

Jan W. Radermacher

Mamma Mia! Revealing Hidden Heterogeneity by PCA-Biplot

MPC Puzzle for Italy's Elderly Poor

SAFE Working Paper No. 382 | February 2023

Leibniz Institute for Financial Research SAFE
Sustainable Architecture for Finance in Europe

Mamma Mia! Revealing Hidden Heterogeneity by PCA-Biplot

MPC Puzzle for Italy's Elderly Poor*

Jan W. Radermacher[†]

February 13, 2023

Abstract

I investigate consumption patterns in Italy and use a PCA-biplot to discover a consumption puzzle for the elderly poor. Data from the third wave (2017) of the Eurosystem's Household Finance and Consumption Survey (HFCS) indicate that Italian poor old-aged households boast lower levels of the marginal propensity to consume (MPC) than suggested by the dominant consumption models. A customized regression analysis exhibits group differences with richer peers to be only half as large as prescribed by a traditional linear regression model. This analysis has benefited from a visualization technique for high-dimensional matrices related to the unsupervised machine learning literature. I demonstrate that PCA-biplots are a useful tool to reveal hidden relations and to help researchers to formulate simple research questions. The method is presented in detail and suggestions on incorporating it in the econometric modeling pipeline are given.

*I thank Andreas Hackethal, Nathanael Vellekoop, Fabian Nemecek, Kevin Bauer, Jan Brinkhuis, Satyajit Dutt, Lukas Fischer, Gyozo Gyöngyösi, Samuel Hartzmark, Florian Hett, Alexander Hillert, Ruggero Jappelli, Lorian Pelizzon, Charlotte Persitzky, Celina Proffen, the participants of the SAFE Brown Bag on 18 May 2022 and Goethe University Finance Brown Bag on 8 June 2022 for helpful comments and suggestions. I thank Erik Reifer and Felix Kuntz for excellent research assistance. Financial support from the Leibniz Institute for Financial Research SAFE is gratefully acknowledged. This paper uses data from the Eurosystem's Household Finance and Consumption Survey (HFCS).

[†]SAFE, GSEFM and Goethe University, radermacher@safe-frankfurt.de

1 Introduction

The demographic situation in many Western countries determines that retirees make up a large and increasing fraction of the population. Understanding the consumption behavior of retirees is therefore important to policy makers; also because this group depends over-proportionally on governmental transfers and old-age poverty is likely to increase in Western economies in the foreseeable future.¹

Since retirement is (in most cases) predictable, Life Cycle/Permanent Income Hypothesis models (Modigliani and Brumberg, 1954; Friedman, 1957; Hall, 1978; Deaton, 1991; Carroll, 1997; Kaplan et al., 2014), which are based on the notion of rational agents, would suggest no change in consumption at or during retirement. Reality, however, looks different: consumption declines and savings increase during retirement (Hamermesh, 1982; Mariger, 1987; Robb and Burbidge, 1989; Banks et al., 1998; Bernheim et al., 2001; Schwerdt, 2005; Haider and Stephens Jr, 2007; Olafsson and Pagel, 2018b).² Why is that so? Are the elderly/retired liquidity constrained? What is their MPC? What motivates their behavior? These questions I seek to address with my analysis.

Using data from the third wave (2017) of the Eurosystem's Household Finance and Consumption Survey (HFCS) (HFCN, 2020b), I document that poor old-aged households³ in Italy boast lower MPC levels than suggested by the concept of liquidity constraints. A customized regression analysis exhibits group differences to richer peers to be only 8.7 percentage points while a traditional linear regression model predicts that difference at 15.5 – almost twice as large. The puzzle is: why are MPC differences among the elderly so small although liquidity differences are so large?

¹The relevance of the MPC for economic policy making is intensively discussed in the economic literature. Parker (1999) highlights the relevance of consumption models for fiscal policy and economic growth. Dolde and Tobin (1971) demonstrate how consumption, the MPC, and liquidity constraints, in particular, are impacted through monetary policy, a link which had also been made by Keynes (1936). This relation could also be seen in the first half of 2022 when EUR interest rate swaps increased sharply because of (expectations about) tighter monetary policy causing higher interest rates and hence tighter liquidity constraints for households. More technically, Broda and Parker (2014) address the relevance of the MPC within other economic models used for policy assessment such as dynamic stochastic general equilibrium (DSGE) models.

²Some authors try to rationalize the observation that consumption decreases upon retirement to be in line with the Life Cycle/Permanent Income Hypothesis framework, e.g., Aguiar and Hurst (2005); Hurst (2007); Aguiar and Hurst (2013).

³A large fraction of poor old-aged households in Italy are single and female - hence the reference to the Italian (grand-)mother in the title: *Mamma mia!*

The finding is at odds with the concept of liquidity constraints and recent empirical evidence (Johnson et al., 2006; Parker et al., 2013). However, it could be in line with buffer-stock models (Deaton, 1991; Carroll, 1997). Also Keynes (1936) suggested several motives that could explain such behavior: (i) wish to build up reserves for hard times, (ii) the desire to leave a bequest, or simply (iii) stinginess. To these motives one could add: (iv) cultural or generational reasons – after all, some of the Italian elderly today are the generation that still experienced fascism, WWII, and the post-war period – and (v) a lack of supply, i.e. things/opportunities to consume.⁴ Simply, there are less activities for elderly due to their age or health. After all, also the simple observation by Olafsson and Pagel (2018a) that only very few households feel or truly are liquidity constrained could explain the finding.

Italian data is taken for several reasons. First, Italy is the third largest country by population and economy in the Eurozone and as such very important. Particularly in household consumption theory, much research has been done by Italian economists or with Italian data (Modigliani and Brumberg, 1954; Guiso et al., 1992, 1996; Jappelli and Modigliani, 1998; Jappelli, 2005; Andini et al., 2018; Jappelli and Pistaferri, 2020). Second, it is a rich set of data in the third wave (2017) of the HFCS (HFCN, 2020b) with several thousand observations. Removing observations with missing responses, the remaining sample has more than 7,000 observations. Additionally, the data comes in a convenient format close to its original form such that there is no need to deal with statistical peculiarities such as multiple imputations (HFCN, 2020c).

Methodologically, I use a visualization technique for high-dimensional matrices based on principal component analysis (PCA) in the descriptive part of my data analysis. PCA-biplots (Gabriel, 1971, 1981; Greenacre and Hastie, 1987; Greenacre, 2010) are a method to depict an approximation of all observations, variables, and interior relations of a high-dimensional data matrix in a two-dimensional space allowing to gain important insights on the (joint) data distribution. This helps me in formulating customized regression specifications for my analysis. To my knowledge, this visualization technique has not been used in economics or finance before. I therefore provide a suggestion on how to generally include PCA-biplots in the econometric modeling pipeline.

The remainder of this paper is structured as follows. Chapter 2 introduces

⁴A similar pattern of low consumption was seen during the Covid-19 lockdowns when travel was impossible and restaurants were closed. Households had fewer opportunities to spend their money and hence saved more.

the literature and sets the theoretical background. Chapter 3 presents the data and methodology. Chapter 4 features the analysis, which contains (i) a simple linear model, (ii) a PCA-biplot analysis, and (iii) resulting customized regression set-ups. Chapter 5 discusses limitations and suggestions for further research. Chapter 6 concludes.

2 Literature Review and Theoretical Background

2.1 Consumption Theory

In 1936, Keynes (1936) introduced the concept of the marginal propensity to consume and defined it as the change of consumption per change of income. It was introduced as an aggregate figure alongside an investment rate in his theory of employment, relating to an earlier contribution by Kahn (1931) on public works. Although Pigou (1936) did not receive Keynes (1936) particularly well - he himself had been criticized and ridiculed by it - the novel thoughts on consumption were regarded by him as "no doubt, in a general way, correct" (p. 123).

Early consumption models in the tradition of Keynes (1936) were on aggregate levels and based on a simple function where consumption was a fraction of disposable income (Kahn, 1931; Keynes, 1936). Although Jappelli and Pistaferri (2017) criticize that past saving and standard of living were neglected, these models have fared relatively well empirically - at least it was hard to come up with something better (Dolde and Tobin, 1971).

The first consumption models on an individual level were introduced in the 1950s and of a Life Cycle/Permanent Income Hypothesis type (Modigliani and Brumberg, 1954; Friedman, 1957; Hall, 1978). Related to the concept of Ricardian equivalence from the early 19th century (Ricardo, 1951), these models are based on rational-agent assumptions and they focus no longer on a connection between consumption and disposable income, but exclusively on a connection between consumption and permanent income.

Because these models struggle empirically, they have been frequently updated and augmented (Shefrin and Thaler, 1988). Most famously by Deaton (1991) to the buffer-stock model in which "assets act like a buffer stock, protecting consumption against bad draws of income" (p. 1221).⁵ This buffer-stock model helped to explain why people save or refrain from

⁵Although Deaton (1991) fails to refer to it, a very similar saving motive has been given by Keynes (1936) in Chapter 9: "to build up a reserve against unforeseen contingencies" (p. 94).

borrowing - in combination with the concept of liquidity constraints (Keynes, 1936; Dolde and Tobin, 1971; Hubbard et al., 1986). Or, to put it in the words of Carroll (1997): "the [buffer-stock] model can explain three empirical puzzles: the 'consumption/income parallel' documented by Carroll and Summers; the 'consumption/income divergence' first documented in the 1930s; and the stability of the household age/wealth profile over time despite the unpredictability of idiosyncratic wealth changes" (p. 1).

But also buffer-stock models did not do too well when tested empirically as shown by Parker (1999); Souleles (1999); Johnson et al. (2006); Parker et al. (2013); Kaplan et al. (2014); Kaplan and Violante (2014b). In response, Kaplan et al. (2014) introduced the concept of the wealthy hand-to-mouth, which were said to make up a considerable fraction in most populations. Interestingly, hand-to-mouth households are solely driven by disposable income. Although the reference is generally omitted, this brings consumption theory back to the Keynesian roots it started from with Keynes (1936).⁶

2.2 Liquidity Constraints

Let us zoom in on the concept of liquidity constraints. Liquidity constraints (Keynes, 1936; Dolde and Tobin, 1971; Hubbard et al., 1986) describe situations where individuals cannot seamlessly borrow money as assumed by perfect intertemporal optimization of rational agents in Life Cycle/Permanent Income Hypothesis models (Modigliani and Brumberg, 1954; Friedman, 1957; Hall, 1978). Recognizing the existence of liquidity constraints has led to the hand-to-mouth model for poor people (Johnson et al., 2006; Kaplan et al., 2014; Kaplan and Violante, 2014a).

Yet, the poor hand-to-mouth concept did not suffice for the policy simulation models to match with real data. In response, Kaplan et al. (2014) introduced the concept of the wealthy hand-to-mouth, which arises when households are wealthy, but their wealth is illiquid, e.g., bound in a large asset like a house or - in the famous case of Shakespeare's Antonio, Merchant of Venice - in a few vessels at sea.

An alternative explanation stems from an observation by Olafsson and Pagel (2018a): only very few households feel or truly are liquidity constrained. These claims are based on a study on payday responses in Iceland conducted with a novel type of panel data set: financial account aggregators.⁷

⁶The same is true for Campbell and Mankiw (1989) who simply suggest that half of the population consume their current, i.e. disposable, income following a "rule of thumb" approach.

⁷Financial account aggregators track almost all financial transactions of an individual

In their data only very few households are liquidity constrained. Interestingly, they find significant payday responses also for liquid households such that liquidity constraint theory can no longer be used as an explanation for the behavior. They conclude that these households are likely driven by heuristics⁸ and refer to them as "liquid hand-to-mouth".

With the same data from the Icelandic financial account aggregator, Olafsson and Pagel (2018b) study how financial behavior changes around retirement. They find that "individuals delever upon retirement by reducing their consumer debt and increasing their liquid savings" (p. 2). This is the exact opposite of what is predicted by Life Cycle/Permanent Income Hypothesis models (Modigliani and Brumberg, 1954; Friedman, 1957; Hall, 1978) as Olafsson and Pagel (2018b) state: "any rational agent who expects a fall in income at retirement will save before retirement and dissave after retirement, rather than the other way round" (p. 2).

2.3 Retirement & Old-Age Poverty

It is worthwhile to focus on the topic of retirement and old-age financial behavior - consumption in particular. The demographic situation in many Western countries determines that retirees make up a large and increasing fraction of the population. Many retirees already today struggle to make ends meet and old-age poverty is a topic likely to increase in most Western countries. Additionally, pension reforms of the past three decades have shifted the funding source of retirements from governments and employers more towards voluntary contributions by the citizens themselves (e.g., 401k in the US or Riester-Rente in Germany). Understanding individual consumption and savings behavior is thus utterly important.

Because retirement is (in most cases) predictable, Life Cycle/Permanent Income Hypothesis models (Modigliani and Brumberg, 1954; Friedman, 1957; Hall, 1978; Deaton, 1991; Carroll, 1997; Kaplan et al., 2014), which are based on the notion of rational agents, would suggest no change in consumption at or during the retirement. But - as stated before - the reality looks different: consumption declines and savings increase (Hamermesh, 1982; Mariger, 1987; Robb and Burbidge, 1989; Banks et al., 1998; Bernheim et al., 2001; Schwerdt, 2005; Haider and Stephens Jr, 2007; Olafsson and

and have been frequently used for research in recent years (Nemeczek and Radermacher, 2022).

⁸Heuristics refer here to the behavioral finance and mental accounting literature around Kahneman and Tversky (1979); Thaler (1985); Shefrin and Thaler (1988); Thaler (1994).

Pagel, 2018b).⁹ My analysis addresses this puzzle.

3 Data, Summary Statistics, and PCA-Biplots

3.1 Data

I use data from the third wave (HFCN, 2020b) of the Eurosystem's Household Finance and Consumption Survey (HFCS). This survey provides microdata on a sample of European households' balance sheets, demographics, preferences, and beliefs. It is conducted by each country's central bank and has been made available by the European Central Bank (ECB). At the point of writing (July 2022), there exist three waves of the survey which have been released in 2013, 2016, and 2020. The third wave (released in 2020) was conducted in 2017 and contains a hypothetical windfall question, which can be used as a proxy for MPC preferences. This question is only available as of the third wave of the survey.

I take Italian data for several reasons. First, it is a rich set of data in the third wave (2017) of the HFCS (HFCN, 2020b) with several thousand observations. Removing observations with missing responses, I have a remaining sample size of more than 7,000 observations. Second, the data comes in a convenient format close to its original form such that there is no need to deal with statistical peculiarities such as multiple imputations (HFCN, 2020c). Third, Italy is the third largest country by population and economy in the Eurozone and as such very important.

Particularly in household consumption theory, much research has been done by Italian economists or with Italian data (Modigliani and Brumberg, 1954; Guiso et al., 1992, 1996; Jappelli and Modigliani, 1998; Jappelli, 2005; Jappelli and Pistaferri, 2020). Exemplary, Jappelli (2005) notes that Franco Modigliani "believed that the generosity of the Italian pension system and the large swings in growth and fiscal variables could be used to study the relation between saving, fiscal policy and social security" (p. 176).

3.2 Summary Statistics

For the data description, I start with summary statistics of the individual variables that I regard in my analysis. Figure A.1 on page 33 (in the appendix)

⁹Some authors try to rationalize the observation that consumption decreases upon retirement to be in line with the Life Cycle/Permanent Income Hypothesis framework, e.g., Aguiar and Hurst (2005); Hurst (2007); Aguiar and Hurst (2013).

shows histograms for the MPC and some socio-economic demographic variables.

The MPC is not uniformly distributed but with strong spikes at 0, 50, and 100. The reason is most likely survey-related: respondents provided rough rule-of-thumb estimates such as: save all (MPC=0), half-half (MPC=50), and spend all (MPC=100). As the MPC question is phrased in a hypothetical manner and I am dealing here with survey data, the estimation set-up is certainly not perfect.¹⁰ Hence, I will not spend too much attention to the exact MPC values, but rather take the values with some slack and group them in three clusters around their respective spikes. Apart from these details, the distribution is fairly symmetric and the average close to 50.

Household size is right-skewed and most households have one or two members. Households with three or more members are significantly less frequent. Wealth is reported by the HFCN in quintiles and therefore, by definition, (almost) following a uniform distribution. Net liquid assets are heavily right-skewed with a long right-tail. A transformation will become necessary for a regression analysis to remove outliers. Since there are some negative values, a log-transformation is not possible. Remaining options are transformation to percentiles or some ad-hoc winsorization, i.e. removing outer percentiles.

Education is a categorical variable, where the first three categories correspond to primary, lower secondary, and upper secondary education, respectively. The last column refers to university education. Only about a tenth of the respondents have university education while the other respondents are fairly evenly distributed among the first three education classes.

The gender of the responding person, i.e. the self-reported household head who answered the survey, is 60-40 split between male and female, respectively. This suggests that for heterosexual households (which form likely the large majority), the response person was more often male than female, which is generally in line with expected traditional gender norms where the household head and treasurer is male. The age of the respondents is somewhat representational for the Italian population yet with relatively few young individuals and many individuals above 50. Note that age information is provided in 5-year brackets.

¹⁰The question in HFCN (2020a) reads: "Imagine you unexpectedly receive money from a lottery, equal to the amount of income your household receives in a month. What percent would you spend over the next 12 months on goods and services, as opposed to any amount you would save for later or use to repay loans?" (p. 74). The surveyors were instructed to use a show-card (ruler from 0 to 100) to receive the responses. Yet, it is hard to track how thoroughly these instructions were obeyed.

After the distributions of the individual variables, I look at their relationships to each other. Figure A.2 on page 33 (in the appendix) shows a correlation table/plot for these variables.

The MPC correlates negatively with wealth, net liquid assets, and age; modestly positively with household size; and slightly negatively with education and gender. Strong positive correlations among the explanatory variables exist between wealth, net liquid assets, and education. Strong negative correlations are between household size, gender and age, and between age and education.

The negative correlation between the MPC and net liquid assets is in line with the concept of liquidity constraints (Dolde and Tobin, 1971; Hubbard et al., 1986; Jappelli and Pistaferri, 2010, 2014): households with low liquidity have high MPCs because they have unsatisfied consumption needs. A sudden increase in cash at hand (liquidity) would allow these households to satisfy their suppressed needs, e.g., to fix the broken washing machine or finally buy that new television. Accordingly, they would spend a larger fraction of a cash windfall on consumption and have therefore a higher MPC than households with more liquid means.

The negative correlation between the MPC and age is, principally, at odds with models in the Life Cycle/Permanent Income Hypothesis tradition (Modigliani and Brumberg, 1954; Friedman, 1957; Hall, 1978; Deaton, 1991; Carroll, 1997; Kaplan et al., 2014): MPC should simply not vary with age - that's the whole idea of consumption smoothing over the life cycle. The Permanent Income Hypothesis postulates that households want to smooth consumption as much as possible and consume a fraction of their life-long wealth proportionate to the duration of the respective time period. Transitory cash inflows would hence be equally distributed over the entire remaining life cycle such that only a small extra fraction would be spent each period. Hence, the MPC out of a cash windfall should be below low (no pun intended).

The descriptive statistic analysis performed until now is standard and a good starting point. The derived econometric models might, however, still miss important points or be misleading, e.g., due to variable interactions, leverage points, missing factors, etc. A well known problem with regression analysis is Simpson's Paradox (Simpson, 1951; Blyth, 1972; Wagner, 1982). But also a thorough analysis of descriptive statistics can come with its flaws as the famous Anscombe's Quartet (Anscombe, 1973) and the (somewhat polemic) "Datasaurus Dozen" from a tweet by Alberto Cairo in Figure 1 on page 10 and Figure A.3 on page 34 (in the appendix) illustrate (Matejka and Fitzmaurice, 2017). All data sets in these two illustrations have the same

summary statistics, but their data generating processes are, obviously, very different.

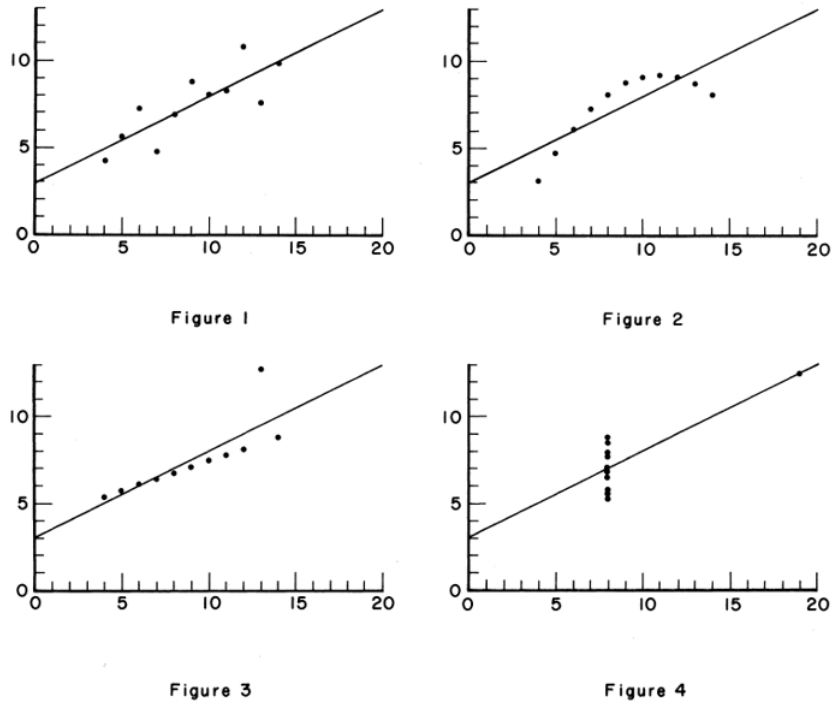


Figure 1: "Anscombe's Quartet" by Anscombe (1973). All data sets have the same summary statistics.

I would therefore like to join statisticians such as Anscombe (1973); Chambers et al. (2018); Matejka and Fitzmaurice (2017) in their case for more and better visual data analysis. In the words of Anscombe (1973): "graphs are essential to good statistical analysis" (p. 17). In particular, he discusses the role of scatterplots in combination with regression analysis. To him, the advantages of graphs are: "(i) to help perceive and appreciate some broad features of the data, [and] (ii) to let us look behind those broad features and see what else is there. Most kind of statistical calculation rest on assumptions about the behavior of the data. Those assumptions may be false, and then the calculations may be misleading" (p. 17).

Fortunately, visualization tools are nowadays way more accessible than in the 1970s. Today, it is possible to visualize almost any type of data with a few lines of (Python) code on a personal computer.

Starting with the suggested analysis by Anscombe (1973), a two-dimen-

sional scatterplot of two variables is a great way to depict both the distribution of each respective variable as well as their relation to each other, i.e. their joint distribution. This is also naturally possible for three variables in three dimensions. Yet, there is no need to stop here with today's visualization and computing power: PCA-biplots allow us to plot (approximations of) high-dimensional scatterplots in two or three dimensions as the next section will show.

3.3 PCA-Biplots

Principal component analysis (PCA) is a dimension reduction technique making use of linear algebra.¹¹ Based on work by Pearson (1901), Hotelling (1933) developed the technique originally in the field of psychometrics. Since then it has been used in other social sciences and has also found its way into economics and finance. Here, it is mostly used for combining several underlying variables into one, e.g., factor analysis or dimensionality reduction, and then using the results in further statistical analysis such as, e.g., regression or correlation analysis (Fifield et al., 2002; Olsson and Hibbs Jr, 2005; Hosseini and Kaneko, 2011; Spolaore and Wacziarg, 2013; Dean and Ortoleva, 2015; Chapman et al., 2018; Falk et al., 2018; Gewers et al., 2018, 2021).¹² Its usage has, in particular, picked up during the last

¹¹For a general introduction to PCA, see, e.g., Hastie et al. (2009); James et al. (2013); StataCorp (2021).

¹²Specifically, Fifield et al. (2002) study returns in emerging stock markets and use PCA to build economic factors. The PCA factors are used as variables in a regression analysis. Olsson and Hibbs Jr (2005) study the effect of biological and geographical conditions on long-run economic development. They use principal component analysis to build categorical variables for biological and geographical conditions from several underlying variables. Their approach is also followed in a similar study by Spolaore and Wacziarg (2013). Hosseini and Kaneko (2011) use PCA on macro country data to develop country sustainability indices. Dean and Ortoleva (2015) use PCA when testing behavioral theories with data from lab experiments. Chapman et al. (2018) identify 21 behaviors which they reduce by PCA to six dimensions to build new parsimonious models for decision making. They point out that dealing with 21 variables independently leads to 378 correlations, which are hard to deal with. However, these can be clustered into six groups by means of PCA. With these groups there are high intra-group correlations and low inter-group correlations. Falk et al. (2018) consider country-variance for economic preferences. They use PCA to combine several underlying variables into broader factors, which they use for subsequent statistical analysis. Gewers et al. (2021) argue for PCA as a natural way to do data exploration. In an older version (Gewers et al., 2018) they also survey also the literature for PCA applications in all fields.

two decades.¹³

In a household finance context, PCA is similarly used to condense several variables into underlying factors or to create an index or other input for a regression analysis (Filmer and Pritchett, 2001; McKenzie, 2005; Vyas and Kumaranayake, 2006; Antonides et al., 2011; Choi and Robertson, 2020).¹⁴

Biplots were introduced by Gabriel (1971) in the journal *Biometrika*. A biplot is a visualization of a matrix of rank two or three. It combines two (lat. bi, hence the name *biplot*) components in the same graph: (i) all rows of the matrix are represented as points and (ii) all columns of the matrix are represented as vectors. Hence, a biplot allows visualization of all data points (rows) while showing correlations between the variables (columns) as angles between vectors. Biplots are useful for depicting high-dimensional matrices, since these matrices can be approximated in two- or three dimensions by PCA using an algorithm called singular value decomposition (SVD).^{15 16}

¹³The application rate of PCA in economics could be similar to how Regression Discontinuity Design (RDD) took off since its introduction by Goldberger (1972) as discussed in Cunningham (2020). Or similar to how more recently Causal Random Forests (Wager and Athey, 2018) have become a popular tool among applied economists.

¹⁴Specifically, Filmer and Pritchett (2001) use PCA on household asset ownership data from India to construct an index, which they use in latter econometric analysis. McKenzie (2005) follows this approach and uses PCA to construct an inequality index for households in Mexico. Vyas and Kumaranayake (2006) generally discuss the use of PCA to create socio-economic status indices from asset data. They argue that PCA on hard data could be a better tool for developing socio-economic-status indices than eliciting soft data from respondents via surveys (because of cost and biases in their answers). Antonides et al. (2011) use PCA in combination with a questionnaire on mental budgeting. PCA allows them to combine many questions of a survey into fewer underlying factors, which is then used for regression analysis. This is nothing else than breaking up higher dimensional data (all original questions) into lower dimensions (only the remaining factors). Choi and Robertson (2020) study attitudes of private investors by means of a survey. They use principal component analysis to break down survey responses for 34 variables into six principal components, which explain 54% of the original variation.

¹⁵SVD (Eckart and Young, 1939; Golub and Reinsch, 1971) is a linear algebra routine allowing to decompose a matrix into ordered factors, where the first factor columns can be used for a low-rank least-squares approximation. See Greenacre (2010) for an introduction.

¹⁶The method is very flexible and can also deal with categorical data. This is called correspondence analysis which Greenacre and Hastie (1987) define as "the principal component analysis of categorical data" (p. 446) and "an explanatory multivariate technique that converts a matrix of nonnegative data into a particular type of graphical display in which the rows and columns of the matrix are depicted as points" (p. 437).

4 Analysis

4.1 Standard Linear Model

I start the econometric analysis by conducting a simple OLS regression broadly inspired by Christelis et al. (2019a) and Christelis et al. (2019b):

$$MPC_i = \beta_0 + \beta_1 HhSize_i + \beta_2 WealthQuint_i + \beta_3 NetLiqAssets_i + \beta_4 Education_i + \beta_5 Gender_i + \beta_6 Age_i + \epsilon_i \quad (1)$$

The results of this simple OLS regression are shown in the first column of Table 3 on page 22. The regression has an adjusted R-squared of 4.2%, which is comparable to Christelis et al. (2019a), even a bit larger. The variables household size (*HhSize*), wealth quintile (*WealthQuint*), percentile rank of net liquid assets (*NetLiqAssets*), and age are significant. Education and gender are not.

The interpretation of the negative coefficients for wealth and net liquid assets is in line with the theory of liquidity constraints (Keynes, 1936; Dolde and Tobin, 1971; Hubbard et al., 1986; Johnson et al., 2006; Parker et al., 2013). But it is at odds with traditional view of the Life Cycle/Permanent Income Hypothesis (Modigliani and Brumberg, 1954; Friedman, 1957; Hall, 1978). The same holds true for the significant coefficient for age. It is at odds with models in the spirit of the Life Cycle/Permanent Income Hypothesis and, additionally, at odds with buffer-stock models (Deaton, 1991; Carroll, 1997; Kaplan et al., 2014).

So far, this classical econometric analysis has enabled us to perform a test around theories such as the Life Cycle/Permanent Income Hypothesis or the concept of liquidity constraints. It has done so by means of a (linear) OLS model. Yet, other hypotheses, take, e.g., the concept of the wealthy hand-to-mouth, we could not simply test with this setup.

Additionally, with our choice of an OLS regression, we have made a number of implicit assumptions that we have not discussed before. One assumption is linearity, i.e. that an increment by one unit at a certain level of an independent variable has the same impact on the dependent variable than the same increment by one unit at any other level of that specific independent variable. To provide a specific example, the age coefficient in the first regression, shown in the first column of Table 3 on page 22, is roughly at -0.2 (per year). The linearity assumption prescribes that the MPC on average and all else being equal (*ceteris paribus*) should be 2 percentage

points lower for every 10 year age difference between households. No matter if the change is from age 20 to 30 or from age 60 to 70.

Another assumption, we are implicitly making is that of independence of effects of the independent variables - also known as the *ceteris paribus* assumption. By this simple OLS setup we rule out that variables could interact with each other and therefore yield different effects. Specifically, poverty could have very different effects for young than for old households and old-age could have very different effects depending on whether a household is rich or poor.

Of course, we need to make simplifying assumptions. But it does not mean that we should not try to somehow test them. For a general suggestion on how to deal with it, I would like to refer to Anscombe (1973): "Most kinds of statistical calculation rest on assumptions about the behavior of the data. Those assumptions may be false, and then the calculations may be misleading. We ought always try to check whether the assumptions are reasonably correct; and if they are wrong we ought to be able to perceive in what ways they are wrong. Graphs are very valuable for these purposes" (p. 17).

Anscombe (1973) also highlights the importance of scatterplots in combination with regression analysis. As discussed earlier, I take this as a starting point and use PCA-biplots as an approximation of multi-dimensional scatter plots. This allows me to work more data-driven and to visually inspect the assumptions I have implicitly made; in particular, with regard to the interaction of variables.

4.2 PCA-Biplot Analysis

Figure 2 on page 16 depicts a PCA-biplot.¹⁷ A PCA-biplot is a graph that contains two (lat. *bi* means two) types of information of a matrix. It depicts the rows of the matrix as points and the columns of the matrix as vectors. The vectors come from the origin and are normalized. The angles of the vectors to each other show the correlations of the variables with each other. The lengths of the vectors indicate how well these variables are mapped in the underlying lower dimensional space, which is spanned by the first two principal components of the matrix.¹⁸

¹⁷PCA-biplot is not the standard name used in the literature. Most authors refer to them simply as biplots (Greenacre, 2010). I introduce the name PCA-biplot since they are a visualization of PCA, a concept which many economists are familiar with.

¹⁸Figure A.4 on page 35 shows the same plot as previously Figure 3. The dots there are a bit smaller and it is zoomed in such that it is easier to see the individual households.

The first principal component (on the X-axis) explains 30.3% of the overall variation in the data. The second principal component (on the y-axis) explains 23.6% of the overall variation in the data. Taken together, the first two principal components explain 53.9% of the overall variation in the data.

Additional to depicting the rows (household information) as dots and the columns (independent variables) as vectors, I use colors for the dots to indicate the value of the dependent variable. Specifically, green dots are households with low MPC scores (below 35); yellow dots have medium scores (from 35 to 65); and red dots have high scores (above 65). The dots (households) can also be counted: Table 1 on 16 lists how many dots (households) are in each quadrant.

Looking at the first quadrant (top right), there are many dots dispersed to the northeastern (top-right) corner. In that direction also point the vectors net liquid assets and wealth quintiles. Households (dots) in that direction/area are very wealthy. If we look at the opposite third quadrant (bottom left): there are poor households and we see a natural sharp frontier. This can be easily explained: there is some lower bound to liquid assets. A household cannot have much less than 0 in liquid assets, due to overdraft limits.

Looking at the second quadrant (top left), I see a similar sharp frontier. This time, it is opposite of the vector household size. This implies that here we have households that have a very low household size. Which easily explains the sharp frontier: there are no households with less than one member. Also, the age vector points in this direction. So, these households are likely quite old: the elderly. Looking at the remaining fourth quadrant (bottom right), I see a slightly similar pattern than before in the first quadrant: some dispersion towards the southeastern corner. This means that there are a few households which are very large: families. Also, they are positioned opposite of the age vector, so they are not very old: young families.

Lastly, education is pointing to the east and is closely linked to the first principal component. This implies that households in the first and fourth quadrant have higher education levels than in the second or third. Moreover, gender features only a relatively short vector, which means that this variable is not mapped well in the two dimensional space spanned by the first two principal components. Nonetheless, a small trend can be seen: a majority of the poor old-aged single households are female. This is plausible if one considers average age differences in marriages (men tend to be older) and average life expectations (women tend to live longer than men).

	Quadrant I	Quadrant II	Quadrant III	Quadrant IV
MPC				
low (0-34)	650 (40%)	835 (42%)	508 (32%)	599 (28%)
medium (35-65)	653 (40%)	684 (34%)	496 (31%)	839 (40%)
high (66-100)	337 (21%)	493 (25%)	593 (37%)	684 (32%)

Table 1: Counts of MPC types for the four quadrants. All data is taken from the third wave (2017) of the HFCS (HFCN, 2020b).

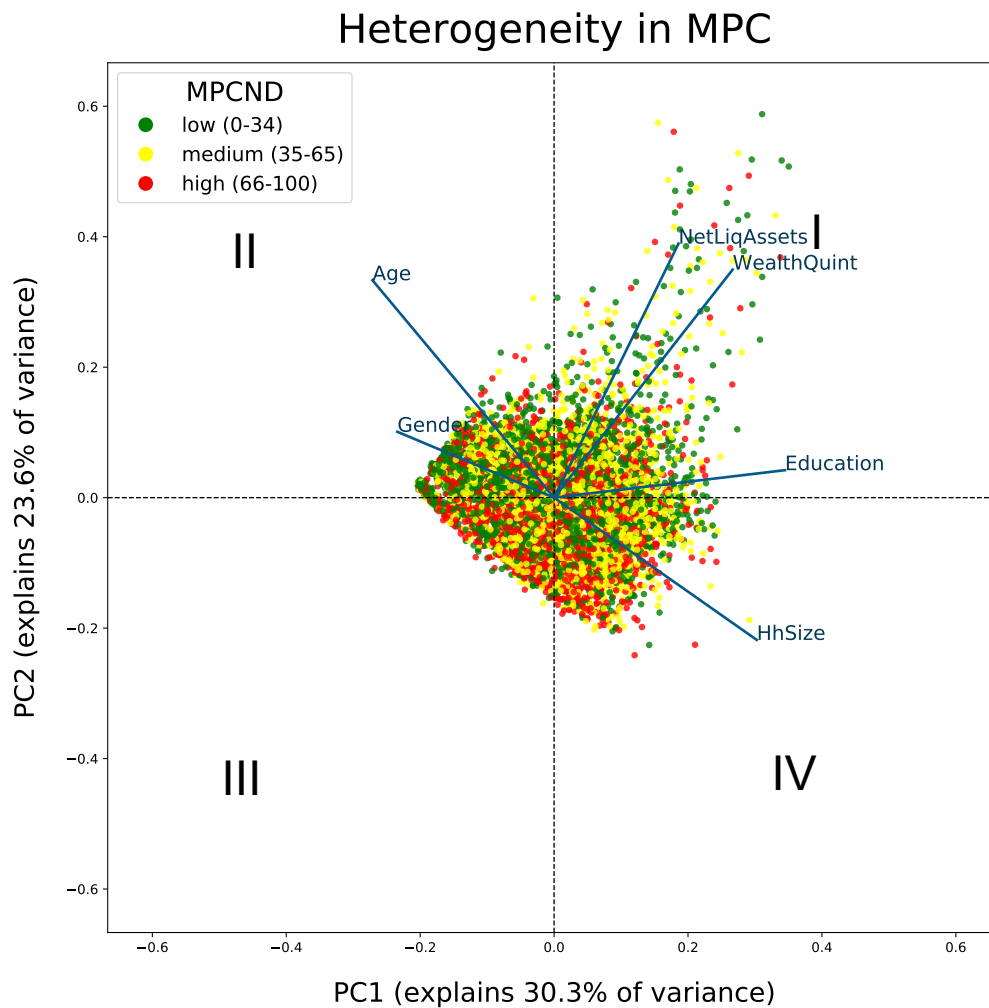


Figure 2: Biplot Italy: colour-coded MPC on small set of demographics. All data is taken from the third wave (2017) of the HFCS (HFCN, 2020b).

A visual inspection of this PCA-biplot gives me a puzzling result: for low-liquidity households (bottom left frontier) there seems to be a structural difference between the sub-groups. Poor young families (bottom right) have majorly high MPCs as most dots are red or yellow and there are only few green dots. But poor elderly (left tip) have majorly low MPCs as there are many green dots and only few yellow or red dots. The concept of liquidity constraints would actually prescribe that all poor households have high levels of MPC, i.e. most dots should be red or yellow. Yet, finding mostly green dots, i.e. low MPC levels, for poor old-aged households is truly puzzling.

As mentioned before, it is possible to exactly count the dots of each color in a specific section. This is what I do next. Figure 3 on page 18 depicts the same plot as Figure 2 on page 16 but the focus is put here on the two different groups which are circled in black: *poor elderly* and *poor families*. As discussed, both groups are on the poor and illiquid frontier of the data set. According to liquidity constraint theory, both groups should therefore have the same high level of MPC because they are constrained in their consumption.

Counting the colors of the dots in the two groups, I can numerically confirm the visual impression from above. As presented in Table 2 on page 18, 43% of the dots in the poor family area have a high MPC (above 65 on a scale of 100) and only 20% have a low MPC (below 35). This is in line with liquidity constraint theory. As these households are poor, they are constrained in their consumption so they cannot afford all the things they would like to buy. For example, if their car breaks down they cannot fix it right away because they lack the financial means. If they now get a monetary windfall they will use a large fraction of that money to purchase the things that they could not afford before. Hence, their marginal propensity to consume is high.

The same one would expect for poor elderly. But the data looks very different. Here, we observe many green dots, i.e. many households with low MPC levels. Particularly, we find that 44% of all the dots in that area have a low MPC score (below 35) and only 29% have a high MPC level (above 65). This is puzzling if one looks at it from a perspective of liquidity constraint theory as discussed previously.

With these insights it is now time to turn back to econometrics. In the next section, I design customized regression specifications to see whether the visual impressions are also numerically significant. Due to the intuition I formed by visually assessing the data, I will include non-linearities and variable-interactions into my linear regression models.

MPC	Poor Elderly ($pc1 \leq -0.15, pc2 \leq 0.10$)	Poor Families ($pc1 \geq 0.00, pc2 \leq -0.10$)
low (0-34)	201 (44%)	109 (20%)
medium (35-65)	120 (26%)	199 (37%)
high (66-100)	134 (29%)	235 (43%)

Table 2: Counts of MPC types for poor elderly and poor families. All data is taken from the third wave (2017) of the HFCS (HFCN, 2020b).

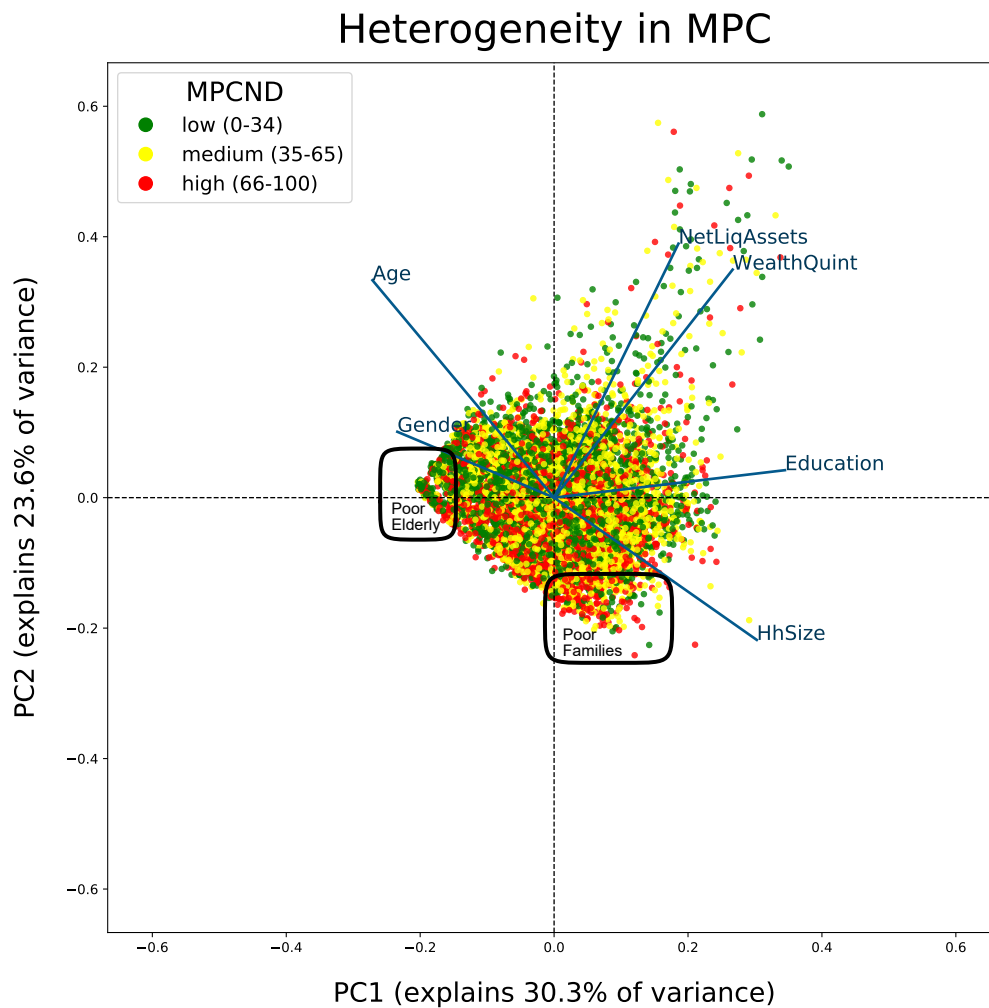


Figure 3: PCA-Biplot Italy: colour-coded MPC on small set of demographics. All data is taken from the third wave (2017) of the HFCS (HFCN, 2020b).

4.3 Customized Regressions

I use the intuition built in the previous section with the PCA-biplot analysis. I discovered a puzzling result regarding liquidity for illiquid/poor elderly as compared to illiquid/poor families. Therefore, I base, in addition to the dependent variable of interest, MPC, the remainder of my analysis on these three explanatory variables: net liquid assets, household size, and age.

I start again with a simple traditional regression for these variables, which I shall use as a benchmark:

$$MPC_i = \beta_0 + \beta_1 NetLiqAssets_i + \beta_2 HhSize_i + \beta_3 Age_i + \epsilon_i \quad (2)$$

The results of this simple OLS regression are shown in the second column of Table 3 on page 22. The regression has an adjusted R-squared of 4.0%. This is very similar to the first regression conducted, which had an adjusted R-squared of 4.2%. I am therefore confident that the three remaining variables are a good choices to describe the variation in the data.

Next, I want to see what happens if I take non-linearities into account. Out of each continuous variable I make a categorical variable for extreme cases. Since I am primarily interested in the effect of net liquid assets to test theories around the concept of liquidity constraints, I create a dummy for the most liquid (top 20%) and most illiquid (bottom 20%) households in my dataset. Additionally, I translate the visual observation from the PCA-biplot in Figure 3 on page 18 as discussed above into econometrics: I create dummies for large households, i.e. five or more members, and old households, i.e. respondents being 70 years or older. This yields the following regression:

$$\begin{aligned} MPC_i = & \beta_0 + \beta_1 NetLiqAssetsQuint_1_i \\ & + \beta_2 NetLiqAssetsQuint_5_i \\ & + \beta_3 HhSize_i \\ & + \beta_4 Age_i + \epsilon_i \end{aligned} \quad (3)$$

The results of this customized OLS regression are shown in the third column of Table 3 on page 22. The regression has an adjusted R-squared of 3.3%. This is similar to but a bit smaller than for the first two regressions conducted. As I dropped all continuous variables and I am essentially working with four dummies only, it is absolutely plausible. To me, this is

still satisfactory as the adjusted R-squared is also well above the R-squared for the regressions in Christelis et al. (2019a), which use a similar setup.

Looking now at the suggested results, I observe that the average MPC-difference between low-liquidity old-age household (51.0) and high-liquidity old-age household (35.6) is at 15.4.¹⁹

In the PCA-biplot in Figure 3 on page 18, I noticed that on the illiquid/poor frontier MPC behavior is very different between (illiquid/poor) families and the (illiquid/poor) elderly. I try to add this information in the econometric specification in form of interaction terms. This yields the following regression:

$$\begin{aligned}
 MPC_i = & \beta_0 + \beta_1 NetLiqAssetsQuint_1_i \\
 & + \beta_2 NetLiqAssetsQuint_5_i \\
 & + \beta_3 HhSize5plus_i \\
 & + \beta_4 Age70plus_i \\
 & + \beta_5 NetLiqAssetsQuint_1_i \times HhSize5plus_i \\
 & + \beta_6 NetLiqAssetsQuint_1_i \times Age70plus_i \\
 & + \beta_7 NetLiqAssetsQuint_5_i \times HhSize5plus_i \\
 & + \beta_8 NetLiqAssetsQuint_5_i \times Age70plus_i + \epsilon_i
 \end{aligned} \tag{4}$$

The results of this customized OLS regression are shown in the fourth column of Table 3 on page 22. The regression has an adjusted R-squared of 3.5% which is similar to the previous regression. The regression now has more terms and not all corresponding coefficients are significant. Importantly, the interaction term for low-liquidity and old-age is significant. But a numeric comparison to richer households as done before is not possible since the required coefficients are not significant. Hence, another regression is performed with fewer regressors:

$$\begin{aligned}
 MPC_i = & \beta_0 + \beta_1 NetLiqAssetsQuint_1_i \\
 & + \beta_2 NetLiqAssetsQuint_5_i \\
 & + \beta_3 NetLiqAssetsQuint_1_i \times HhSize5plus_i \\
 & + \beta_4 NetLiqAssetsQuint_1_i \times Age70plus_i \\
 & + \beta_5 NetLiqAssetsQuint_5_i \times HhSize5plus_i \\
 & + \beta_6 NetLiqAssetsQuint_5_i \times Age70plus_i + \epsilon_i
 \end{aligned} \tag{5}$$

¹⁹The calculations were made by adding the relevant coefficients for the respective groups. Illiquid elderly: $47.9 + 9.2 - 6.1 = 51.0$. Liquid elderly: $47.9 - 6.2 - 6.1 = 35.6$

The results of this customized OLS regression are shown in the fifth column of Table 3 on page 22. The regression has an adjusted R-squared of 3.1% which is similar to the previous regressions. Now, all coefficients relevant for old-age and illiquidity are significant.

The average MPC-difference between illiquid/poor old-aged households (46.2) and liquid/wealthy old-aged households (37.4) is at 8.7 (due to rounding).²⁰ This difference is 6.7 percentage points less than the difference of 15.4 resulting from the regression described by equation 3 and depicted in the third column of Table 3 on page 22. Hence, This regression result suggests that the average MPC difference between rich and poor elderly is only about half as large as initially anticipated, assumed by theories of liquidity constraints, and suggested by the regression results to equation 1, 2, and 3, respectively.

On the other hand, the average MPC-difference between illiquid/poor (young) families (68.3) and illiquid/poor elderly (46.2) is at 22.1 percentage points. This difference is substantially larger than 13.5 percentage points which is suggested by equation 3. The illiquid/poor elderly therefore have on average MPCs that are more similar to the liquid/wealthy elderly and less similar to illiquid/poor families. The opposite was suggested by specifications that did not account for these variable interactions.

In summary, the usage of PCA-biplots has given me intuition about the presence of non-linearities and relevant interactions of variables. It has allowed me to formulate customized regression specifications in which I could numerically show the presence of these issues. The results to these regressions have shown that MPC levels for illiquid/poor families are higher than suggested by a traditional simple OLS model. Finally, it has shown that the MPC levels for illiquid/poor elderly is lower than suggested by a traditional simple OLS model. This finding is also hard to square with the concept of liquidity constraints and prior empirical evidence (Johnson et al., 2006; Parker et al., 2013). This forms yet another liquidity puzzle. The implications and potential explanations I discuss in the next section.

²⁰The calculations were made by adding the relevant coefficients for the respective groups. Illiquid elderly: $46.0 + 12.8 - 12.6 = 46.2$. Liquid elderly: $46.0 - 5.5 - 3.1 = 35.6$

	(1)	(2)	(3)	(4)	(5)
	MPC	MPC	MPC	MPC	MPC
NetLiqAssets	-15.357*** (1.655)	-19.24*** (1.393)			
HhSize	1.170*** (0.368)	0.938*** (0.346)			
Age	-0.199*** (0.031)	-3.144*** (0.293)			
WealthQuint	-1.330*** (0.354)				
Education	-0.490 (0.866)				
Gender	-0.473 (0.866)				
NetLiqAssets_Quint1			9.227*** (1.012)	11.064*** (1.242)	12.756*** (1.168)
NetLiqAssets_Quint5			-6.243*** (1.030)	-7.168*** (1.339)	-5.476*** (1.271)
HhSize_5plus			7.445*** (1.975)	5.874** (2.655)	
Age_70plus			-6.064*** (0.843)	-5.191*** (1.081)	
NetLiqAssets_Quint1 x HhSize_5plus				3.653 (4.391)	9.527*** (3.505)
NetLiqAssets_Quint1 x Age_70plus				-7.419*** (2.251)	-12.610*** (1.978)
NetLiqAssets_Quint5 x HhSize_5plus				2.218 (6.105)	8.092 (5.508)
NetLiqAssets_Quint5 x Age_70plus				2.141 (2.129)	-3.050* (1.838)
Intercept	69.812*** (2.890)	65.940*** (2.096)	47.941*** (0.616)	47.679*** (0.616)	45.988*** (0.519)
Obs.	7371	7371	7371	7371	7371
Adj. R-squared	0.042	0.040	0.033	0.035	0.031

Table 3: Results for traditional (1-2) and customized (3-5) regressions. All data is taken from the third wave (2017) of the HFCS (HFCN, 2020b).

5 Limitations and Discussion

5.1 Survey Data

My analysis is based entirely on survey data. But I am in good company. Also very famous contributions relied exclusively on surveys for their "empirical evidence", e.g., Carroll (1997); Parker (1999); Shapiro and Slemrod (2003); Johnson et al. (2006); Parker et al. (2013); Kaplan et al. (2014); Christelis et al. (2019a); Jappelli and Pistaferri (2020). Surveys, however, are prone to measurement errors as pointed out by Olafsson and Pagel (2018a). Some of these errors can be regarded as random noise, e.g., (i) no incentive to answer correctly, (ii) trouble understanding phrasing of question, and (iii) reported

behavior differing from real behavior. Other problems, however, can lead to non-random biases, e.g., (iv) social stigma or (v) mistrust regarding data privacy (Karlán and Zinman, 2008). All problems might apply with the data I use for my analysis. It is therefore important to regard them and the results with the usual caution.

5.2 Yet Another Consumption Puzzle

I document that illiquid/poor old-aged households in Italy boast lower MPC levels than suggested by the concept of liquidity constraints. These poor elderly (possibly affected by old-age poverty) have significantly lower MPCs than younger poor households and rather similar MPCs with richer elderly.

This is at odds with the concept of liquidity constraints and empirical evidence by Johnson et al. (2006); Parker et al. (2013). It could, however, be in line with buffer-stock models (Deaton, 1991; Carroll, 1997). Also Keynes (1936) suggested several motives that could explain such behavior: (i) "[t]o build up a reserve against unforeseen contingencies" (p. 94), (ii) "[t]o bequeath a fortune" (p. 95), and (iii) "[t]o satisfy pure miserliness, i.e. unreasonable but insistent inhibitions against acts of expenditure as such" (p. 95).

To these motives one could add: (iv) cultural or generational reasons - after all the Italian elderly today are the generation that still experienced fascism and WWII - and (v) a lack of supply, i.e. things/opportunities to consume.²¹ Simply, there are less activities for elderly because of their age and health. After all, also the simple observation by Olafsson and Pagel (2018a) that only very few households feel or truly are liquidity constrained could explain the finding.

Additionally, we can look to Thaler (1994) who summarizes three general problems of Life Cycle/Permanent Income Hypothesis models: (i) unrealistic optimization assumption, (ii) self control problems among humans, and (iii) unrealistic fungibility of wealth assumption. He suggests to augment the theory around Life Cycle/Permanent Income Hypothesis models by "a set of 'mental accounts' with varying marginal propensities to consume" (p. 188) and refers to earlier work by Shefrin and Thaler (1988); Thaler (1985, 1990)

After all, the puzzle remains interesting. Several potential explanations have been discussed, but a conclusive answer cannot be given. Further

²¹A similar pattern of low consumption was seen during the Covid-19 lockdowns when travel was impossible and restaurants were closed. Household had fewer opportunities to spend their money and hence saved more.

research remains necessary and might benefit also from other methods. More qualitative or mixed-method approaches might be deemed appropriate. Interviewers could look with interviewees through their spending data and ask detailed questions about their behavior.

5.3 Integrating PCA-Biplots into Econometric Analysis

I suggest a tool that extends descriptive statistic analysis to provide better visualizations of original data in its entire complexity. This is in line with the suggestions by Anscombe (1973) who argues for a combination of numerical calculations and graphs. It is also in line with how other authors have introduced novel statistical tools into the field of economics such as regression discontinuity design (Goldberger, 1972), machine learning techniques (Athey and Imbens, 2017), or causal random forests (Wager and Athey, 2018).

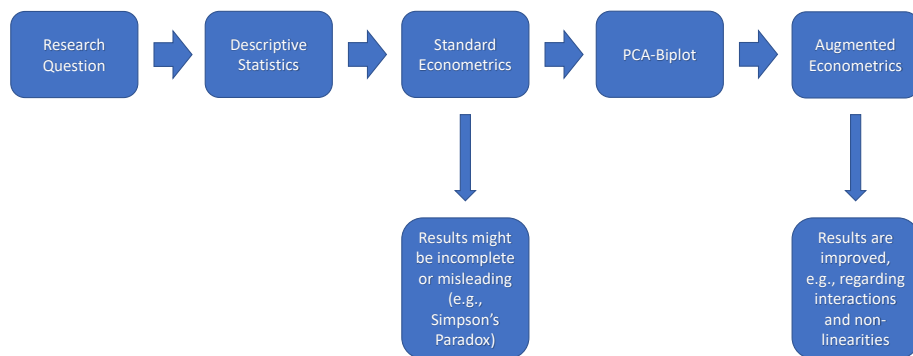


Figure 4: Flow-chart of econometric analyses with PCA-Biplots as visual aid to regression specification.

A general problem with regression analysis is that it is additive and interactions and non-linearities are not naturally dealt with. PCA-biplots are particularly useful to address the problem of interactions as has been emphasized in this paper. Non-linearities could also be addressed, but this can also be done well by decision trees and random forests as other authors have shown.

PCA-biplots help practitioners to quickly come up with more customized regression specifications. They enable researchers to consider distributions and correlations of several variables at the same time. Additionally, resulting plots can be (with some explanation) easily understood by non-technical individuals. They are consequently also a great communication tool.

PCA-biplots can be successfully integrated into the econometric modeling pipeline as shown in this analysis. Figure 4 on page 24 shows schematically how this could be done in general.

In summary, I suggest a tool that helps researchers draw causal inference based on visual inspection. This allows to make better regression specifications by finding the optimal combination of variables. Policy makers will get better findings and can therefore target actions better to sub-groups to have more customized policies and better results.

6 Conclusion

I analyze consumption patterns in Italy using survey data from the third (2017) wave of the HFCS and apply a PCA-biplot, a method from unsupervised machine learning, to reveal hidden heterogeneity. I discover that poor elderly households have lower MPC scores than other poor households – a finding at odds with the concept of liquidity constraints and the dominating views in the consumption literature – and confirm the observation by means of customized regression specifications which include interaction terms.

I interpret the findings as a suggestion that age-related, generational, and country-specific factors impact the MPC and are a source of heterogeneity that is not captured by standard consumption models. This finding is particularly relevant to policy makers who are interested in the purchasing power and aggregate demand effects of income changes. In light of growing concerns about old-age poverty as well as an increasing share of retirees in the overall Italian population, this topic becomes very relevant.

Apart from consumption theory, this study introduces an intuitive visualization from unsupervised machine learning, the PCA-biplot. This plot enables the visualizations of (approximations of) multiple correlation and scatter plots at the same time. It enables thus a holistic view of the data. As it enabled me in this application to detect the old-age consumption puzzle, I also provide a more general suggestion on how to combine PCA-biplots with standard econometric modeling.

References

- Aguiar, M. and Hurst, E. (2005). Consumption versus expenditure. *Journal of Political Economy*, 113(5):919–948.
- Aguiar, M. and Hurst, E. (2013). Deconstructing life cycle expenditure. *Journal of Political Economy*, 121(3):437–492.
- Andini, M., Ciani, E., de Blasio, G., D’Ignazio, A., and Salvestrini, V. (2018). Targeting with machine learning: An application to a tax rebate program in Italy. *Journal of Economic Behavior & Organization*, 156:86–102.
- Anscombe, F. J. (1973). Graphs in statistical analysis. *The American Statistician*, 27(1):17–21.
- Antonides, G., De Groot, I. M., and Van Raaij, W. F. (2011). Mental budgeting and the management of household finance. *Journal of Economic Psychology*, 32(4):546–555.
- Athey, S. and Imbens, G. W. (2017). The state of applied econometrics: Causality and policy evaluation. *Journal of Economic Perspectives*, 31(2):3–32.
- Banks, J., Blundell, R., and Tanner, S. (1998). Is there a retirement-savings puzzle? *American Economic Review*, pages 769–788.
- Bernheim, B. D., Skinner, J., and Weinberg, S. (2001). What accounts for the variation in retirement wealth among US households? *American Economic Review*, 91(4):832–857.
- Blyth, C. R. (1972). On Simpson’s paradox and the sure-thing principle. *Journal of the American Statistical Association*, 67(338):364–366.
- Broda, C. and Parker, J. A. (2014). The economic stimulus payments of 2008 and the aggregate demand for consumption. *Journal of Monetary Economics*, 68:S20–S36.
- Campbell, J. Y. and Mankiw, N. G. (1989). Consumption, income, and interest rates: Reinterpreting the time series evidence. *NBER Macroeconomics Annual*, 4:185–216.
- Carroll, C. D. (1997). Buffer-stock saving and the life cycle/permanent income hypothesis. *The Quarterly Journal of Economics*, 112(1):1–55.

- Chambers, J. M., Cleveland, W. S., Kleiner, B., and Tukey, P. A. (2018). *Graphical methods for data analysis*. Chapman and Hall/CRC.
- Chapman, J., Dean, M., Ortoleva, P., Snowberg, E., and Camerer, C. (2018). *Econographics*. Technical report, National Bureau of Economic Research.
- Choi, J. J. and Robertson, A. Z. (2020). What matters to individual investors? evidence from the horse's mouth. *The Journal of Finance*, 75(4):1965–2020.
- Christelis, D., Georgarakos, D., Jappelli, T., Pistaferri, L., and Van Rooij, M. (2019a). Asymmetric consumption effects of transitory income shocks. *The Economic Journal*, 129(622):2322–2341.
- Christelis, D., Georgarakos, D., Jappelli, T., Pistaferri, L., and Van Rooij, M. (2019b). Wealth shocks and mpc heterogeneity. Technical report, National Bureau of Economic Research.
- Cunningham, S. (2020). Causal inference. *The Mixtape*, 1.
- Dean, M. and Ortoleva, P. (2015). Is it all connected? a testing ground for unified theories of behavioral economics phenomena. *A Testing Ground for Unified Theories of Behavioral Economics Phenomena (August 13, 2015)*.
- Deaton, A. (1991). Saving and liquidity constraints. *Econometrica: Journal of the Econometric Society*, pages 1221–1248.
- Dolde, W. and Tobin, J. (1971). Wealth, liquidity, and consumption. Technical report, Cowles Foundation for Research in Economics, Yale University.
- Eckart, C. and Young, G. (1939). A principal axis transformation for non-hermitian matrices. *Bulletin of the American Mathematical Society*, 45(2):118–121.
- Falk, A., Becker, A., Dohmen, T., Enke, B., Huffman, D., and Sunde, U. (2018). Global evidence on economic preferences. *The Quarterly Journal of Economics*, 133(4):1645–1692.
- Fifield, S. G., Power, D. M., and Sinclair, C. D. (2002). Macroeconomic factors and share returns: an analysis using emerging market data. *International Journal of Finance & Economics*, 7(1):51–62.
- Filmer, D. and Pritchett, L. H. (2001). Estimating wealth effects without expenditure data—or tears: an application to educational enrollments in states of india. *Demography*, 38(1):115–132.

- Friedman, M. (1957). *Theory of the consumption function*. Princeton University Press.
- Gabriel, K. R. (1971). The biplot graphic display of matrices with application to principal component analysis. *Biometrika*, 58(3):453–467.
- Gabriel, K. R. (1981). Biplot display of multivariate matrices for inspection of data and diagnosis. *Interpreting Multivariate Data*, pages 147–173.
- Gewers, F. L., Ferreira, G. R., Arruda, H. F. D., Silva, F. N., Comin, C. H., Amancio, D. R., and Costa, L. D. F. (2021). Principal component analysis: A natural approach to data exploration. *ACM Computing Surveys (CSUR)*, 54(4):1–34.
- Gewers, F. L., Ferreira, G. R., de Arruda, H. F., Silva, F. N., Comin, C. H., Amancio, D. R., and Costa, L. d. F. (2018). Principal component analysis: A natural approach to data exploration. *Working Paper*.
- Goldberger, A. S. (1972). Selection bias in evaluating treatment effects: Some formal illustrations. Technical report.
- Golub, G. H. and Reinsch, C. (1971). Singular value decomposition and least squares solutions. In *Linear algebra*, pages 134–151. Springer.
- Greenacre, M. and Hastie, T. (1987). The geometric interpretation of correspondence analysis. *Journal of the American statistical association*, 82(398):437–447.
- Greenacre, M. J. (2010). *Biplots in practice*. Fundacion BBVA.
- Guiso, L., Jappelli, T., and Terlizzese, D. (1992). Saving and capital market imperfections: The italian experience. *The Scandinavian Journal of Economics*, pages 197–213.
- Guiso, L., Jappelli, T., and Terlizzese, D. (1996). Income risk, borrowing constraints, and portfolio choice. *The American economic review*, pages 158–172.
- Haider, S. J. and Stephens Jr, M. (2007). Is there a retirement-consumption puzzle? evidence using subjective retirement expectations. *The review of economics and statistics*, 89(2):247–264.

- Hall, R. E. (1978). Stochastic implications of the life cycle-permanent income hypothesis: theory and evidence. *Journal of political economy*, 86(6):971–987.
- Hamermesh, D. S. (1982). Consumption during retirement: The missing link in the life cycle. Technical report, National Bureau of Economic Research.
- Hastie, T., Tibshirani, R., and Friedman, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction*, volume 2. Springer.
- HFCN (2020a). Euro area questionnaire on household finance and consumption plus paradata. Technical report, ECB Statistics Paper.
- HFCN (2020b). The household finance and consumption survey: methodological report for the 2017 wave. Technical report, ECB Statistics Paper.
- HFCN (2020c). The household finance and consumption survey: user guide. Technical report, ECB Statistics Paper.
- Hosseini, H. M. and Kaneko, S. (2011). Dynamic sustainability assessment of countries at the macro level: A principal component analysis. *Ecological indicators*, 11(3):811–823.
- Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of educational psychology*, 24(6):417.
- Hubbard, R. G., Judd, K. L., Hall, R. E., and Summers, L. (1986). Liquidity constraints, fiscal policy, and consumption. *Brookings Papers on Economic Activity*, 1986(1):1–59.
- Hurst, E. (2007). Understanding consumption in retirement: recent developments. *Working Paper*.
- James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013). *An introduction to statistical learning*, volume 112. Springer.
- Jappelli, T. (2005). The life-cycle hypothesis, fiscal policy and social security. *PSL Quarterly Review*.
- Jappelli, T. and Modigliani, F. (1998). The age-saving profile and the life-cycle hypothesis. *Long-run growth and short-run stabilization*, page 12.
- Jappelli, T. and Pistaferri, L. (2010). The consumption response to income changes. *Annual Review of Economics*, 2:479–506.

- Jappelli, T. and Pistaferri, L. (2014). Fiscal policy and mpc heterogeneity. *American Economic Journal: Macroeconomics*, 6(4):107–36.
- Jappelli, T. and Pistaferri, L. (2017). *The economics of consumption: theory and evidence*. Oxford University Press.
- Jappelli, T. and Pistaferri, L. (2020). Reported mpc and unobserved heterogeneity. *American Economic Journal: Economic Policy*, 12(4):275–97.
- Johnson, D. S., Parker, J. A., and Souleles, N. S. (2006). Household expenditure and the income tax rebates of 2001. *American Economic Review*, 96(5):1589–1610.
- Kahn, R. F. (1931). The relation of home investment to unemployment. *The Economic Journal*, 41(162):173–198.
- Kahneman, D. and Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2):263–292.
- Kaplan, G. and Violante, G. L. (2014a). A model of the consumption response to fiscal stimulus payments. *Econometrica*, 82(4):1199–1239.
- Kaplan, G. and Violante, G. L. (2014b). A tale of two stimulus payments: 2001 versus 2008. *American Economic Review*, 104(5):116–21.
- Kaplan, G., Violante, G. L., and Weidner, J. (2014). The wealthy hand-to-mouth. Technical report, National Bureau of Economic Research.
- Karlan, D. and Zinman, J. (2008). Lying about borrowing. *Journal of the European Economic Association*, 6(2-3):510–521.
- Keynes, J. M. (1936). *The general theory of employment, interest, and money*. MacMillan.
- Mariger, R. P. (1987). A life-cycle consumption model with liquidity constraints: theory and empirical results. *Econometrica: Journal of the Econometric Society*, pages 533–557.
- Matejka, J. and Fitzmaurice, G. (2017). Same stats, different graphs: generating datasets with varied appearance and identical statistics through simulated annealing. In *Proceedings of the 2017 CHI conference on human factors in computing systems*, pages 1290–1294.

- McKenzie, D. J. (2005). Measuring inequality with asset indicators. *Journal of population economics*, 18(2):229–260.
- Modigliani, F. and Brumberg, R. (1954). Utility analysis and the consumption function: An interpretation of cross-section data. *Franco Modigliani*, 1(1):388–436.
- Nemeczek, F. and Radermacher, J. W. (2022). Personality-augmented mpc: Linking survey and transaction data to explain mpc heterogeneity by big five personality traits.
- Olafsson, A. and Pagel, M. (2018a). The liquid hand-to-mouth: Evidence from personal finance management software. *The Review of Financial Studies*, 31(11):4398–4446.
- Olafsson, A. and Pagel, M. (2018b). The retirement-consumption puzzle: New evidence from personal finances. Technical report, National Bureau of Economic Research.
- Olsson, O. and Hibbs Jr, D. A. (2005). Biogeography and long-run economic development. *European Economic Review*, 49(4):909–938.
- Parker, J. A. (1999). The reaction of household consumption to predictable changes in social security taxes. *American Economic Review*, 89(4):959–973.
- Parker, J. A., Souleles, N. S., Johnson, D. S., and McClelland, R. (2013). Consumer spending and the economic stimulus payments of 2008. *American Economic Review*, 103(6):2530–53.
- Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin philosophical magazine and journal of science*, 2(11):559–572.
- Pigou, A. C. (1936). Mr. j m keynes' general theory of employment, interest and money. *Economica*, 3(10):115–132.
- Ricardo, D. (1951). *Principles of Political Economy and Taxation, Works and Correspondence*. Cambridge University Press. edited by Pierro Sraffa.
- Robb, A. L. and Burbidge, J. B. (1989). Consumption, income, and retirement. *Canadian Journal of Economics*, pages 522–542.
- Schwerdt, G. (2005). Why does consumption fall at retirement? evidence from germany. *Economics Letters*, 89(3):300–305.

- Shapiro, M. D. and Slemrod, J. (2003). Consumer response to tax rebates. *American Economic Review*, 93(1):381–396.
- Shefrin, H. M. and Thaler, R. H. (1988). The behavioral life-cycle hypothesis. *Economic inquiry*, 26(4):609–643.
- Simpson, E. H. (1951). The interpretation of interaction in contingency tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 13(2):238–241.
- Souleles, N. S. (1999). The response of household consumption to income tax refunds. *American Economic Review*, 89(4):947–958.
- Spolaore, E. and Wacziarg, R. (2013). How deep are the roots of economic development? *Journal of economic literature*, 51(2):325–69.
- StataCorp, L. (2021). Stata statistical software: Release 17 (2021). *College Station, TX: StataCorp LP*.
- Thaler, R. (1985). Mental accounting and consumer choice. *Marketing science*, 4(3):199–214.
- Thaler, R. H. (1990). Anomalies: Saving, fungibility, and mental accounts. *Journal of economic perspectives*, 4(1):193–205.
- Thaler, R. H. (1994). Psychology and savings policies. *The American Economic Review*, 84(2):186–192.
- Vyas, S. and Kumaranayake, L. (2006). Constructing socio-economic status indices: how to use principal components analysis. *Health policy and planning*, 21(6):459–468.
- Wager, S. and Athey, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, 113(523):1228–1242.
- Wagner, C. H. (1982). Simpson's paradox in real life. *The American Statistician*, 36(1):46–48.

Appendices

A Descriptive Statistics



Figure A.1: Histograms of key variables. All data is taken from the third wave (2017) of the HFCS (HFCN, 2020b).

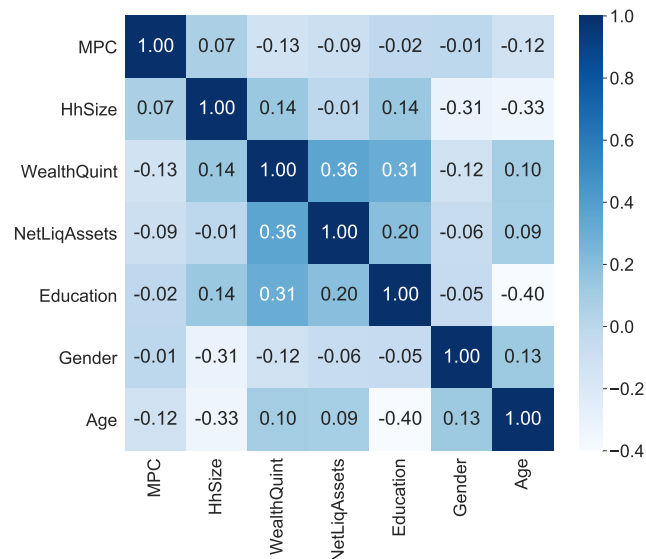


Figure A.2: Correlations of key variables. All data is taken from the third wave (2017) of the HFCS (HFCN, 2020b).

B Datasaurus Dozen



Figure A.3: "Datasaurus Dozen" by Alberto Cairo. All datasets have the same summary statistics (Matejka and Fitzmaurice, 2017).

C Additional PCA-Biplot

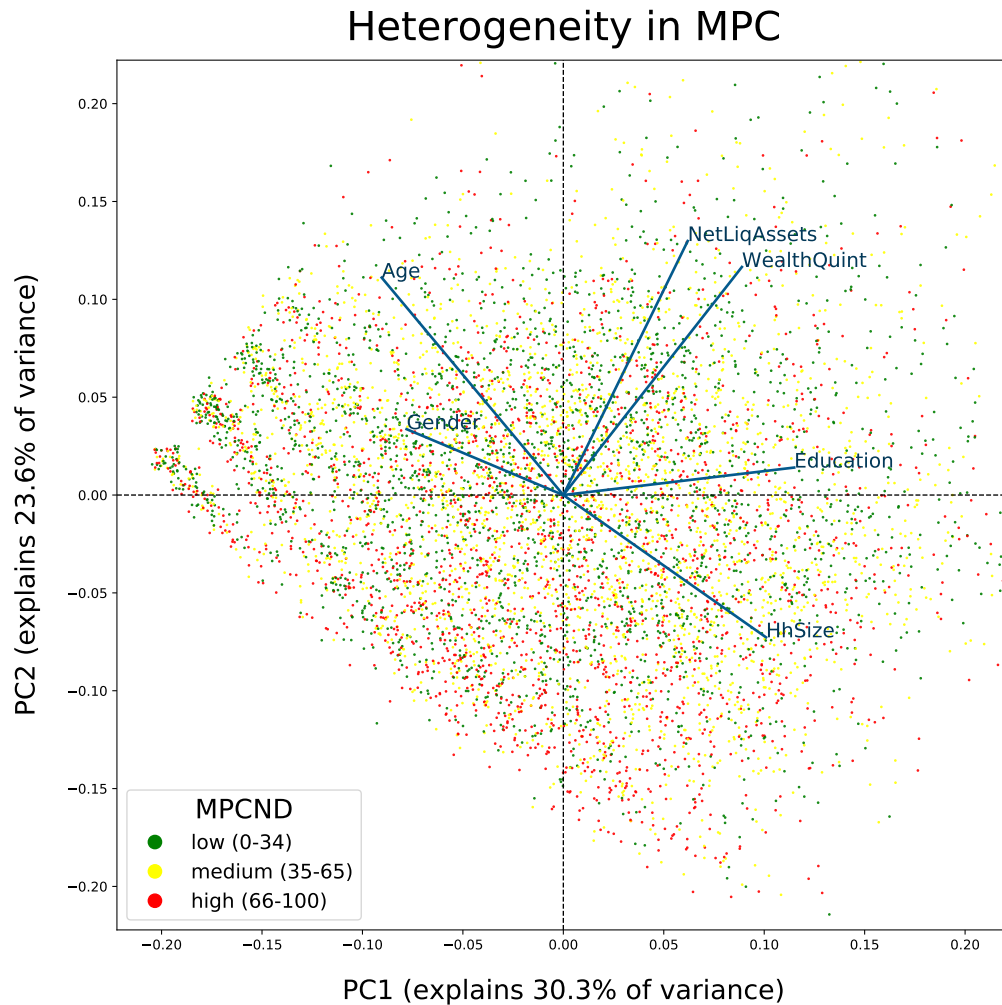


Figure A.4: PCA-Biplot Italy: colour-coded MPC on small set of demographics. All data is taken from the third wave (2017) of the HFCS (HFCN, 2020b).

Recent Issues

No. 381	Olga Goldfayn, Nathanael Vellekoop	The Role of Personality Traits in Household Loan Expectations and Borrowing Constraints
No. 380	Alexander Hillert, Alexandra Niessen-Ruenzi, Stefan Ruenzi	Mutual Fund Shareholder Letters: Flows, Performance, and Managerial Behavior
No. 379	Irene Mecatti, Tobias Tröger	Who Should Hold Bail-Inable Debt and How can Regulators Police Holding Restrictions Effectively?
No. 378	Sabine Bernard, Benjamin Loos, Martin Weber	How Speculative Asset Characteristics Shape Retail Investors' Selling Behavior
No. 377	Konstantin Egorov	Optimal Policy under Dollar Pricing
No. 376	Raphael Abiry, Marien Ferdinandusse, Alexander Ludwig, Carolin Nerlich	Climate Change Mitigation: How Effective is Green Quantitative Easing?
No. 375	Olivia S. Mitchell, Orly Sade, Abigail Hurwitz	Testing Methods to Enhance Longevity Awareness
No. 374	Franco Fiordelisi, Giulia Fusi, Angela Maddaloni, David Marqués-Ibáñez	Pandemic Lending: Micro and Macro Effects of Model-Based Regulation
No. 373	Emily Kormanyos, Tobin Hanspal, Andreas Hackethal	Do Gamblers Invest in Lottery Stocks?
No. 372	Alessandro Pollastri, Paulo Rodrigues, Christian Schlag, Norman Seeger	A Jumping Index of Jumping Stocks? An MCMC Analysis of Continuous-Time Models for Individual Stocks
No. 371	Mariano Massimiliano Croce, Tatyana Marchuk, Christian Schlag	The Leading Premium
No. 370	Benjamin M. Abdel-Karim, Kevin Bauer, Oliver Hinz, Michael Nofer	The Effects of Discontinuing Machine Learning Decision Support
No. 369	Katja Langenbucher	Consumer Credit in The Age of AI – Beyond Anti-Discrimination Law
No. 368	Vanya Horneff, Raimond Maurer, Olivia S. Mitchell	How Would 401(k) 'Rothification' Alter Saving, Retirement Security, and Inequality?
No. 367	Jens Lausen, Benjamin Clapham, Peter Gomber, Micha Bender	Drivers and Effects of Stock Market Fragmentation - Insights on SME Stocks