The Dissertation Committee for Rostislav Bogoslovskiy Certifies that this is the approved version of the following dissertation:

Essays in Macroeconomics

Committee:

Olivier Coibion, Supervisor

Andrew Glover

Saroj Bhattarai

Ricardo Fernholz

Matthias Kehrig

Essays in Macroeconomics

by

Rostislav Bogoslovskiy, B.S.; M.S.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

December 2016

Dedication

To my parents and wife

Acknowledgements

I would like to thank my committee members Olivier Coibion, Andy Glover, Saroj Bhattarai, Matthias Kehrig, and Rick Fernholz for helpful feedback and suggestions. I am grateful to my advisor Olivier Coibion for his guidance. Special thanks to my wife Natalia Em and my friend Anna Popova for their support. I appreciate the financial support from the Department of Economics at the University of Texas at Austin in the form of scholarships, tuition assistantships, and TA/RA assignments.

Essays in Macroeconomics

Rostislav Bogoslovskiy, PhD

The University of Texas at Austin, 2016

Supervisor: Olivier Coibion

Chapter 1 “Income Inequality, Income Mobility, and the Supply of Credit” studies a mechanism that can potentially explain the negative correlation between income inequality and intergenerational mobility in the cross-section of US commuting zones. The mechanism operates through human capital accumulation and unsecured credit. Parent income is modeled as dependent on innate ability and a random income shock. When a parent takes out a loan to finance their child's college degree, the interest rate faced by the parent depends on the default risk perceived by banks, which is greater in areas with higher local inequality due to a higher variance of the income shock. This higher cost of credit reduces investment in child human capital on the part of low-income parents and thereby can lower intergenerational income mobility. My model of human capital accumulation with unsecured credit predicts about half of the effect of income inequality on income mobility observed in the data.

Chapter 2 “Classification Models: Predicting Moves with USAA Customer Data” compares three statistical models to predict residential moves using USAA customer data. Logistic regression, LASSO, and Random Forest all correctly predict moving outcome

about 69% of the time in the out-of-sample test. While this accuracy is achieved with hundreds of available variables, using only five most powerful predictors reduces out-of-sample prediction accuracy by only about 1 percentage point. Age is the single strongest predictor of moving, followed by rental/homeownership information and military status. Additional local factors, such as MSA-level unemployment rate or population are found to have no impact on the likelihood of moving.

To quantify the welfare effects of prohibiting universal default - one of the key provisions of the Credit CARD Act of 2009 – in Chapter 3 “Is Universal Default Socially Desirable?”, I present a model of unsecured consumer credit, with borrowers taking out two defaultable loans from different creditors, and compare social welfare when the creditors can practice universal default and when they cannot. Prohibiting universal default makes it cheaper to default on one credit card, increases bankruptcy rate, makes creditors raise interest rates, and lowers social welfare by 1.78% of lifetime consumption.

Table of Contents

List of Tables

List of Figures

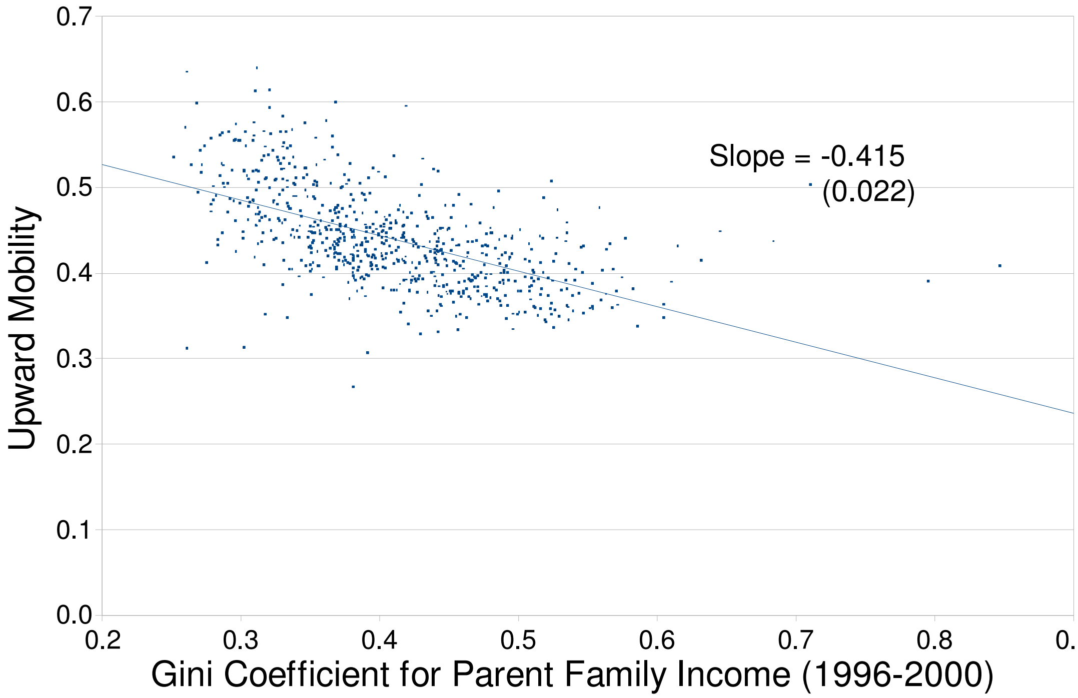
Chapter 1: Income Inequality, Income Mobility, and the Supply of Credit

1.1 INTRODUCTION

A strong negative correlation between income inequality and measures of intergenerational income mobility has recently been documented both across countries (Corak 2013, Andrews and Leigh 2009, Blanden 2013), and across commuting zones[1] within the U.S. (Chetty et al. 2014, see Figure 1.1), but the source of this correlation remains unexplored, with potentially important policy implications. For example, understanding the underlying mechanism could suggest new policies aimed at improving the chances of children born into low-income families to climb up the social ladder. Some possible explanations have been suggested. Intragenerational income inequality and intergenerational income mobility are both affected by the degree of heritability of income-related traits, the degree of parents' propensity to invest in their children's human and non-human capital (Becker and Tomes 1976), as well as the efficacy of human capital investment, the earnings return to human capital, and the progressivity of public investment in human capital (Solon 2004). Differences across countries or CZs in both intergenerational mobility and income inequality could arise from differences in any of these factors.

[1] Commuting zones (CZs) are geographical aggregations of counties that are similar to metropolitan areas but cover the entire U.S. (Tolbert and Sizer 1996).

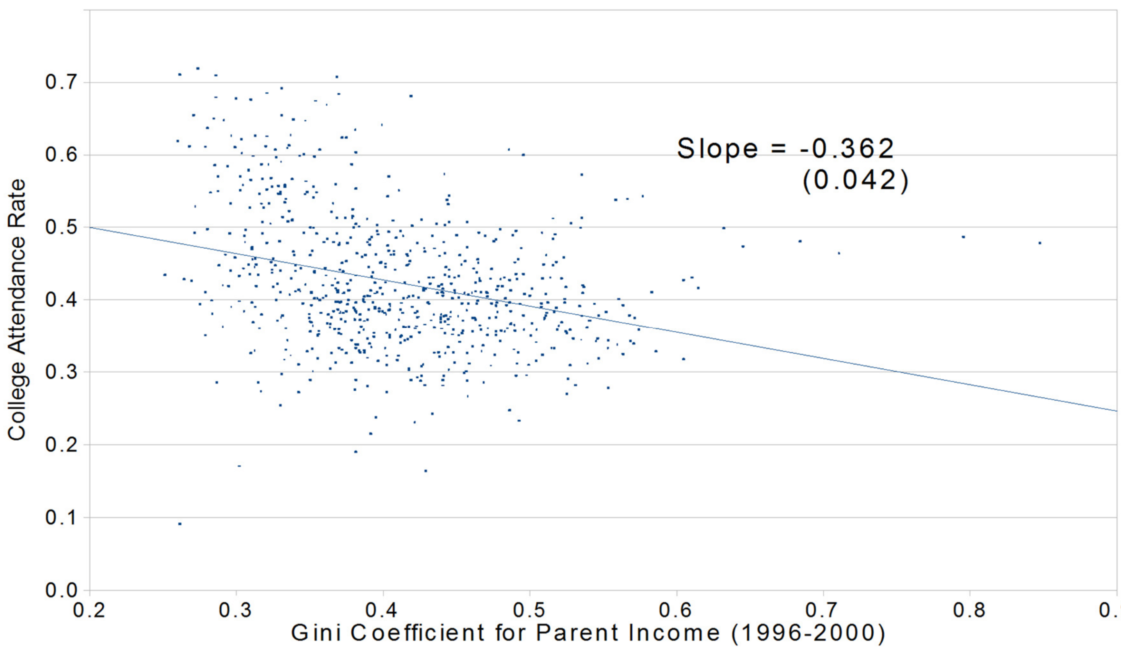
Figure 1.1: Upward mobility and inequality in the US commuting zones



Source: Equality of Opportunity Project, http://www.equality-of-opportunity.org, and author's calculations. Note: Gini coefficient is computed based on mean parent income for 1996-2000. For a given CZ, upward mobility is computed as the mean income rank of children in that CZ with parents in the bottom half of the national parent income distribution. For definitions of parent income and child income, see Chetty et al. 2014.

While each of these factors could account for a link between higher inequality and lower income mobility, none explains how the level of parent income inequality per se can affect income mobility directly. This paper proposes such a causal relationship and explores under what conditions it can arise. If parent income is a function of parent ability and a random income shock, then the level of income inequality across parents depends on the distributions of ability and the random component. In a setting where parents rely in part on unsecured credit to invest in their child's education before the income shock in realized, higher variance of the income shock, conditional on parent's ability, implies a greater likelihood that parents will be unable to, or choose not to, repay the loan. This higher probability of default induces banks to charge higher interest rates to relatively low-income applicants in higher-inequality regions, and may even completely shut down access to credit for sufficiently low-income applicants, thus making the education loan less accessible to low-income applicants. In such a setting, educational attainment will be lower for low-income applicants in higher-inequality regions, thereby generating a negative correlation between income inequality and intergenerational mobility as observed in the data.

Figure 1.2: Income inequality vs college attendance rate for children of parents with income below nationwide median



Source: Equality of Opportunity Project, http://www.equality-of-opportunity.org, and author's calculations. Note: each data point represents a US commuting zone. College attendance rate is defined as the fraction of children having one or more 1098-T forms filed on one's behalf when the child is aged 18-21, see Chetty et al. 2014. College attendance rate is computed for children whose parent's income over 1996-2000 was below the median in the nationwide distribution of parent income. Gini coefficient is computed based on mean parent income for 1996-2000.

Figure 1.2 plots the cross-CZ relationship between the Gini coefficient of parent income over 1996-2000 and mean college attendance rate for children of parents whose income over 1996-2000 was in the bottom half of the nationwide parent income distribution. My mechanism is consistent with the strong negative correlation between parent income inequality and college attendance rate among children from poor families.

A key insight of the paper is that the source of regional differences in inequality plays a central role in the potential importance of this channel. If differences in inequality are entirely due to the differences in the dispersion of abilities, parents with the same ability in high inequality and low inequality regions pose the same default risk to the lender and hence are offered the same credit terms, which makes them equally likely to invest in child's college. However, the thicker left tail of the distribution of abilities in the high inequality region would imply a larger fraction of parents who cannot afford a loan resulting in lower mobility in the high inequality region. If, on the other hand, differences in inequality are attributed to differences in the variance of the random component of parental income, then there are two effects on mobility at play. First, a low-ability parent in a high inequality region is regarded as a higher risk than a parent with the same ability in the low inequality region and hence is offered worse credit terms making him less likely to invest in child's college. This effect lowers income mobility. Second, higher variance of the random (or "luck") component of income mechanically raises the odds of children of low-ability parents of moving up the social ladder. This effect increases income mobility. The overall effect of the variance of the random component on mobility is a priori ambiguous and needs to be explored within a quantitative model.

The strength of the proposed channel that causes mobility to be lower in regions with higher inequality depends critically on the source of regional variation in inequality of parent income. More empirical work is needed to understand what drives this regional variation in inequality. Although some studies have performed the decompositions of the increase in nationwide income inequality over several past decades into changes in persistent and transient components of income (DeBacker et al. 2013, Heathcote, Perri, and Violante 2010, Kopczuk, Saez, and Song 2010 among others), no studies have attempted this decomposition at the cross-sectional level (e.g. across commuting zones).

While the literature on economic and political consequences of income inequality at the national level and its increase over the past three decades is quite large (see, for example, Rajan 2010, Stiglitz 2013, Piketty 2014), relatively little attention has been given to the effects of local income inequality. According to the standard model of household debt, the life-cycle/permanent income model of Ando and Modigliani 1963 and Friedman 1957, as long as the variation in local income inequality across regions is at least partially due to the variation in the volatility of the transitory component of income, household borrowing should be higher in regions with higher inequality since households use debt to smooth consumption over temporary income declines. Another strand of literature that goes back to Veblen 1899 and Duesenberry 1949 investigates how relative income considerations and social comparisons affect consumption and borrowing decisions. In a recent notable paper, Bertrand and Morse 2013 study whether consumption of middle-income households is affected by the consumption or the income of the rich (80th percentile of income distribution)[2]. They use the Consumer Expenditure Survey and explore the state-year variation in the consumption of the rich and show that consumption of the non-rich is higher when consumption of the rich is higher, holding everything else constant including own income. Bertrand and Morse attribute this positive association to the "keeping up with the Joneses" effect, with an important prediction that demand for credit by low and middle income households must be higher in regions with higher income inequality. Coibion et al. 2014 test this prediction using data on household debt at different levels of aggregation (zip code, county, and state) and find that holding everything else constant, households in higher inequality regions accumulated less debt over the course of early-mid 2000s – the opposite of what the "keeping up with the Joneses" hypothesis would suggest. Moreover, using data on mortgage applications, Coibion et al. 2014 find that low income rank households have a higher probability of getting their mortgage application denied in high inequality regions than do households with the same income rank in low inequality regions, holding everything else constant. This finding emphasizes the importance of local inequality for the supply of credit. Coibion et al. 2014 present a theoretical model in which banks use the level of local inequality along with the borrower's income to infer the borrower's unobserved type and likelihood of default. Since the predictability of unobserved type based on observed income increases with inequality, in

[2] See also Neumark and Postlewaite 1998, Zizzo and Oswald 2001, Christen and Morgan 2005, Luttmer 2005, Daly and Wilson 2006, Maurer and Meier 2008, Charles, Hurst and Roussanov 2009, Kuhn et al. 2010, Heffetz 2011, Guven and Sorensen 2012, and Bricker et al. 2014 on relative income concerns and social comparisons.

equilibrium low income households face a higher interest rate or a higher probability of getting the loan application denied, the higher the level of inequality, all else equal.

In this paper, I take the insights of Coibion et al. 2014 one step further and introduce an intergenerational component to study how income inequality can affect intergenerational upward income mobility. Specifically, I assume that the borrowing is done by the parents for the sole purpose of financing their child's college education. To keep the model simple, I assume that the parent's ability type is perfectly observed by the bank. Parent income depends on ability and an income shock. The income shock is realized after the borrowing is done. The level of local inequality plays a crucial role in the determination of the probability of default perceived by the bank and thus is incorporated in the interest rate and the loan acceptance decision. As a result, the fraction of low ability parents (who on average also tend to have low income) investing in child's college education is lower in high inequality regions than in low inequality regions, which makes poverty more persistent across generations in higher inequality regions.

This paper also contributes to the growing literature on intergenerational mobility. Studies of the causes of spatial variation in income mobility, which is the focus of this paper, were limited to cross-country differences in mobility and attempts to interpret them as evidence of cross-country differences in the efficacy of institutions for human capital investment (Bjorklund and Jantti 1997, Checchi et al. 1999). A few exceptions are worth noting. Cutler and Glaeser 1997 use cross-MSA level data to study the effect of spatial segregation of blacks on schooling and employment and find that blacks in more segregated areas have significantly worse outcomes than blacks in less segregated areas. Graham and

Sharkey 2013 find that greater economic segregation is associated with lower mobility. Chetty et al. 2014, use US population tax data, which allow to reliably link children born in 1980 or later to their parents, to identify five factors that strongly correlate with intergenerational income mobility. Areas with high income mobility tend to have less income inequality, less residential segregation, better primary schools, greater social capital, and greater family stability. Chetty et al. 2014 do not propose any causal mechanisms that determine income mobility but provide publicly available aggregated data to "facilitate research on such mechanisms". This paper explores one such causal mechanism.

The paper is organized as follows. Section 1.2 describes the model and the calibration process. Section 1.3 reports the results and estimates of the goodness of fit of the model. Section 1.4 concludes.

1.2 MODEL

1.2.1 Environment and Parent's Problem

There is a unit mass of parents whose ability type μ is drawn from lognormal distribution, $\ln \mu \sim N(m, \sigma_\mu^2)$. Each parent has a child, whose ability type μ_c is related to the parent's according to

$$\ln \mu_c = \zeta \ln \mu + (1-\zeta) \ln m + \ln \eta \, ,$$

where $\eta \sim N(0, (1-\zeta^2)\sigma_\mu^2)$ is independent of μ and iid across children, and $0 < \zeta < 1$. Child ability follows the same lognormal distribution as parent's, $\ln \mu_c \sim N(m, \sigma_\mu^2)$. Parents care both about their own as well as their children's consumption, discounting the latter by a factor $\alpha \leq 1$ which captures the degree of parental altruism.

Each parent makes one or two binary decisions. The first decision is whether or not to upgrade their child's ability by investing in additional schooling. To do so, the parent needs to take out a loan of a fixed size T from a bank at the gross interest rate R which depends on parent's ability μ. If a loan is taken out by the parent, the child's ability becomes $\mu_c + \gamma$. Parameter γ captures the expected return to additional schooling or training on the part of the child.

After the borrowing decision is made, the parent's income shock ε and the child's income shock ε_c are realized. Both shocks are multiplicative and are independently drawn from lognormal distribution $\ln \varepsilon,\ \ln \varepsilon_c \sim N(0, \sigma_\varepsilon^2)$. If the parent has not taken out a loan, then parent income is equal to $y = \mu\varepsilon$, which the parent consumes, and the child income is equal to $y_c = \mu_c \varepsilon_c$, which the child consumes, and there is no additional decision for the parent to make.

If, on the other hand, the parent has taken out a loan to upgrade child's ability, the parent decides whether or not to repay the loan to the bank. If the parent pays back the loan, RT is subtracted from his income, which leaves $\mu\varepsilon - RT$ for the parent to consume. The parent has the option to default on the loan, in which case the parent gets to consume the entire income $\mu\varepsilon$ but is punished by a fixed utility cost D. I assume that the parent automatically defaults, if $\mu\varepsilon - RT < 0$, i.e. if repaying the loan would make consumption negative.

Mathematically, a parent with ability type μ solves the following maximization problem

$$\max_{e,d\in\{0,1\}} \mathbb{E}\left[\ln(\mu\varepsilon - e(1-d)RT) - Dd + \alpha\ln((\mu_c + \gamma e)\varepsilon_c)\right]$$

$$s.t. \ln\mu_c = \zeta\ln\mu + (1-\zeta)m + \ln\eta,$$

where the expectation is taken over the joint distribution of $(\varepsilon, \varepsilon_c, \eta)$.

It is possible that no bank is willing to lend to a certain type μ. In that case, the parent's problem is degenerate: e and d are automatically set to zero and the parent simply consumes his or her income $\mu\varepsilon$ at the end of the period.

Variables e and d denote indicators for the child ability upgrade decision and the default decision, respectively and α is the degree of parental altruism.

1.2.2 Bank's Problem

Assume a perfectly competitive banking industry populated by infinitely many banks. Each bank B observes parent ability type μ and decides on the interest rate $R_b(\mu)$. Conditional on taking out a loan at the interest rate $R_b(\mu)$ offered by the bank, the parent will default with some probability, so when choosing the interest rate, the bank has to balance the benefit of higher revenue with the cost associated with the higher probability of default. Bank B's profit maximization problem is given μ,

$$\max_{R_b(\mu)} \mathbb{E}\left[\pi | e = 1; \mu\right] = -T(1+\tau) + P\{d = 0 | e = 1; \mu\}R_b(\mu)T$$

$$s.t. R_b(\mu) \leq \min\{R_{-b}(\mu)\}$$

where τ is the bank's exogenous cost of funds. Parents borrow from the bank offering the lowest interest rate, so each bank's constraint states that own interest rate should be the lowest, should the bank be willing to offer a loan at that interest rate. Implicit in the formulation of the bank's maximization problem is the assumption that the bank can always choose to offer no credit. By doing so, a bank can secure zero profit.

Due to the perfect competition assumption, in equilibrium, the expected profit equals zero for any ability type μ. To prove this, suppose there is an equilibrium in which there is a bank that earns a positive expected profit on loans made to some ability type μ_0. Then another bank can offer a slightly lower interest rate to μ_0, win the business of all parents with μ_0 and still have positive profit. By way of contradiction, this shows that there can not be an equilibrium in which there exists a bank that earns a positive profit on loans to some ability type. Therefore, expected profit from lending to any ability type must be equal to zero in equilibrium. Moreover, if for a given μ there are multiple values of the interest rates $R(\mu)$ at which the bank breaks even, the equilibrium interest rate is the smallest of such values, due to the perfect competition forces.

Since all banks are symmetric and perfectly competitive, the model can be simplified by assuming there is instead a representative bank that offers the smallest possible interest rate at which it breaks even or offers no credit, if breaking even is not feasible.

1.2.3 Equilibrium

1.2.3.1 Definition

An equilibrium is a collection of functions $\{R(\cdot), e(\cdot), d(\cdot)\}$ defined on $(0, \infty)$ – the support of μ – such that

1) given $e(\cdot)$ and $d(\cdot), R(\cdot)$ is the smallest interest rate at which the representative bank breaks even, or $R(\cdot)$ is not defined, if breaking even is not feasible, and

2) given $R(\cdot)$, functions $e(\cdot)$ and $d(\cdot)$ solve the parent's maximization problem.

1.2.3.2 Solution

Start with the bank's problem. The zero-expected-profit condition implies that for any μ,

$$R(\mu) = \frac{1+\tau}{P\{d=0|e=1;\mu\}}.$$

The equilibrium interest rate is proportional to the bank's costs of funds, which it passes on to borrowers. It is also inversely proportional to the probability of no default conditional on taking out a loan and on borrower's type, meaning that the bank demands higher interest rate from riskier borrowers.

Now solve for the parent's default decision. An ability type μ parent who has taken out a loan at the interest rate $R(\mu)$ defaults on the loan if and only if $\varepsilon < R(\mu)T/\mu$ (in this case repaying the loan would cause consumption to go negative) or else if utility of default exceeds utility of repayment:

$$\ln(\mu\varepsilon) - D > \ln(\mu\varepsilon - R(\mu)T),$$

which implies

$$\varepsilon < \frac{R(\mu)Te^D}{\mu(e^D - 1)}.$$

Define $\varepsilon^*(\mu)$ to be

$$\varepsilon^*(\mu) = \frac{R(\mu)Te^D}{\mu(e^D - 1)}.$$

So the parent's default decision follows a simple threshold rule: default if and only if $\varepsilon < \varepsilon^*(\mu)$. Note that for any μ, the default region $\varepsilon < \varepsilon^*(\mu)$ covers the set of ε values that trigger "automatic" default, $\varepsilon < R(\mu)T/\mu$, since $\varepsilon^*(\mu) > R(\mu)T/\mu$. Note that $\varepsilon^*(\mu)$ increases in R and T and decreases in D. Higher interest rate or loan amount shift $\varepsilon^*(\cdot)$ curve up and make default relatively more attractive, and thus more likely, for all μ types, in comparison to no default option. On the other hand, higher default punishment D lowers $\varepsilon^*(\mu)$ for all μ and makes default relatively less attractive and therefore less likely.

Feeding this default decision back in the bank's problem implies that the equilibrium interest rate is the solution to the following fixed-point problem:

$$R(\mu) = \frac{1+\tau}{P\{\varepsilon > \varepsilon^*(\mu)\}} = \frac{1+\tau}{1 - F_\varepsilon(\varepsilon^*(\mu))} = \frac{1+\tau}{1 - F_\varepsilon(\frac{R(\mu)Te^D}{\mu(e^D - 1)})}$$

where F_ε is the cumulative distribution function of the distribution of ε. Note that this equation may not have a solution for some values of μ, in which case the bank offers no loan to those parents.

Using backward induction, at the point when parents get to make the child ability upgrade decision, they know $\varepsilon^*(\mu)$, i.e. they know when they are going to default. Assuming $\mu\varepsilon > R(\mu)T$, parent's expected utility of upgrading child ability is equal to

$$\int_0^{\varepsilon^*(\mu)} (\ln \mu\varepsilon - D) dF_\varepsilon + \int_{\varepsilon^*(\mu)}^{\infty} \ln(\mu\varepsilon - R(\mu)T)\, dF_\varepsilon$$

$$+ \alpha \int_0^{\infty} \int_0^{\infty} (\ln\left(\mu^\zeta e^{(1-\zeta)m}\eta + \gamma\right) + \ln \varepsilon_c) dF_{\varepsilon_c} dF_\eta,$$

while expected utility of not upgrading child ability is equal to

$$\int_0^{\infty} \ln(\mu\varepsilon)\, dF_\varepsilon + \alpha \int_0^{\infty} \int_0^{\infty} (\ln\left(\mu^\zeta e^{(1-\zeta)m}\eta\right) + \ln \varepsilon_c) dF_{\varepsilon_c} dF_\eta.$$

Using the fact that ε and ε_c have the same distribution, one can find that the parent will upgrade child ability if and only if

$$\alpha \int_0^{\infty} \ln\left(\frac{\mu^\zeta e^{(1-\zeta)m}\eta + \gamma}{\mu^\zeta e^{(1-\zeta)m}\eta}\right) dF_\eta > \int_{\varepsilon^*(\mu)}^{\infty} \ln\left(\frac{\mu\varepsilon}{\mu\varepsilon - R(\mu)T}\right) dF_\varepsilon + DF_\varepsilon(\varepsilon^*(\mu)).$$

On the left-hand side of this equation is the expected benefit on investment in child ability upgrade, which comes from higher child's expected consumption, while on the right-hand side is the expected cost of such investment, which is the sum of expected reduction in parent's consumption should the parent repay the loan and the expected utility cost of default should the parent default on the loan.

1.2.3.3 Calibration

The model has one period and since the borrowing decision falls on the beginning of the period and the repayment decision falls on the end of the period, the length of the period is set to 10 years which is a typical duration of a second mortgage or home equity line of credit, often used by parents to finance their children's college education. The period that the model is calibrated to is 2000-2010.

Mean logarithm of ability type, m is set equal to 13.1062, so that $e^{13.1062}$ is equal to \$492,000 - ten times the median parent income in 2000 (Chetty et al. 2014). The variances of μ and ε are calibrated using the following procedure.

The mean Gini coefficient of parent income in the cross section of U.S. commuting zones in 2000 was 0.398. I assume that the actual distribution of parent income in every year from 2000 to 2010 in the data followed the lognormal distribution $\ln y = \ln \mu + \ln \varepsilon$, where $\ln \mu \sim N(m, \sigma_{\mu,1}^2)$ and $\ln \varepsilon \sim N(0, \sigma_{\varepsilon,1}^2)$. I fixed the ratio $\sigma_{\varepsilon,1}/\sigma_{\mu,1}$ at 1/2 based on the estimate of the ratio of variances of one-year temporary and permanent components of income $\sigma_{\varepsilon,1}^2/\sigma_{\mu,1}^2 = 1/4$ in DeBacker et al. 2013, and I chose $\sigma_{\varepsilon,1}$ to match the median one-year Gini of parent income in 2000. The resulting values of $\sigma_{\varepsilon,1}$ and $\sigma_{\mu,1}$ are 0.3296 and 0.6592, respectively. Next, using these values of the variances of one-year μ and ε, I simulated parent income for a large number of parents over 10 years by drawing the values of μ from the lognormal distribution $N(m, \sigma_{\mu,1}^2)$ once and keeping them constant over the ten-year period, and redrawing ε from $N(0, \sigma_{\varepsilon,1}^2)$ in every year for every parent. I then computed the ten-year income for every parent and found the Gini coefficient in the resulting distribution, it was equal to 0.3635. Then I set σ_μ equal to $\sigma_{\mu,1}$ and chose the standard deviation of the ten-year income shock σ_ε so that the lognormal distribution of ten-year parent income $\ln y = \ln \mu + \ln \varepsilon$, where $\ln \mu \sim N(m, \sigma_\mu^2)$ and $\ln \varepsilon \sim N(0, \sigma_\varepsilon^2)$, produces Gini of parent income equal to 0.3635. The resulting values are $\sigma_\varepsilon = 0.1072$ and $\sigma_\mu = 0.6592$.

The value of γ is set equal to the college wage premium over ten years in 2000, \$220,000 (Daly and Bengali 2014). The value of T is set equal to \$58,000 which was the year 2000 present value of average four-year cost of tuition, room, and board over 2000-2004. The bank's cost of funds τ is set equal to $(1.031)^{10} - 1 = 0.357$, where 0.031 is the geometric average federal funds rate over 2000-2010.

The parent altruism rate α is set equal to 0.69, which is the estimate of parent altruism rate in Nishiyama 2000. The degree of heritability of ability ζ is set equal to 0.32, which is the estimate of the intergenerational elasticity of IQ score estimated by Black et al. 2009.

The utility cost of default D is the only parameter that cannot be tied to data or existing estimates. It is calibrated to match the college entry rate for the birth cohort of 1980, which is 0.60 (from Bailey and Dynarski 2011). Annualized parameter values are summarized in Table 1.1.

Table 1.1. Calibrated parameters

Parameter	Definition	Value	Target
m	Mean of $\ln \mu$	13.1	Median parent income (\$49,200)
σ_μ	St. dev. of $\ln \mu$	0.6592	Parent income Gini (0.398)
σ_ε	St. dev. of $\ln \varepsilon$	0.1072	See text
γ	College premium	22,000	Average college wage premium
T	Cost of college	5,800	Average tuition, room, and board cost
τ	Cost of funds	0.031	Average Fed. funds rate over 2000-10
α	Altruism rate	0.69	Nishiyama (2000) estimate
ζ	Ability heritability	0.32	Black et al. (2009) estimate
D	Cost of default	0.27	College entry rate = 0.6

The ultimate goal of this paper is to evaluate the effect of parent income inequality on intergenerational income mobility. I first solve the model using the benchmark calibration described above and then holding all other parameters unchanged, I consider a new set of values of σ_μ and σ_ε so that the resulting Gini coefficient of parent income is one standard deviation higher than the mean in the cross-section of U.S. commuting zones. In the data (Chetty et al. 2014), one standard deviation is 0.081, so I target the Gini of parent income of 0.398+0.081=0.479. The new values of σ_μ and σ_ε that hit the target are 0.8104 and 0.1339, respectively, both values are about 23% higher than the original values.

1.3 RESULTS

1.3.1 Equilibrium

1.3.1.1 Interest Rates

Figure 1.3 visualizes solution to the bank's problem for the benchmark parameterization. Each curve shows bank profit as a function of the offered gross interest rate (compounded over 10 years) for a particular level of parent ability. For any given interest rate, bank profit increases in ability type, so profit curves rise with ability. For a relatively low ability level $\mu_1 = \$40{,}600$, the bank can not earn a positive profit, so a parent with this ability will not be offered a loan. For a relatively high ability level $\mu_2 = \$44{,}300$, the bank can earn a positive profit, but in equilibrium, the perfect competition will push the interest rate to the lowest possible value at which the bank is willing to offer a loan. That equilibrium interest rate is the smaller interest rate at which the profit equals zero and it is labeled $R(\mu_3)$ on the graph. Ability level $\mu_1 = \$42{,}500$ is the marginal parent ability: for this ability level, the profit curve touches the zero-profit horizontal line. The bank will be willing to lend to this parent at $R(\mu_2)$. All parents with ability type above μ_2 will be offered a loan and all parents with ability type below μ_2 will be denied a loan. It is also apparent from the graph that the equilibrium interest rate falls as ability level rises, which confirms the intuition.

Figure 1.4 shows the same profit curves as in Figure 1.3 (three solid lines) and adds three profit curves for the same levels of ability μ_1, μ_2, μ_3 for the "+1SD" CZ (dashed lines). Since a parent from the "+1SD" CZ poses a greater default risk to the bank than a parent with the same level of ability from the median CZ, we would expect the bank to offer worse loan terms to the parents from the high inequality CZ. Indeed, note that each dashed profit curve lies below the solid profit curve for the same μ and this creates a wedge between interest rates across the two CZs ($R(\mu_2)$ in "+1SD" CZ is greater than $R(\mu_2)$ in the median CZ) and also creates a margin of parent ability types who do not receive credit in the high inequality CZ but do receive credit in the median inequality CZ (parents with ability greater than μ_2 but less than μ_3 are offered a loan in the median CZ but are denied a loan in "+1SD" CZ).

Figure 1.3: Profit curves for some values of μ for MEDIAN CZ

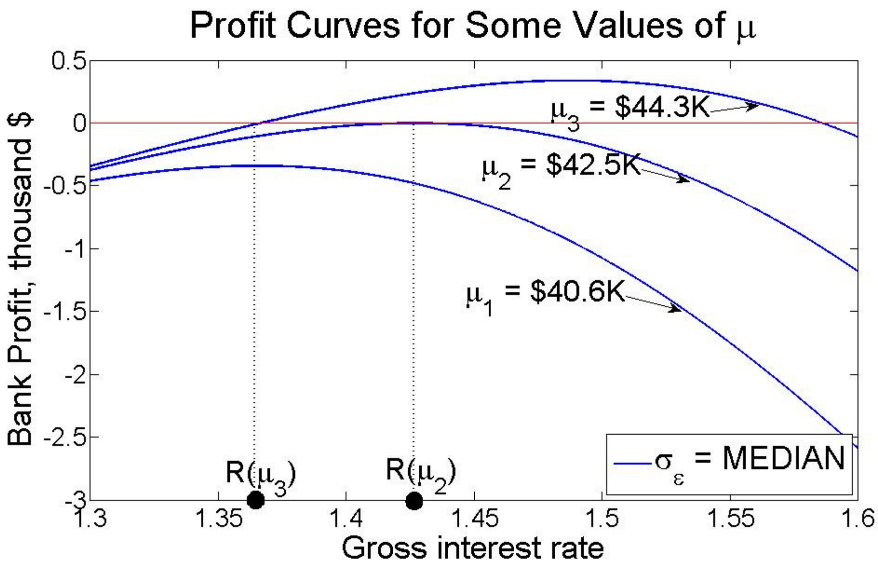
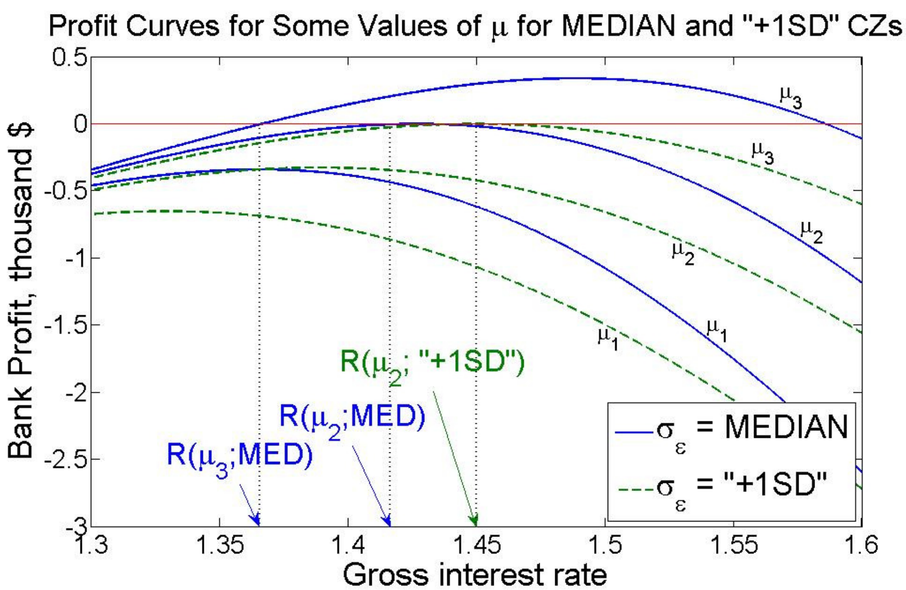


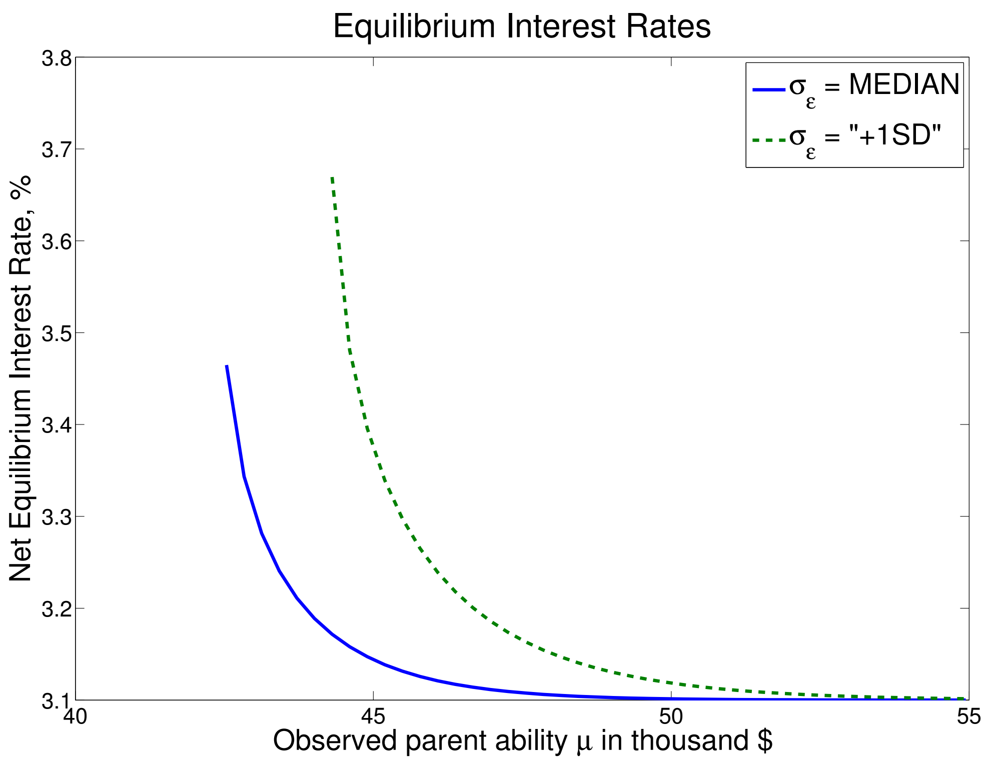
Figure 1.4: Profit curves for some values of μ for "+1SD" CZ



Note: In Figure 1.3, each curve shows bank profit as a function of the offered gross interest rate (compounded over 10 years) for a particular level of parent ability. Figure 1.4 plots the same profit curves as in Figure 1.3 (three solid lines) and adds three profit curves for the same levels of ability μ_1, μ_2, μ_3 for the "+1SD" CZ (dashed lines).

Figure 1.5 plots equilibrium net annual interest rates as functions of observed annual ability μ for the median and the "+1SD" commuting zones. Both interest rate curves converge to 3.1%, which is equal to the cost of funds to the bank. This indicates that as the parent ability increases, the probability of default tends to zero, so does the default interest rate premium, and the perfectly competitive bank offers a loan almost at cost.

Figure 1.5: Equilibrium interest rates

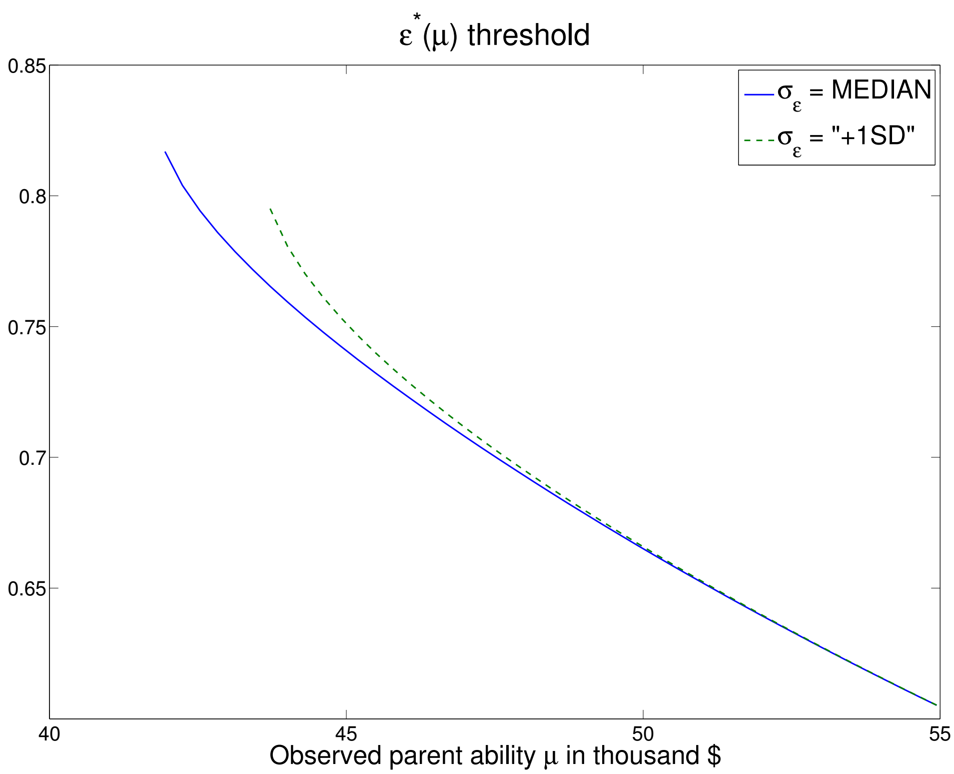


Note: Annual equilibrium net interest rates in percentage terms are plotted against observed annual parent ability μ for the median and the "+1SD" CZs.

1.3.1.2 Parent Default Decision d and College Decision e

A parent with ability type μ defaults on a loan whenever $\varepsilon < \varepsilon^*(\mu)$. Figure 1.6 shows $\varepsilon^*(\mu)$ for both commuting zones. The threshold $\varepsilon^*(\mu)$ is a decreasing function of μ since parents default only when debt repayment would leave them with too low level of consumption causing a sharp drop in utility. As μ increases, it takes increasingly worse income shock to prompt default.

Figure 1.6: Threshold $\varepsilon^*(\mu)$ in the MEDIAN and "+1SD" commuting zones



The model predicts that in equilibrium, parents always take out a loan to send the child to college as long as the bank offers the loan. Parent's benefit of child attending college manifested in higher level of expected child's utility from more consumption appears to outweigh the expected cost of reduced own consumption or default.

1.3.1.3 Measure of upward income mobility

Since the mechanism that links higher income inequality to lower income mobility operates at the margin in the bottom half of the parent income distribution, intergenerational income elasticity may not properly capture the effect of income inequality on income mobility and a measure of upward income mobility should be used instead. I use a measure similar to Chetty et al. 2014 absolute upward mobility. Specifically, using the benchmark parameter values, I first simulate a large array of parent's ability types, income shocks (thus obtaining ten-year income), and decisions as to whether send their child to college and whether to default on a loan. I regard this large array of parents as the U.S. population of parents in 2000. For each parent, I simulate the income shock for his child and using the simulated college attendance indicator, I compute income for each child. The resulting distribution of child income is regarded as the U.S. distribution of child income in 2010. Next, I simulate two much smaller groups of parents using the benchmark parameter values for the first group and those higher values of σ_μ and σ_ε that produce one standard deviation higher level of parent inequality. I treat these two groups as commuting zones with the median level of parent income inequality and one standard deviation above the median, respectively (the latter CZ will be referred to as "+1SD"). For each parent income in each of the CZs, I compute the percentile income rank in the simulated nationwide parent income distribution. Each parent's percentile income rank is a number between 0 and 1 and it shows the fraction of parents in the nationwide distribution with smaller income. I also simulate these parents' child income and assign to each child his or her percentile income rank in the nationwide distribution of child income. Then I define upward mobility in a given CZ as the mean child income rank among children in the given CZ whose parents income rank is between 0 and 0.5, i.e. parents in the bottom half of the nationwide parent income distribution.

1.3.1.4 Predicted Changes in Upward Income Mobility and College Attendance Rate

Table 1.2 below is the main focus of the paper.

Table 1.2: Upward mobility for MEDIAN and "+1SD" values of σ_μ and σ_ε

	σ_ε MEDIAN CZ	σ_ε "+1SD" CZ
σ_μ MEDIAN CZ	0.4221	0.4178
σ_μ "+1SD" CZ	0.4080	0.4062

The elements on the main diagonal are the model upward income mobility measures in the median CZ and the "+1SD" CZ. The commuting zone with the Gini of parent income one standard deviation higher than the median has higher variances of both μ and ε and lower upward income mobility. I decompose the transition from the median CZ to the "+1SD" CZ into separate increases in σ_μ and σ_ε. The results of this decomposition are shown in the off-diagonal cells of the table. The increase in σ_μ alone has about the same effect on upward mobility as the combined increase in σ_μ and σ_ε.

An increase in σ_ε alone has two effects on upward mobility. The first effect is due to the credit supply: fewer parents are offered credit to finance their children college degree and children of those marginal parents who can no longer obtain a loan will see their income rank fall significantly due to the loss of the college wage premium. This effect per se lowers upward mobility. The second effect is purely mechanical: with higher variance of the income shock, children of all parents in the bottom half of the distribution have a disproportionally higher chance of getting their income rank increased than decreased since their income rank is relatively low by construction and is much closer to the zero lower bound than the upper bound of one. This effect per se improves upward mobility. The net effect of an increase in σ_ε works towards reducing upward mobility (from 0.4221 to 0.4178).

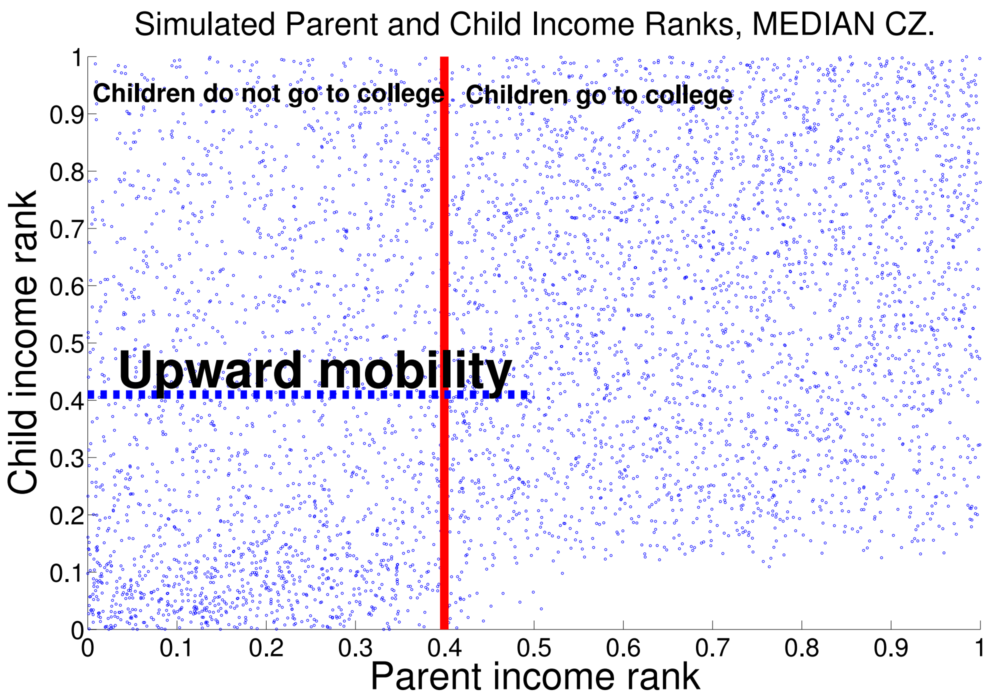
An increase in σ_μ alone has an unambiguously adverse effect on upward income mobility. The cut-off level of parent ability that separates parents who finance their child college degree and those who don't does not depend on σ_μ. However, an increase in σ_μ reallocates some of the mass of the parents in the bottom half of the distribution who send their child to college across the cut-off to the no-college zone, which results in a strictly lower average child income rank. This is evident from Table 1.3 showing child college attendance rate, i.e. the fraction of parents in a CZ who send their kids to college.

Table 1.3: College attendance rate for MEDIAN and "+1SD" values of σ_μ and σ_ε

	σ_ε MEDIAN CZ	σ_ε “+1SD” CZ
σ_μ MEDIAN CZ	0.6068	0.5836
σ_μ “+1SD” CZ	0.5882	0.5686

Higher σ_μ or higher σ_ε unilaterally have a negative effect on upward mobility. The total effect on upward mobility is more pronounced yet.

Figure 1.7: Simulated parent and child income ranks in the MEDIAN CZ.



Note: Each dot corresponds to a simulated pair of parent and child income ranks for the median CZ. Parent income rank is defined as the fraction of parents with a lower income in the simulated nationwide distribution of parent income. Child income rank is defined as the fraction of children with a lower income in the simulated nationwide distribution of child income. The horizontal line is the level of upward mobility, defined as the mean of the income ranks of children born to parents with income rank less than 0.5. The vertical line separates all parents into two groups: parents with rank less than 0.3932 don't send their child to college and parents with rank greater than 0.3932 do.

Figure 1.7 plots simulated parent and child income ranks (in the simulated distributions of parent and child income, respectively) for the median CZ. The horizontal line is the level of upward mobility, defined as the mean of the income ranks of children born to parents with income rank less than 0.5. The vertical line separates all parents into two groups: parents with rank less than 0.3932 (1 - 0.6068) don't send their child to college and parents with rank greater than 0.3932 do. An increase in σ_μ alone does not change the threshold value for parent's μ ($42,500) that separates parents who send their child to college from those who do not. However, the higher variance of μ distributes more parents below the threshold, thus reducing the fraction of parents who get a loan and send their child to college. An increase in σ_ε alone, on the one hand, increases the threshold value for parent's μ (to $44,300) leaving more parents below the threshold thus reducing mobility, but on the other hand, results in a higher weight of the luck component in the determination of the child income and rank, thus increasing mobility. The net effect of the unilateral increase in σ_ε on mobility is negative.

1.3.1.5 Goodness of Fit of the Model

It is now possible to compare the model prediction to the data. Specifically, the difference in Gini of parent income between the median and high inequality commuting zones is equal to one standard deviation of parent Gini, which is equal to 0.081. The predicted difference in upward mobility is the difference between the diagonal cells in Table 1.2, 0.4062 – 0.4221 = -0.0159. The implied slope of the relationship between parent income Gini (on the horizontal axis) and upward mobility (on the vertical axis) is equal to -0.0159/0.081 = -0.1963. In the data, this slope is -0.415. Therefore, the model can explain -0.1963/(-0.415) = 47.3% of the relationship between Gini of parent income and intergenerational income mobility.

The model predicts the college attendance rate to be 0.6068 – 0.5686 = 0.0382 lower in the CZ with parent income inequality one standard deviation higher than the median. The implied slope of the relationship between parent income Gini (on the horizontal axis) and college attendance rate (on the vertical axis) is equal to -0.0382/0.081 = -0.472, while this slope is only -0.362 in the data. So my mechanism explains well the variation in the college attendance rates and explains about half of the relationship between parent income inequality and upward income mobility across US commuting zones.

1.4 CONCLUSION

This paper proposes a causal mechanism that could potentially explain negative correlation between income inequality and intergenerational income mobility across US commuting zones observed in the data. The mechanism is also consistent with the fact that in the cross section of commuting zones, college attendance rate for the 1980-82 birth cohort is negatively correlated with the Gini coefficient of their parents' income in 1996-2000. In my model, altruistic parents need to borrow, if they wish to invest in children's human capital accumulation and they have an option to default on the loan. In equilibrium, parents execute this option if the value of the stochastic transitory component of income, which realizes after the loan is made, is sufficiently low. Along with observed permanent component of parent income, the level of local parent income inequality and the source of its regional variation both play a role in the determination of the risk of default. If regional variation in parent income inequality is mostly due to the variation in the transitory component of income, then a low-income parent in a high-inequality commuting zone is perceived by the banks as a higher risk than a parent with the same income in a low-inequality commuting zone, other things equal. As a result, for low-income parents in high-inequality regions, banks price loans higher and deny a higher fraction of loan applications than in low-inequality regions, which explains the cross-sectional negative correlation between inequality in parent income and both children's college attendance rate and upward income mobility. If, on the other hand, regional variation in parent income inequality is mostly due to the variation in the permanent component of income, college attendance rate will also be lower in higher inequality commuting zones, but mostly because there will be more parents on the low end of income distribution who cannot afford a loan, and less so because of the variation in the loan interest rates and availability of credit.

Chapter 2: Classification Models: Predicting Moves with USAA Customer Data

2.1 INTRODUCTION

Companies try to predict customers' behavior to improve business processes and maximize profits. Online retailers, for example, recommend products based on search and purchase history[3]. Netflix recommends you movies based on what you have watched and how you liked it. Banks build models to predict default or prepayment on a mortgage.

In this paper, I use USAA customer data to build a predictive model of moving. There are two reasons why a financial services company may be interested in predicting when customers are moving. First, each move is stressful and the company could take some of that stress off by automatically updating customer's information. And second, people often shop for a car or home insurance when they move, so knowing when a customer is going to move could lower attrition rate by means of proactive marketing efforts.

I used logistic regression, LASSO (L^1-regularized logistic regression), and Random Forest to analyze USAA data and build the most accurate prediction model. All the three classification models yield similar performance, with LASSO providing the smallest misclassification error on the held-out test set. That LASSO model uses 143 features.

[3] For an extreme example, see the story of how Target figured out a teen girl was pregnant before her father did: http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html

I also explored the tradeoff between model accuracy and model simplicity. I found that a parsimonious logistic regression model with only five features performs only slightly worse than the accuracy-maximizing LASSO with 143 features.

Among the features that I found to be the most important in explaining moves, age has by far the greatest explanatory power, followed by features associated with home ownership and military status. All of these findings are intuitive: people tend to move more when they are younger, when they rent, and when they are in the military.

Using additional data, I have found that neither local (MSA-level) unemployment rate nor MSA population has a statistically significant effect on the likelihood of moving in the logistic regression model that includes all original data. MSA dummies, except one, are not statistically significant either, which points to no effect of location-specific factors on the likelihood of moving.

2.2 DATA

The data were provided at the DataHack UT predictive modeling competition held on the campus of the University of Texas in November 2014. Two data sets were provided. The first set, which I will be referring to as set A, contained a snapshot of records of 50,000 USAA customers as of September 30, 2013 and the dependent variable, the indicator of moving during the month of October of 2013. The other data set, call it set B, was another sample of 50,000 USAA customers as of September 30, 2013 in which the outcome (moving) was not known to the participants. The goal was to predict the outcome in set B and the metric used in the competition to determine the winner was the percentage of correctly predicted outcomes, or equivalently, the percentage of incorrectly classified observations, referred to as the misclassification error.

Each data set had an equal number of customers who moved during the reference period and those who did not move. 600 variables were provided for each customer that can be split into two major categories, personal and USAA-related information. Notable examples of personal information were age, education level, number of children, military status, homeowner/renter status, residence state and zip code. Among USAA-related variables were the number of USAA products the customer had, account balances, activity on those accounts, as well as 3, 6, and 12 month lagged values of some of these and other variables.

Since the outcome in set B is unknown, in this paper I only analyze set A. To prepare the data from set A for analysis, a few data cleaning procedures were in order. After deleting variables with missing rate exceeding 50% or with a single value for all observations, there were 152 variables left. The remaining missing values were imputed as the mean value for numerical variables and the mode value for categorical variables. Some categorical variables had too many levels, with some levels containing only a few observations. In that case, I grouped levels with too few observations into a new category 'other'. I created a dummy variable for each level (or a group of levels if grouped) for each categorical variable. After this data cleaning, I had 462 independent numerical variables; some of them were original numerical variables and others were numerical variables derived as dummies off of levels of categorical variables. To differentiate these 462 variables from the original set of variables, I will refer to the former as *features*, as is common in statistics and machine learning literature.

Next, I split set A at random into two disjoint subsets. The first set, containing 80% of observations, I used for training and cross-validating different models and the remaining 20% observations I used for model testing. The goal in the competition was to minimize the misclassification error on set B, but since the outcome in set B is unknown, I will approximate this by minimizing the misclassification rate on the testing subset of set A. Another question I posed in this analysis is to build a simple parsimonious model that would utilize a few features yet have high predictive power.

Toward these goals, I analyzed three classification methods that are commonly used in predictive modeling: logistic regression, LASSO regularized logistic regression, and Random Forest. An important benefit of LASSO regression and Random Forest that drove my choice of models is that they have convenient "built-in" feature selection facilities that are helpful in achieving the second goal of building a highly predictive model with a small set of features. At the same time, logistic regression is probably the most common classification model used in a variety of fields, so I included logistic regression as a benchmark model.

2.3 OVERVIEW OF CLASSIFICATION MODELS USED

In this section, I will describe each of the three classification methods I used to predict moves. Each of the three models were estimated on the training sample containing

N=40,000 observations. Let y_i denote the outcome for observation *i*: $y_i = 1$, if person i moved and $y_i = 0$ otherwise. For each *i*, let *x* be a *p*x1 vector of features, where *p*=462.

2.3.1 Logistic Regression

The conditional probability of moving given a vector of features *x*, $P\{y = 1|x\}$, is modeled as the logistic function of the linear combination of the components of vector *x*:

$$P\{y = 1|x\} = \frac{e^{\beta_0 + x^T\beta}}{1 + e^{\beta_0 + x^T\beta}}$$

where β is a *p*x1 vector of parameters and β_0 is a scalar parameter.

The log-likelihood function to be maximized over (β_0, β^T) is

$$\sum_{i=1}^{N} \left[y_i(\beta_0 + x_i^T\beta) - \log\left(1 + e^{(\beta_0 + x_i^T\beta)}\right)\right].$$

2.3.2 Regularized Logistic Regression and LASSO as the Special Case

Regularized logistic regression differs from the simple logistic regression in that it adds a penalty term to the log-likelihood function:

$$\sum_{i=1}^{N} \left[y_i(\beta_0 + x_i^T\beta) - \log\left(1 + e^{(\beta_0 + x_i^T\beta)}\right)\right] - \lambda[(1-\alpha)\|\beta\|_2^2 + \alpha\|\beta\|_1],$$

where $\lambda > 0$ is the penalty parameter and $0 \le \alpha \le 1$ is the relative weight of L^1-norm penalty versus L^2-norm penalty.

The case $\alpha = 0$, when the penalty is entirely L^2-norm, is called the Ridge regression, whereas the case $\alpha = 1$, when the penalty is entirely L^1-norm, is called the

LASSO Regression. When the penalty parameter λ increases, both Ridge and LASSO regressions tend to "shrink" beta coefficients, i.e. the components of the parameter vector β tend to decrease in absolute value. In addition, in the LASSO regression (but not in the Ridge regression), the maximum likelihood solution will assign zero values to some betas, and the number of zeroed out betas tends to increase as the penalty parameter λ increases.

The best value of λ is typically found via k-fold cross-validation (CV). The entire training set T gets split into k disjoint sets of equal size, denote them $CV_1,\ldots,CV_k$. For each value of λ on a grid and for each fold i=1,…,k, the model is estimated k times, each time holding out a different cross-validation set CV_i (i.e. for each fold i=1,…,k, the model is estimated on $T\backslash CV_i$) and at each fold i, for all observations in CV_i, moving predictions are obtained and the misclassification error is computed. The mean of those k misclassification errors is the cross-validation error for a particular value of λ.

2.3.3 Random Forest

Random Forest is an ensemble learning method for classification and regression. In case of classification, it operates by constructing ("growing") a multitude of independent decision trees at training time and outputting the class that is the mode of the individual trees' class predictions. Given the original training set T of size N, each tree is grown as follows:

1. Form a bootstrap sample by sampling N cases at random, with replacement, from T. This sample will be the training set for growing the tree.

2. Let *M* be the number of input features, then a number *mtry* << *M* is specified such that at each node, *mtry* variables are selected at random out of the *M* and the best split on these *mtry* features is used to split the node. The value of *mtry* is held constant during the forest growing.

3. Each tree is grown to the largest extent possible.

The accuracy metric for the Random Forest algorithm is the so-called out-of-bag (OOB) error. For a given original training dataset *T*, a given value of *mtry*, and a given number of trees, the out-of-bag error is computed as follows. Each observation *i* in *T* is put down each tree for which observation *i* is not in the bootstrap sample for that tree (i.e. "out of bag"). Each such tree "votes" on observation *i*'s class and the class that gets most of the votes is the predicted observation *i*'s class. The proportion of times that the predicted class is not equal to the true class averaged over all observations in the training set is the OOB error. In Breiman 1996 and Breiman 2001, the author showed that the OOB error is an unbiased estimate of the classification error and therefore, there is no need for cross-validation or a separate test set to get an unbiased estimate of the test set error.

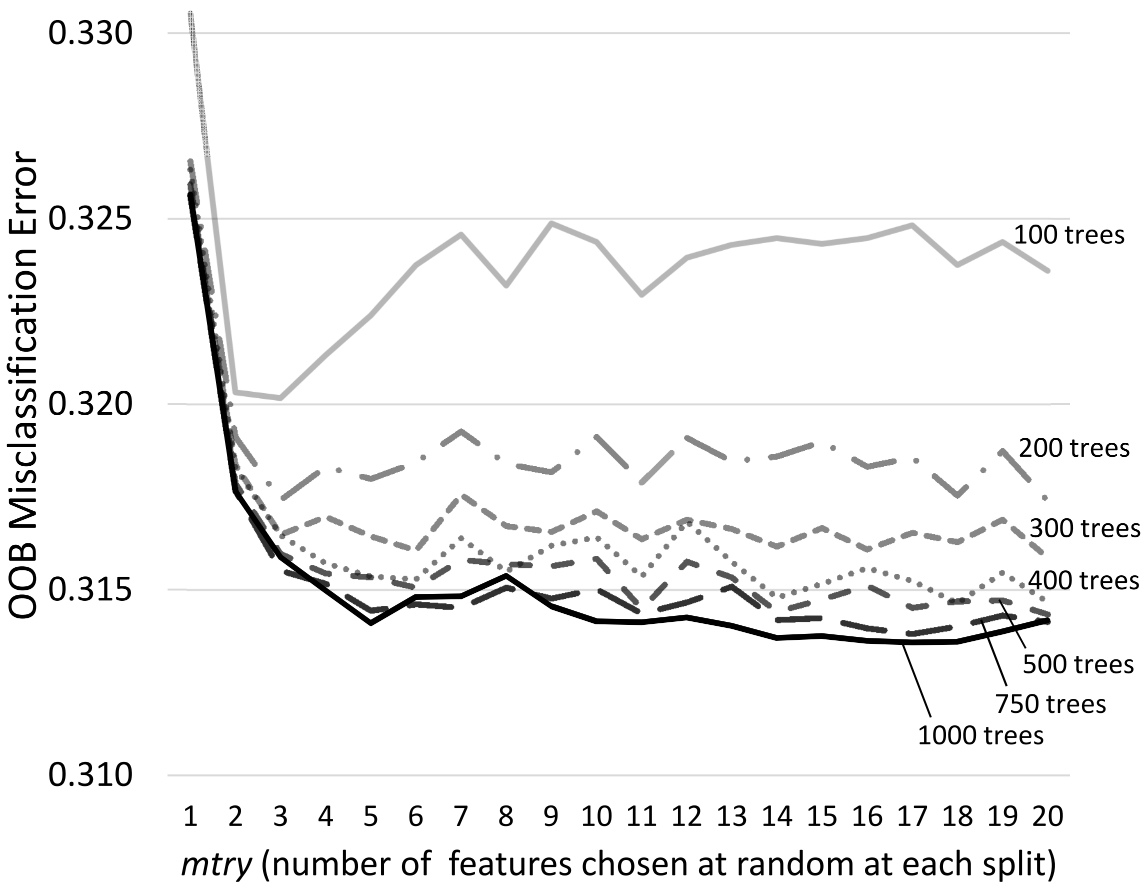
2.4 RESULTS

2.4.1 Random Forest

Figure 2.1 plots the out-of-bag error of the Random Forest algorithm as a function of parameter *mtry*, which is the number of features chosen at random at each decision node of the tree. The seven lines correspond to different numbers of trees ranging from 100 to

1000 trees. For each number of trees and *mtry*, the out-of-bag (OOB) error was computed as the average OOB error over five independent runs.

Figure 2.1: Random Forest out-of-bag misclassification error



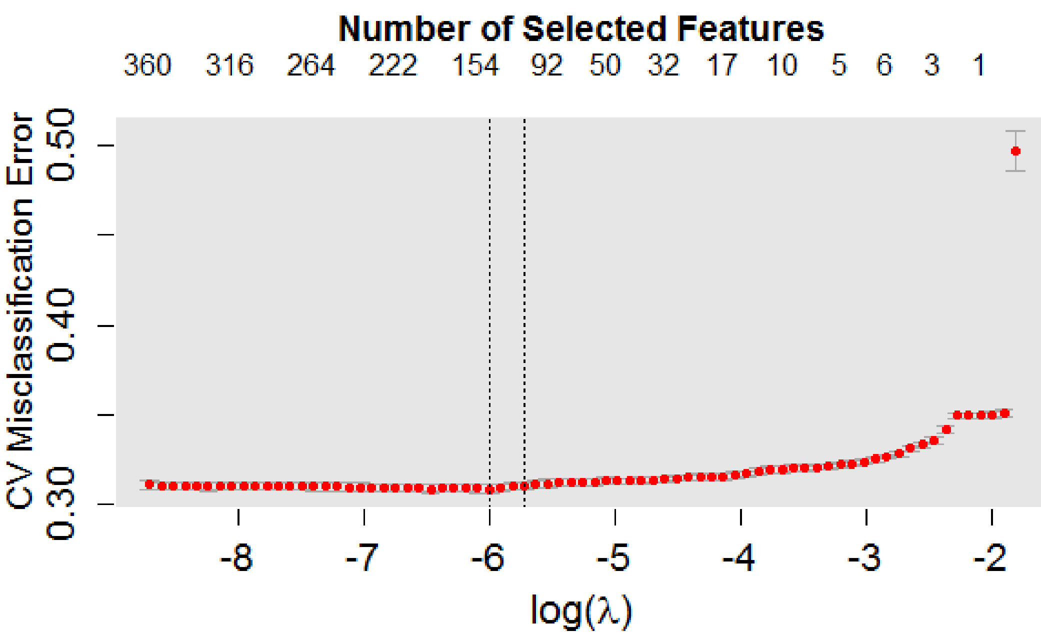
The performance of Random Forest improves with the number of trees, although at a diminishing rate. As *mtry* increases, the OOB error initially falls but quickly levels off. Random Forest shows the best performance when it uses 1000 trees and 17 candidate features at each decision node. The OOB misclassification error is then minimized at 0.3136 and the error on the test set is 0.3068.

2.4.2 LASSO

Figure 2.2 plots out-of-sample misclassification error for the LASSO regression against the logarithm of λ, the weight of the penalty portion of the objective function.

For each value of λ, some of the features are assigned a zero coefficient because their marginal contribution to maximizing the objective function is smaller than the marginal penalty associated with keeping them. The number of remaining features that survive penalization is shown on the upper axis.

Figure 2.2: Lasso CV misclassification error and the number of selected features



On the right side of the plot, for larger values of λ, more weight is placed on the penalty portion of the objective function and fewer features get selected. In the extreme case, a large enough value of λ leaves no features at all, resulting in the misclassification error close to 0.5 (the dot in the upper right corner). The misclassification error decreases as λ decreases, because more features are added to the regression, which tends to reduce the error.

Conversely, on the left side of the plot, penalty has a small weight, so hundreds of features get selected. Remaining features have very little additional explanatory power, so including them by decreasing λ does not improve model fit (and even somewhat worsens it), which explains the slightly upward-bending left tail. The best model fit in terms of minimizing cross-validation misclassification error is achieved with 143 selected features (marked with the left vertical dashed line), and the CV error is 0.3087, which is slightly better than the Random Forest OOB error.

2.4.3 Model Comparison

Table 2.1 compares the results of three models: logistic regression with all the features included, LASSO, and Random Forest. For LASSO, I show the results of the best model specification with 143 selected features. LASSO yields the smallest misclassification error on the held-out test set (testing error), closely followed by Random Forest and the unregularized logistic regression. In Table 2.1, I also report other errors for the three models, although these are not all directly comparable. The OOB error only applies to Random Forest. The unregularized logistic regression does no cross-validation because there no parameters to tune. LASSO has the penalty parameter λ, which is tuned by means of 10-fold cross-validation. Mean (over 10 folds) training, CV, and testing errors for the optimal λ are reported in Table 2.1.

Table 2.1: Model comparison

	Model		
Error Type	Logistic	LASSO	Random Forest
Training	0.3026	0.3041	N/A
CV	N/A	0.3087	N/A
OOB	N/A	N/A	0.3136
Testing	0.3070	0.3053	0.3068

Since LASSO with the value of λ tuned to minimize CV error also yields the smallest misclassification error on the test set, this model would be the model of choice for the best prediction of the moves. However, from the business perspective, it may not be practical to use a model with hundreds of explanatory variables or features. In the next section, I will build a parsimonious model of moves and analyze the trade-off between model simplicity and accuracy.

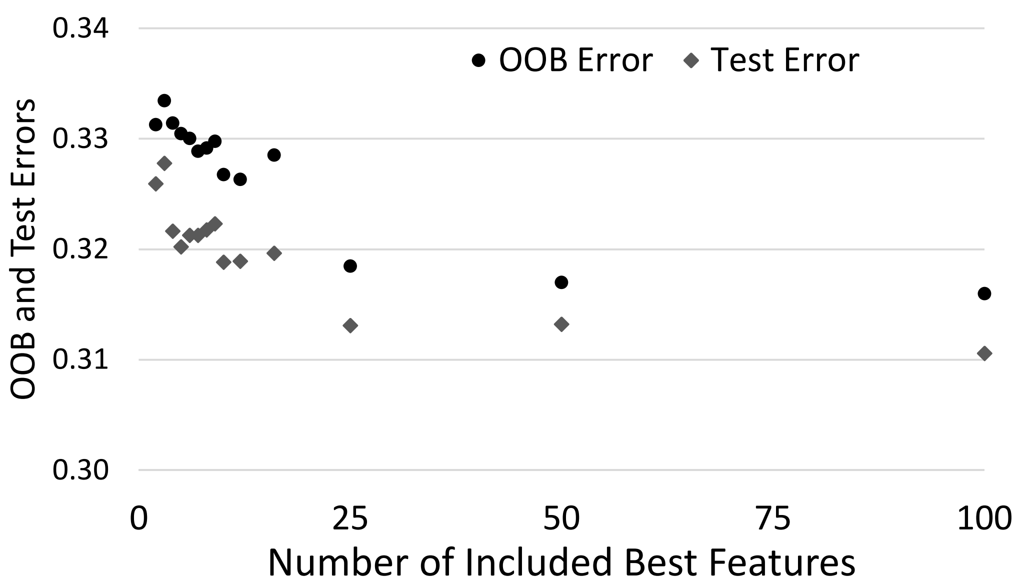
2.4.4 Building a Parsimonious Model

To build a simple yet highly predictive model, I used LASSO and Random Forest for variable selection. As seen in Figure 2.2, it is possible to include almost any desired number of the most important features by adjusting the LASSO penalty parameter λ. On the other hand, Random Forest algorithm has a "built-in" function of ranking features by their relative importance[4] which I used to iteratively reduce the number of features.

[4] This algorithm works as follows. In every tree grown in the forest, put down the OOB observations and count the number of votes cast for the correct class. Randomly permute the values of feature m in the OOB pool and put these observations down the tree. Subtract the number of votes for the correct class in the feature-m-permuted OOB data from the number of votes for the correct class in the untouched OOB data. The average of this number over all trees in the forest is the raw importance score for feature m.

Specifically, I first ran Random Forest using the full set of features and LASSO with λ set to zero to include all the features. Then I kept 100 most important features per Random Forest ranking and also adjusted λ in LASSO so it selected most important 100 features. I further halved the number of selected features to 50, then to 25 and continued reducing the number of selected features, eventually bringing the number of selected features down to two (I tried 100, 50, 25, 16, 12, 10, 9, 8, 7, 6, 5, 4, 3, and 2 features). At each iteration, I ran both Random Forest and LASSO 5 times and recorded mean out-of-bag and testing errors.

Figure 2.3: Random Forest based iterative feature selection



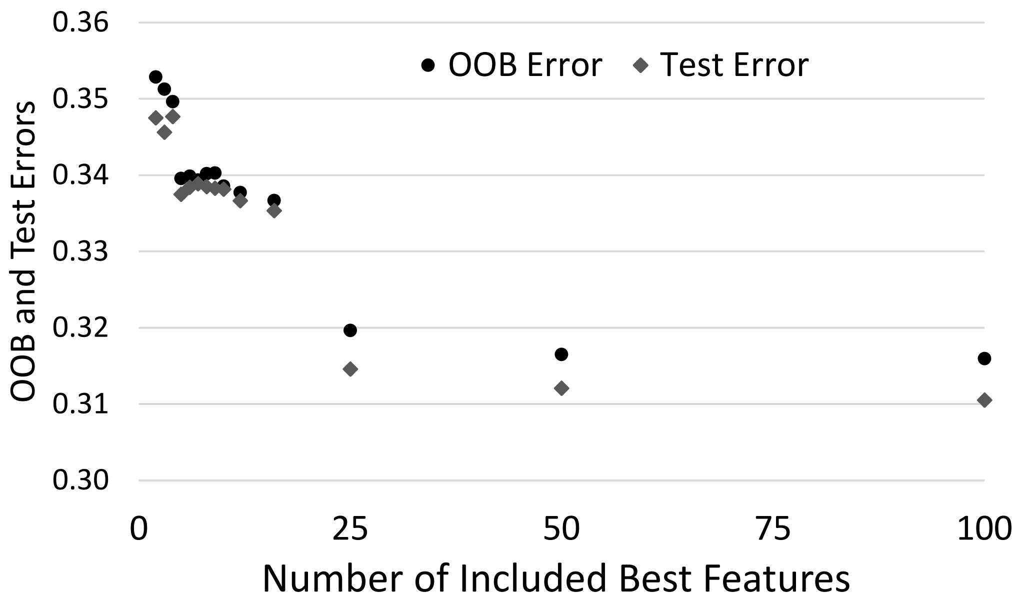
Note: Starting from the most important 100 features of the Random Forest with all 462 features and *mtry* = 17, at every iteration, a new Random Forest is run, features get ranked by importance and a smaller subset of the most important features is selected for the next iteration.

Results for Random Forest are shown in **Error! Reference source not found.**Figure 2.3. As the number of features is reduced (which corresponds to moving from right to left along the horizontal axis), both the OOB and the testing errors generally increase. Starting from the best selected 100 features, iterative elimination of the least important 50 and then another 25 features leaves 25 most important features that yield only slightly higher OOB and testing errors than the original 100 best features. This indicates that out of hundreds of available features only a handful contributes to predicting moves and also confirms the necessity of feature selection for practical business purposes.

An interesting question for Random Forest based feature selection is whether it is important to reduce the number of features iteratively, which can be viewed as an additional complication relative to a simple one-step extraction of a few (say, five) features ranked highest in terms of importance by the Random Forest run on the full set of features.

To evaluate the prediction accuracy gain of the iterative feature reduction procedure versus the non-iterative one, I ran Random Forests with the same number of features (100, 50, 25, 16, 12, 10, 9, 8, 7, 6, 5, 4, 3, and 2) using the non-iterative feature selection procedure. Specifically, keeping the list of all 462 features sorted by importance by the Random Forest run on all of them fixed, I ran Random Forests with the best 100, 50, 25, 16, …, 2 features picked from that list. Again, for each number of selected features, I did five runs and computed the average OOB and test errors. The results are shown in Figure 2.4**Error! Reference source not found.**. The comparison of Figure 2.3 and Figure 2.4 shows very similar performance of the iterative and the non-iterative procedures for 100, 50, and 25 features. However, for 16 and fewer features, the iterative procedure outperforms the non-iterative one by about 0.01-0.02. This confirms the importance of using the iterative procedure to select a handful of the most predictive features, if the goal is to build a parsimonious model that would include only a few most predictive features.

Figure 2.4: Random Forest based non-iterative feature selection



Note: A pair of a black and grey dots represents averaged over 5 runs OOB and test errors for a specific numbers of included features. These features are obtained as the most important features of the Random Forest with all 462 features and the optimal value of the *mtry* parameter, 17.

For LASSO, the results are similar. As shown in Figure 2.2, prediction accuracy gradually declines as more features get dropped from the model, and similarly to Random Forest, LASSO with five most important features yields misclassification error that is only about 0.01 larger than the best LASSO specification with 143 features. Beyond these five most important features, additional improvements in model performance come at a high cost of including a lot more features.

Table 2.2: Five most predictive features selected by LASSO and Random Forest models

LASSO		Random Forest	
Feature	Description	Feature	Description
AGE	Numeric variable AGE is CUSTOMER CURRENT AGE	AGE	CUSTOMER CURRENT AGE
RENTACT=2 dummy	Categorical variable RENTACT is “RENT ACTIVITY STATUS”	RENTACT=2 dummy	Categorical variable RENTACT is “RENT ACTIVITY STATUS”
MILST=1 dummy	Categorical variable MILST is “MILITARY STATUS OF CUSTOMER”	RENTER=1 dummy	Categorical variable RENTER is “RENTERS POLICIES IN FORCE”
HSNGCLAS =2 dummy	Categorical variable HSNGCLAS is “HOUSING CLASSIFICATION”	CNTCORP	Numeric variable CNTC ORP is the total count of USAA products
HOME=1 dummy	Categorical variable HOME is “HOMEOWNERS ONLY”	CNTCORP_6 MB	This is the value of CNT CORP 6 months back

In Table 2.2, the sets of five most important features for LASSO and Random Forest are not exactly the same but they both include age and rent activity status. LASSO also has features associated with home ownership, housing, and military status. Table 2.3 reports results of a logistic regression of moving on these five features[5]. Other things being equal, younger people, military personnel, and renters are more likely to move. Figure 2.5 shows the fraction of movers at every age in the training set. The fraction of movers increases with age over the range of 18 to 22 years old, starting at about 0.57 and reaching the maximum of 0.73 at age 22. From 23 to about 80 years old, the relationship between age and the likelihood of moving is almost monotonically decreasing (due to very thin data after about 80 years old, the fraction of movers shows erratic behavior but stays below 50%). Early in their career, people tend to move often and are more likely to stay at the

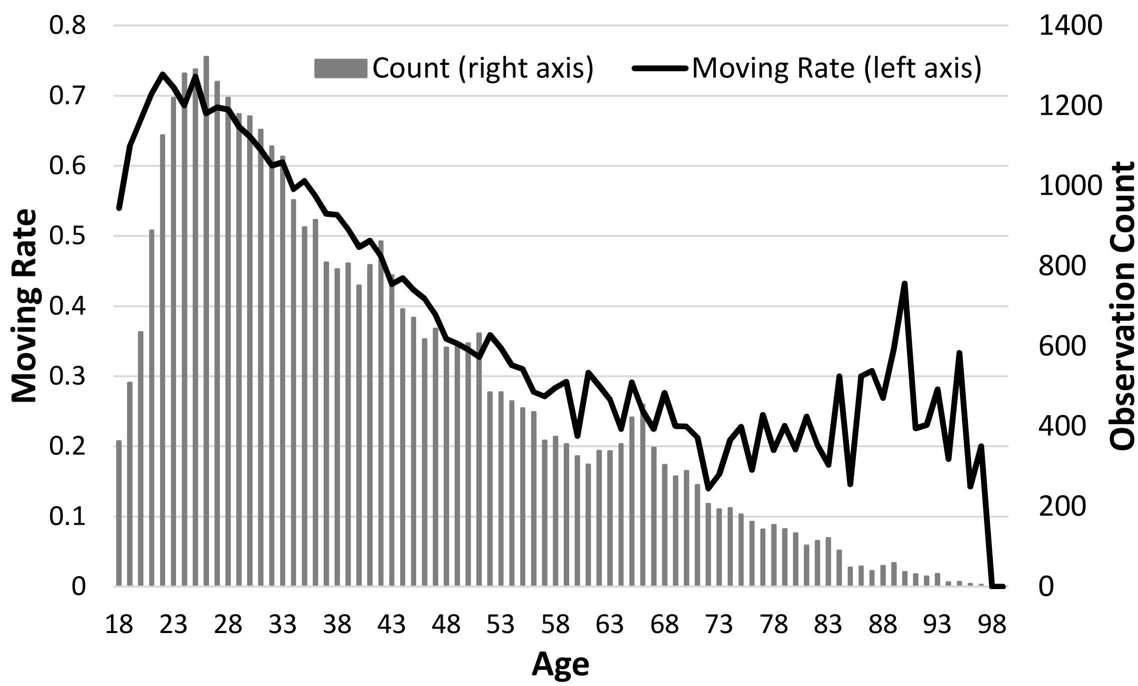
[5] I report the results of the logistic regression rather than LASSO here because LASSO shrinks coefficients, making difficult both the interpretation of the coefficients and inference.

same residence longer as they get older and have children. Age alone yields an out-of-sample misclassification error of about 0.35 in LASSO, which can be seen in Figure 2.2.

Table 2.3: Results of logistic regression estimation with five best features selected by LASSO

	Estimate	Std. Error	t-statistic
Intercept	1.0239	0.0372	27.51
AGE	-0.0325	0.0008	-42.42
HOME=1 dummy	-0.4083	0.0314	-12.99
MILST=1 dummy	0.5360	0.0296	18.12
RENTACT=2 dummy	0.5944	0.0298	19.94
HSNGCLAS=2 dummy	0.3128	0.0263	11.87
Obs. in training set:	40,000	*Training error:*	0.3200
Obs. in testing set:	10,000	*Testing error:*	0.3154

Figure 2.5: Age is the single most important feature



2.4.5 Additional Features

So far, I have established that age, military status, and housing status are the most predictive factors. One can argue that geographical location of residence may play an important role for the probability of moving. In particular, the size of the city and the local unemployment rate may affect the propensity to move. For instance, other things equal, a local job change is more likely to result in a longer commute that necessitates moving in a big city than in a small one. Local unemployment rate may affect the propensity to move in a variety of ways. For instance, other things equal, a city with higher local unemployment rate will likely have higher rate of push migration (moving to a different city) but also lower rate of intra-city migration.

To study possible effects of local factors on moving propensity, I matched zip codes available in the data to metropolitan statistical areas (MSA) and created dummy variables for the biggest (by observation count in the training dataset) 28 MSAs, grouping all other MSAs into the 'other' category. I further obtained data on population and unemployment rate at MSA level in 2013. I then ran logistic regression, LASSO, and Random Forest with all these new features on top of the original 462 features. Surprisingly, neither MSA-level unemployment rate nor population showed significant in the logistic regression. Furthermore, out of 29 MSA dummies, only one was significant at 1% level and all the others had p-values exceeding 0.1. None of the three models saw any improvements in prediction accuracy.

2.5 CONCLUSION

In this paper, I used logistic regression, LASSO, and Random Forest to predict residential moves. All the three classification models yield similar performance, with LASSO providing the smallest misclassification error on the held-out test set. The LASSO model uses 143 features.

A parsimonious LASSO model with only five features performs only slightly worse than the LASSO with 143 features tuned to minimize the cross-validation error. In a business environment where model complexity is costly, a parsimonious model with a few features may be preferred over a complex model with more than a hundred features because of relatively small differences in model performance.

Among the features that I found to be the most important in explaining moves, age has by far the greatest explanatory power, followed by features associated with home ownership and military status. All of these findings are intuitive: people tend to move more when they are younger, when they rent, and when they are in the military.

Using logistic regression, I have found that neither local (MSA-level) unemployment rate nor MSA population has a statistically significant effect on the likelihood of moving. MSA dummies, except one, are not statistically significant either, which points to no effect of location-specific factors on the likelihood of moving.

Chapter 3: Is Universal Default Socially Desirable?

3.1 INTRODUCTION

The Credit Card Accountability, Responsibility and Disclosure Act (or Credit CARD Act) of 2009, that went into effect in February 2010, essentially outlaws universal default in the United States.[6] Before that, a credit card issuer had the right to raise the interest rate on an existing credit card balance to the default rate should the borrower default with another creditor; this practice is known as universal default. According to the Consumer Action 2008 Credit Card Survey, 77% of surveyed credit card issuers (17 of 22) answered "yes" to the question "Can you increase my APR or change my terms 'any time for any reason'?" This includes all top ten issuers, which held an 87.55% market share of almost a trillion dollars in general purpose card outstandings in 2008.[7]

Supporters of universal default argue that lenders should be able to use all available information at all times to price their credit products based on the risk of default and thus avoid adverse selection. Prohibiting universal default creates similar economic incentives as unobserved borrower's characteristics. In both cases, lenders, being unable to discriminate between high and low risk types, will have to charge a flat interest rate, which will adversely shift the pool of borrowers. Prohibiting universal default reduces the cost to the borrower when he does not make a payment when the payment is due. Prior to the Credit CARD Act, default on one credit card would usually trigger default interest rates on all other credit cards. Since default interest rates are much higher than regular credit card

---

[6] H.R. 627: Credit Card Accountability Responsibility and Disclosure Act of 2009, Section 171.
[7] Source: Nilson Report, April 2009

rates, the cost of defaulting on a single credit card would be substantial if the borrower were to carry large balances on multiple cards.

On the other hand, universal default is an example of the "tragedy of the commons" negative externality, arising when each credit card issuer tries to maximize its collection from the "common pool" - a financially distressed borrower. Chapter 7 of the US bankruptcy law allows a debtor to file for bankruptcy and get all their unsecured debt discharged in exchange for non-collateralized assets above a certain exemption level that are allocated pro rata among all lenders. So once a holder of multiple credit cards gets into financial troubles, each credit card issuer faces the prospect of receiving a small, if any, fraction of the amount owed to them. As a result, each creditor has an incentive to become the first to collect since the benefits of collection accrue to this creditor alone. If the financially-troubled debtor continues to pay her debts, she will try to pay off the balances with the highest interest rate first. This creates an incentive for each lender to unilaterally raise the interest rate thus making the debt burden even harder to bear, which in turn increases the probability of bankruptcy.

In this paper, I make an attempt to quantify welfare implications of prohibiting universal default practice signed into law by the Credit CARD Act of 2009. How much does society gain or lose from the new law that prohibits universal default? Specifically, I build a model of unsecured consumer debt in which borrowers hold two credit cards and have an option to default on either or both cards. I calibrate the model under the assumption that creditors can practice universal default (to mimic the state of the credit card industry before the CARD Act took effect) and compute social welfare in equilibrium. Then I switch to the no universal default regime, recompute the equilibrium and social welfare and compare social welfare in the two regimes. I find that prohibiting universal default lowers social welfare by 1.78% of lifetime consumption. Prohibiting universal default makes it more attractive for liquidity-constrained borrowers to default on one card. Bankruptcy rate goes up causing lenders to raise interest rates. With higher interest rates, the opportunity cost of bankruptcy falls, further reducing borrowers' incentives to pay off debts. In the end, equilibrium interest rates are much higher when universal default is prohibited, which has a strong negative impact on borrowers' ability to smooth consumption and results in a substantial social welfare loss.

3.2 RELATED LITERATURE

Universal default and in particular quantitative analysis of welfare implications appears to be an underdeveloped area of research. More broadly however, the issue of the externality imposed by creditors seeking individual collection and related policies have been studied both from theoretical and empirical perspectives. The welfare effects of this externality and bankruptcy as a way to correct it were examined in a series of papers by Thomas H. Jackson, along with Douglas Baird and Robert Scott.[8] They found that bankruptcy can actually increase the welfare of creditors by eliminating wasteful creditor collections, and consumers receive part of the social gains through lower interest rates. Other remedies to combat the externality from competitive collections studied in the

[8] See, for example, Jackson 1985 and 1986, Baird and Jackson (2001) and Jackson and Scott (1989).

literature are credit counseling services in the US and bank pools in Germany, which coordinate and therefore decrease creditors' collection efforts.

This study is most closely related to the working paper by Ausubel, Baranov, and Dawsey 2008 that attempts to quantitatively compare universal default regime to what the authors refer to as the own default regime, in which universal default clause does not apply. They build a simple three-period model in which a consumer receives stochastic income in the second and the third periods but is willing to consume in all the three, so he has to borrow on credit cards in period one to smooth consumption. In their model the consumer can default in period 2 but must repay the debt with accrued interest in period 3 up to his income in period 3. The authors arbitrarily choose parameter values and simulate the economy under universal default and own default. Their preliminary results show that under universal default, the penalty (default) rate is higher and social welfare is slightly lower. They conclude that their "simulations appear to favor limitations on the practice of universal default". The crucial limitation of their approach, however, is that they use an overly simplistic model and no real data to study a quantitative policy question, so the policy implications of their study are questionable.

Unlike some studies that assume perfect competition among lenders and risk-based pricing, in which the price of a loan depends on its size (e.g. Livshits, MacGee and Tertilt 2007, Chatterjee, Corbae, Nakajima and Rios-Rull 2006), this paper focuses on strategic interaction between lenders in offering non-exclusive contracts. Several other papers study non-exclusive contracts. Parlour and Rajan 2001, for example, show that the competitive zero-profit outcome may not obtain in the limit, but in their paper, default is strategic and never occurs in equilibrium. In this paper, default is used as an insurance against bad realizations of income and may occur in equilibrium.

3.3 MODEL

3.3.1 Environment

3.3.1.1 Households

Time is discrete, infinite and starts from period 1. There is one non-storable consumption good which is also the unit of account. There is a unit measure of expected utility maximizing households each of whom receives in each period an idiosyncratic stochastic endowment (income) that follows a finite-state one-period Markov process with the transition matrix π. In each period, a household can save by buying one-period discounted non-contingent bonds with fixed return equal to the world risk-free interest rate or borrow by selling one-period discounted non-contingent bonds to at most two credit card companies (henceforth CCC, creditors, or lenders) A and B at regular prices q_A^R and q_B^R set by the creditors. There are two disjoint groups of creditors and each household must choose one creditor from each group (more on that in the subsection "Credit Card Companies"). Other than the regular bond prices, credit card contracts also include default bond prices q_A^D and q_B^D, and credit limit l (the same across all lenders within a group). A household starts a period in a state that includes the current endowment shock e, default indicators δ_A, δ_B, a pair of credit card debt amounts a and b that are due today, the amount of savings s the household is receiving today, and the bankruptcy indicator f. The condition $\delta_x = 0$ means that the household is in good standing with creditor $x \in \{A, B\}$; $\delta_x = 1$ means that the household is in default with creditor *x*. If the household defaulted on its debt with creditor *x* in the previous period, it can stay in default without filing for bankruptcy for at most one period, i.e. today the household either repays the debt with accrued interest to creditor *x* or else has to file for bankruptcy (in which case the bankruptcy indicator *f* will change its value from zero to one). When a household files for bankruptcy, it gets all its debts (to both creditors) discharged immediately, the bankruptcy indicator *f* switches to one and stays at one forever, and the household stays in financial autarky from then on (no borrowing or saving is allowed from the period of bankruptcy filing onward). In each period and each state where *f*=0 (no bankruptcy), a household receives an endowment *e* and savings *s*, and gets to decide whether to repay his credit card debts *a* and *b* (by choosing δ_A' and δ_B'), how much to borrow, if possible, on credit cards (by choosing a' and b') and save (by choosing s'). If saving, a household buys one-period non-contingent risk-free bond priced at q^{RF} that pays one unit of consumption good next period. At the end of each period, the household consumes. In a state where *f*=1 (bankruptcy), the household simply receives its endowment *e* and consumes it. Partial default, i.e. repayment of a fraction of credit card debt is not allowed. Defaulting on a loan from creditor *x* prohibits additional borrowing from *x* in the same period. Once a household files for bankruptcy, it continues to receive an endowment in every period thereafter but in the same period, a "newborn" household is added. The newborn household immediately draws an endowment shock and, starting off with zero credit card debt and no savings, gets to decide how much to borrow or save. The total population therefore may grow over time but the mass of households able to participate in financial markets is unity in every period. At the beginning of period zero, all households have zero credit card debts and no savings and draw an income shock from the invariant (ergodic) income distribution.

The model economy operates under one of the two legal regimes. The only difference between the two regimes is in the conditions when the default bond price applies. In one regime, called "NUD", when a borrower has not repaid the loan to creditor *x*, the amount on that loan rolls over at the default interest rate defined as $1/q_x^D$. This is the only condition when the default bond price appears under NUD regime. All new loans are made at the regular interest rate defined as the inverse of the regular bond price. In another regime, called "UD", the default interest rate is charged not only on all defaulted loans as in the NUD regime, but also applies to new loans when the borrower is in default with *the other* creditor.

Under universal default, the household's optimal default and borrowing/saving decisions of the household in each period can be characterized by (the solution to) the following Bellman equation

$$V^{UD}(e,\delta_A,\delta_B,a,b,s,f|\sigma)=\max_{\delta'_A,\delta'_B,a',b',s',c} u(c)+\beta\mathbb{E}_{e'|e}V^{UD}(e',\delta'_A,\delta'_B,a',b',s',f'|\sigma)$$

subject to

$$\delta'_A\in\begin{cases}\{0,1\} & if \quad a<0\ AND\ f=0\\ \{0\} & if \quad a\geq 0\ AND\ f=0,\\ \{1\} & if \quad f=1\end{cases}\quad \delta'_B\in\begin{cases}\{0,1\} & if \quad b<0\\ \{0\} & if \quad b\geq 0\\ \{1\} & if \quad f=1\end{cases}$$

$$a'\in\begin{cases}\left\{\frac{a}{q_A^D}\right\} & if \quad \delta'_A=1\\ [l,0] & if \quad \delta'_A=0\end{cases},\quad b'\in\begin{cases}\left\{\frac{b}{q_B^D}\right\} & if \quad \delta'_B=1\\ [l,0] & if \quad \delta'_B=0\end{cases}$$

$$q_A=\begin{cases}q_A^D & if \quad \delta'_A+\delta'_B>0\\ q_A^R & if \quad \delta'_A+\delta'_B=0\end{cases},\quad q_B=\begin{cases}q_B^D & if \quad \delta'_A+\delta'_B>0\\ q_B^R & if \quad \delta'_A+\delta'_B=0\end{cases}$$

$$f' = \begin{cases} 0 & if \quad \delta_A\delta'_A = 1\ AND\ \delta_B\delta'_B = 0 \\ 1 & otherwise \end{cases},$$

$$c = \begin{cases} e + s + (1-\delta'_A)(a - a'q_A) + (1-\delta'_B)(b - b'q_B) - q^{RF}s', c \geq 0 & if \quad f' = 0 \\ e & if \quad f' = 1 \end{cases}$$

where

$$u(c) = \frac{c^{1-\rho}}{1-\rho}$$

is the period utility function and $\sigma = (q_A^R, q_B^R, q_A^D, q_B^D)$ is the credit card terms from both creditors.

Under no universal default, the problem differs only in the bond prices:

$$V^{NUD}(e, \delta_A, \delta_B, a, b, s, f|\sigma) = \max_{\delta'_A, \delta'_B, a', b', s', c} u(c) + \beta \mathbb{E}_{e'|e} V^{NUD}(e', \delta'_A, \delta'_B, a', b', s', f'|\sigma)$$

subject to

$$\delta'_A \in \begin{cases} \{0,1\} & if \quad a < 0\ AND\ f = 0 \\ \{0\} & if \quad a \geq 0\ AND\ f = 0 \\ \{1\} & if \quad f = 1 \end{cases}, \quad \delta'_B \in \begin{cases} \{0,1\} & if \quad b < 0 \\ \{0\} & if \quad b \geq 0 \\ \{1\} & if \quad f = 1 \end{cases}$$

$$a' \in \begin{cases} \left\{\frac{a}{q_A^D}\right\} & if \quad \delta'_A = 1 \\ [l, 0] & if \quad \delta'_A = 0 \end{cases}, \quad b' \in \begin{cases} \left\{\frac{b}{q_B^D}\right\} & if \quad \delta'_B = 1 \\ [l, 0] & if \quad \delta'_B = 0 \end{cases}$$

$$q_A = \begin{cases} q_A^D & if \quad \delta'_A = 1 \\ q_A^R & if \quad \delta'_A = 0 \end{cases}, \quad q_B = \begin{cases} q_B^D & if \quad \delta'_B = 1 \\ q_B^R & if \quad \delta'_B = 0 \end{cases}$$

$$f' = \begin{cases} 0 & if \quad \delta_A\delta'_A = 1\ AND\ \delta_B\delta'_B = 0 \\ 1 & otherwise \end{cases},$$

$$c = \begin{cases} e + s + (1-\delta'_A)(a - a'q_A) + (1-\delta'_B)(b - b'q_B) - q^{RF}s', c \geq 0 & if \quad f' = 0 \\ e & if \quad f' = 1 \end{cases}$$

3.3.1.2 Credit Card Companies

There are two groups of credit card companies, which I will also refer to as creditors, lenders, or banks, interchangeably. With a slight abuse of notation, these two groups are called A and B, exactly the same as individual members of respective groups. In each group, there are infinitely many identical credit card companies that compete for borrowers in bond prices with other creditors within the group in a perfectly competitive fashion. Creditors never observe individual household income but know the possible realizations of income and also know that the distribution of income is time invariant. At the beginning of period 1, prior to households making their default and borrowing/saving decisions, both creditors choose their regular bond prices q_A^R and q_B^R, once and for all.

Creditor A chooses regular bond price q_A^R so as to maximize per-period profit in the invariant distribution of borrowers, p^A defined as

$$p^A = \int_0^1 (p_i^A) di,$$

where p_i^A is creditor A's profit from household $i \in [0,1]$ which is determined by its state and default and borrowing/saving decisions given the credit card terms. Specifically, under UD regime,

$$p_i^A = \begin{cases} -a_i + q_A^R a_i' & if \quad \delta_i^A = 0, \delta_i'^A = 0, \delta_i^B = 0 \\ -a_i + q_A^D a_i' & if \quad \delta_i^A = 0, \delta_i'^A = 0, \delta_i^B = 1 \\ -\frac{a_i}{q_A^D} + q_A^R a_i' & if \quad \delta_i^A = 1, \delta_i'^A = 0, \delta_i^B = 0 \\ -\frac{a_i}{q_A^D} + q_A^D a_i' & if \quad \delta_i^A = 1, \delta_i'^A = 0, \delta_i^B = 1 \\ 0 & if \quad \delta_i'^A = 1 \end{cases}.$$

Conditions in the above formula leave out several bankruptcy scenarios. A household declaring bankruptcy in some period *t* brings zero profit to creditors in that period, but it is assumed that the newborn added instead of the bankrupt is assigned the bankrupt's index *i*, so the above formula will be applied to the newborn, who, of course, is not going to file for bankruptcy in period *t*.

Similarly, under NUD regime,

$$p_i^A = \begin{cases} -a_i + q_A^R a_i' & if \quad \delta_i^A = 0, \delta_i'^A = 0 \\ -\frac{a_i}{q_A^D} + q_A^R a_i' & if \quad \delta_i^A = 1, \delta_i'^A = 0 \\ 0 & if \quad \delta_i'^A = 1 \end{cases}.$$

Note the difference between profits in the two regimes: whether household *i* has or has not made a repayment on her debt to creditor *B* is irrelevant for creditor *A*'s profit under NUD regime but it does matter under UD regime because it affects which bond price applies to new loans from creditor *A*. The maximization problem for creditor *B* is analogous to that for creditor *A*.

3.3.1.3 Equilibrium

Definition. An equilibrium for regime UD or NUD is a collection of bond prices $(q_A^R, q_B^R, q_A^D, q_B^D)$ and household's policy functions $\delta_A'(e, \delta_A, \delta_B, a, b, s, f|\sigma)$, $\delta_B'(e, \delta_A, \delta_B, a, b, s, f|\sigma)$, $a'(e, \delta_A, \delta_B, a, b, s, f|\sigma)$, $b'(e, \delta_A, \delta_B, a, b, s, f|\sigma)$, and $s'(e, \delta_A, \delta_B, a, b, s, f|\sigma)$ such that

1. policy functions $\delta_A'(e, \delta_A, \delta_B, a, b, s, f|\sigma)$, $\delta_B'(e, \delta_A, \delta_B, a, b, s, f|\sigma)$, $a'(e, \delta_A, \delta_B, a, b, s, f|\sigma)$, $b'(e, \delta_A, \delta_B, a, b, s, f|\sigma)$, and $s'(e, \delta_A, \delta_B, a, b, s, f|\sigma)$ solve the household's maximization problem in every period;

2. given policy functions $\delta_A'(\cdot)$, $\delta_B'(\cdot)$, $a'(\cdot)$, $b'(\cdot)$, and $s'(\cdot)$, creditor *A*'s bond price q_A^R maximizes its per-period profit in the invariant distribution of households μ given *B*'s bond price q_B^R and *B*'s bond price q_B^R maximizes *B*'s per-period profit in the invariant distribution of households μ given *A*'s bond price q_A^R;
3. no-entry condition: no other credit card company finds it profitable to enter the market and offer a bond price that will deliver positive profit, and
4. invariant distribution of households μ: for any state $(e, \delta_A, \delta_B, a, b, s, f)$, the mass (fraction) of households in that state is constant over time.

3.3.1.4 Computational Strategy

I start with UD regime, since universal default was a common practice in the credit card industry before the Credit CARD Act, and then repeat the process under NUD regime with some modifications. The computational algorithm that I use in Matlab to find an equilibrium consists of the following steps.

0. Create a fine enough grid for debt amounts *a* and *b* and savings *s*. Create a grid for q_A^R and q_B^R such that the respective interest rates (inverse of bond prices) are spaced 0.0025 apart.
1. Simulate income process for *N*=200,000 households and large enough number of time periods *T* (100 is usually enough).
2. Pick a pair of regular bond prices q_A^R and q_B^R on a grid.
3. Given q_A^R and q_B^R, solve the household's problem to find policy functions $\delta_A'(\cdot)$, $\delta_B'(\cdot)$, $a'(\cdot)$, $b'(\cdot)$, and $s'(\cdot)$.

4. Obtain the invariant distribution $\mu(e, \delta_A, \delta_B, a, b, s, f)$ as follows.

4.1. In period 1, draw income for each household from the ergodic distribution of income.

4.2. Set their debts, savings, and bankruptcy flag equal to zero.

4.3. Apply household's policy functions and the income transition matrix to find the state $(e', \delta_A', \delta_B', a', b', s', f')$ for each household in period 2.

4.4. Compute the difference between the distributions of households in periods 1 and 2 as the maximum difference between the masses of households in periods 1 and 2 across all states.

4.5. Keep simulating the economy from one period to the next as in step 4.3 until the difference between distributions in the two consecutive time periods computed as in step 4.4 falls below a preset tolerance level.

5. Compute one-period profits of credit card companies *A* and *B* in the invariant distribution μ.

6. If the profit of either creditor is negative, this cannot be an equilibrium. In this case, go back to step 2 and try a different pair of bond prices. If profits are positive, check for potential entry, i.e. given the candidate pair of bond prices (q_A^R, q_B^R), check if a new entrant from group *A* can offer a higher bond price $q_C^R > q_A^R$ and enjoy a positive profit or a new entrant from group *B* can offer a higher bond price $q_C^R > q_B^R$ and enjoy a positive profit. If either such entry attempt yields a positive profit it means (q_A^R, q_B^R) is not an equilibrium pair of bond prices, so one needs to go back

to step 2 and try a different pair of bond prices. If, on the other hand, such entry attempt is unprofitable, the pair of bond prices (q_A^R, q_B^R) is a part of equilibrium.

7. When an equilibrium pair of bond prices is found, starting with the invariant distribution found at step 4.5, simulate the economy for 45 more periods (covering the age range of 20-64, typically the most economically active years of a person's lifetime) and compute consumption of each simulated household in each period.
8. For each household, compute per-period consumption equivalent, which is defined as the fixed dollar value of consumption such that if consumed repeatedly over 45 years, would yield the same total utility as the simulated time-variant consumption stream. At the end of this step, for each simulated agent, consumption equivalent is computed.
9. Compute mean consumption equivalent over all simulated agents. This, combined with per-capita banks' profits, is the measure of per-capita social welfare. Since banks' profits in equilibrium will be zero or near zero, social per-capita welfare will be almost entirely made up of mean consumption equivalent.

3.3.2 Mapping the Model to U.S. Data

To answer the question put in the title of this paper, I collect US data regarding income, credit cards, and bankruptcy. I use annual data and assume cardholders can stay in default for a year without formal bankruptcy filing.[9]

[9] There is no hard data on the typical time it takes from the first missed payment on a credit card until bankruptcy filing, but from anecdotal evidence I found on the Internet, this is a reasonable assumption. The

Table 3.1 provides values for some model parameters.

Table 3.1: Parameter values

Parameter	Description	Value	Source
ρ	Coefficient of risk aversion	2	Standard in macro literature
$1/q^{RF}$	Gross risk-free interest rate	1.04	McGrattan and Prescott 2000
π	Income transition matrix	5x5 matrix	Corbae et al. 2009

To pin down the five states of income, I use the 2008 Consumer Expenditure Survey where I take the data for the after-tax household income by quintile (average within each quintile) from which I subtract quintile-averaged annual expenditures on mortgage or rental payments and utilities. The rationale for subtracting those expenditures is three-fold. First, these are usually required to be made in cash, which reduces the amount of disposable cash left from each paycheck to pay credit cards. Second, these expenditures constitute a large fraction of household income, ranging from 8% for the top-income quintile to 41% for the lowest-income quintile. And third, according to the CreditCards.com survey of December 2008, when finances are tight, 90% of people would pay the mortgage or utilities before paying other obligations. To calculate quintile-average housing payments, for each quintile, I attach the percentage of households with mortgage to the average mortgage payment, the percentage of renters to the average rental payment and take the sum (assuming housing expenditure of those homeowners without mortgage is zero).

The median credit card limit in 2007 was $18,000 (SCF 2007). According to the survey conducted by CreditCards.com, the median default annual interest rate on credit

amount of time between first missed payment and bankruptcy filing varies and it depends on the creditor's collection strategy, amount of debt and individual borrower's circumstances.

cards in January 2008 was 28.99%. So I set q^D = 1/1.2899 and l= -18,000 as the default interest rate and the credit limit. Note that $18,000 is the maximum outstanding debt, so the maximum amount borrowed is $18,000 times the appropriate bond price, which will always be less than $18,000. The discount rate β is calibrated under UD regime with the above credit card terms to match the ratio of non-business bankruptcy filings in 2008 due to earnings shock to the total U.S. population age 18+ in 2008. According to the U.S. Courts, there were 1.074 million bankruptcy filings in 2008. According to the Census Bureau, the total population age 18 and above in 2008 was 230.1 million. Therefore, the ratio of people who filed to total population age 18 and above is 0.47%. Based on the data on bankruptcy filings over 1984-95 from PSID, Chakravarty and Rhee 1999 find that 12.2% of filing people cite job loss as the reason for filing and 41.3% cite credit misuse. I associate these reasons with the income shock in my model. The other three reasons cited – marital disruption, health-care bills, and lawsuit/harassment – are expenditure shocks, not income shocks. Since in my model, I do not consider expenditure shocks, I take as a target the fraction of all non-business bankruptcy filings associated only with an income shock to population age 18+. This number is (12.2%+41.3%)·0.47% = 0.25%.

It may be interesting to note that the number of credit cards that each household is allowed to have in the model, two, is not only convenient from the computational viewpoint, but is also the median number of credit cards among all cardholders in 2007 (SCF 2007).

3.4 SIMULATIONS AND RESULTS

3.4.1 UD regime

The result of the grid search for the UD regime is shown in Figure 3.1 and Figure 3.2 (the entire grid cannot be shown for lack of space, so I zero in on the part of the grid that contains the equilibrium). Each row in both tables corresponds to $1/q_A^R$ (the interest rate of Bank A), each column corresponds to $1/q_B^R$ (the interest rate of Bank B), and inside are the total profits of *A* and *B* over 200,000 simulated households in thousand dollars. For both banks, profit tends to rise with own interest rate, although this relationship is non-monotonic due to the discrete nature of the household's problem (they also choose the borrowing amount on a grid). The cell with unique equilibrium interest rates, $(q_A^R, q_B^R) =$ (1/1.08,1/1.2575), is marked with a border. The fraction of households filing for bankruptcy in each period is equal to 0.222%, which is slightly under the target of 0.25%. Potential entrants from neither group can get a positive profit by offering a lower interest rate, which manifests itself in all the cells above the equilibrium in Figure 3.1 and all cells to the left of the equilibrium in Figure 3.2 being negative.

Figure 3.1: Profit of Bank *A* under UD regime

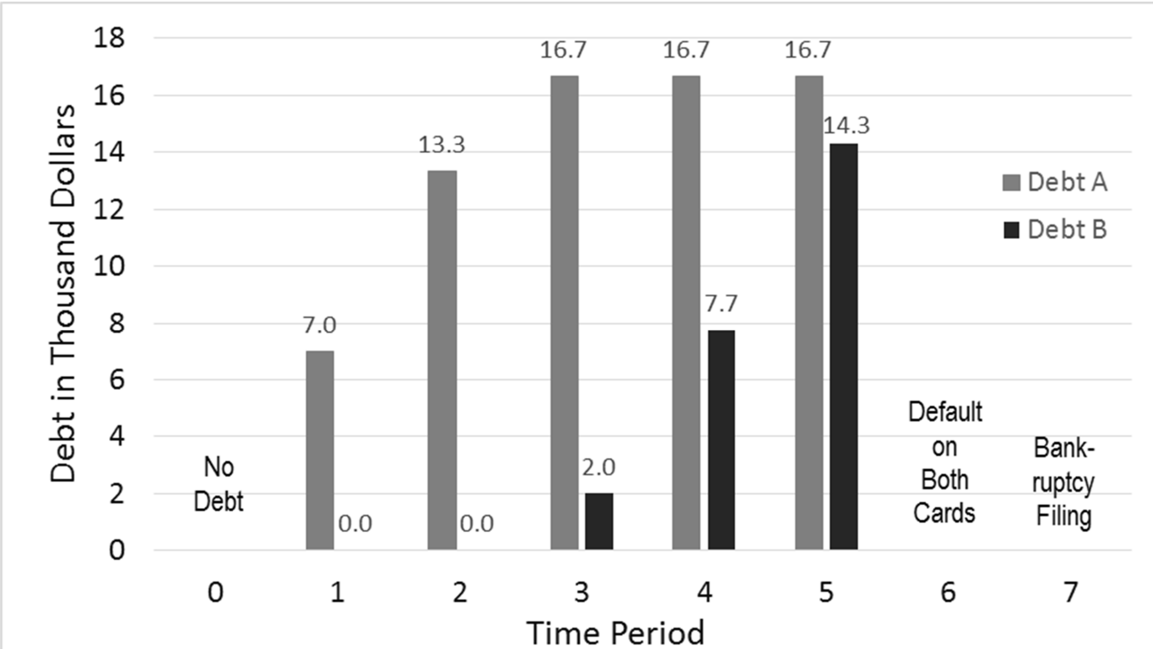
Interest Rate of Bank A	Interest Rate of Bank B												
	1.245	**1.2475**	**1.25**	**1.2525**	**1.255**	**1.2575**	**1.26**	**1.2625**	**1.265**	**1.2675**	**1.27**	**1.2725**	**1.275**
1.07	-719	-958	-1310	-1336	-1334	-1172	-1148	-1178	-1197	-1031	-900	-976	-970
1.0725	-792	-907	-784	-824	-841	-624	-617	-624	-836	-825	-643	-676	-768
1.075	-550	-534	-500	-472	-482	-544	-528	-355	-397	-417	-474	-674	-689
1.0775	-103	-123	-169	-145	-132	-223	-185	-307	-287	-374	-408	-274	-271
1.08	100	27	143	219	238	**47**	-36	-8	46	41	-60	-41	84
1.0825	472	452	316	274	451	432	476	488	-2	-19	-35	-2	328
1.085	238	762	783	846	826	862	808	808	600	594	541	282	231
1.0875	1070	1055	1065	1073	1089	537	550	558	396	367	700	725	748
1.09	758	765	745	752	861	852	837	987	1005	838	798	839	875

Figure 3.2: Profit of Bank *B* under UD regime

Interest Rate of Bank A \ Interest Rate of Bank B	1.245	1.2475	1.25	1.2525	1.255	1.2575	1.26	1.2625	1.265	1.2675	1.27	1.2725	1.275
1.07	384	13	44	15	38	145	263	185	256	465	553	568	650
1.0725	-75	16	-208	-205	-221	281	333	428	-123	-87	479	476	479
1.075	-277	-235	-256	-151	-203	-216	-150	338	406	439	450	-82	15
1.0775	-315	316	316	274	326	-169	-151	-141	-143	34	119	112	175
1.08	-312	-216	-194	-95	-71	**18**	20	100	135	165	177	196	278
1.0825	-36	-17	-61	-53	72	96	108	120	-260	-312	-267	-223	344
1.085	-578	-53	-52	17	26	94	139	158	164	182	242	-94	-101
1.0875	-123	-105	-36	2	23	-316	-309	-258	-361	-356	-106	-42	-62
1.09	-612	-523	-462	-473	-289	-272	-173	-139	-154	-156	-152	-125	-114

Under the UD regime, simulated agents never choose to default on just one credit card and borrow on the other card because in that case they will be subject to the default interest rate on both cards. Instead, when experiencing a series of low income realizations, agents first accumulate debt on the low interest rate credit card and only after reaching credit limit, start borrowing on the higher interest rate card. If the series of bad income shocks is long enough, eventually agents will max out the second card as well, default (miss a payment) on both cards at the same time, and declare bankruptcy in the next period. Figure 3.3 shows the dynamics of an agent's debt amounts on the two cards, starting with no debt or savings, when the agent repeatedly receives the worst income draw.

Figure 3.3: Accumulation of debt on cards *A* and *B* under UD regime when a borrower repeatedly receives the worst income draw

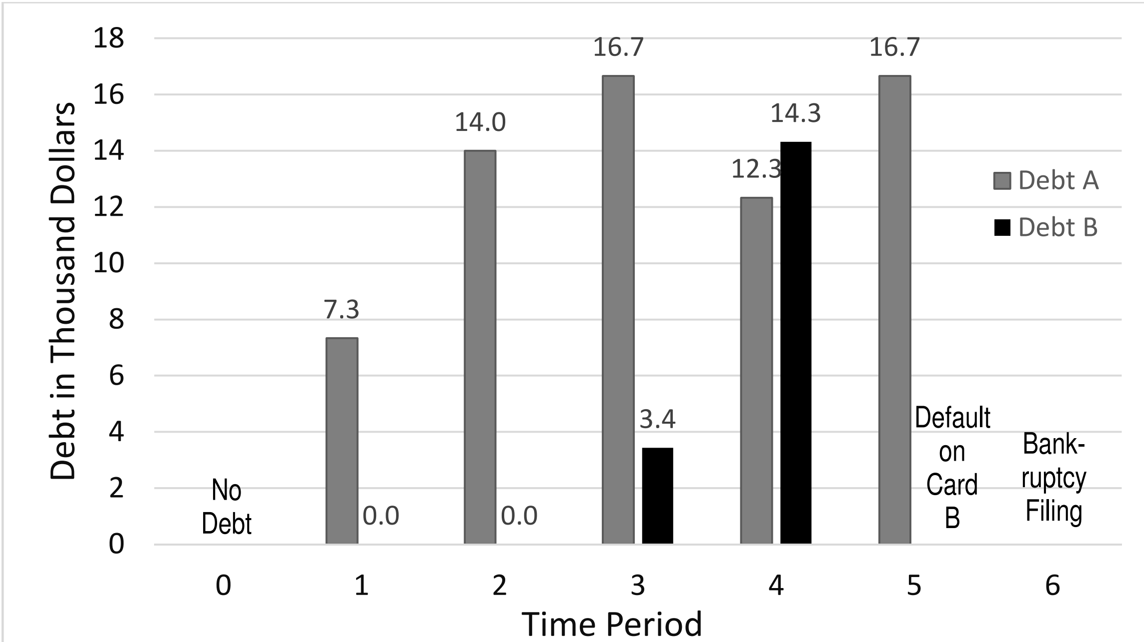


Social welfare per-capita equals \$38,503. The next step is to find an equilibrium under NUD regime, compute social welfare per-capita, and compare it to that under the UD regime.

3.4.2 NUD regime

Simply prohibiting banks from charging default interest rate on their non-defaulted credit card results in borrowing behavior that brings negative profits to bank *B*. For example, assuming the same pair of bond prices as in UD equilibrium results in the following debt dynamics when a borrower is hit with a series of lowest income shocks.

Figure 3.4: Accumulation of debt on cards A and B under NUD regime when a borrower repeatedly receives the worst income draw, assuming regular bond prices as in UD equilibrium



Agents max out card *B* after only two periods of using it (periods 3 and 4), and then default on it in period 5. Under the UD regime however, it takes three consecutive periods of worst income realizations from the period in which an agent starts using card *B* and the period in which default occurs, which is one more period than under UD regime. This one-period difference in time-to-default between the two regimes plays a very important role, since credit card issuers make money off of repaying borrowers but lose money when agents default. It turns out that even the highest possible regular interest rate on card *B*, 28.75%, is not sufficient for bank *B* to break even. Under the assumption that the default interest rate is exogenously fixed at 28.99% (which caps regular interest rate at 28.75%), an equilibrium under the NUD regime with two groups of banks fails to exist, because bank *B* will always make negative profit.

So I relax the assumption that the default interest rate is fixed under the NUD regime and allow it to vary in a way consistent with the UD equilibrium interest rates. Specifically, for NUD regime, I assume that the difference between the default interest rate and the regular interest rate on card *A* ("default premium") is fixed at 20.99 percentage points (which is equal to the difference between the default interest rate on card *A* and regular interest rate on *A* under UD regime, 28.99-8.00), and likewise, the default premium on card *B* to be fixed at 28.99-25.75 = 3.24 percentage points.

The comparison between UD and NUD regime outcomes should be done in a model environment in which the only difference between the two regimes is the conditions that trigger penalty interest rates. So far I have computed the UD equilibrium keeping default interest rates constant at 28.99%, but under the NUD regime, I'm going to assume this away and instead keep default premiums constant. To eliminate this difference between the two regimes, I verify that the previously found UD equilibrium still holds, if the UD economy operated under the constant default premiums assumption. That is, I find that in the UD regime, under the assumption that default interest rates are equal to regular interest rates plus a constant premium (which is bank specific: 20.99 percentage points for Bank *A* and 3.24 p.p. for Bank *B*), the original UD equilibrium, in which Bank *A* charges 8% regular interest rate and Bank *B* charges 25.75% rate, still holds indeed, i.e. should Bank *A* offer a regular interest rate lower than 8% (while Bank *B*'s regular interest rate is still 25.75%), it will earn negative profit and likewise, should Bank *B* offer a regular interest rate lower than 25.75% (while Bank *A*'s regular interest rate is still 8%), it will, too, make negative profit.

Now that the default interest rates in both regimes are determined in exactly the same way, let's move on to the discussion of equilibrium under NUD regime. Figure 3.5 and Figure 3.6 below show profits for Banks *A* and *B* using a very coarse grid. This bird's eye view shows negative profits for Bank *B* across the board except (1.1,1.44) and the region where the interest rate on *B* is 1.92-1.96.

Figure 3.5: Profit of Bank *A* under NUD regime, coarse grid

Interest Rate of Bank B

Interest Rate of Bank A	1.28	1.32	1.36	1.4	1.44	1.48	1.52	1.56	1.6	1.64	1.68	1.72	1.76	1.8	1.84	1.88	1.92	1.96
1.1	-523	-731	-669	-760	-1187	-3646	-3661	-3722	-3548	-3604	-3458	-3118	-3129	-2985	-3054	-2625	-2593	-2795
1.12	36	163	179	-290	-2002	-2185	-2189	-2276	-2367	-2540	-2176	-2054	-2094	-2379	-2326	-1458	-1555	-1340
1.14	1098	1427	1067	832	-830	-818	-1430	-1358	-1677	-1505	-1350	-1234	-687	-541	-98	-134	-97	-436
1.16	2004	1846	1657	1914	61	-89	-532	-574	-559	-593	-409	-396	-460	234	357	525	501	417
1.18	1911	1978	1772	918	1342	392	-487	143	257	233	492	218	175	-179	-175	-81	-1953	-1784
1.2	2790	2795	2358	694	-423	5	350	283	-180	-417	-662	-720	-813	-900	-1527	-1676	-1908	-2345
1.22	3309	3152	3107	980	824	606	434	66	-106	-154	-265	-727	-681	-1518	-1667	-1757	-1449	-1899
1.24	3918	3759	1476	1494	1193	896	384	233	13	-77	-498	-897	-919	-1459	-1410	-1548	-865	-1096
1.26	3830	4063	1452	1459	1116	849	936	699	681	328	-716	-771	-106	-208	-354	-458	-1155	-1299

Figure 3.6: Profit of Bank *B* under NUD regime, coarse grid

Interest Rate of Bank B

Interest Rate of Bank A	1.28	1.32	1.36	1.4	1.44	1.48	1.52	1.56	1.6	1.64	1.68	1.72	1.76	1.8	1.84	1.88	1.92	1.96
1.1	-1954	-1500	-740	-161	58	-2430	-2376	-2353	-2284	-2090	-2129	-1629	-937	-708	-488	-408	-165	-11
1.12	-2211	-2165	-1566	-447	-3256	-3037	-2905	-2735	-2555	-2061	-2063	-1740	-1511	-1114	-700	-474	-217	-71
1.14	-2249	-2331	-1860	-1407	-3556	-2988	-3185	-2806	-2573	-1998	-1644	-1358	-1256	-1077	-778	-529	-254	99
1.16	-2474	-2061	-1617	-1579	-3553	-3362	-2747	-2596	-2349	-1928	-1887	-1612	-1256	-1353	-880	-167	205	292
1.18	-2698	-1879	-1403	-3842	-2690	-3139	-3979	-2902	-2061	-1803	-2006	-1596	-1392	-1164	-1265	-884	-2304	-2188
1.2	-2842	-2252	-1498	-4567	-4328	-3931	-3692	-3092	-2645	-2244	-2720	-2613	-2470	-2396	-2195	-2255	-2509	-2861
1.22	-3020	-2422	-1812	-4732	-4098	-3821	-3496	-3359	-2902	-2747	-2433	-2612	-2531	-2811	-2862	-2766	-2873	-3040
1.24	-3126	-2435	-5221	-4891	-4184	-4174	-3613	-3423	-3009	-2721	-2824	-2923	-2867	-3574	-3088	-3137	-2856	-2845
1.26	-2853	-2583	-5453	-4902	-4509	-4055	-3608	-3313	-3188	-2912	-3821	-3256	-3005	-2977	-2801	-2761	-3023	-2955

Search on a fine grid around (1.1,1.44) does not yield an equilibrium. Search on a fine grid where the interest rate on *B* is 1.92-1.96 does yield an equilibrium with regular bond prices (1/1.1475,1/1.93):

Figure 3.7: Profit of Bank *A* under NUD regime, fine grid

Interest Rate of Bank B

Interest Rate A	1.9200	1.9225	1.9250	1.9275	1.9300	1.9325	1.9350	1.9400
1.1400	-97	-97	-100	-100	-174	-175	-175	-168
1.1425	-125	-125	-128	-211	-210	-230	-225	-225
1.1450	-176	-167	-170	-173	-173	-172	-187	-190
1.1475	-23	112	112	105	**102**	94	94	99
1.1500	24	24	15	13	14	113	111	76

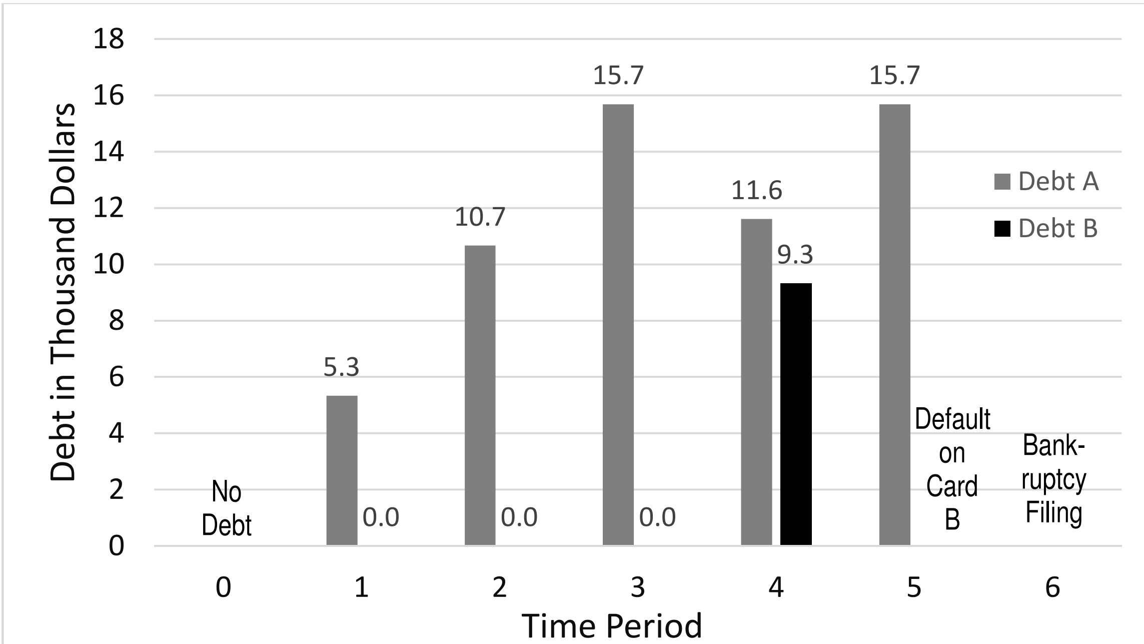
Figure 3.8: Profit of Bank *B* under NUD regime, fine grid

Interest Rate of Bank B

Interest Rate A	1.9200	1.9225	1.9250	1.9275	1.9300	1.9325	1.9350	1.9400
1.1400	-254	-242	-222	-209	-177	-165	-152	-86
1.1425	-125	-113	-94	-102	-90	-71	-94	-70
1.1450	-145	-159	-140	-126	-114	-103	-90	-65
1.1475	-93	-149	-137	-105	**0**	14	26	57
1.1500	-11	2	24	45	58	116	126	153

With no cap on regular interest rates, an equilibrium under NUD is obtained when bank *B* charges an extremely high interest rate to make up for the losses from more frequent defaults. There are more bankruptcies here than under the UD regime: 0.254% compared to 0.222%. Social per-capita welfare is equal to $37,817, which is 1.78% less than under the UD regime. The debt dynamics chart below helps shed light on why bank *B* has to charge such a high interest rate: when the rate on the high interest rate card is too high, agents minimize the use of the high interest rate card and only carry debt on it for one period before defaulting when hit by a series of the worst income shocks.

Figure 3.9: Accumulation of debt on cards A and B under NUD regime when a borrower repeatedly receives the worst income draw, NUD equilibrium bond prices

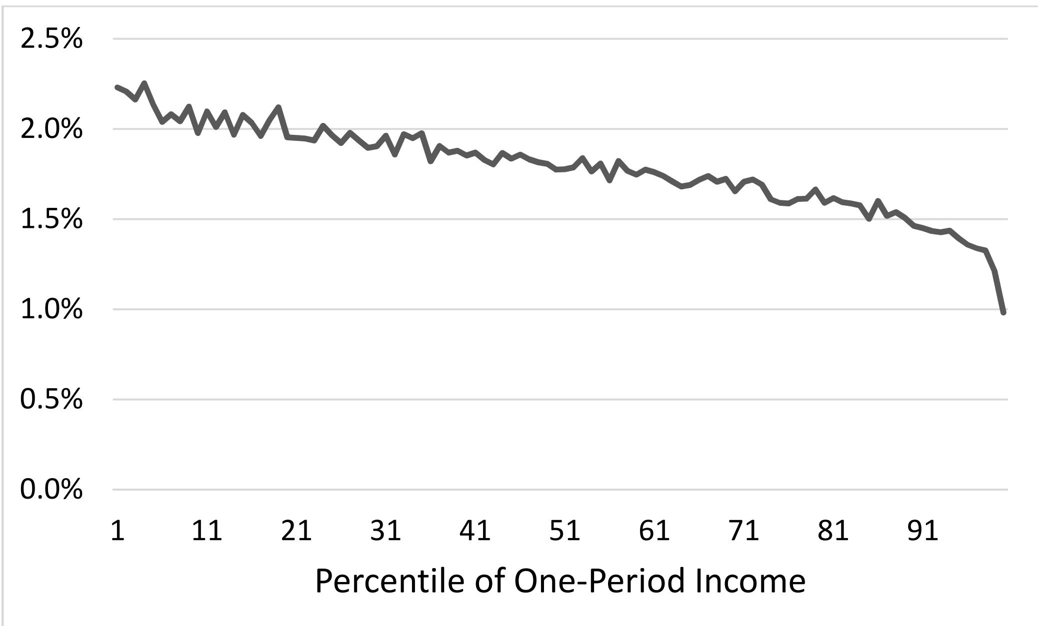


Bank *B* will lose a lot due to bankruptcies but with a high enough rate will make enough in interest from agents who get a series of bad income shocks long enough to start using card *B* but not too long to trigger bankruptcy. Those agents that get a good enough income realization in the period after they start using card *B* (or in the following period after they default on card *B*) pay a lot in interest, which eventually allows bank *B* to break even.

3.4.3 Distributional effects of regime switching

Although the society as a whole is clearly worse off when universal default in prohibited, we would expect poorer households to suffer the most since it's the poor who use debt the most. To study how regime switching affects different income groups, I broke down all 200,000 simulated households into 100 income bins (percentiles of per-period income) and for each income percentile group, I computed mean percentage reduction in consumption equivalent caused by regime switching. Figure 3.10 shows 100 income percentiles from the poorest to the richest households and the associated percentage drop in consumption equivalent for each income percentile.

Figure 3.10: Percentage loss in consumption equivalent by income percentile



As expected, poor households get affected by regime switching the most while rich households (who nevertheless may occasionally have low income and need to borrow) experience the smallest impact on their consumption equivalent in percentage terms.

3.5 CONCLUSION

In this paper I make an attempt to quantify the welfare effect of prohibiting universal default, which is one of the key aspects of the Credit CARD Act of 2009. I built a model in which consumers can borrow on two credit cards and compute it under two legal regimes. Under universal default regime, a creditor can charge default interest rate when the borrower is current with payments to them but defaults with another creditor, just like before the Credit CARD Act of 2009 took effect. Under no universal default regime, such practice is outlawed. I calibrated the model using US data on income, expenditures, and credit cards statistics under the UD regime and computed it under both regimes.

The novelty of this paper is the market structure in which two groups of banks offer different interest rates on credit cards that account for different levels of risk. This framework allows to study welfare implications of the provision of the CARD Act that prohibits universal default. When universal default is not allowed, borrowers take advantage of it by defaulting on the high interest rate card more often, still keeping the rate on the low interest rate card low. This change in borrower's behavior makes high interest rate card issuers raise the interest rate, which induces even more defaults. It requires a very high interest rate of 93% APR on the high rate card to catch up with the increasing risk of default and the losses that come with it. As a result, households' ability to smooth consumption is significantly diminished, which results in a sizeable reduction of per-capita welfare of 1.78% in terms of average consumption equivalent. Low-income households, who use credit cards to smooth out income fluctuations the most, take the biggest hit: the bottom 10 percent of the income distribution have their consumption equivalent reduced by more than 2%.

Bibliography

Ando, Albert and Franco Modigliani. 1963. "The "Life Cycle" Hypothesis of Saving: Aggregate Implications and Tests." American Economic Review, 53(1), 55-84.

Andrews, Dan and Andrew Leigh. 2009. "More inequality, less social mobility." Applied Economics Letters, 16(15): 1489-92.

Ausubel, Lawrence M., Oleg Baranov, and Amanda E. Dawsey. 2009. "Penalty Interest Rates, Universal Default, and the Common Pool Problem of Credit Card Debt." Working paper

Bailey, Martha J. and Susan M. Dynarski. 2011. "Gains and Gaps: Changing Inequality in U.S. College Entry and Completion." NBER Working Paper 17633, National Bureau of Economic Research, Inc.

Baird, Douglas G. and Thomas H. Jackson. 2001. "Cases, Problems and Materials on Bankruptcy." Foundation Press.

Becker, Gary S. and Tomes, Nigel. 1979. "An equilibrium theory of the distribution of income and intergenerational mobility." Journal of Political Economy, 87(6): 1153-89.

Bertrand, Marianne, and Adair Morse. 2013. "Trickle-Down Consumption." NBER Working Paper No. 18883.

Björklund, Anders, Jesper Roine, and Daniel Waldenström. 2012. "Intergenerational top income mobility in Sweden: Capitalist dynasties in the land of equal opportunity?" Journal of Public Economics, 96(5): 474-484.

Blanden, Jo. 2013. "Cross-Country Rankings in Intergenerational Mobility: A Comparison of Approaches from Economics and Sociology." Journal of Economic Surveys, 27(1): 38-73.

Breiman, Leo. 1996. “Out-of-bag estimation.” Manuscript, ftp.stat.berkeley.edu/pub/users/breiman/OOBestimation.ps

Breiman, Leo. 2001. “Random Forests.” Machine learning, 45(1): 5-32.

Bricker, Jesse, Rodney Ramcharan, and Jacob Krimmel. 2014. "Signaling Status: The Impact of Relative Income on Household Consumption and Financial Decisions." FEDS Working Paper No. 2014-76.

Chakravarty, S., and E.-Y. Rhee. 1999. "Factors Affecting an Individual's Bankruptcy Filing Decision." Mimeo, Purdue University.

Charles, Kerwin K., Erik Hurst, and Nikolai Roussanov. 2009. "Conspicuous Consumption and Race." Quarterly Journal of Economics, 124(2): 42-67.

Chatterjee, Satyajit, Dean Corbae, Makoto Nakajima, and Jose-Victor Rios-Rull. 2007. "A Quantitative Theory of Unsecured Consumer Credit with Risk of Default." Econometrica, 75(6): 1525-1589.

Checchi, D., Andrea Ichino, and Aldo Rustichini. 1999. "More equal but less mobile? Education financing and intergenerational mobility in Italy and in the US." Journal of Public Economics, 74(3): 351-93.

Chetty, Raj, Nathaniel Hendren, Patrick Kline, and Emmanuel Saez. 2014. "Where is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States." Quarterly Journal of Economics, 129(4): 1553-1623.

Christen, Markus and Ruskin M. Morgan. 2005. "Keeping Up with the Joneses: Analyzing the Effect of Income Inequality on Consumer Borrowing," Quantitative Marketing and Economics 3: 145-173.

Coibion, Olivier, Yuriry Gorodnichenko, Marianna Kudlyak, and John Mondragon. 2014. "Does Greater Inequality Lead to More Household Borrowing? New Evidence from Household Data". NBER Working Paper No. 19850

Corak, Miles. 2013. "Income Inequality, Equality of Opportunity, and Intergenerational Mobility." Journal of Economic Perspectives, 27(3): 79-102.

Corbae, Dean, Pablo D'Erasmo, and Burhanettin Kuruscu. 2009. "Politico-Economic Consequences of Rising Wage Inequality." Journal of Monetary Economics, 56(1): 43-61.

Cutler, David M. and Edward L. Glaeser. 1997. "Are Ghettos Good or Bad?" Quarterly Journal of Economics, 112 (3): 827-72

Daly, Mary and Leila Bengali. 2014. "Is It Still Worth Going to College?" FRBSF Economic Letter, 2014-13

Daly, Mary C. and Daniel J. Wilson. 2006. "Keeping Up with the Joneses and Staying Ahead of the Smiths: Evidence from Suicide Data." Federal Reserve Bank of San Francisco WP 2006-12.

DeBacker, Jason, Brad Heim, Vasia Panousi, Shanthi Ramnath, and Ivan Vidangos. 2013. "Rising Inequality: Transitory or Persistent? New Evidence from a Panel of US Tax Returns." Brookings Papers on Economic Activity, Spring.

Duesenberry, James S. 1949. “Income, Saving, and the Theory of Consumer Behavior.” Harvard University Press, Cambridge, MA.

Friedman, Milton. 1957. “A Theory of the Consumption Function.” Princeton University Press, Princeton, NJ.

Graham, Bryan and Patrick Sharkey. 2013. "Mobility and the Metropolis: How Communities Factor Into Economic Mobility." Washington D.C.: Economic Mobility Project, The Pew Charitable Trusts.

Guven, Cahit and Bent E. Sorensen. 2012. "Subjective Well-Being: Keeping Up with the Joneses. Real or Perceived?" Social Indicators Research, 109: 439-469.

Heathcote, Jonathan, Fabrizio Perri, and Gianluca L. Violante. 2010. "Unequal We Stand: An Empirical Analysis of Economic Inequality in the United States, 1967-2006." Review of Economic Dynamics, 13: 15-51.

Heffetz, Ori. 2011. "A Test of Conspicuous Consumption: Visibility and Income Elasticities." Review of Economics and Statistics 93(4): 1101-1117.

Iceland, John. 2004. "Beyond Black and White: Metropolitan residential segregation in multiethnic America." Social Science Research, 33 (2): 248-271.

Jackson, Thomas H. 1985. "Translating Assets and Liabilities to the Bankruptcy Forum." Journal of Legal Studies, 14, 73, 74-77.

Jackson, Thomas H. 1986. "The Logic and Limits of Bankruptcy Law." Harvard University Press, Cambridge, MA.

Jackson, Thomas H. and Robert E. Scott. 1989. "On the Nature of Bankruptcy: An Essay on Bankruptcy Sharing and the Creditors' Bargain." Virginia Law Review, 75(2): 155-204.

Kopczuk, Wojciech, Emmanuel Saez, and Jae Song. 2010. "Earnings Inequality and Mobility in the United States: Evidence from Social Security Data since 1937." Quarterly Journal of Economics, 125: 91-128.

Krueger, Dirk and Fabrizio Perri. 2006. "Does Income Inequality Lead to Consumption Inequality? Evidence and Theory." Review of Economic Studies, 73: 163-93.

Kuhn, Peter, Peter Kooreman, Adriaan R. Soetevent, and Arie Kapteyn. 2010. "The Effects of Lottery Prizes on Winners and their Neighbors: Evidence from the Dutch Postcode Lottery." IZA Discussion Paper 4950.

Livshits, Igor, James MacGee, and Michele Tertilt. 2007. "Consumer Bankruptcy: A Fresh Start." American Economic Review 97(1): 402-18.

Luttmer, Erzo F. P. 2005. "Neighbors as Negatives: Relative Earnings and Well-Being," Quarterly Journal of Economics 120(3): 963-1002.

Maurer, Jurgen and Andre Meier 2008. "Smooth It Like the Joneses? Estimating Peer-Group Effects in Intertemporal Consumption Choice," The Economic Journal 118: 454-476.

McGrattan, Ellen R. and Edward C. Prescott. 2000. "Is the Stock Market Overvalued?" Federal Reserve Bank of Minneapolis Quarterly Review, 24(4): 20-40.

Neumark, David and Andrew Postlewaite. 1998. "Relative Income Concerns and the Rise in Married Women's Employment." Journal of Public Economics, 70: 157-183.

Nishiyama, Shinichi. 2000. "Measuring Time Preference and Parental Altruism." Congressional Budget Office, Working Paper.

Parlour, Christine A. and Uday Rajan. 2001. "Competition in Loan Contracts." American Economic Review, 91(5): 1311-1328.

Piketty, Thomas. 2014. "Capital in the Twenty First Century." Harvard University Press, Cambridge, MA.

Rajan, Raghuram. 2010. "Fault Lines: How Hidden Fractures Still Threaten the World Economy." Princeton University Press, Princeton, NJ.

Solon, Gary. 2004. "A Model of Intergenerational Mobility Variation over Time and Place." In Miles Corak (ed.), “Generational Income Mobility in North America and Europe.” Cambridge: Cambridge University Press, pp. 38-47.

Stiglitz, Joseph. 2013. "The Price of Inequality: How Today's Divided Society Endangers Our Future" W. W. Norton and Company.

Tolbert, Charles M. and Molly Sizer. 1996. "U.S. Commuting Zones and Labor Market Areas: A 1990 update." Economic Research Service Staff Paper, 9614.

Veblen, Thorstein. 1899. “The Theory of the Leisure Class: an Economic Study of Institutions.” Reprinted 1994, Penguin Books: New York, NY.

Zizzo, Daniel J. and Andrew J. Oswald. 2001. "Are People Willing to Pay to Reduce Others' Income." Annales d'Economie et de Statistique 63/64: 39-65.