PRELIMINARY AND DRAFT VERSION

Can Economically Intuitive Factors Improve Ability of Proprietary Algorithms to Predict Defaults of Peer-to-Peer Loans?

Atay Kizilaslan

Aziz Lookman

Abstract: We examine whether economically intuitive loan and borrower characteristics and macroeconomic factors predict peer-to-peer loan defaults beyond what proprietary algorithms predict. Using county-level unemployment data, we find that loans originated in high unemployment areas are more likely to default. In addition, we develop a novel measure of job stability, and show that borrowers with greater job stability are less likely to default on their loans. Our results are robust to controlling for other loan and borrower characteristics previously used to model default risk.

I. Introduction

Peer to peer lending (P2P) is an alternative way of debt financing in which individuals and firms lend and borrow directly among each other rather than having a financial institution as an intermediary. Web-based P2P platforms have experienced exponential growth in recent years and each P2P platform emphasizes its use of sophisticated, algorithm-driven underwriting to analyze data provided by the borrower in the loan application as well as data from multiple other sources to assess a borrower's credit risk.

Since the algorithms used by the P2P platforms are proprietary, it is an open question as to how well they work in practice to predict default risk. In this paper, we ask the following question: Do these algorithms incorporate the discriminative power of macroeconomic factors, loan and borrower characteristics that economic theory suggests should be predictors of default risk?

In particular, we focus on cross-sectional variation in unemployment rates at the time of loan issuance and the job stability of the borrower. Based on our analysis of data from LendingClub, a prominent P2P platform in the US, we find that inclusion of macroeconomic factors and borrower-specific characteristics can improve the discriminative power of algorithms. For example, in the subset of loans with high interest rates ("high risk loans"), those issued in high unemployment areas default two percent more than the loans in low unemployment areas and borrowers with greater job stability default five percent more than borrowers with lower job stability. Furthermore, our proposed factors remain highly significant even after controlling for a host of other borrower and loan characteristics, such as the borrower's debt to income ratio ("DTI"), and the frequency of recent credit inquiries also have incremental predictive power.

Moreover, the impact of macroeconomic conditions and borrower characteristics is additive

Earlier literature on P2P lending focus on these online platform's ability to mitigate information asymmetry. Using Prosper.com data, Lin, Prabhala, and Viswanatan (2009), for instance, examine the effect of social networks on lending outcomes. The authors focus on the relational aspect of social networks, that is roles and identities of individuals in the network (e.g., borrower and lender), and find that borrowers with stronger and more verifiable social networks are more likely to obtain a loan with a lower interest rate. In addition, the authors show that these lenders are less likely to default. Similarly, Duarte, Siegel and Young (2009) examine the impact of trust on lending outcomes using Prosper.com's data. They find that borrowers perceived as less trustworthy are less likely to get their loan funded. Meanwhile, borrowers perceived as trustworthy enjoy lower interest rate on their loans.

Recent literature focuses on defaults in P2P loans. For instance, Iyer, Khwaja, Luttmer and Shue (2015) shows that P2P lenders can achieve 45 percent greater accuracy in predicting default using their own subjective information instead of credit score, which is a traditional measure of creditworthiness used by financial institutions. Meanwhile, Serrano-Cinca, Gutiérrez-Nieto, and López-Palacios (2015) show that loan purpose, annual income, current housing situation, and indebtedness are statistically significant predictors of P2P loan default. In addition, the authors show that credit grade assigned by P2P platform combined with borrower's debt level improve the performance of the default model.

Our paper contributes to this recent literature on P2P loan defaults, focusing on the impact of local macroeconomic conditions and job stability on loan default. A recent paper by Li, Yao, Wen and Yang (2016) also examines whether macroeconomic conditions and borrower characteristics are determinants of prepayment and default in P2P loans. However, they focus on general macroeconomic trends, such as national-level changes in GDP. Since underwriting standards have also changed frequently, it is difficult to separate out the impact of changes in macroeconomic conditions from changes in underwriting quality. We, on the other hand, focus on how cross-sectional variation in local macroeconomic factors, such as zip-code level unemployment rates, predicts loan default. Various studies show the importance of local unemployment on loan default. For instance, Elul, Souleles, Chomsisengphet, Glennon, and Hunt (2010) show that county-level unemployment shocks are associated with high mortgage default. Similarly, studies on student loans show that unemployment is positively associated with student loan default (See Gross, Cekic, Hossler and Hillman (2009) for literature review).

We also construct a novel predictor of consumer loan defaults, the borrower's job stability. We construct this predictor based on whether the borrower works for a government agency. To our knowledge, we are the first to demonstrate that job stability is associated with a lower default rate of P2P loans.

The remainder of the paper is organized as follow. In Section 2, we describe the P2P marketplace. In Section 3, we describe our datasets. In section 4, we discuss our methodology, and Section 5 presents our results. Finally, section 6 presents concluding remarks.

II. Description of the P2P Marketplace

Web-based P2P lending first started in U.K. in 2005, and the first U.S. P2P marketplace was established in 2006. As of 2017Q1, LendingClub, the largest P2P platform in U.S., had facilitated about $26 billion in loans and Prosper, the second largest P2P platform, had facilitated about $9 billion in loans. While the level of P2P debt is still small relative to the total consumer debt of $3 trillion, P2P lending has grown dramatically in recent years and it is expected to continue to grow dramatically. For example, Transparency Market Research predicts that P2P lending will grow at a compound annualized rate of approximately 50% between 2016 and 2024.[1] While P2P lending started with consumer loans, today, various P2P platforms also offer small business, student, and real estate loans as well.

In P2P lending, a borrower applies for a loan by filing an online application that collects data about the borrower's income, employment history, purpose of the loan and other information that the P2P platform uses to determine a borrower's creditworthiness. Using this information as well as data from various other sources, the P2P platform assigns a credit risk rating to the borrower and determines the interest rate of the loan. However, unlike bank, a P2P platform doesn't fund the loan. Instead, the financing of the loan comes from individual investors who decide for themselves which loans they want to finance.[2] Also, unlike a bank, the revenues for a P2P platform comes primarily from fees, such as a placement fee charged to borrowers, and a servicing fee charged to the investors.

From a borrower's perspective, as compared to banks or other lending institutions, a P2P platform provides a fast and streamlined loan application process. Additionally, a P2P platform is likely to offer financing to borrowers who are deemed too risky for banks and offer better financing terms to risky borrowers who otherwise would only be able to borrow from alternative lenders such as payday loan lenders. For investors, a P2P platform provides an investment opportunity with a potentially attractive rate of return and the option to invest in loans with varying degrees of risk and return.

[1] TMR reports a compound annual growth rate.
[2] In recent years, P2P loans have also been financed via securitization.

III. Data Sources

Our two main sources are loan data from LendingClub and macroeconomic data from the Bureau of Labor Studies. We describe the data and our two main predictors in greater detail below.

a. Lending Club data

We obtain our loan data from LendingClub.com, the largest P2P platform in United States. LendingClub provides unsecured financing of up to $50,000 for personal loans, small business loans, auto refinancing and financing for certain medical procedures. We obtained the data from LendingClub's website at: (https://www.lendingclub.com/info/download-data.action).[3] This data contains information on several loan and borrower characteristics, such as loan amount, loan purpose, debt-to-income ratio, as well as information on the status of the loan (i.e., whether the loan is current, late, matured or in default.)

We focus on loans with a 36-month term that originated between 2011 and December 2015. We excluded loans issued in earlier years as the data is very sparse for these loans. In addition, we excluded loans originated after 2015 as these loans are relatively young and may not have enough default data to draw meaningful results. As part of our data cleaning process, we also drop observations with the following characteristics:

i. Missing information on variables required for the analysis[4]
ii. DTI in excess of 50%
iii. Loan amount or principal repaid is negative
iv. Last payment received more than 42 months after the loan was issued.

Our final sample consists of 603,701 loans.

Using loan-level data, we construct a novel predictor of defaults based on the job stability of the borrower. We construct this variable as an indicator based on whether or not the borrower works for a Federal, State or local government agency. We identify the borrower's employer by performing a text string search on the employer field (i.e., emp_title) in the LendingClub data. We flag a borrower as a government employee if this field contains terms that would be

---

[3] The company also provides information on declined loans on their website.
[4] These variables are loan amount, issue date, interest rate, payments received to date for total amount funded, principal received to date, interest received to date, maturity, unemployment rate for the state that the borrower is in as of the month of loan issue, annual county-level unemployment rate, as a fraction of the national unemployment rate.

indicative of a government employer.[5] We hypothesize that loans made to borrowers with greater job stability are less likely to default.

Unemployment Data

We obtain unemployment data from Bureau of Labor Statistics' website.[6] Our main results are based on annual county level unemployment data. Merging this county-level data with the LendingClub loan data is not a straightforward exercise because LendingClub data only provides geographic information about a borrower in the form of the first 3 digits of the borrower's zip code ("zip-3"). Accordingly, we develop a translation table to zip-3 codes to county. For the zip-3 codes where we are do not have a matching county, we use the state unemployment rate. Since our analysis is based on cross-sectional variation in unemployment rates, we construct a relative measure as our proxy of unemployment intensity. We define this measure as the employment rate for a county in a given year, divided by the average unemployment rate across all counties for that year. We were unable to obtain county-level unemployment for 2015. Since we found that the relative unemployment rate for a zip-3 code is stable through time, we use the 2014 relative unemployment rates to fill in the 2015 relative unemployment rates.

IV. Summary Statistics

a. Growth in Loan Origination

In Figure 1a and 1b, we present number and dollar amount of loans originated from 2011 through 2015. As both figures show, there growth in loan origination has been dramatic over our sample period. While only approximately 14,000 loans were originated in 2011, by the end of our sample period, the number of originations has increased almost 20 times to about 280,000 loans by 2015. Similarly, the dollar amount of loans has also dramatically increased throughout our sample period, and has increased almost 27 times from $132 millions in 2011 to $3.6 billions in 2015. While this is a very rich sample to study defaults, given the tremendous growth in

[5] The exact string definition is provided in Appendix A.

[6] For a limited number for years, we collect several socio-economic macro variables at the county level from the countyhealthrankings.org website, such as: homicide rates, educational level, incidences of drug overdose and number of primary care physicians. We also collect county-level income information from the IRS website. However, our analysis to-date suggests that none of these factors are significant after controlling for unemployment.

originations, a significant number of loans in our sample (about 50%) are still maturing. Recall that we focus on loans with 3-year terms and therefore loans issued in 2014 or later are still maturing and the default outcomes for these cohorts is unknown.

b. Loan and Unemployment Characteristics

We present summary statistics of our LendingClub and unemployment data in Table 1. As seen, the average loan in our sample has an interest rate of approximately 12%, a principal amount of approximately $12,500, and a default rate of 10.02%. While the default rate is similar to 10.9% reported by Serrano-Cinca, Gutiérrez-Nieto, and López-Palacios (2015), the interest rate and loan size are somewhat higher than what Serrano-Cinca, Gutiérrez-Nieto, and López-Palacios (2015) report, possibly because their analysis uses data from earlier years.

As Table 1 presents, 3.21 percent of the borrowers is government employees.[7] Table 1 also shows that the inter-decile relative unemployment factor varies between 0.85 and 1.16 from first quartile to the last quartile. Furthermoere, our sample exhibits considerable variation in other factors used to explain defaults so we can examine the discriminative power or our measures is robust to inclusion of these other factors. For example, the DTI varies between 0 and 48% and the number of loan inquiries within last six month varies between 0 and 8.

V. Results

In this section, we describe our results. Since our focus is on whether factors other than the interest rate charged on a loan are predictive of a loan's default rate, we pay particular attention to controlling for the impact of interest rates in our analysis. First, we demonstrate the discriminative power of our proposed variables with univariate descriptive statistics. Next, we perform a formal statistical analysis of the significance of our factors using a multivariate Cox proportional hazard model. As a robustness test, we also use a machine learning algorithm to estimate a non-parametric baseline model for the impact of interest rates on loan default.

[7] Note that a borrower can take more than one loan during our sample period. However, our current analysis doesn't allow us to determine unique borrowers. Thus, our borrower-level statistics presents upper bound for that particular variable.

a. Univariate Analysis

i. Relative Unemployment Rate

To illustrate the impact of unemployment rates on default, while controlling for variation in interest rates, we compare default rates for loans issued in high versus low unemployment areas using a two by two sort. First, we sort loans based on their interest rate in each origination year. Then, we categorize them as being "low risk", "moderate risk" or "high risk" loans based on whether their interest rate is in the 0-25th percentile, 25th – 75th percentile or 75th -100th percentile of interest rate for loans issued in that year. Next, in each loan origination year, we sort loans based on the relative unemployment rate, that is, the unemployment rate relative to the national average unemployment rate. Then, we flag each loan as being in a low or high employment area based on whether their relative unemployment rate is in the 0-25th percentile or in the 75th -100th percentile. Next, for each loan origination year, we calculate the default rate for each loan-unemployment category. Finally, we average the yearly rates to obtain the average differential across 2011-2015.

The default rates for different categories of loans based on their risk and unemployment category is shown in Figure 2a. Loans issued to borrowers residing in high unemployment rates default at markedly higher rates. For example, for high-risk loans, the default rate in high unemployment areas is about two percentage points higher in that of low unemployment areas.[8] Loans in high unemployment areas default at a higher rate across all years. In Figure 2b, we plot the differential in default rates for the high-risk loans. As seen, the differential is positive across all years in the sample. For example, for the 2012 cohort, loans in high unemployment areas have a default rate that is 3% higher than loans issued in low unemployment areas.

ii. Job Stability

To illustrate the impact of job stability on default, we perform a similar analysis using a two by two sort, except that we categorize the loan by the job stability of the borrower (instead of the

[8] Since the 2014 and 2015 cohort are still maturing, in order to facilitate aggregation and comparison of different cohorts, we scale up the still-maturing default rates for these cohorts by factors of 1.1 and 1.3, respectively to make them comparable to the default experience of the fully matured earlier cohorts. We estimate these scaling factors by comparing the interim to final default rates of the 2011-2013 cohorts at similar maturities as the 2014 and 2015 cohort.

relative unemployment rate). We categorize a borrower as having high job stability if the borrower works for the government and as having low job stability otherwise.

The default rates for different risk-job stability categories of loans is shown in Figure 3a. Loans issued to borrowers with low job stability default at markedly higher rates. For example, for high-risk loans, the default rate of loans taken out by employees with lower job stability is almost five percentage points higher than loans taken out by employees with greater job stability. Loans taken out by borrowers with more stable jobs default at a lower rate in all years. In Figure 3b, we plot the differential in default rates for the high-risk loans. As seen, the differential is positive across all years in the sample. For example, for the 2011 cohort, loans taken out by employees with lower job stability have a default rate that is almost 5.5% higher than loans taken out by employees with greater job stability.

b. Multivariate Analysis

To examine whether relative unemployment rates and job stability offer incremental predictive power over and above other factors considered in the literature, we estimate a Cox proportional hazard rate model. Cox proportional hazard models are commonly used in lending literature. For example, Demyanyk and Van Hemert (2011) uses Cox proportional hazard model to examine the drivers of mortgage default. Similarly, Li, Yao, Wen and Yang (2016) uses competing hazard models to study the drivers of P2P loan defaults and prepayments. These results are summarized in Table 2.

In model 1, we use the interest rate as the sole predictor; we also include the year of origination as a categorical variable to control for variation in underwriting standards over time. As expected the interest rate is economically and statistically significant. In model 2, we add our relative unemployment rate and job stability variables as additional predictors. Both have the expected sign and are highly statistically significant. Additionally, the magnitude of the interest rate coefficient is virtually unchanged, suggesting that our variables are segmenting defaulted and non-defaulted loans along dimensions not considered by algorithms used to compute the interest rate.

Finally in model 3, we add a host of control variables considered in the literature. Our main variables remain statistically significant. The magnitude of the relative unemployment rate factor is virtually unchanged (it decreases by only about 5%), suggesting none of the other factors used to predict P2P loan defaults account for the impact of cross-sectional variation in unemployment rates. The magnitude of the coefficient on the job stability variable drops only marginally (decreases by about 20%), suggesting that this variable too is segmenting loans along a dimension substantially different from the previously-considered factors.

An alternative explanation for our results is that the interest rate has a non-linear relationship with defaults, and our factors are simply a proxy for the non-linearities in the relationship between interest rates and defaults. To examine this hypothesis, we used a non-parametric transform of the interest rate as the control variable in Model 1. To construct this transform, we first used a machine learning algorithm to model the relationship between interest rates and defaults using a machine learning algorithm.[9] Next, for each observation, we computed the hazard ratios predicted by this machine learning algorithm. By construction, this predicted hazard ratio captures the predictive power of the interest rate variable, regardless of whether the predictive power arises from a linear or non-linear transformation of the interest rate. We then use these predicted hazard ratios as our control for interest rate in Model 1. In unreported results, we find that our results are robust to using such non-linear monotone transformations of the interest rate as the control.

c. Economic Significance

To examine the economic significance of our results, we compare loans issued to borrowers with low job stability and residing in high unemployment areas ("bad loans") to loans issued to borrowers with high job stability and residing in low unemployment areas ("good loans"). We repeat the two by two step process described in Section V.a. , except that now within each subset of loans with similar interest rates, we compare the performance of "good" versus "bad" loans as defined above.

We compare the default rates in Figure 4a. As seen, the economic impact is significant. For example, even for the low risk loans, the default rate on the "bad" loans is more than three percentage points higher than for the "good" loans. In fact, the default rate in "bad" loans is more than double that of the default rate in "good" loans, which is only 2.4%. Differences in default rates for loans in the moderate and high risk categories are also economically significant.

---

[9] Specifically, we used a tree-based boosting algorithm and constrained the relationship between interest rates and defaults to be monotone and increasing. We use the gbm package in R for our analysis.

Next, we examine the corresponding impact on returns to investors in these loans. Our returns metric for a loan is the profit (or loss) for an investor in that loan, as a fraction of the loan amount. Since these are three-year loans, we divide the profitability ratio by three to approximate annualized returns. In addition, as this metric is only meaningful for matured loans, we only use loans from the matured cohorts, i.e., 2011-2013 cohorts. As seen in Figure 4b, the "good" loans significantly outperform the "bad" loans. For example, for the high risk loans, the returns on the "good" loans are 125 bps higher than for the "bad" loans. Given that the "bad" loans have a return of only about 400 bps, this differential reflects an improvement in excess of 30%.

VI. Conclusions

In this paper, we show that incorporating loan and borrower characteristics as well as macroeconomic factors can improve the prediction power of proprietary algorithms. Specifically, we show that changes in local macroeconomic conditions and job stability predict default rates above and beyond what proprietary algorithms predict. Our research is still ongoing; we are also examining whether additional factors that can improve prediction power of proprietary algorithms.

Appendix A. Variable Definitions

Debt-to-Income: A ratio calculated using the borrower's total monthly debt payments on the total debt obligations, excluding mortgage and the requested LC loan, divided by the borrower's self-reported monthly income.

Default: An indicator variable that is equal to one if the "loan_status " field in the LendingClub data contains the phrase "charged off" and zero otherwise.
Employment Duration: Employment length in years. Possible values vary between zero and ten. Zero means an employment less than a year and ten means ten or more years of employment.
Interest Rate: Interest Rate on the loan.
Home Ownership: The home ownership status provided by the borrower during registration or obtained from the credit report.
Job Stability: An indicator variable that is equal to one if the "emp_title" field in the LendingClub data includes one of the following terms: "^US ","US+Attorney", "CDC", "United States", "USAF", "air force", "police", "fire dep*","air force","army", "^department of ", "USCG", "USPS", "US Treasury", "Government", "city of ", "State of ", "County ", "*department of ", "town of","township", "dept? of ", "federal ","prison", "corrections" ,"sherriff*", "school","of education", "commonwealth of", and zero otherwise.
Loan Amount: The listed amount of the loan applied for by the borrower.
Number of Loan Inquiries within the last 6 months: The number of inquiries in past 6 months. It excludes auto and mortgage inquiries.
Unemployment: Annual county level unemployment rate for the year of loan origination. When county level data is unavailable, we use state level data for the year of loan origination.
Verification Status: indicates either LendingClub verified the income or income source is verified or income is not verified.

References

Demyanyk, Y. and O. Van Hemert (2011), "Understanding the Subprime Mortgage Crisis," *The Review of Financial Studies* 24(6), pp. 1848–1860.

Duarte J., S. Siegel and L. Young (2012), "Trust and Credit: The role of Appearance in Peer-to-peer Lending," *The Review of Financial Studies* 25(8), pp. 2455–2484.

Elul, R., N. S. Souleles, S. Chomsisengphet, D. Glennon, and R. Hunt (2010), "What "Triggers" Mortgage Default?," *American Economic Review*, 100(2), pp. 490-94.

Gross, J. P.K.. O. Cekic, D. Hossler, and N. Hillman (2010), "What Matters in Student Loan Default: A Review of the Research Literature," *Journal of Student Financial Aid* 39(1), Article 2.

R. Iyer, A. I. Khwaja, E. F. P. Luttmer, and S. Kelly (2016) "Screening Peers Softly: Inferring the Quality of Small Borrowers," *Management Science* 62(6), pp. 1554-1577.

Li, Z. and X. Yao, Q. Wen, and W. Yang (2016), " Prepayment and Default of Consumer Loans in Online Lending," Working Paper.

Lin, M., N. Prabhala, and S. Viswanathan (2012),"Judging Borrowers by the Company They Keep: Friendship Networks and Information Asymmetry in Online Peer-to-Peer Lending," *Management Science* 59(1), pp.17 -35.

Serrano-Cinca C, B. Gutiérrez-Nieto, L. López-Palacios L (2015), "Determinants of Default in P2P Lending," *PLoS ONE* 10(10).

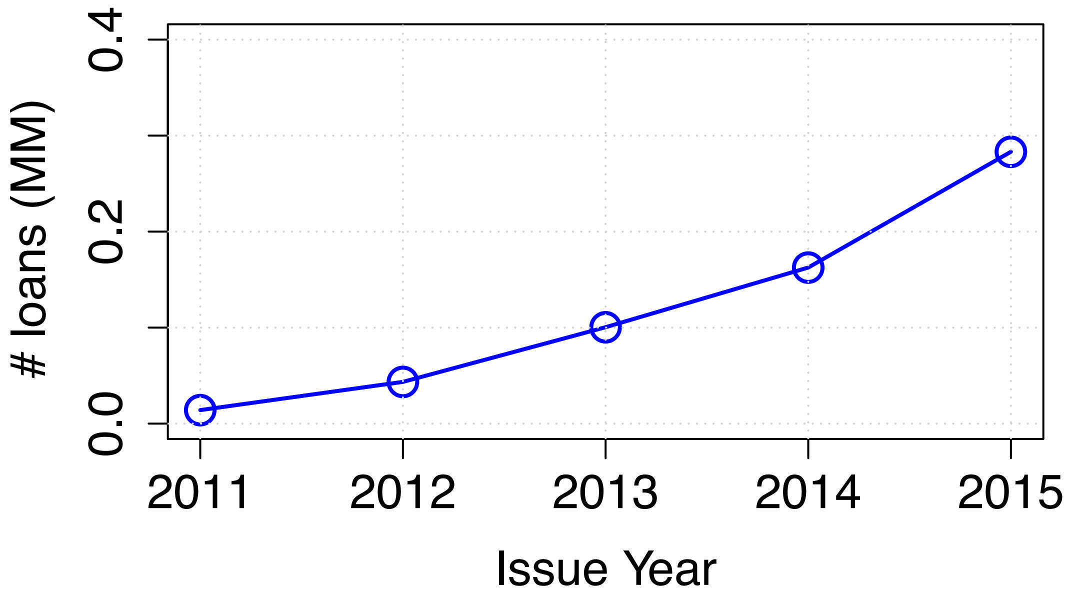

Figure 1.a. Total number of loans issued between 2011 and 2015.

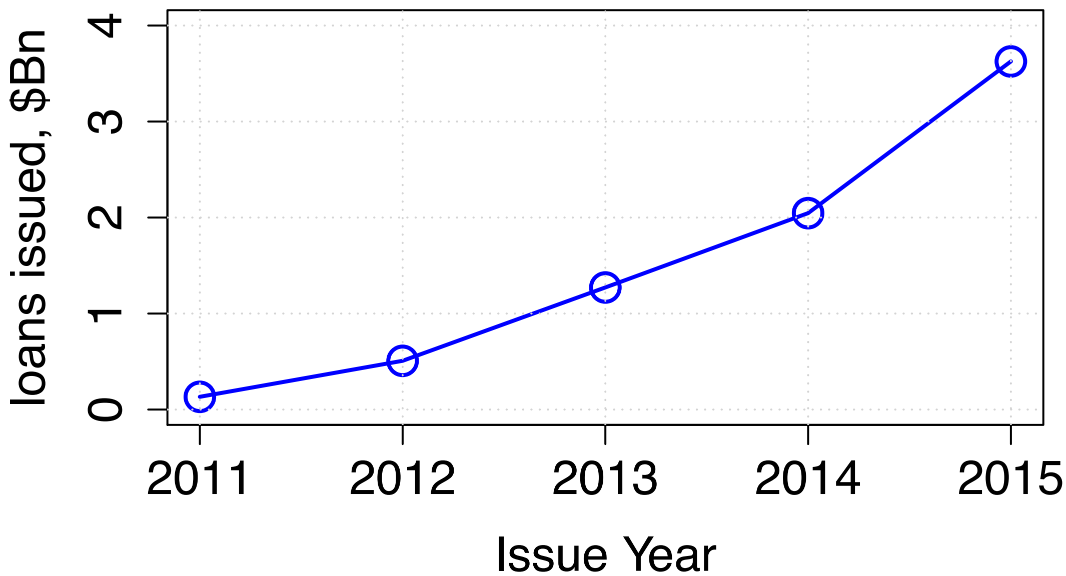

Figure 1.b. Total dollar value of loans originated between 2011 and 2015.

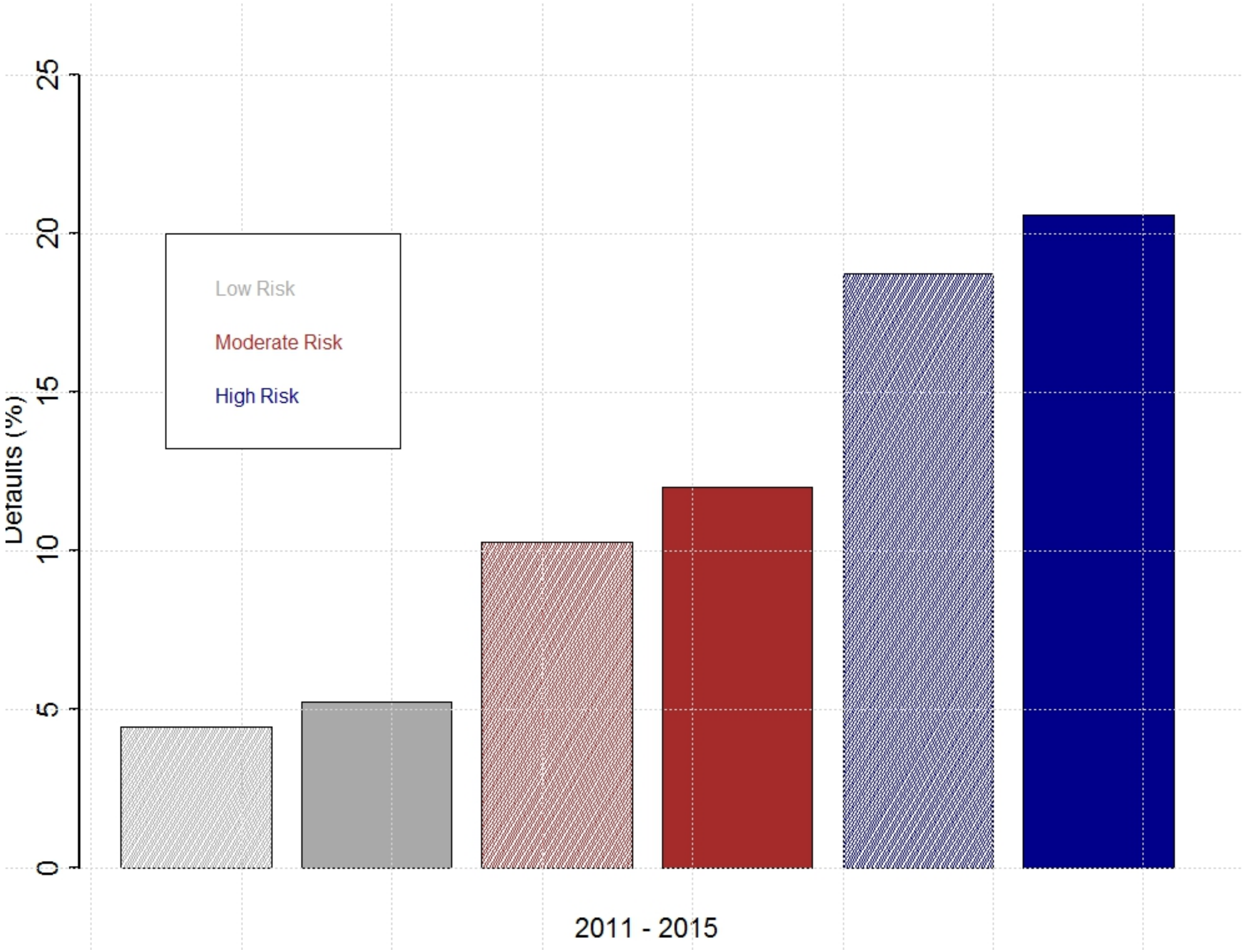

Figure 2.a. The average percentage of loan defaults in low and high-unemployment areas by three risk categories. Shadowed bars show default rate in low-unemployment areas and solid bars present default rates in high-unemployment area.

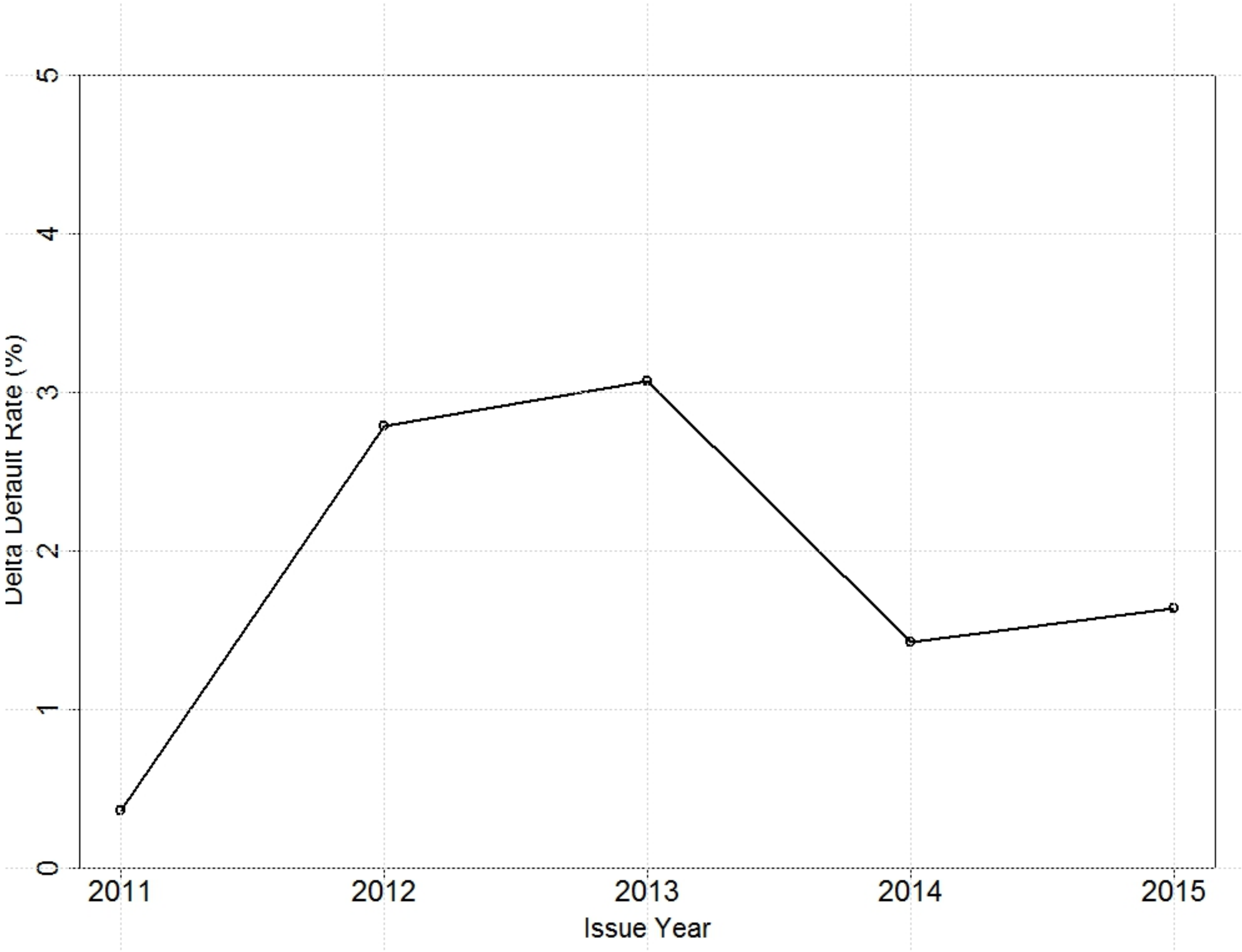

Figure 2.b. The difference in loan default rate between high and low-unemployment areas for high risk loans.

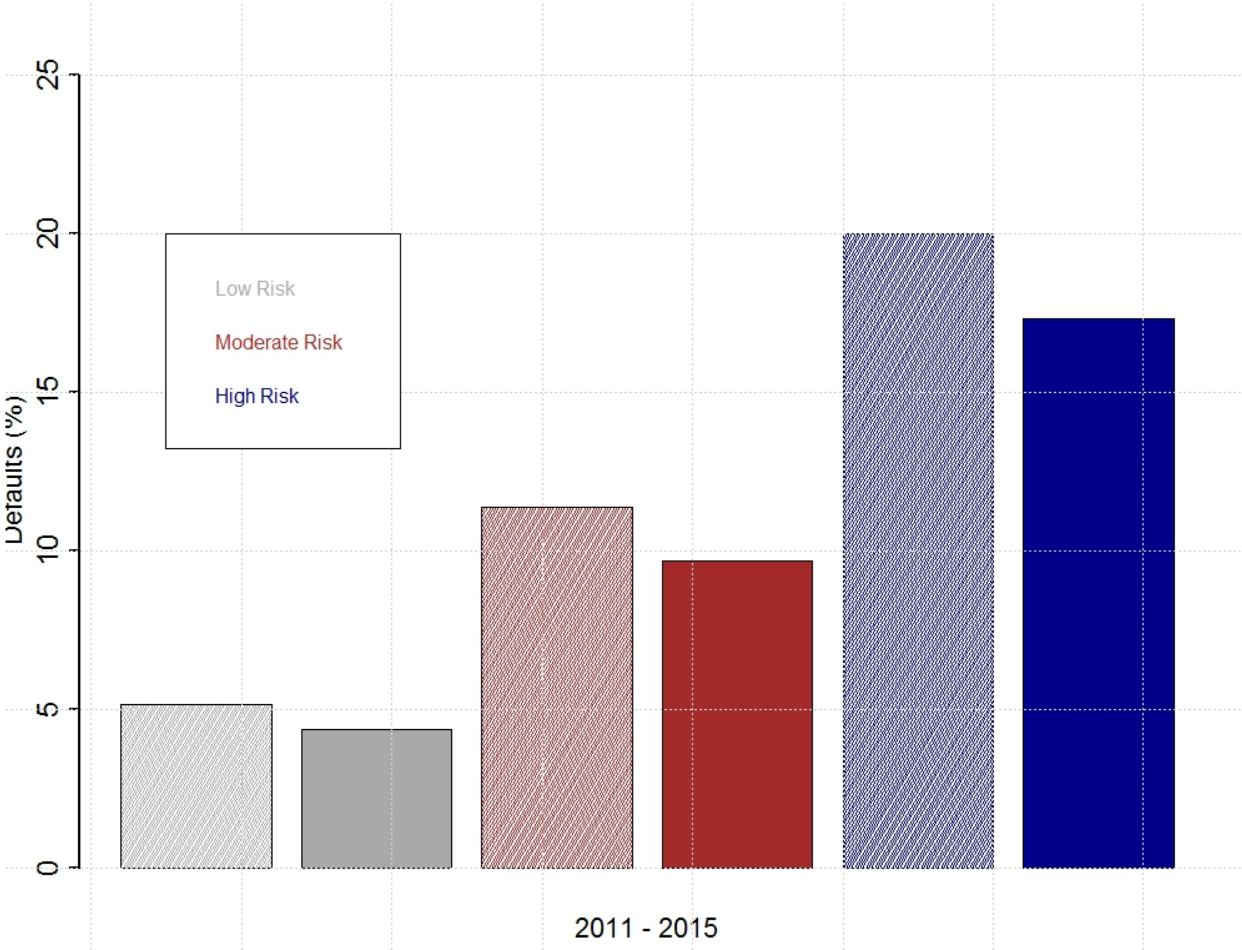

Figure 3.a. The average percentage of loan defaults by government-employment and risk category. Shadowed bars show default rate in loans obtained by government employees while solid bars show default rate in loans obtained by non-government employees.

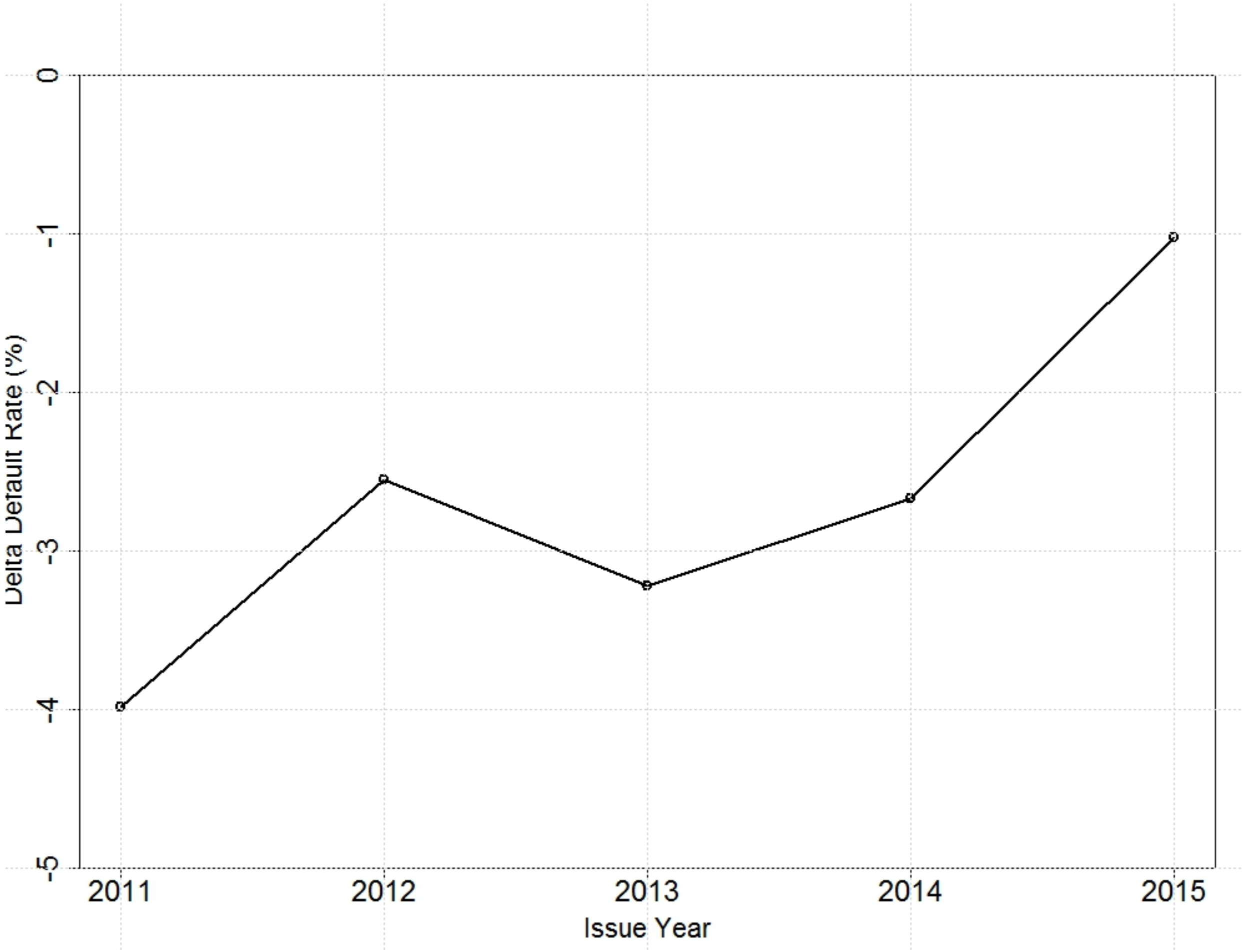

Figure 3.b. The difference in loan default rate between non-government and government employee borrowers for high risk loans.

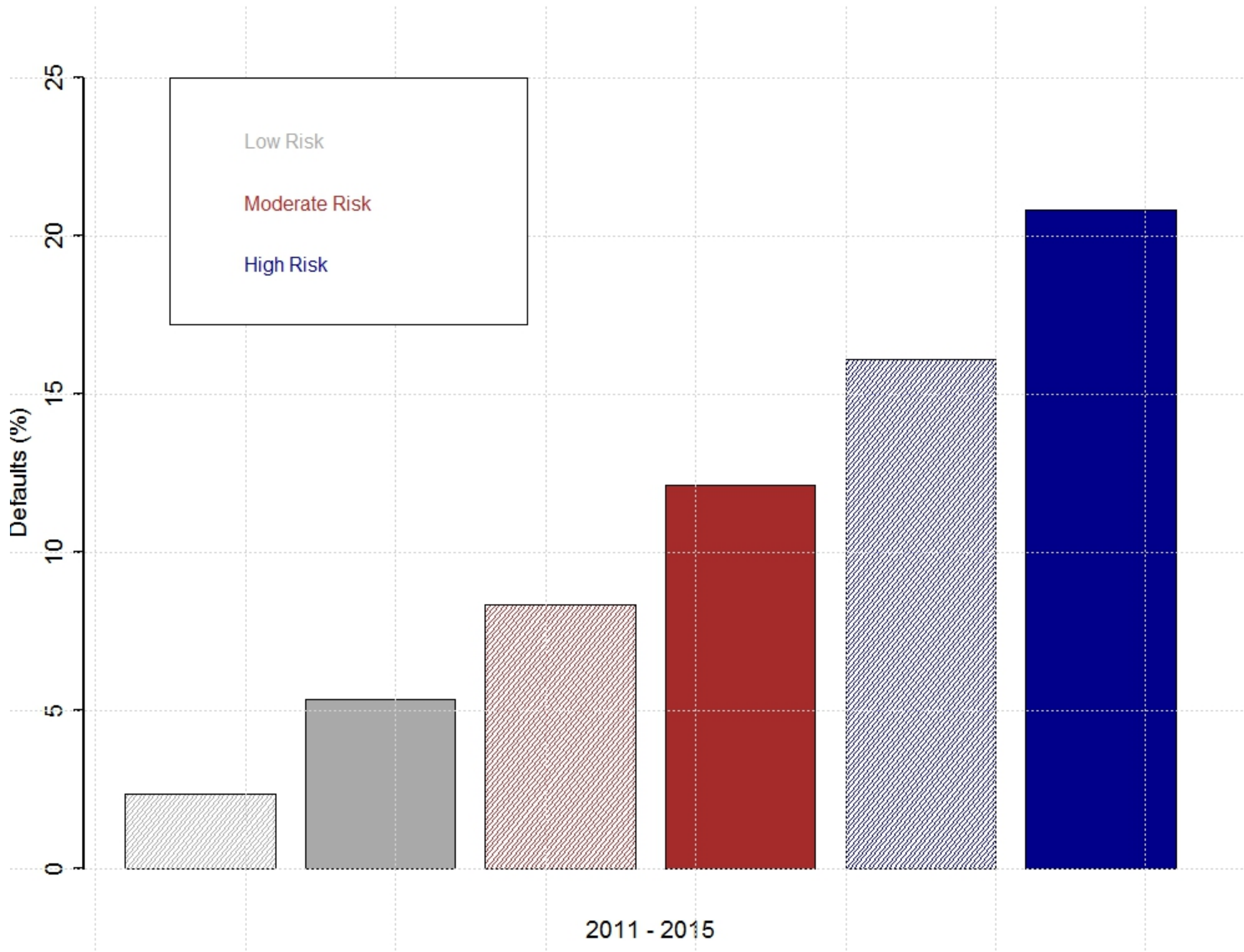

Figure 4.a. The average percentage of loan defaults in "good" and "bad" loans. Shadowed bars show default rate in "good" loans and solid bars present default rate in "bad" loans.

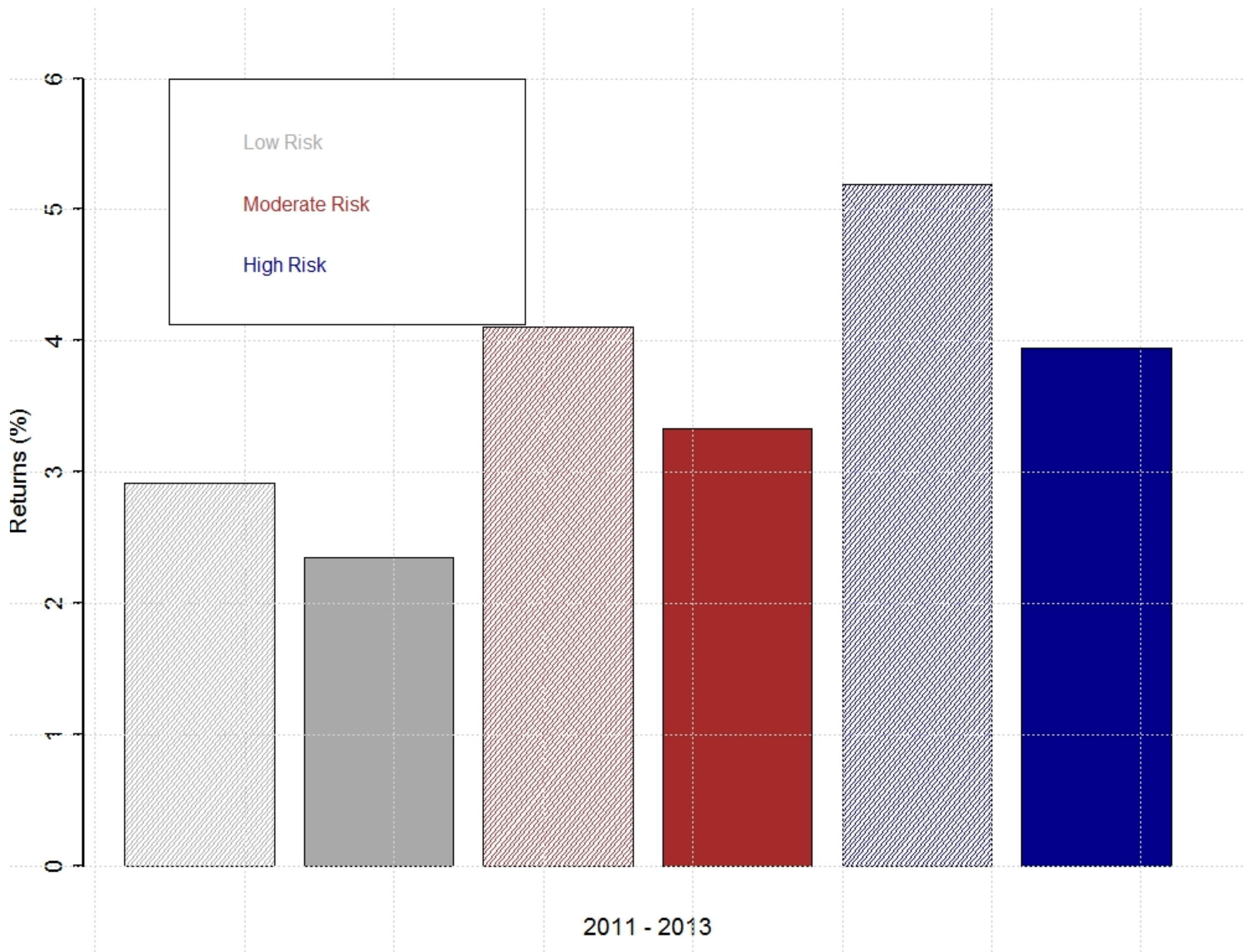

Figure 4.b. Annualized returns on “good” and “bad loans”. Shadowed bars show returns on “good” loans and solid bars show returns on “bad” loans.

Table 1. Summary Statistics

Table 1 presents summary statistics for the variables used in the empirical analysis. Information on loans are obtained from LendingClub.com and unemployment data is obtained from Bureau of Labor Statistics. Variable definitions appear in Appendix A.

Summary Statistics	N	Mean	Min	1st Quartile	Median	3rd Quartile	Max	Standard Deviation
Defaut (%)	603701	10.02	--	--	--	--		--
Interest Rate (%)	603701	12.03	5.32	8.90	11.99	14.47	28.99	3.87
Loan Amount (in $)	603701	12,563.95	1,000.00	6,850.00	10,000.00	16,100.00	35,000.00	7,806.73
Debt-to-Income (%)	603701	17.69	0.00	11.46	17.16	23.47	48.48	8.24
Unemployment Rate at loan issuance	603701	1.01	0.33	0.85	0.98	1.16	3.41	0.25
Dummy: Job Stability (%)	603701	3.21	0.00	--	--	--	--	--
Number of Loan inquiries within the last 6 months	603701	0.68	0.00	0.00	0.00	1.00	8.00	0.96
Issue Year	603701	--	2011	2013	2014	2015	2015	--
Employment Duration (%)								
Less than a year	49,284	8.16%		--	--	--		--
1 year	40,194	6.66%		--	--	--		--
2 years	54,905	9.09%		--	--	--		--
3 years	48,809	8.08%		--	--	--		--
5 years	38,355	6.35%		--	--	--		--
More than 10 years	187,637	31.08%		--	--	--		--
Other ()	184,517	30.56%		--	--	--		--
Home Ownership								
Mortgage	281,340	46.60%		--	--	--		--
Own	61,647	10.21%		--	--	--		--
Rent	260,637	43.17%		--	--	--		--
Verification Status								
Source Verified	210,106	34.80%		--	--	--		--
Verified	181,548	30.07%		--	--	--		--
Not Verified	212,047	35.12%		--	--	--		--

Table 2. Regression Results

Table 2 presents regressions that examine the likelihood of loan default. Information on loans are obtained from LendingClub.com and unemployment data is obtained from Bureau of Labor Statistics. For the proportional Cox hazard estimation, we report coefficients and p-values. The period of analysis is from 2011 to 2015. We use ***,**, and * to denote that the coefficient estimate is different from zero at the 1%, 5%, and 10% levels (two-tailed), respectively. Model 1 is the baseline case that includes interest rate and year fixed effects. Model 2 includes unemployment and job security dummy. Model 3 includes DTI, dummies for number of credt inquiries in the last six month, dummies for home ownership, and dummies for employment duration. Variable definitions appear in Appendix A.

	Dependent Variable: Loan Default		
	Estimation Method: Cox Proportional Hazard Model		
Variables	Model 1	Model 2	Model 3
Interest Rate	0.142 ***	0.142 ***	0.127 ***
	0.001	0.001	0.001
Unemployment		0.141 ***	0.133 ***
		0.016	0.016
Dummy: Job Stability		-0.174 ***	-0.143 ***
		0.024	0.024
DTI			0.015 ***
			0.000
Year Fixed Effects	Yes	Yes	Yes
Controls for Credit Inqiuries in last 6 months	No	No	Yes
Controls for Home Ownership	No	No	Yes
Controls for Employment Duration	No	No	Yes
Number of Observations	603,701	603,701	603,701
Pseudo R-Square	0.031	0.032	0.036