1 Multiple correspondence analysis (MCA)

There is a variety of approaches used to distill measures of associations between variables and individuals for large datasets; one of the most familiar methods is principal component analysis (PCA). This approach usually deals with continuous variables and consists of choosing the main orthogonal vectors of a data matrix, or principal components, that explain most of the total variability in the data, or alternatively those with eigenvalues greater than one. Multiple correspondence analysis (MCA) is concerned with displaying the categories of more than two discrete variables, and is useful for categorical data. This is achieved by defining dummy variables for each category and re-expressing the data in the form of a cases-by-variables indicator matrix.

The MCA technique uses a distance measure rather than the orthogonalization technique which underlies PCA. MCA transforms the association between categories of discrete variables into coordinates in a multidimensional space. It assigns scale values to the categories of the discrete variables and maximizes the variance of those scores to find: (1) the associations between the variables, and (2) the proximity between individuals. Points in the same direction from the origin are highly associated. Points around the origin represent the mean, while points away from the origin deviate from the mean.

Formally, consider *n* individuals with *p* characteristics $X = \{X_1, \dots, X_p\}$. Each characteristics X_j, $j = 1, \dots, p$, has k_j categories. The total number of categories is $k = \sum_{j=1}^{p} k_j$. Define x_{i,k_j} such that:

$$x_{i,k_j} = \begin{cases} 1 \; if \; individual \; i, i=1,\dots,n, is \; in \; category \; k_j, \\ 0 \; otherwise; \end{cases} \qquad (1)$$

and let $X_k = \left[X_{k_1}, \dots, X_{k_p}\right] \in \mathbb{R}^{nxk}$. Let $\phi_{k_j} = (\phi_{jl})_{l=1,\dots,k_j}$ denote the scale value vector of category $l = 1, \dots, k_j$ and $\tilde{X}_k = \sum_{j=1}^{p} X_{k_j} \phi_{k_j}$ be the scaled variable induced by scaling X_k. Let $\phi_k = (\phi'_{k_1}, \dots, \phi'_{k_p})'$ be the *k* dimension vector of scale values and $\tilde{X} = \left[\tilde{X}_1, \dots, \tilde{X}_p\right]$ the matrix of scaled variables.

The MCA principle solves:

$$max_{\phi_k} Var\left[\frac{1}{p}\sum_{j=1}^{p}\sum_{l=1}^{k}\phi_{jl}x_{jl}\right] \qquad s.t. \quad e'_k D\phi_k = 0 \quad , \quad \phi'_k D\phi_k = np \qquad (2)$$

ACCEPTED MANUSCRIPT

$$\Leftrightarrow max_{a_{jl}} Var\left[\frac{1}{p}\sum_{j=1}^{p}\sum_{l=1}^{k}\sqrt{\frac{a_{jl}^{2}n}{pn_{jl}}}x_{jl}\right] \quad s.t. \quad \sum_{j=1}^{p}\sum_{l=1}^{k}\sqrt{n_{jl}}a_{jl} = 0 \quad , \quad \sum_{j=1}^{p}\sum_{l=1}^{k}\sqrt{a_{jl}} = 1 \quad (3)$$

Where D is a k dimensional diagonal matrix constructed with the non-full frequencies n_{jl}, e_k is a k-dimensional vector of ones, and $a_k = \sqrt{\frac{n_{jl}}{np}}\phi_{jl}$. The solution ϕ^* of (2) and a^* of (3) are linked by the relationship:

$$\phi^* = (np)^{1/2}D^{-1/2}a^* \qquad (4)$$

which tells us that the solution value of the scaling parameter, ϕ^* and a^*, depends on the number of characteristics, the number of individuals and the frequency of occurrence, D.

The k components of ϕ^* are the category factors (principal components), $\hat{\psi}^* = (1/p)X\phi^*$, and the subject factors (normalized variables) $\psi^* = \hat{\psi}^*/(\lambda_h)^{1/2}$. The eigenvalue λ_h is the variance explained by the hth principal component and $p\lambda_h/(k - p)$ is the proportion of the variance explained by the hth component. The contribution of category $l, l = 1, \dots, k$, to the hth principal component, c_l^h, and the correlation between $\tilde{X}_j^h$ and ψ^*, are then given by:

$$c_{jl}^h = \frac{n_{jl}(\phi_{jl}^*)^2}{np}, \qquad corr\left(\tilde{X}_j^h, \psi^*\right) = (\lambda_h p \sum_{l=1}^{k} c_{jl}^h)^2 \qquad (5)$$

Contributions c_{jl}^h help locate the observations or variables important for a given factor, while correlations $corr(\tilde{X}_j^h, \psi^*)$ help locate the factors important for a given observation or variable.

MCA codes data by creating several binary columns for each variable with the constraint that only one of the columns takes the value 1. This creates artificial additional dimensions because one categorical variable is coded with several columns. As a result, the inertia (i.e. variance) of the solution space is artificially inflated, hence the percentage of inertia explained by the first dimension is severely underestimated. This problem can be alleviated by using, for example, the correction of the data matrix eigenvalues in Greenacre (1993).

3.2 Cluster Analysis

A cluster is usually defined as a collection of data objects (similar to one another within the same cluster and dissimilar to the objects in other clusters). Cluster analysis is a technique of grouping a set of data objects into clusters. It can for example help: (1) marketers discover distinct groups in their customer bases, and then use this knowledge to develop targeted marketing programs; and (2) insurers identifying groups of motor insurance policy holders

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

with a high average claim cost. The clustering methodology in this paper builds on the factors obtained from the MCA. The results of the MCA for our sample application suggest three principal factors (dimensions) as shown in Section 4 of results. The need to know the number of factors which span the data and the initial number of clusters to be formed under the clustering analysis when working with large datasets underscores the importance of conducting the MCA analysis as the first step.

We perform k-means clustering due to the size of the dataset. This method maximizes the between-cluster variance and minimizes the within-cluster variance relative to the mean of the cluster. The within-cluster variation forms homogeneous clusters. The algorithm initially assigns objects to a pre-assigned number of clusters, and these observations are successively reassigned between clusters by minimizing the within-cluster variation. Observations are reassigned to new clusters only when the within-cluster variation is reduced by that reallocation. In our application the initial distribution of observations into the clusters is random, although robustness tests to alternative initial conditions (first observation, last observation and observation by predefined category) produced qualitatively similar results.

To be more precise, let $\mathbb{F} = (F_1, F_2, F_3)$ be the space spanned by the 3 factors obtained from MCA. Each individual $i, (i = 1, \dots, n)$ in the sample is associated with its coordinate $f_i = (f_{ji})_{1\le j\le 3}$. Let $\mathbb{S} = \{S_k; k = 1, \dots, 6\}$ be a partition of the *n* individuals into six sub-categories, $n_k = |S_k|$, and $u_k = (\bar{f}_{jk})_{1\le j\le 3}$, $\bar{f}_{jk} = \frac{1}{n_k}\sum_{i\in S_k} f_{jk}^i$. The algorithm for each individual *i,* solves:

$$min_{S_k\in\mathbb{S}} \|f_i - u_k\|^2 , \qquad (6)$$

where $\|.\|$ is the Euclidean norm. The algorithm has three steps:

1. Specify the initial clusters, $\mathbb{S}_0 = \{S_k^0; k = 1, \dots, 6\}$, and define the centroids $u_k^0 = (\bar{f}_{jk}^0)_{1\le j\le 3}$;
2. For each observation *i*, compute $\|f_i - u_k\|^2$ and set $i \in S_{\bar{k}}^{(t)}$ if $\|f_i - u_{\bar{k}}\|^2 \le \|f_i - u_k\|^2$ for all $k \neq \bar{k}$. Recalculate the centroids $u_k^{(t)} = (\bar{f}_{jk}^{(t)})_{1\le j\le 3}$;
3. Iterate step 2 until $\|f_i - u_{\bar{k}}\|^2 < \epsilon$, for some $\epsilon > 0$.

The number of clusters needs to be pre-assigned in step 1. The initial distribution of observations into the number of clusters indicates how the initial groups centers are to be obtained underscoring the importance of conducting MCA first.

4 Example: Mortgage product choice

The literature on mortgage product choice is primarily focused on whether households select fixed- or adjustable-rate mortgages. The main characteristics which determine this choice are relative interest rates and other mortgage costs, income

ACCEPTED MANUSCRIPT

and wealth constraints and borrower mobility.[5] For example, Campbell and Cocco (2003) find that risk-averse households with a large mortgage size, risky income, high default cost, or low probability of moving tend to prefer fixed-rate mortgages (FRMs). Dhillon et al. (1987), Brueckner and Follain (1988), and Phillips and VanderHoff (1991) conclude that the mortgage costs and interest rates are the most important inputs to mortgage product choice. However, other studies, particularly outside the U.S., underscore the relevance of borrower characteristics,

[5]See Baesel and Biger (1980), Statman (1982), Brueckner (1992), and Posey and Yavas (2001).

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Capone Jr and Cunningham (1992), Sa-Aadu and Sirmans (1995), Coulibaly and Li (2009), Cocco (2013), Dungey et al. (2015) and Dungey et al. (2018).

We consider households who choose one of four different contract styles for an owner-occupied mortgage; variable-rate mortgages (VRMs) which are dominant in the Australian market, short-term fixed-rate mortgages (SFRMs), home equity loans (HEs), and discounted variable-rate ('teaser' or 'honeymoon') loans (HMs). VRMs offer a flexible interest rate that varies with time, following the cost of funds of the bank and the interbank cash rate reported by the Reserve Bank of Australia (RBA); however, rates change at the discretion of the lender.[6] SFRMs offer certainty of repayment for a fixed period shielding borrowers from interest rate risk. In Australia, most SFRMs are fixed for less than 5 years. The Australian market does not feature a long-term fixed interest rate contract such as the ones offered in the U.S. and Denmark. Early repayments are highly penalized for SFRMs; the penalty varies with the outstanding balance, the remaining term, the size of the payment and the current market interest rate.[7] A popular alternative is a discounted variable-rate mortgage, also known as a 'teaser' or 'honeymoon' mortgage (HM), which offers an initial discount from the variable rate for a fixed period of time and also has high early prepayment costs associated. At the other end of the spectrum, to allow home-owners to take advantage of the equity accrued in their home due in part to the large rises in house prices in Australia over the

[6]In particular, the interest rate may be adjusted by the bank at any time (typically with a 14 day notice period to the customer), and it does not follow any particular indicator rate, as is the case with the popular tracker-loans in Europe. For the majority of the sample period in this paper the mortgage rate was closely related to a mark-up over the interbank cash rate reported by the Reserve Bank of Australia (RBA), although relationship was broken as cost of funds became disassociated from official interest rates during the global financial crisis.

[7]That is, the Australian market does not give advantageous terms to refinancing at lower fixed rates as is common in the U.S. with the 'points' system.

ACCEPTED MANUSCRIPT

past 30 years, home equity loans (HEs) offer funds secured against the equity of a home, with funds available for any personal use and accessible in diverse ways. A number of other loans exist which are not studied here due to their very small representation.

The major Australian banks source most of their funding from domestic deposits and wholesale debt, and in this sense, most of the risks associated with mortgages are held on banks' balance sheets. Four major players dominate the Australian financial landscape.[8] During the sample period (January 2003 to August 2008), 60 percent of owner-occupier loan approvals were made by the four major banks, and 20 percent by smaller banks; see Davies (2009) who also document the relatively low level of mortgage securitization. Mortgage insurance is private in Australia; it covers the lender in the event of borrower default but the cost is borne by the borrower. Australian owner-occupied homes are free from capital gains tax, and since 2000 first-time home buyers (FHBs) receive federal and state cash-grant incentives to purchase their first house; these incentives vary across time and region as shown in Dungey et al. (2011).[9]

4.1 Empirical Results

The dataset used in this study is unique in that it contains detailed information on 577,444 mortgage applications generated by a major bank in Australia between January 2003 and August 2008.[10]. The bank, which remains anonymous, provided

[8] In alphabetical order: Australia and New Zealand Bank (ANZ), Commonwealth Bank of Australia (CBA), National Australia Bank (NAB), and Wetspac Banking Corporation (WBC).

[9] More details on the Australian housing and mortgage market environment can be found in Dungey et al. (2015, 2018); Yanotti and Dungey (2014); Yates (2011).

[10] This type of database is hard to come by, and although it does not cover the most recent period it provides a clear exemplar of the possibilities of MCA. It is not able to be updated

ACCEPTED MANUSCRIPT

all information collected and used in the mortgage contracting process, see Dungey et al. (2015) for details. The descriptive statistics are given Table 1 and a brief description of the variables used is provided in Table 2.

One group of variables concerns household structure and demographics, and includes borrower age, gender, marital status, number and age of dependent children. Loan servicing capacity is captured by income quartiles and employment status; in robustness tests we included occupational categories. Additional information on their financial position is provided by net wealth quartiles. As a proxy for mobility we use the time in years spent at the current address. We also discern borrowers who are first-time home buyers (FHBs) from those who are repeated buyers, and those who apply alone or with a co-borrower. The borrower characteristic variables used in our analysis are defined in Table 2.

The average borrower is male, married, aged 41 years old, with gross monthly income of AUD$7,266 and a stock of net wealth of AUD$416,183 at time of application, and he chooses a VRM mortgage product; see Table 1. The location of this average applicant is shown at the intersection of the axes in Figure1 which plots out all the different characteristics located in the Dimension 1 and Dimension 2 space from the MCA.

Based on the locational information in Figure 1 the MCA co-ordinate plot in Figure 2 shows more specifically how the values of different characteristics are located in space relative to the average. Plotting the different locations of age groups, numbers of dependents, income and wealth quartiles shows clearly how dimension 1 is strongly determined by young applicants who have no dependents, have low income and low wealth.

The second dimension, on the vertical axis in Figure 2, contrasts income and

ACCEPTED MANUSCRIPT

wealth. The second dimension contrasts young families (married borrowers under 40 years old), who have low net wealth but high income, have dependents under 5 years old and tend to be employees – with old or senior households – borrowers over 40 years old, who are highly immobile, have low income but high net wealth; see details in Figure 1.

Figure 3 highlights the dimension 1 and dimension 2 spaces and suggests 4 possible groupings, and also locates the available mortgage products in the space. This suggests for example, that the group at the right of the figure captures single applicants, low wealth, low income, young, first-home buyers. It should also be apparent that in these two dimensions there are no products that are contained in the grouping for these applicants. In contrast, the bottom left grouping are older, with dependents and higher income and wealth and self-employed. The HE loans are located in this dimension space.

Figure 4 shows how the groups rearrange when considered in a different space, between dimension 1 and dimension 3. In particular the older, self-employed, higher wealth and income group are now high on the dimension 3 axis (but still located in a similar place on dimension 1). The placement of the groupings in Figure 4 strongly suggest that dimension 3 is related to mobility. In addition to the four groups located in Figure 3, it suggests two further possible groupings.

4.2 The typologies and product location

The results of the MCA suggest three principal factors (dimensions) - income and wealth, income versus wealth, and mobility - and six distinct clusters with which to build a borrower typology. The borrower characteristics are distinguished in ways

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

that clearly identify them as follows (Table 3 presents the descriptive statistics for each group):

1. **Constrained (female) households** comprise mainly female borrowers (64%), aged between 30 and 60 years old, two thirds of whom are single, and the overwhelming majority with no dependents (81%). They have lived at their current address on average for 7.6 years, have income in the lowest quartile, are mainly employees, with medium wealth.

2. **Risk averse, constrained families** are borrowers between 30 and 40 years old, mostly married with a co-borrower, and have dependents under 5 years old, who have lived on average at their current address for only 3 years, have medium income and medium wealth, are mainly employees.

3. **Seniors** are mostly over 50-year old, married borrowers who apply with a co-borrower, have lived at their current address on average for 12 years, have high income and large wealth, and have no dependents.[11]

4. **Young first-home buyers** are primarily under 30 years old borrowers, who have lived at their current address on average for 4 years, have medium income and the lowest wealth levels, are mostly employees, and have no dependents; only 20% of them are married and 45% are first-time home buyers.

5. **Settled families** are mainly married borrowers between 40 and 50 years old, who have a co-borrower, have lived in their current address on average

[11] Children or adults living at home and receiving income support.

ACCEPTED MANUSCRIPT

for 7 years, with medium income and medium wealth (13% of them are self-employed), and dependents between 5 and 15 years old.

6. **Low risk households** describes married borrowers in their 30s and 40s, who have lived in their current address on average for 5 years, have no dependents or have dependents over 15 years old, and in particular have very high income and wealth (27.5% of them are self-employed).

The results are summarised in Table 4. They reveal that medium income, medium wealth borrowers are clustered in an area with more product choices, perhaps because this market is most attractive to the bank. Higher income, wealthier, lower risk borrowers take VRMs. The low-risk families and settled families – clusters 5 and 6 – are well served by this product. Settled families access three products which are either contained within or on the very edge of their characteristics cluster – the position of the HM shows that most settled families are already well established financially and do not wish to take advantage of the more restrictive financial terms of a HM.

Constrained (female) households – cluster 1 – are mainly served by HMs designed to attract newer and more constrained borrowers. These borrowers are the most income constrained in the sample, suggesting they are seeking the lower initial repayments in HMs. These borrowers are expected to be highly sensitive to interest rate movements.

Young risk averse families – cluster 2 – prefer SFRMs, as predicted by the mortgage choice literature; see Campbell and Cocco (2003), Coulibaly and Li (2009) and Dungey et al. (2015). These loans offer households certainty in their repayments for a period between 3-5 years on average. We observe in this cluster

ACCEPTED MANUSCRIPT

a group of borrowers (with revealed characteristics to the left of the vertical line in Figure 5) which appear to be under-served; they do not access mortgage products associated with their characteristics, and may be well suited for a new type of mortgage product. This sub-set of borrowers are slightly older, with higher income but lower wealth. Longer-term FRMs, such as those 30-year FRMs offered in U.S., Denmark and New Zealand, may be a better fit for this sub-set of more risk-averse households; see also Followill (1998).

Mobile FHBs – cluster 4 – take HMs and SFRMs, which offer them lower initial payments and short-term certainty in repayments; in line with the findings in Sa-Aadu and Sirmans (1995) where young borrowers with income growth expectations prefer short-term interest rate fixity. The option of HMs may be allowing these borrowers to access housing credit earlier than otherwise. This cluster also exhibits a sub-set with characteristics spread away from the mortgages offered, which suggests that there may be potential for designing new mortgage products to better serve these FHBs. Miles and Pillonca (2008) suggests an indexed mortgage with a flexible equity-share component: a mortgage product that sets repayments by reference to a real interest rate and not a nominal one covering FHBs from shifts in nominal rates and a mortgage with features where borrowers and lenders share the equity of a house to decrease exposure to house price movements. Other alternatives could be stepped-up fixed-rate mortgages and interest-only periods; see Miles (2004).

HEs are designed to allow homeowners to smooth consumption through the equity of their property, particularly in periods of strong house price growth. Borrowers over 50 with high income and wealth – cluster 3 – find these products suitable. It is apparent that the HE products on offer are still placed on the edge

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

of the characteristics of cluster 3, and not near any of the other groups.

In contrast to the predictions in the mortgage choice literature, the two profiles with borrowers who have spent less time at their current address, a variable expected to proxy for mobility, do not select VRMs, but rather SFRMs and HMs. A number of characteristics of the Australian markets may drive these results. On the one hand, the mobility motive on VRMs may not be strong in the Australian mortgage market because SFRMs offer fixed rates only for a short term, and there are substantial costs to moving associated with stamp duty charges; see Dungey et al. (2011). On the other hand, although it has been used previously in the literature, the time at current address may not be a good proxy for mobility as it is correlated with borrower age – although this correlation is not particularly strong within this sample data at 0.29.

In all, the comparison of the characteristics of the data-generated clusters and the characteristics of the loans help us to identify groups which are well serviced and those which may not be. Risk-averse young families and mobile first home buyers are currently serviced by HM and SFRM products. They are trading off their relative lack of wealth for the more restrictive terms. However, the mobile FHBs have a substantial bulk of characteristics to the right of these products in dimension 1, reflecting that they would prefer products with even lower wealth requirements. Risk-averse young families show characteristics that indicate they have a preference for keeping the loan payments down, with more bulk of the cluster showing higher income constraint than the product offerings allow. The HE product is particularly interesting in that it is clearly designed to meet some of the needs of the seniors. However, even then it is not centrally located in their preferences. Despite their revealed lower mobility in the descriptive statistics

ACCEPTED MANUSCRIPT

they choose the products which are most suited for mobile borrowers (VRMs). This mainly reflects that the VRM is the loan with the most favourable terms in general, and thus the seniors will be prepared to use a product on the edge of their preference cluster.

We acknowledge that these results at least partly reflect the bank's business and marketing strategy, and is only representative of part (60%) of the Australian mortgage market. The results also show that more risk averse borrowers select SFRMs, but we do not find evidence of a strong mobility motive to take VRMs. In this sense borrowers seem to be taking the mortgage products closest to their characteristics based on the current offerings. We do not observe evidence of any predatory lending, but very well targeted product differentiation and price discrimination by lenders.

Our results are robust to different period sub-samples and different choices within the application of both the MCA and cluster analysis implementations. Similar results were obtained for a sample between January 2003 and May 2009, and sub-samples in January 2003 and January 2008, using the Burt, adjusted Burt and Indicator matrices in the MCA, and by applying joint correspondence analysis (JCA). We tried alternative combinations of active and supplementary variables. We defined clustering processes using k-means and k-medians algorithms and experimented with alternate assigned number of clusters and different distance measures and starting conditions. The only distance measure definition that appears to provide some change in result is that of absolute distance. The qualitative results discussed in this paper are maintained throughout this battery of robustness tests.

ACCEPTED MANUSCRIPT

5 Conclusion

This paper introduces the use of MCA techniques to the financial product choice literature as a means of providing evidence to the debate on financial inclusion. The MCA technique identifies groups from within a sample containing categorical data with common characteristics.

Our contention is that identifying the defining consumer characteristics, and aligning them with the product choices will reveal empirical evidence on gaps in the suite of product offerings which may contribute to financial exclusion. Strategies to improve financial inclusion may well include reducing the degree of compromise required from those currently under-serviced by expanding product offerings. We demonstrate how this information can be revealed using MCA and clustering analysis for an example based on Australian mortgage market data. This reveals that customers who are attractive to the financial institution (in terms of low risk and good return) are well serviced with a relatively large range of choices. Some customer groups, however, are under-serviced, female headed households desire greater recognition of wealth constraints and there is a relatively poor match between the product and borrower characteristics for young first-home buyers in each of the wealth, income and mobility dimensions. Product development which more closely meets consumer revealed needs is likely to be as important to extending financial inclusion as technology acceptance, financial literacy and overcoming lack of infrastructure. MCA provides an excellent tool for analyzing the often categorical data available on households and consumers and assessing the alignment of products which may best meet their financial needs.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

References

Asselin, L. and V. Anh (2008): "Multidimensional poverty measurement with multiple correspondence analysis," in *Quantitative Approaches to Multidimensional Poverty Measurement*, ed. by N. Kakwani and J. Silber, Palgrave.

Atkinson, A. and F. Messy (2013): "Promoting Financial Inclusion through Financial Education: OECD/INFE Evidence, Policies and Practice," OECD Working Papers on Finance, Insurance and Private Pensions No. 34, OECD.

Baesel, J. and N. Biger (1980): "The allocation of risk: Some implications of fixed versus index-linked mortgages," *Journal of Financial and Quantitative Analysis*, 15, 457–68.

Black, H. A., B. L. Robinson, A. M. Schlottmann, and R. L. Schweitzer (2003): "Is race an important factor in bank-customer preferences? The case of mortgage lending," *Journal of Real Estate Finance and Economics*, 26, 5–26.

Boyd, W. L., M. Leonard, and C. White (1994): "Customer preferences for financial services: An analysis," *International Journal of Bank Marketing*, 12, 9–15.

Brueckner, J. (1992): "Borrower mobility, self-selection, and the relative prices of fixed-and adjustable-rate mortgages," *Journal of Financial Intermediation*, 2, 401–421.

Brueckner, J. and J. Follain (1988): "The rise and fall of the ARM: An econometric analysis of mortgage choice," *Review of Economics and Statistics*, 70, 93–102.

Campbell, J. Y. and J. F. Cocco (2003): "Household risk management and optimal mortgage choice," *Quarterly Journal of Economics*, 118, 1449–1494.

Capone Jr, C. A. and D. F. Cunningham (1992): "Estimating the marginal contribution of adjustable-rate mortgage selection to termination probabilities in a nested model," *Journal of Real Estate Finance and Economics*, 5, 333–357.

Cocco, J. F. (2013): "Evidence on the benefits of alternative mortgage products," *Journal of Finance*, 68, 1663–1690.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Coulibaly, B. and G. Li (2009): "Choice of mortgage contracts: Evidence from the survey of consumer finances," *Real Estate Economics*, 37, 659–673.

Davies, M. (2009): "Household debt in Australia," in *Household debt: Implications for monetary policy and financial stability*, Bank for International Settlements, 46, 19–30.

Demirguc-Kunt, A. and L. Klapper (2013): "Measuring financial inclusion: Explaining variation in use of financial services across and within countries," *Brookings Papers on Economic Activity*, Spring, 279–340.

Devlin, J. F. (2002): "An analysis of choice criteria in the home loans market," *International Journal of Bank Marketing*, 20, 212–226.

Dhillon, U., J. Shilling, and C. Sirmans (1987): "Choosing between fixed and adjustable rate mortgages: Note," *Journal of Money, Credit and Banking*, 19, 260–267.

Dungey, M., F. Doko Tchatoka, G. Wells, and M. B. Yanotti (2015): "Mortgage choice determinants: The role of risk and bank regulation," *Economic Record*, 91, 417–437.

Dungey, M., F. Doko Tchatoka, and M. B. Yanotti (2018): "Endogeneity in Household Mortgage Choice," *Economic Modelling*, 73, 30–44.

Dungey, M., G. Wells, and S. Thompson (2011): "First home buyers' support schemes in Australia," *Australian Economic Review*, 44, 468–479.

Dunn, K. B. and C. S. Spatt (1988): "Private information and incentives: Implications for mortgage contract terms and pricing," *Journal of Real Estate Finance and Economics*, 1, 47–60.

Finkelstein, A., E. F. P. Luttmer, and M. J. Notowidigdo (2012): "What good is wealth without health? The effect of health on the marginal utility of consumption," *Journal of the European Economic Association*, 11, 221–258.

Followill, R. (1998): "An analytical comparison of the durations and price sensitivities of fixed-rate, constant payment and constant amortization mortgages," *International Review of Financial Analysis*, 7, 51–64.

ACCEPTED MANUSCRIPT

Gabaix, X. and D. Laibson (2006): "Shrouded attributes, consumer myopia, and information suppression in competitive markets," *Quarterly Journal of Economics*, 121, 505–540.

Gary-Bobo, R. J. and S. Larribeau (2004): "A structural econometric model of price discrimination in the French mortgage lending industry," *International Journal of Industrial Organization*, 22, 101–134.

Gower, J. C. (1966): "Some distance properties of latent root and vector methods used in multivariate analysis," *Biometrika*, 53, 325–338.

Green, R. K. and S. M. Wachter (2005): "The American mortgage in historical and international context," *Journal of Economic Perspectives*, 19, 93–114.

Greenacre, M. J. (1988): "Correspondence analysis of multivariate categorical data by weighted least-squares," *Biometrika*, 75, 457–467.

——— (1993): "Biplots in correspondence analysis," *Journal of Applied Statistics*, 20, 251–269.

Guttman, L. (1944): "A basis for scaling qualitative data," *American Sociological Review*, 9, 139–150.

Hayashi, C. (1950): "On the quantification of qualitative data from the mathematico-statistical point of view," *Annals of the Institute of Statistical Mathematics*, 2, 35–47.

Hill, M. O. (1974): "Correspondence analysis: A neglected multivariate method," *Applied Statistics*, 23, 340–354.

Hirschfeld, H. (1935): "A connection between correlation and contingency," in *Proceedings of the Cambridge Philosophical Society*, Cambridge University Press, vol. 31, 520–524.

Hoffman, D. and G. Franke (1986): "Correspondence analysis: Graphical representation of categorical data in marketing research," *Journal of Marketing Research*, 23, 213–227.

Jagtiani, J. and C. Lemieux (2017): "Fintech lending: Financial inclusion, risk pricing, and alternative information," Federal Reserve Bank of Philadelphia Working Paper No. 17-17, Federal Reserve Bank of Philadelphia.

ACCEPTED MANUSCRIPT

Koijen, R., S. Van Nieuwerburgh, and M. Yogo (2016): "Health and Mortality Delta: Assessing the Welfare Cost of Household Insurance Choice," *Journal of Finance*, 71, 957–1009.

Kolenikov, S. and G. Angeles (2009): "Socioeconomic status measurement with discrete proxy variables: Is principal component analysis a reliable answer?" *Review of Income and Wealth*, 55, 128–165.

Le Roux, B. and H. Rouanet (2004): *Geometric data analysis: From correspondence analysis to structured data analysis*, Dordrecht, Kluwer.

Luarn, P. and H.-H. Lin (2005): "Toward an understanding of the behavioral intention to use mobile banking," *Computers in Human Behavior*, 21, 873–891.

Manté, C., G. Bernard, P. Bonhomme, and D. Nerini (2013): "Application of ordinal correspondence analysis for submerged aquatic vegetation monitoring," *Journal of Applied Statistics*, 40, 1619–16388.

Mendes, S., M. J. Fernández-Gómez, M. Pereira, U. Miranda Azeitero, and P. Galindo Villardón (2012): "An empirical comparison of canonical correspondence analysis and STATICO in the identification of spatio-temporal ecological relationships," *Journal of Applied Statistics*, 5, 979–994.

Miles, D. (2004): "The UK mortgage market: Taking a longer-term view. Final report and recommendations," HMSO Treasury, London.

Miles, D. and V. Pillonca (2008): "Financial innovation and European housing and mortgage markets," *Oxford Review of Economic Policy*, 24, 145–175.

Morawczynski, O., D. Hutchful, E. Cutrell, and N. Rangaswamy (2010): "The Bank Account is Not Enough: Examining Strategies for Financial Inclusion in India," in *Proceedings of the 4th ACM/IEEE International Conference on Information and Communication Technologies and Development*, New York, NY, USA: ACM, ICTD '10, 24:1–24:11.

Mylonakis, J. (2007): "A bank customer analysis and mortgage services evaluation: Implications of market segmentation policies," *Banks and Bank Systems*, 2, 157–157.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Phillips, R. A. and J. VanderHoff (1991): "Adjustable- versus fixed-rate mortgage choice: The role of initial rate discounts," *Journal of Real Estate Research*, 6, 39–52.

Piskorski, T. and A. Tchistyi (2010): "Optimal mortgage design," *Review of Financial Studies*, 23, 3098–3140.

Posey, L. and A. Yavas (2001): "Adjustable and fixed rate mortgages as a screening mechanism for default risk," *Journal of Urban Economics*, 49, 54–79.

Sa-Aadu, J. and C. Sirmans (1995): "Differentiated contracts, heterogeneous borrowers, and the mortgage choice decision," *Journal of Money, Credit and Banking*, 27, 498–510.

Sourial, N., C. Wolfson, B. Zhu, J. Quail, J. Fletcher, S. Karunananthan, K. Bandeen-Roche, F. Béland, and H. Bergman (2010): "Correspondence analysis is a useful tool to uncover the relationships among categorical variables," *Journal of Clinical Epidemiology*, 63, 638–646.

Spearman, C. (1904): "'General intelligence', objectively determined and measured," *The American Journal of Psychology*, 201–292.

Stafford, M. R. (1996): "Demographic discriminators of service quality in the banking industry," *Journal of Services Marketing*, 10, 6–22.

Stanton, R. and N. Wallace (1998): "Mortgage choice: What's the point?" *Real Estate Economics*, 26, 173–205.

Statman, M. (1982): "Fixed rate or index-linked mortgages from the borrower's point of view: A note," *Journal of Financial and Quantitative Analysis*, 17, 451–457.

Talaga, J. A. and J. Buch (1998): "Consumer trade-offs among mortgage instrument variables," *International Journal of Bank Marketing*, 16, 264–270.

Tenenhaus, M. and F. Young (1985): "An analysis and synthesis of multiple correspondence analysis, optimal scaling, dual scaling, homogeneity analysis and other methods for quantifying categorical multivariate data," *Psychometrika*, 50, 91–119.

ACCEPTED MANUSCRIPT

Warnock, V. C. and F. E. Warnock (2008): "Markets and housing finance," *Journal of Housing economics*, 17, 239–251.

Yanotti, M. and M. Dungey (2014): "Mortgage Product Choice in Australia: The Impact of Market Stress," *JASSA, The Finsia Journal of Applied Finance*, 44–52.

Yates, J. (2011): "Housing in Australia in the 2000s: On the agenda too late?" in *The Australian economy in the 2000s*, Reserve Bank of Australia, 261–296.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Table 1: Summary statistics for loan and borrower characteristics across loan types

Variables	VRM	HM	SFRM	HE	All
Loan Amount	$231,297	$167,263	$190,021	$179,584	$207,398
	(169,116)	(92,977)	(106,661)	(160,778)	(147,988)
Repayment	$1,585	$1,114	$1,354	$0	$1,360
	(1,255)	(676)	(789)	(0)	(1,104)
Interest Rate	7.13%	6.63%	7.34%	7.82%	7.08%
	(0.84)	(0.10)	(0.78)	(1.20)	(0.94)
LTV	61.60%	63.00%	66.20%	56.15%	62.37%
	(21.90)	(20.78)	(20.21)	(19.75)	(21.39)
PTIR	14.40%	10.51%	14.38%	25.23%	14.02%
	(16.16)	(13.58)	(14.88)	(20.86)	(15.95)
Valuation	$397,979	$293,490	$352,778	$496,864	$371,348
	(931,230)	(152,611)	(207,613)	(302,141)	(711,659)
Age	41 yrs	40 yrs	38.5 yrs	47 yrs	41 yrs.
	(10.60)	(11.06)	(10.13)	(10.75)	(10.77)
# Dependent	0.80	0.82	0.79	0.73	0.79
	(1.08)	(1.11)	(1.09)	(1.08)	(1.09)
Age Youngest Dependent	6.5 yrs	6.5 yrs	6 yrs	7 yrs	6.5 yrs
	(5.08)	(4.94)	(4.79)	(5.20)	(5.01)
Time at current address	6.3 yrs	6.3 yrs	5.2 yrs	8.2 yrs	6.2 yrs
	(6.87)	(6.77)	(6.15)	(7.66)	(6.80)
Gross Monthly Income	$7,862	$5,614	$6,927	$9,392	$7,266
	(5,371)	(3,585)	(4,549)	(6,847)	(5,079)
Net Wealth	$447,658	$320,495	$344,003	$744,069	$416,183
	(409,657)	(305,372)	(332,739)	(550,495)	(395,967)
Females	30%	34%	34%	25%	31%
Married	70%	64%	65%	78%	68%
FHB	8.5%	11%	11%	1.2%	9%
Co-borrower	71.5%	62%	67%	69%	68%
Self-employed	18%	15%	14%	33%	17%
Total	324,431	134,609	90,370	28,034	577,444

Means, (standard deviations) and proportions. Sample Jan2003-Aug2008. Prices are in real terms Q1 2006.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Table 2: Borrower characteristic variables used in MCA.

Description	Categories
Age of the main borrower.	Age under 30 yrs. (< 30); age between 30-39 yrs. ($30-39$); age between 40-49 yrs. ($40-49$); age between 50-59 yrs. ($50-59$); age 60 yrs and over (≥ 60).
Gender of the main borrower	Male, Female.
Marital status of the main borrower	Single; Married (married or in a de-facto relationship).
Presence of dependents according to their age	No dependents (No_Dpndnt); dependents under 5 yrs. ($Dpndnt\_U5$); dependents between 5 and 15 yrs. ($Dpndnt\_O5U15$); dependents older than 15 yrs. ($Dpndnt\_O15$).
Main borrower income quartile.	$I1$; $I2$; $I3$; $I4$.
Employment status for the main borrower	Employee (Emp); self-employed ($Self\_Emp$).
Main borrower net wealth quartile.	$W1$; $W2$; $W3$; $W4$.
Borrower's reported years spent at their current address.	$t < 2$ ($M1$); $2 \leq t < 4$ ($M2$); $4 \leq t < 6$ ($M3$); $6 \leq t < 8$ ($M4$); $t \geq 8$ ($M5$).
First-time home buyer	repeat buyer (non_FHBs); first-time home buyer ($FHBs$).
Presence of a co-borrower.	Single applicant ($Single\_App$); Co-borrower ($CoBorrwr$).
Mortgage type	adjustable-rate mortgage (VRM); short-term fixed-rate mortgage ($SFRM$); 'honeymoon' loans (HM); home equity loans (HE).

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Table 3: Clusters Overview

Variables	Clusters					
	1	2	3	4	5	6
Average age	44.30 yrs	34.34 yrs	54.13 yrs	29.00 yrs	42.01 yrs	42.77 yrs
(standard deviation)	(10.56)	(4.68)	(8.73)	(6.74)	(5.59)	(8.89)
Proportions						
Age<30	6.96%	12.68%	1.12%	67.55%	0.47%	4.68%
Age 30-39	25.75%	77.16%	2.88%	26.76%	29.32%	34.32%
Age 40-49	37.26%	9.91%	19.19%	3.26%	64.19%	38.47%
Age 50-59	22.65%	0.25%	52.71%	1.93%	5.71%	19.23%
Age ≥60	7.38%	0.01%	24.10%	0.50%	0.30%	3.30%
Risk Aversion						
Average number of dependents	0.31	1.43	0.29	0.09	1.62	0.78
(standard deviation)	(0.73)	(1.11)	(0.73)	(0.39)	(1.11)	(1.11)
Average age youngest dependent	9.87	2.83	13.27	2.80	8.61	6.88
(standard deviation)	(4.62)	(2.53)	(5.17)	(2.94)	(4.24)	(4.84)
Proportions						
No dependents	80.91%	24.96%	83.35%	94.05%	19.84%	61.18%
Dependents < 5 yrs	2.47%	64.23%	0.59%	5.06%	12.53%	15.32%
Dependents 5-15 yrs	14.38%	10.77%	9.09%	0.84%	64.01%	21.69%
Dependents > 15 yrs	2.25%	0.04%	6.97%	0.05%	3.62%	1.81%
Female	64.34%	16.86%	22.32%	29.88%	38.30%	15.40%
Married	32.86%	87.73%	83.43%	21.08%	93.54%	87.21%
Co-borrower	13.55%	95.34%	76.76%	44.01%	85.85%	92.51%
FHB	2.08%	5.80%	1.21%	45.38%	0.36%	1.90%
Mobility						
Average time at cur address	7.66 yrs	2.84 yrs	11.80 yrs	4.32 yrs	6.99 yrs	5.16 yrs
(standard deviation)	(7.63)	(2.68)	(9.07)	(6.55)	(5.40)	(5.25)

Means and (standard deviations).

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Table 3 – continued from previous page

Variables	**Clusters**					
	1	**2**	**3**	**4**	**5**	**6**
Proportions						
< 2yrs cur addrss	16.88%	36.32%	8.34%	42.98%	11.89%	23.30%
2-4 yrs cur addrss	19.11%	32.37%	10.72%	29.31%	15.47%	25.27%
4-6 yrs cur addrss	17.20%	18.82%	10.57%	9.84%	20.00%	19.56%
6-8 yrs cur addrss	10.13%	7.97%	6.02%	2.61%	17.29%	10.03%
> 8 yrs cur addrss	36.67%	4.52%	64.35%	15.26%	35.35%	21.83%
Income Risk						
Average gross monthly income	$4,202	$7,335	$9,403	$6,076	$6,252	$10,145
(standard deviation)	(2,147)	(3,711)	(7,366)	(3,522)	(2,960)	(6,382)
Proportions						
1st income quartile	62.27%	10.27%	18.53%	32.42%	22.20%	5.67%
2nd income quartile	24.13%	31.31%	14.86%	29.27%	36.38%	12.80%
3rd income quartile	7.46%	36.45%	21.09%	23.87%	29.88%	28.14%
4th income quartile	2.14%	21.97%	45.49%	14.45%	11.54%	53.39%
Self-employed	9.84%	7.40%	48.74%	5.64%	12.93%	27.50%
Wealth Risk						
Average wealth stock	$287,377	$282,028	$900,930	$144,920	$343,811	$623,562
(standard deviation)	(203,929)	(202,832)	(565,512)	(162,346)	(200,222)	(428,681)
Proportions						
1st wealth quartile	26.08%	29.87%	2.67%	77.98%	8.66%	8.00%
2nd wealth quartile	38.15%	35.56%	4.36%	12.53%	40.96%	11.16%
3rd wealth quartile	27.81%	26.60%	14.35%	6.43%	41.25%	26.70%
4th wealth quartile	7.96%	7.97%	78.61%	3.06%	9.13%	54.14%
Total	88,514	101,631	60,449	76,341	84,445	91,216

Means and (standard deviations).

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

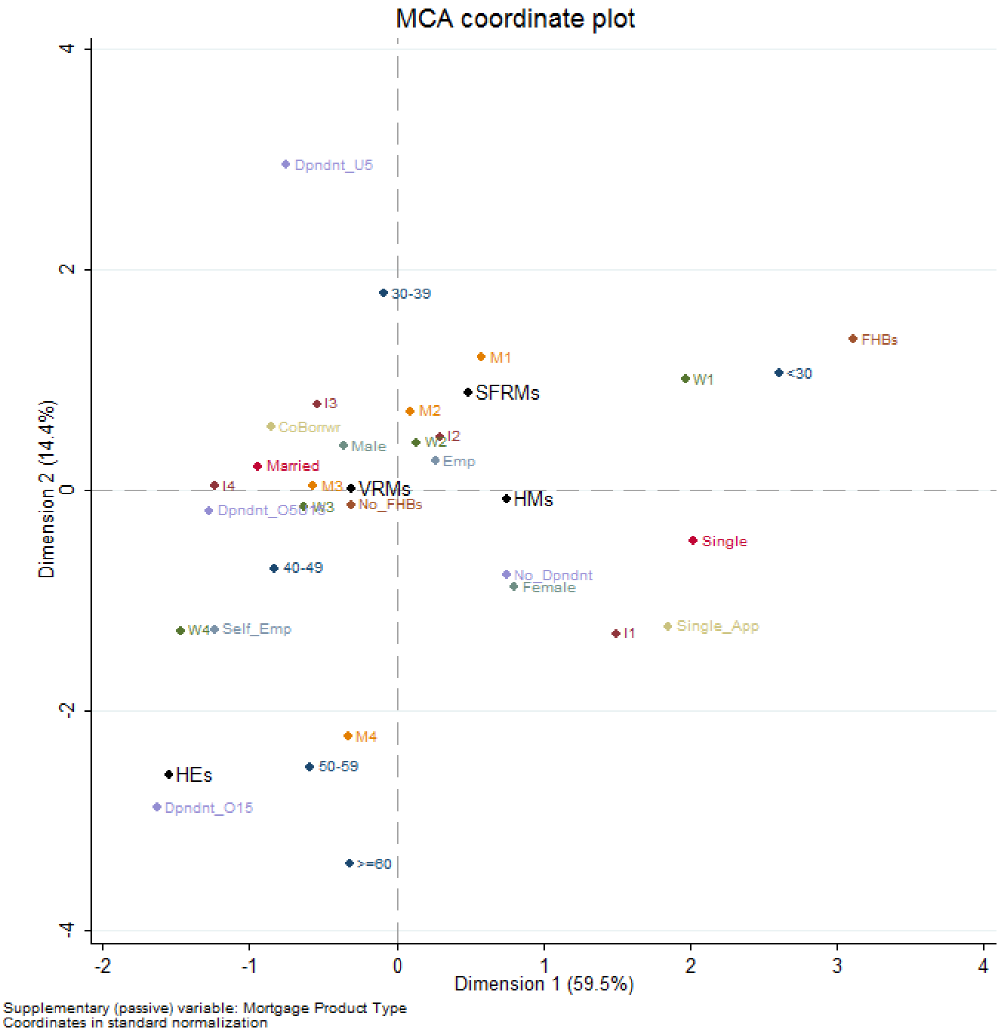
Table 4: Borrowers typologies

Cluster	Profile	Borrower Characteristics	Mortgage type
1	Constrained (female) household	Females, low income, medium wealth, 40-50 yrs., no dependents, not mobile.	HMs
2	Young, risk averse, constrained families	Married, co-borrowers, 30-40 yrs., dependents under 5 yrs., medium income, medium wealth, mobile.	SFRMs
3	Senior	Self-employed, 50-60 yrs., married, high income, high wealth, co-borrower, no dependents, not mobile.	HEs, VRMs
4	Young FHBs	Under 30 yrs., FHBs, low wealth, medium income, mobile, no dependents.	SFRMs, HMs
5	Settled families	40 yrs., medium income, medium wealth, married, co-borrower, dependents 5-15 yrs., mobile.	HMs, SFRMs, VRMs
6	Low-risk households	30-50 yrs., high income, high wealth, self-employed, married, co-borrower.	VRMs

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

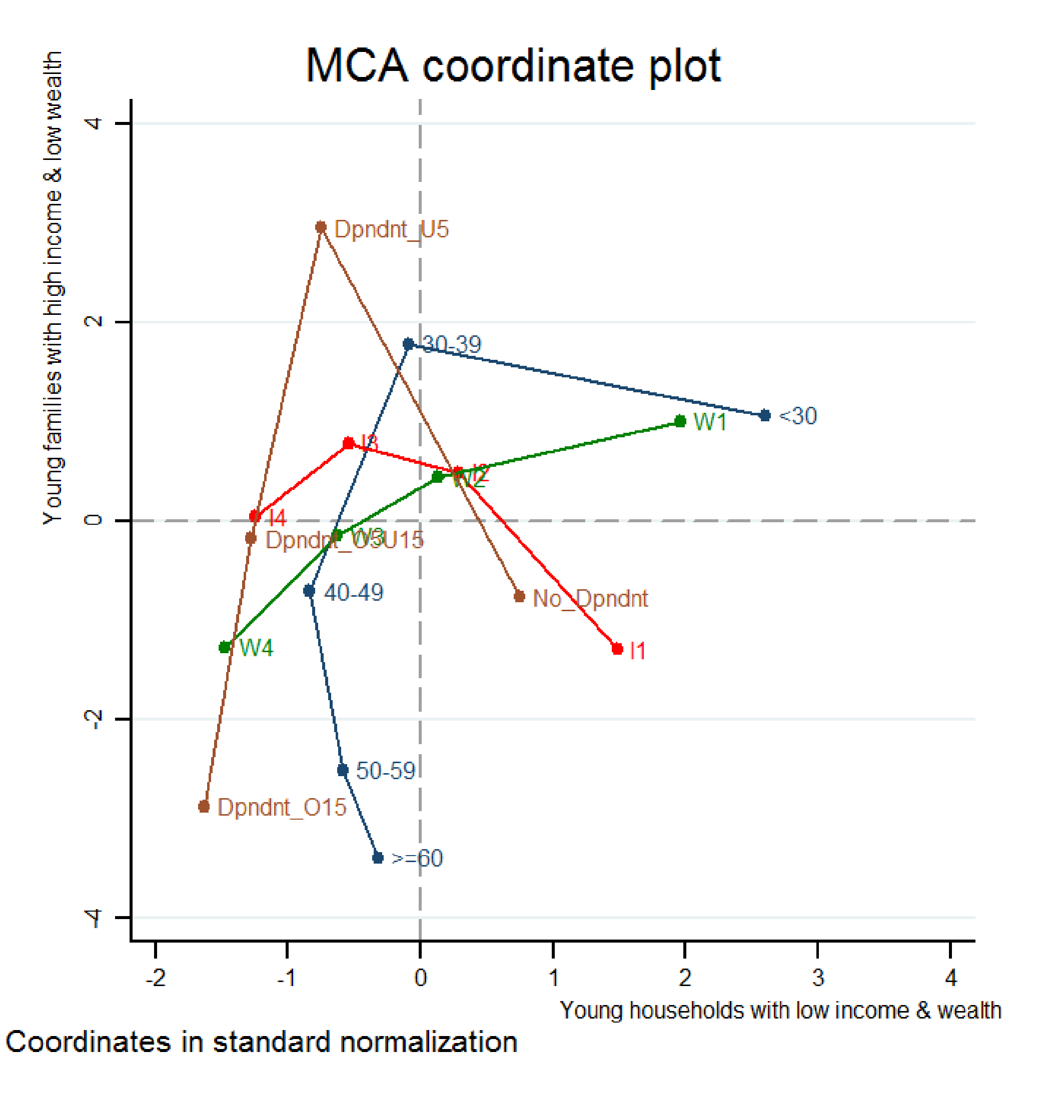
Figure 1: Multiple correspondence analysis



Each point represents the coordinates for a category in dimensions 1 and 2. Points away from the origin show categories that deviate from the mean. The four main mortgage types are depicted in black: *VRM* (variable-rate mortgages), *SFRM* (short-term fixed-rate mortgages), *HM* ('honeymoon' mortgages) and *HE* (home equity loans). The rest of the categories for borrower characteristic are depicted in colors. Age categories: <30, $30-39$, $40-49$, $50-59$, $>=60$. Income quartiles: $I1$, $I2$, $I3$ and $I4$. Wealth quartiles: $W1$, $W2$, $W3$ and $W4$. Marital status: *Single* and *Married*. Mobility categories go from more mobile to less mobile: $M1$, $M2$, $M3$, $M4$, $M5$. Gender: *Female* and *Male*. Dependents: *No_Dpndnt*, *Dpndnt_U*5, *Dpndnt_O*5*U*15, and *Dpndnt_O*15. Fist-time home buyers: *FHBs*, *non_FHBs*. Co-borrowers: *CoBorrwr*, *Single_App*. Employment status: *Self _Emp* and *Emp*.

ACCEPTED MANUSCRIPT

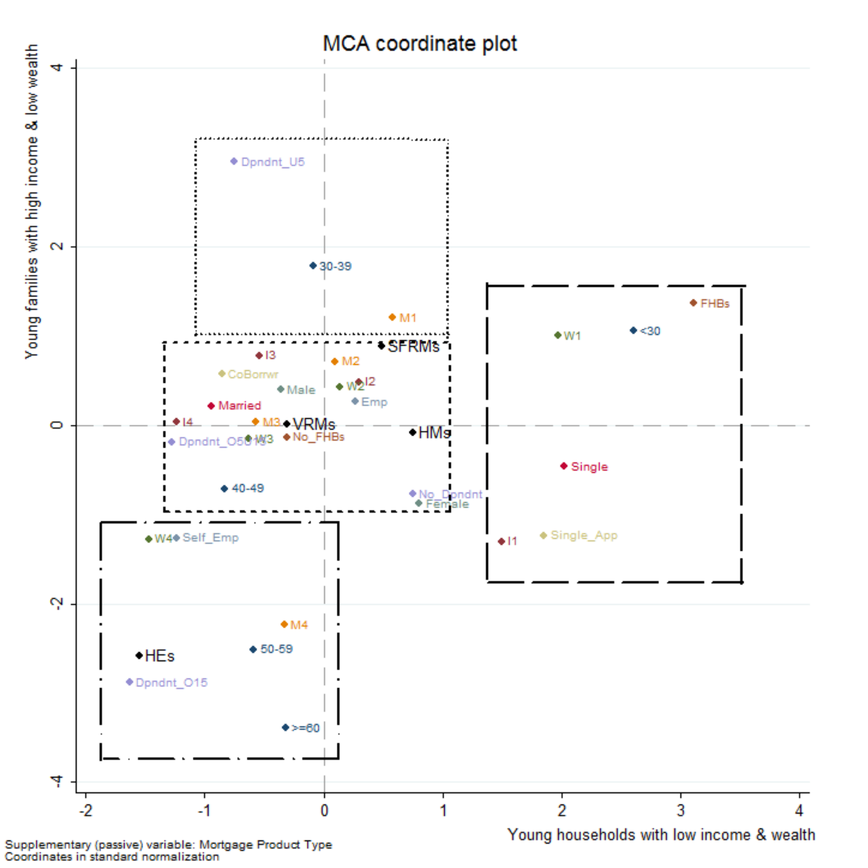
Figure 2: Multiple correspondence analysis: dimensions 1 and 2



Dimension 1 contrasts low income and wealth levels with high income and wealth levels. The positive values of dimension 1 show borrowers under 30 years old, with no dependents and income and wealth levels in the 1st and 2nd quartiles. The negative side of the horizontal axis shows borrowers over age 30, with dependents and income and wealth levels in the 3rd and 4th quartiles. The positive side of dimension 1 reveals young households with low income and wealth, while the negative side shows families with high income and wealth. **Dimension 2** contrasts high income and low wealth levels with low income and high wealth levels. The positive values of dimension 2 show borrowers under 40 years, with dependents under 5 years old, wealth in the 1st and 2nd quartile, but income above the 1st quartile. These borrowers represent young families with high income but low wealth levels. The negative values of the vertical axis reveals borrowers who are over 40 years old, have dependents over 5 years old or no dependents, and have wealth in the 3rd and 4th quartiles but income in the 1st quartile.

ACCEPTED MANUSCRIPT

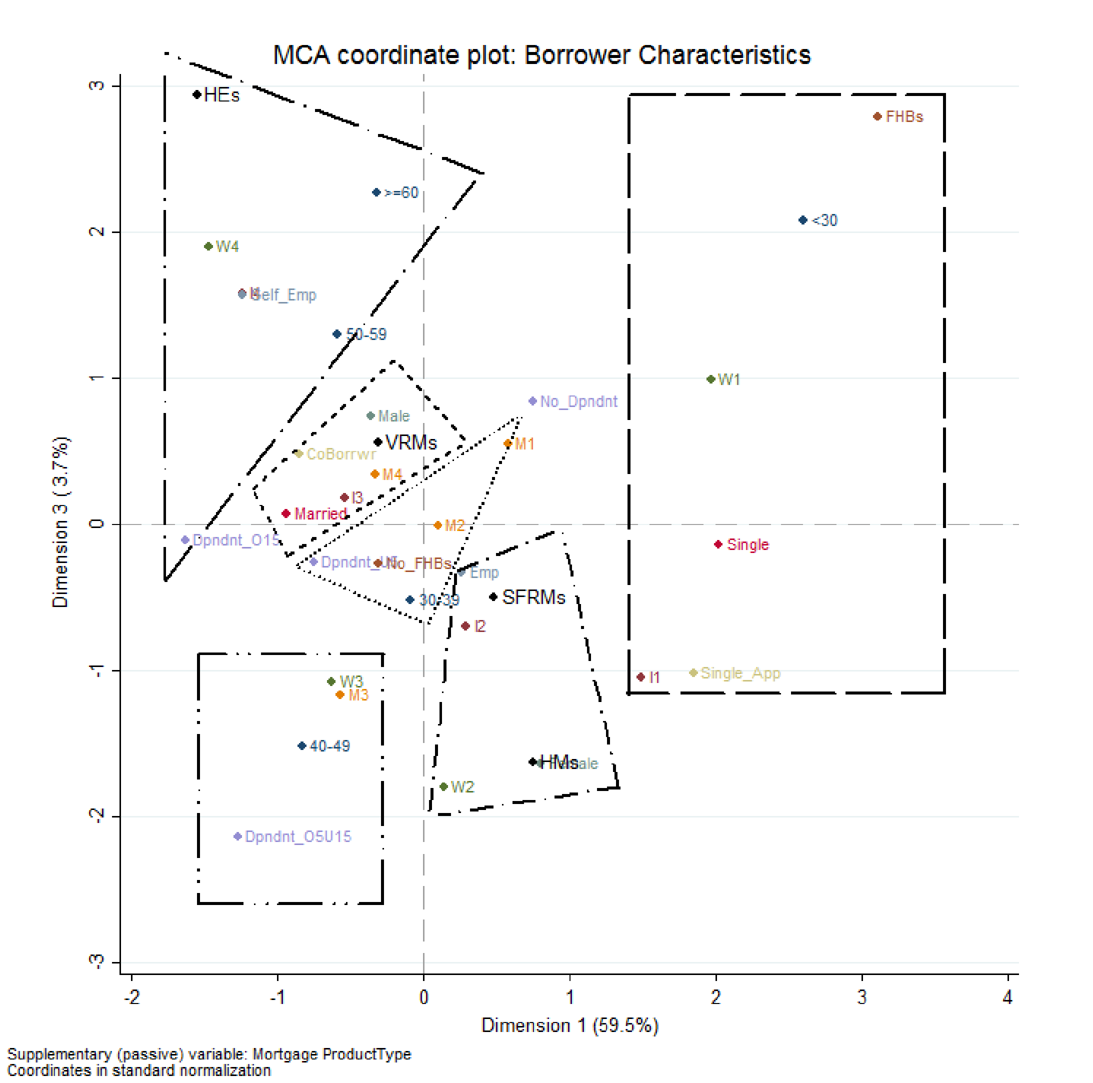
Figure 3: Multiple correspondence analysis: sub-sets in dimensions 1 and 2



The four main mortgage types are depicted in black: *VRM* (variable-rate mortgages), *SFRM* (short-term fixed-rate mortgages), *HM* ('honeymoon' mortgages) and *HE* (home equity loans). The rest of the categories for borrower characteristic are depicted in colors. Age categories: < 30, $30-39$, $40-49$, $50-59$, $>= 60$. Income quartiles: $I1$, $I2$, $I3$ and $I4$. Wealth quartiles: $W1$, $W2$, $W3$ and $W4$. Marital status: *Single* and *Married*. Mobility categories go from more mobile to less mobile: $M1$, $M2$, $M3$, $M4$, $M5$. Gender: *Female* and *Male*. Dependents: *No_Dpndnt*, *Dpndnt_U5*, *Dpndnt_O5U15*, and *Dpndnt_O15*. Fist-time home buyers: *FHBs* and *non_FHBs*. Coborrowers: *CoBorrwr* and *Single_App*. Employment status: *Self_Emp* and *Emp*.
This figure reproduces Figure 1, however here we visually identify 4 distinctive groups of borrowers with different characteristics. Each group is contained in a box with borders in different patterns. The dashed box groups single borrowers under 30 yrs., who are first-time buyers and single applicants, and have income and wealth levels in the 1st quartile. A similar interpretation can be given to the other boxes.

ACCEPTED MANUSCRIPT

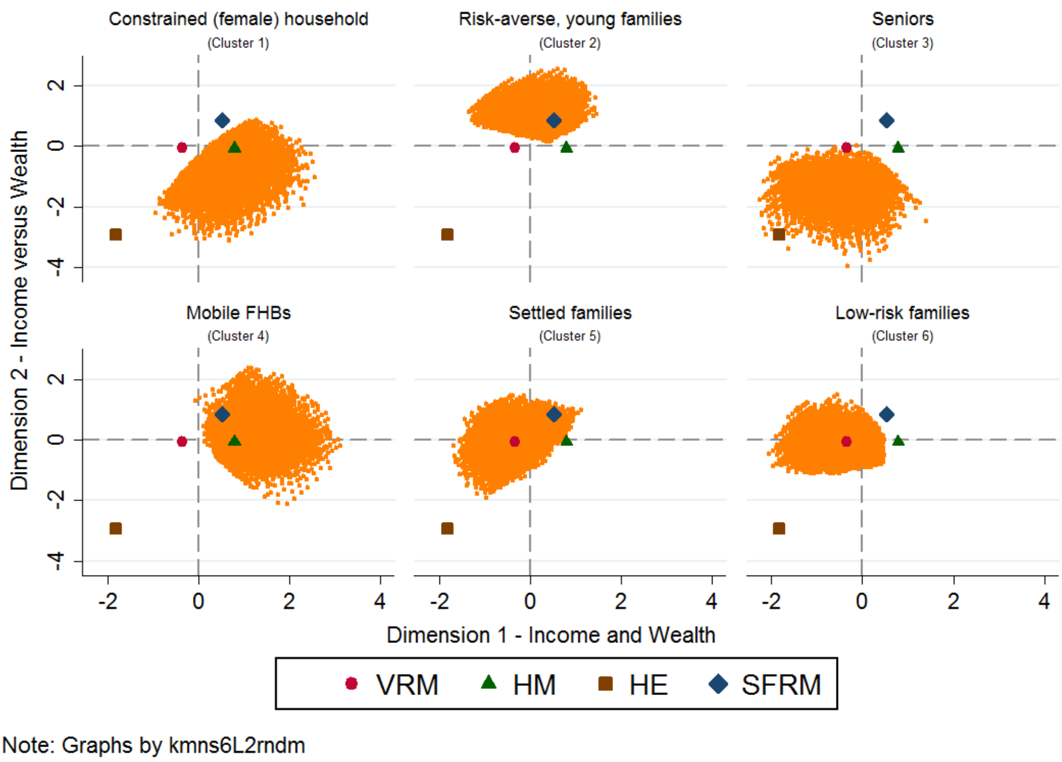
Figure 4: Multiple correspondence analysis: sub-sets in dimensions 1 and 3



The four main mortgage types are depicted in black: *VRM* (variable-rate mortgages), *SFRM* (short-term fixed-rate mortgages), *HM* ('honeymoon' mortgages) and *HE* (home equity loans). The rest of the categories for borrower characteristic are depicted in colors. Age categories: < 30, $30-39$, $40-49$, $50-59$, $>= 60$. Income quartiles: *I*1, *I*2, *I*3 and *I*4. Wealth quartiles: *W*1, *W*2, *W*3 and *W*4. Marital status: *Single* and *Married*. Mobility categories go from more mobile to less mobile: *M*1, *M*2, *M*3, *M*4, *M*5. Gender: *Female* and *Male*. Dependents: *No_Dpndnt*, *Dpndnt_U*5, *Dpndnt_O5U*15, and *Dpndnt_O*15. First-time home buyers: *FHBs* and *non_FHBs*. Coborrowers: *CoBorrwr* and *Single_App*. Employment status: *Self_Emp* and *Emp*.
This figure presents the MCA coordinates in dimensions 1 and 3, and adds two more groups to the four groups in Figure 3. Each group is contained in a box with borders in different patterns, as before. We find the four groups already presented in dimensions 1 and 2 in Figure 3, but now we detect 2 more groups.

ACCEPTED MANUSCRIPT

Figure 5: Cluster analysis



The horizontal axis depicts dimension 1 and the vertical axis depicts dimension 2, as already defined in the previous figures. Each graph represents a cluster formed with the individual observations. The four main mortgage types are identified: *V RM* (variable-rate mortgages), *SFRM* (short-term fixed-rate mortgages), *HM* ('honeymoon' mortgages) and *HE* (home equity loans). The clustering has been performed by using the k-means algorithm, with random initial observations, and Euclidean distance.

ACCEPTED MANUSCRIPT

Highlights

Application of multiple correspondence analysis (MCA) to the literature on financial product choice.

MCA is a useful way of assessing the typology of actual or potential consumers.

Our MCA application provides a useful means of detecting areas of financial under-servicing.

Customers suffer from mismatch between characteristics of available and desirable products.

ACCEPTED MANUSCRIPT