People Helping (White) People:

Evidence of Racial Discrimination in Peer-to-Peer Lending

Robert Cecot

Presented to the Department of Economics

in partial fulfillment of the requirements

for a Bachelor of Arts degree with honors

Harvard College

Cambridge, Massachusetts

March 20, 2008

Abstract

Previous studies have found discrimination against minorities in the market for home mortgage loans. This paper contributes to the literature by examining over 55,000 hand-coded images of unsecured loan applications in an online lending marketplace to measure the effect of racial discrimination on lenders' funding decisions. This analysis takes advantage of special properties of the platform as well as full access to available information. Using econometric techniques and a robust dataset, the reported results support a finding of significant racial discrimination. Blacks and Hispanics are less likely to be funded when requesting loans compared to equivalently qualified whites, a result that echoes the previous literature on mortgage lending. Furthermore, Blacks are subjected to an additional interest rate premium compared to other borrowers. An analysis of individual loan performance reveals the race of Black borrowers as a useful predictor of subsequent default rates. The relationship between these results is analyzed using a model of loan pricing as a function of default rate. This analysis suggests the level of discrimination measured against black borrowers is too low to be economically efficient. Theories that reconcile the differences in observed results for Hispanics and Blacks, the implications of these theories, and caveats of the study are presented.

Table of Contents

1 Introduction

Defining discrimination is one of the most disagreed upon issues between lawmakers and economists. Although legally defined as the treatment of an individual based on his or her membership in a protected class, economists make a finer distinction. For a policy to be discriminatory, it must have an unsubstantiated, and therefore unjustified, difference in treatment for a group member. The latter type of discrimination is particularly interesting as it often leads to economic inefficiency. This paper evaluates both types of racial discrimination in the online lending marketplace created by Prosper.com. While discrimination is found for both Blacks and Hispanics, some discrimination appears to be justified by estimates of loan default rates. Specifically, discrimination against Blacks can be explained by an increased risk of default, but discrimination against Hispanics cannot.

Accurate measures of racial discrimination in the marketplace have eluded researchers for many years. Traditional experimental methods usually involve matched pair audit studies that attempt to equate all observable characteristics of paired individuals from two different races. These pairs are then sent to attempt some task, and any observed heterogeneity in the treatment of one confederate over another is ascribed to racial discrimination. In this way, racial discrimination has typically been measured in labor markets (Firth, 1981; Kenney and Wissoker, 1994), product markets (Ayres and Siegelman, 1995), and credit markets (Yinger, 1986). These studies have been met with significant criticism from academics, calling into question their premises (Heckman and Siegelman, 1993). More creative experimental methodologies have had better success in avoiding these conflicts in labor markets (Bertrand and Mullainathan, 2003) but have not been generalized to other contexts. This paper focuses on the other established method for detecting discrimination: robust econometric analysis.

Econometric techniques have been commonly utilized to test for discrimination in labor markets (Blinder, 1973; Oaxaca, 1973) but have received skepticism for their inability to provide reliable results (Gunderson, 1989). Critics point to the difficulties related to controlling for a broad set of characteristics in regressions that may still be important factors in an applicant's perceived productivity. Application of these methods has had better success in measuring discrimination in credit markets (Munnell, et al., 1996), but has received significant criticism along similar grounds (Horne, 1997).

This paper resolves many of the outstanding issues regarding the use of econometric techniques to measure racial discrimination by focusing on a new credit market pioneered by the online lending company Prosper. Although the use of online applications in testing for discrimination has typically been to restrict the observation of race and utilize difference measurements (Morton et al., 2003), Prosper encourages borrowers to include images of themselves when requesting loans. These images, along with detailed information on credit background, are viewed by multiple lenders who individually decide whether to fund part of the loan request. The self-contained nature of a web application, along with the public availability of Prosper's data, allows an econometric model to control for exactly all the characteristics that are visible to lenders, while its inclusion of images can shed light on the effects of racial discrimination.

Prosper and similar sites heavily tout the benefits of social lending with slogans such as, "Someone out there believes in you."[1] They do not consider the potential downside of revealing personal characteristics, as would be the case for a member of a group that is discriminated against. I believe it is important to analyze the potential pitfalls of this model. Furthermore, the Equal Credit Opportunity Act clearly states that it "shall be unlawful for any creditor to discriminate against any applicant, with respect to any aspect of a credit transaction ... on the basis of race, color, religion, national origin, sex or marital status, or age…"[2] Since Prosper is officially the originator of every loan, and then sells the claim on payments to the online "lender," the legal ramifications of this analysis are considerable.

In this study, I utilize Prosper's full list of over 17,000 funded loans and an equivalently sized random selection of denied applications as the basis for my sample. I then observe over 55,000 images to code each borrower for race and gender. Using this information, I answer three clear questions:

(1) Do lenders discriminate based on observed race when choosing to fund loans?

(2) For funded loans, does the Prosper lending market's interest rate outcome depend on the race of the borrower?

(3) Do differences in default risk exist for borrowers of different races?

I also evaluate whether any measured discrimination is economically efficient, and interpret my results under a consistent model of discriminatory lending.

---

[1] Prosper.com homepage slogan, 03/19/2008: http://www.prosper.com
[2] 15 U.S.C. § 1691(a)(1)

In line with previous research, I find that racial discrimination exists in the funding stage of a loan. Furthermore, I find the result to be significantly greater than that measured in Munnell's research on mortgage lending, suggesting that regulated institutions may be better at mitigating the presence of discrimination than individual lenders. In addition, I am able to isolate discrimination's effect on each minority group, a result not possible with prior mortgage lending studies because of their small sample size, finding additional denial rates of approximately 12 percentage points for Blacks and Hispanics with no apparent discrimination against Asians.

Interrelated with the decision to fund a loan, I find that Blacks are subject to an interest rate premium not detected for any other minority group. Having been limited in their ability to be funded, Blacks face another layer of discrimination when their loans are priced higher than those for equivalent loan candidates of other races.

I then present estimates of marginal default likelihood for all races, determining that Blacks appear to exhibit a significantly greater default risk than other races, controlling for all observable data. I test whether this difference in risk is absorbed by racial discrimination by estimating default likelihood while controlling for interest rate, and determine that a higher default rate for Blacks remains. This suggests that the market may not be statistically discriminating enough for Blacks, while the taste-based discrimination in funding for Hispanics appears economically efficient.

Finally, I present plausible arguments to explain the existence of the aforementioned effects. I extend Becker's (1971) fundamental model of discrimination originally published in 1957 to hypothesize that there may be two types of discrimination at work simultaneously: taste-based and statistical. The interaction between these two types yields the different effects for Blacks and Hispanics. Additionally, I suggest general market failure, loan tenure, and race-related shocks as potential explanations for the default measures associated with Blacks.

This study proceeds as follows: Section II includes a review of the literature concerning discrimination in mortgage lending, the most analogous work to my current analysis. Section III gives background information on Prosper's online lending platform and the nature of its funding decisions. Section IV describes the data collection process and provides a summary of the data and inter-group differences in composition. Section V presents the methodology used and the results of the study. Section VI discusses theories that explain the results and avenues of investigation into their verity. Section VII presents caveats of my analysis, and Section VIII concludes.

2 Literature Review

Substantive econometric research on racial discrimination in the decision to grant credit began with the Federal Reserve Bank of Boston's review of mortgage lending (Munnell et al., 1996). Previous literature had attempted to draw conclusions from data made available by the Home Mortgage Disclosure Act of 1975 (King, 1980; Schafer and Ladd, 1981). The results, although conveying a considerable amount of discrimination, were not conclusive because the data lacked key variables regarding an applicant's credit history. Researchers at the Federal Reserve Bank of Boston worked directly with lenders to gather enough data regarding applicant background for a sophisticated econometric analysis. They surveyed lenders to deduce the information that played a role in making mortgage decisions, and collected an additional set of 38 variables. Their final sample consisted of approximately 3000 applications, only 700 of which were from Blacks or Hispanics.

Munnell et al. ran a series of logit regressions using the binary funding decision as their dependent variable and race as their variable of interest. They included controls for default risk, loan, and personal characteristics. Although an uncontrolled regression found denial rates of 28% for minorities and 10% for Whites, the fully controlled specification found an 8% difference in approval between the two groups. My work applies this analysis to lending on Prosper, but extends it two levels further to analyze effects on interest rates and defaults, and adopts a more controlled framework to address the two types of criticisms attributed to the Federal Reserve Bank of Boston's results.

The first type of criticism regards the actual findings of discrimination in funding. Horne (1997) argues that attempts to replicate the Fed's results with slightly different model specifications yields insignificant results. He argues that due to the inability of Munnell et al. to understand exactly which variables were used by lenders in making their decisions, the method of researching 'relevant' data and including it in a regression is flawed and subject to bias in the choice of controls. My research nullifies this argument due to the defined nature of Prosper lending decisions: lenders observe only a predetermined list of variables, all of which are included in my analysis. If a competent race-blind lender makes a lending decision using only the controls that are fully incorporated into the regression, the estimated coefficients on race dummy variables should be zero.

The second type of criticism reflects a deeper concern regarding the definition of discrimination. Legally, it appears that any use of race in making a credit decision, even in a purely profit-maximizing fashion as a determinant of a statistical model, is prohibited. However, economists differentiate between the use of race as an effective signal of an unobservable characteristic (statistical discrimination) and its use purely as a result of personal prejudice (taste discrimination). While the first type of discrimination is profit-maximizing, Becker (1971) models the second as a cost the discriminator bears to satisfy a personal distaste for members of one group. Applying this model to the market for mortgage lending, critics of the Federal Reserve Bank of Boston's study argue that if minorities are systematically discriminated against, only especially qualified candidates would be funded for loans, and, all else equal, default rates would be lower for discriminated groups. Because the authors of these studies find minorities to have higher default rates than Whites, they claim that the findings of discrimination at the funding stage are inconsistent with loan performance (Van Order et al., 1993; Berkovec et al., 1994, 1996a,b). Researchers have been unable to reconcile these two findings without additional insight into the interaction between funding, loan pricing, and default. The simultaneous nature of these outcomes makes previous analyses of the economic efficiency of discrimination suspect (Yezer, Phillips, and Trost, 1994).

I build upon the previous literature by utilizing Prosper's unique platform as well as a larger collection of samples for each minority group. With this larger sample, any variation in discriminatory behavior at the three stages of inference can be exploited to generalize a common theory that reconciles this seemingly paradoxical evidence on discrimination in funding and differences in default rates.

3 Prosper Background

Prosper was launched in 2006 with a simple mission: to disintermediate the market for consumer lending, thereby increasing economic efficiency, and taking a part of the pie in the process. The value proposition for lenders is simple: instead of risking your money in the market or depositing it into a bank, invest in people directly. According to its website, Prosper also seeks to make lending more "socially rewarding", a reference to the ability for lenders to choose who they fund and see the effect their dollar has. For borrowers, Prosper offers the opportunity to pay lower interest rates directly to lenders rather than to institutions, another "social" process, and ensure they get the lowest interest rate through a reverse auction system.

In its short two year lifespan, Prosper has experienced tremendous growth. It currently records over 217,000 members, some of which may be borrowers, lenders, or both. Historical listings are in excess of 216,000 with 17,402 being funded for a total loan portfolio of over $110,800,000 as of February 1st, 2008. Prosper was the first peer-to-peer lending site in the United States, although the following exist worldwide: IOU Central and CommunityLend in Canada, Fair Rates in Denmark, and Zopa in the UK. A similar service is offered in the US by LendingClub, although Prosper is the dominant player. All of these sites emphasize the social and helping component of their work, as seen in their various slogans: "Better Rates. Together.", "People lending to people.", "The best rates. The nicest people."[3]

[3] An interesting non-profit variant of this model has been developed by Kiva.org that offers interest-free microloans (6-12 months) to entrepreneurs in the developing world in exchange for updates on their work. To this date, Kiva claims a repayment rate of 99.9% (source: http://www.kiva.org/about/risk/overview)

Visitors arriving at Prosper's homepage (Appendix A, Figure 1) are shown a preselected group of borrower photos as well as summary statistics regarding interest rates and returns. A borrower can register for an account and, after verifying their identity and authorizing a query of their credit report, can post a listing for a loan. Loans are restricted to be three years in length. The maximum amount requested depends on the borrower's individual state's consumer protection laws. Borrowers are encouraged to start their listing at the highest interest rate they are able to afford and within the limits of their state laws, while Prosper provides guidance as to the likelihood of being funded at that rate given the borrower's credit profile. Borrowers then are given the opportunity to give their listing a personal touch by uploading one member photo that is associated with their profile, which would be common to all listings the borrower creates, and four images associated with their listing. Borrowers commonly include pictures of themselves, their family, and their pets. These images are captioned, and a title is chosen by the borrower for his or her listing. Borrowers are also able to create a listing description, for which Prosper provides a template.

Lenders are able to sign up and begin funding loans as soon as they verify their bank account with Prosper. At that point, lenders can look through listings by specifying search criteria and browsing the results; a sample search result screen is included as Figure 2 in Appendix A.[4] If they decide to investigate a listing further, they are shown a

[4] Since this dataset was retrieved, Prosper has allowed lenders to create their own portfolio plans, in addition to the four plans provided by the company, that will automatically invest in multiple listings at low amounts. This is intended to diversify a lender's portfolio. Previously, lenders were able to create "standing orders" that would automatically invest in borrowers that met a certain criteria. Prosper was unwilling to provide data for the utilization of these orders, but robustness checks for effects caused by the introduction of these plans produced no marginal difference in results.

page listing details about the borrower and his or her credit history and loan terms (Figure 3). If a loan is unfunded, a lender can make an offer to lend money. For borrowers that opt out of the auction process, once a listing is fully funded it is automatically closed. A majority of borrowers, however, opt to allow the listing to run its full duration. At this point, lenders can “outbid” one another by to accepting slightly lower rates. Prosper uses a system of proxy bidding, allowing a lender to set his or her lowest acceptable rate. Once a listing closes, the highest winning rate becomes the interest rate of the loan.

Loans are paid via direct withdrawal from a borrower’s bank account. Each missed payment is afforded a 15 day grace period, after which a loan is classified as late and turned over to a collection agency for their assistance. Once a loan is 4 months late, it is classified as uncollectable and prepared for sale to a debt buyer.[5] Lenders are entitled to the proceeds of this sale.

4 Data

Prosper’s management team believes strongly in the value of transparency and easy access to information.[6] In that spirit, Prosper makes available for download a full dataset of all historical data that was viewable on its website to registered members. According to their website,[7] the data is intended to aid academic research as well as to empower lenders to do their own analysis of market conditions. The availability of information has since been upgraded to include an API, Application Program Interface,

---

[5] Prosper’s dataset makes a distinction between loans that are “4+ months late” and “defaulted” as the former has not yet been included in a debt sale. Because debt sales are a manually negotiated process where Prosper sells hundreds of delinquent loans at a time, and an irregular time, the two conditions are treated as equivalent signals of default in the remainder of this analysis.

[6] http://www.prosper.com/about/academics.aspx

[7] http://www.prosper.com/tools/

which allows applications to interact with Prosper's database in a live ad-hoc fashion. Potential implications of this data availability are discussed in the concluding section.

4.1 Prosper's Dataset

Prosper's full dataset is updated daily; the version used in this paper was downloaded on February 1, 2008. It contains information for 216,390 listings, 217,324 members, and 17,402 funded loans. I utilize all funded loans and randomly select an equal number of unfunded listings to comprise the full sample of 34,804 listings discussed in the remainder of this paper.[8] Appendix B contains descriptions for each of thirty-three variables utilized from Prosper's dataset.

Table 1 describes the variables that are common to all 34,804 listings, the first seven of which are collectively referred to as "Listing Controls" in this paper. These include the amount of money a borrower has requested and six binary variables denoting group membership, bank account verity, homeowner status, endorsement by other members, whether the listing becomes a reverse auction upon funding or just closes, and whether it is a borrower's first listing on Prosper. The remaining four variables common to all listings include a dummy representing the outcome of the funding decision, a sequential public identifier for all listings,[9] the initial interest rate that a borrower chooses when making his or her request, and the borrower's state of residence.[10]

---

[8] Each listing is assigned a random hexadecimal key in additional to a sequential id number when it is generated. I used the random key to select the first 17,402 unfunded listings. Correlation analysis confirmed the observations were randomly selected across all time periods.

[9] This is later used as a useful control for time, more specifically for the growth of loan volume on Prosper which can cause changes in funding decisions due to increased competitiveness on both sides of the market.

[10] Prosper loans are subject to the usury laws of each individual state and, because minority concentration varies by state, state fixed effects are included in most regressions. For more information please see http://www.prosper.com/legal/states_and_licenses.aspx

On April 18, 2006[11] Prosper began including additional information regarding a borrower's credit history. Table 2 describes these seven variables, collectively referred to as "Core Credit Controls", which represent a subsample of 33,884 listings (97.4%). These basic data points include a borrower's debt to income ratio, their credit grade among AA, A, B, C, D, E, and HR, as well as the number of current delinquencies, historical delinquencies across seven years, public records for the past ten years, total credit lines, and credit inquiries within six months of the report being generated.[12]

Prosper once again increased the amount of borrower information they provided on February 12, 2007.[13] These are described in Table 3, the first six of which are referred to as "Extended Credit Controls" and next four as "Employment Controls", which represent a subsample of 22,227 listings (63.9%). The additional credit information available to lenders includes the total amount of delinquent debt, the number of public records in the last year, current credit lines, revolving credit balance, and bank card utilization.[14] The employment information collected includes the borrower's current employment status, the length he or she has maintained aforementioned status, the income quintile the borrower inhabits, and the borrower's selection of occupation from a list of sixty-seven predefined options.

Lastly, once a listing is funded and becomes a loan, it is associated with an additional set of variables. These include the age of the loan in months, the interest rates paid by the borrower and received by the lender, and the current status of the loan.

---

[11] http://www.prosper.com/help/topics/whats_new.aspx
[12] Reports are generated whenever a listing is created.
[13] http://www.prosper.com/help/topics/whats_new.aspx
[14] Amount of revolving credit that is currently being used by the borrower.

4.2 Coding Race and Gender

An additional data point that Prosper includes in their export is a list of URLs that reference the set of images a borrower included with his or her listing. Each borrower is allowed to include one "Member" image as well as four "Listing" images. Borrowers include a variety of different images: pictures of themselves, family members, pets, cars, homes, cartoons, etc. Including borrowers that decided not to show any picture at all, a listing was associated with 1.6 images on average. I created a web-based applet that presented the full set of images associated with a listing as well as their respective captions and viewed approximately 55,000 images to determine the race and gender of each borrower. Each listing was coded in the following way: If the listing had no images, or only included one of five randomly selected default member images, it was marked *Control* = 1. If, based on the image-caption pairs, I was not reasonably sure of the race and gender of the individual borrower, the listing was marked *Unknown* = 1. This was typically used with images of family pets, inanimate objects, and when race and gender were not clearly observable. Finally, assuming neither of the other two cases, an image was successfully coded for both race (*White* | *Black* | *Hispanic* | *Asian* = 1) and gender (*Male* | *Female* = 1).

4.3 Summary Statistics

Summary statistics for the full sample of 34,804 listings are listed in Figure 4.1.

Figure 4.1: Summary Statistics

Averages (*binary)	**Control**	**Unknown**	**Identified**	**White**	**Black**	**Hispanic**	**Asian**
Num. Observations	13109	11517	10178	7068	2047	654	409
Male*				0.594	0.431	0.491	0.577
Funded*	0.364	0.578	0.586	0.633	0.444	0.456	0.709
Opening Rate	0.180	0.193	0.196	0.194	0.202	0.197	0.200
Amount	6997.47	7033.67	6534.87	6698.00	5865.13	6430.91	7234.04
Verified Bank Account*	0.491	0.729	0.723	0.758	0.626	0.609	0.775
Home Owner*	0.372	0.403	0.306	0.326	0.276	0.242	0.225
Endorsed*	0.041	0.123	0.156	0.165	0.131	0.116	0.193
Close Immediately*	0.426	0.347	0.324	0.304	0.382	0.381	0.271
Debt to Income Ratio	0.532	0.449	0.461	0.464	0.474	0.397	0.463
Current Delinquencies	3.805	3.052	3.037	2.662	4.428	3.324	2.151
Delinquencies 7 Yrs	11.381	8.956	9.233	8.605	11.929	8.764	7.473
Public Records 10 Yrs	0.691	0.598	0.543	0.543	0.648	0.390	0.262
Total Credit Lines	25.137	23.969	23.391	23.949	22.180	21.903	22.156
Inquiries 6 Mos	4.154	3.858	3.751	3.421	4.473	4.637	4.448
Amount Delinquent	3512.94	2541.31	2600.00	2253.89	3975.23	2345.99	2848.44
Revolving Credit Bal	11537.69	13392.69	10846.67	11781.05	6936.11	10658.97	12232.92
Bankcard Utilization	0.610	0.582	0.593	0.588	0.595	0.632	0.621
Length of Status	48.7	43.5	37.6	36.4	45.8	31.4	29.1
Credit Rating	**Control**	**Unknown**	**Identified**	**White**	**Black**	**Hispanic**	**Asian**
AA	4.80%	8.31%	6.40%	7.90%	1.79%	2.80%	5.66%
A	5.71%	8.00%	7.07%	8.06%	4.09%	3.92%	7.55%
B	7.69%	10.96%	10.21%	10.82%	6.73%	8.40%	17.36%
C	13.54%	15.31%	15.14%	15.89%	12.18%	16.25%	13.58%
D	16.20%	17.37%	18.58%	18.85%	17.29%	19.89%	17.74%
E	16.15%	13.60%	14.42%	13.93%	16.44%	14.29%	13.96%
HR	35.90%	26.44%	28.19%	24.54%	41.48%	34.45%	24.15%
NC	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Employment Status	**Control**	**Unknown**	**Identified**	**White**	**Black**	**Hispanic**	**Asian**
Full-time	86.47%	85.51%	85.38%	84.98%	86.37%	86.27%	86.79%
Self-employed	5.23%	8.21%	7.28%	7.79%	5.79%	5.88%	6.79%
Part-time	3.41%	3.44%	4.40%	4.22%	4.17%	6.16%	6.04%
Retired	3.29%	1.78%	1.88%	1.89%	2.47%	1.12%	0.00%
Not employed	1.60%	1.06%	1.06%	1.11%	1.19%	0.56%	0.38%
Income	**Control**	**Unknown**	**Identified**	**White**	**Black**	**Hispanic**	**Asian**
Not Employed	1.25%	0.79%	0.81%	0.85%	0.94%	0.28%	0.38%
Not Displayed	1.59%	1.54%	1.02%	1.02%	0.77%	1.96%	0.75%
$1-24,999	17.36%	15.96%	16.72%	15.81%	19.34%	21.57%	14.34%
$25,000-49,999	41.82%	40.19%	42.84%	41.89%	46.25%	44.26%	42.26%
$50,000-74,999	22.27%	23.07%	22.77%	23.17%	21.72%	20.17%	24.15%
$75,000-99,999	8.31%	9.88%	9.22%	9.54%	7.58%	8.40%	12.08%
$100,000+	7.40%	8.58%	6.62%	7.73%	3.41%	3.36%	6.04%

It is informative to compare the control sample of borrowers with no images at all against the sample of borrowers who definitively identify themselves, which is the restricted set analyzed in the remainder of this paper. Identified borrowers are funded more often (58.6% to 36.4%) although they tend to set higher starting interest rates for themselves (19.6% to 18.0%). They tend to invest more work into their Prosper listings, as seen in the number that have verified their bank accounts (72.3% to 49.1%) or received endorsements from other members (15.6% to 4.1%). These borrowers generally have a better credit history, with a lower debt to income ratio (46.1% to 53.2%), fewer historical delinquencies (9.23 to 11.38), a lesser delinquent balance ($2600 to $3513), and a lower Revolving Credit Bal ($10,847 to $11,538). They are more commonly in the AA or A credit grade (13.48% to 10.51%) and less commonly in the High Risk grade (28.19% to 35.90%). Finally, they tend to disallow their listings from operating as auctions less commonly[15] (32.4% to 42.6%), are more typically self-employed (7.28% to 5.23%) but have maintained the length of their employment status for a shorter duration (37.6 months to 48.7 months).

The restricted sample of 10,178 positively identified listings can now be analyzed. This sample is mostly composed of Whites (69.4%) and Blacks (20.1%) with fractional representation by Hispanics (6.4%) and Asians (4.0%). Given the similarities in statistics among Blacks and Hispanics, as well as Whites and Asians, and the proportional differences in representation among these two pairings, the remaining summary analysis

[15] Depending on the starting interest rate, listings that enter into reverse auction mode should be less appealing to lenders than those that close when funded—lenders are more likely to be able to take advantage of overpriced loans (from the borrowers perspective) when they do not need to compete with other lenders.

will compare Whites and Blacks, although Asians and Hispanics, respectively, could be virtually substituted for either. Whites are more commonly male (59.4% to 43.1%) and more likely to be funded (63.3% to 44.4%), although they request more money ($6698 to $5865) and are less likely to disallow their listings from being auctions (30.4% to 38.2%), but set a slightly lower starting interest rate (19.4% to 20.2%). They also tend to have a stronger presence on Prosper, more often having verified their bank account (75.8% to 62.6%) and having been endorsed by another member (16.5% to 13.1%). They generally have a better credit profile, with a slightly lower debt to income ratio (46.4% to 47.4%), less historical delinquencies (8.60 to 11.93), a smaller delinquent balance ($2254 to $3975), and less credit inquiries (3.42 to 4.47). A larger percentage of Whites are in the AA or A credit grades (15.96% to 5.88%) and significantly less have High Risk credit (24.54% to 41.48%). Although the employment levels are approximately equivalent, Whites more commonly have salaries in the top two quintiles (17.27% to 7.58%) but have maintained their employment status for less time (36.4 months to 45.8 months). Whites, however, tend to have a significantly higher Revolving Credit Bal ($11,781 to $6,936) and slightly more credit lines (23.9 to 22.2).

Appendix C contains additional summary statistics for all listings, as well as a new comparison of certain subsamples in Tables 3, 4, and 5. Table 3 compares data averages for funded and unfunded listings among the different races. Funded listings share many of the same inter-race differences as in the larger sample discussed above: Blacks have weaker credit histories across all categories than Whites and receive loans that are approximately $800 less on average. One point of interest regards the values of *Opening Rate*: although Whites and Blacks share the average starting rate for unfunded listing (18.2%), the starting rate for funded listings is lower for Whites (20.1%) than it is for Blacks (22.5%), leaving open the possibility that there may be some bias affecting the decision of who to fund. Table 4 presents summary statistics of credit and employment data for a parallel analysis. The composition of funded Black borrowers is generally riskier than funded White borrowers; for example, 19.2% of Black loan recipients were classified in the High Risk credit grade, compared to 11.61% of Whites. Employment and income distributions are similar for both groups. Table 5 provides summary statistics for current and defaulted loans by race. Some statistics are now more similar for Whites and Blacks that have defaulted, although defaulted Blacks generally still have weaker credit histories than Whites, as seen in comparisons of current delinquencies (6.242 to 4.575) and public records in the last year (0.156 to 0.071). However, defaulted Blacks generally also have a much lower revolving credit balance ($2327 to $10,407) and a lower delinquent balance ($2775 to $3208) while generally defaulting on a lower average loan balance ($4430 to $4742). It is worth highlighting that the age at which Black loans default is the lowest of any of the races, and about half a month earlier than that of Whites (9.63 months to 10.02 months).

5 Methods and Results

The following sections utilize econometric analysis to answer three key questions about the sample of positively identified listings when controlling for all observables:

(1) Do lenders discriminate based on race when choosing to fund loans?

(2) For funded loans, does the Prosper lending market's interest rate outcome depend on the race of the borrower?

(3) Do differences in default risk exist for borrowers of different races?

The first section measures discrimination at the funding stage to determine whether race has a marginal effect in the likelihood of being funded, controlling for all other observable characteristics. Then, for loans that are successfully funded, the next section explores the existence of discrimination in setting interest rates for these groups—discrimination would imply a marginal change in the interest rate based solely on race. The final section tests for any actual difference in loan performance based on race (after controlling for observables) and analyzes the effect of any previously identified discrimination on default rate. All of the analysis is restricted to the set of listings where both race and gender could be positively identified.[16] Summary results are presented in the body of the paper with detailed regression results available in Appendix D.

5.1 Funding Decisions

In order to determine whether a borrower's race has an effect on the likelihood that his or her listing will be funded, I estimate a maximum-likelihood probit regression on the restricted dataset of positively identified listings of the general form:

$$\Pr(\mathit{Funded}_i) = \beta_0 + \beta_1 \mathit{Black}_i + \beta_2 \mathit{Hispanic}_i + \beta_3 \mathit{Asian}_i + \beta_4 \mathit{Female}_i + [\mathit{Interaction\ Effects}]_i + [\mathit{Listing\ Controls}]_i + [\mathit{Core\ Credit\ Controls}]_i + [\mathit{Extended\ Credit\ Controls}]_i + [\mathit{Employment\ Controls}]_i + \varepsilon_i$$ [17]

[16] Additional research could explore the effect of choosing to include an image.

[17] To be precise, errors are not truly i.i.d. because I do not drop multiple listings from the same borrower. I believe this is still a very realistic approximation because the total number of repeat borrowers is small in comparison to the full sample size.

The dependent variable investigated is *Funded*, a binary variable that equals 1 in the case of a fully funded loan and 0 if the listing expires without being funded. I compute and report the marginal effects. Therefore, the coefficients presented for the binary variables of interest represent the marginal change in probability of being funded resulting from a discrete change in the dummy variable from 0 to 1.[18]

FIGURE 5.1 – MODELS OF LOAN FUNDING (PROBIT DERIVATIVES; STANDARD ERRORS IN PARENTHESES)

Specification: Additional Variables Included	Black	Hispanic	Asian	Female	Obs. / Pseudo R^2
(1) Race and Gender Interactions	-0.204866**	-0.159991**	0.062216$^+$	-0.065113**	10178
	(0.018176)	(0.028636)	(0.033059)	(0.012039)	0.0245
(2) Model 1 plus Listing Controls	-0.203961**	-0.154876**	0.091063*	-0.041971**	10178
	(0.023564)	(0.035928)	(0.042239)	(0.015266)	0.429
(3) Model 2 plus Core Credit Controls	-0.178234**	-0.191535**	0.048828	0.028710	9910
	(0.047599)	(0.059213)	(0.088507)	(0.032448)	0.615
(4) Model 3 plus Extended Credit Controls and Employment Controls	-0.125009**	-0.123735*	0.059362	0.026310	6371
	(0.041311)	(0.062294)	(0.080406)	(0.026908)	0.669

** $p<0.01$, * $p<0.05$, $^+$ $p<0.1$

Figure 5.1 presents the impact of a borrower's race and gender on the outcome of his or her loan request. Model (1) displays the uncontrolled regression results, signaling strong negative correlations on the funding decision for being Black, Hispanic, or Female, as well as slightly positive effects for being Asian. Of course, most of this effect is due to other factors correlated with membership in these groups and the funding decision; model (2) replicates these results with the minimum Listing Controls added and provides some added detail about the full sample of 10,178 coded listings. Additional controls in this model include *Opening Rate* which controls for systematic differences in the starting rate certain groups may choose and *Listing Number* which serves as a proxy

[18] This is equivalent to running a *probit* regression in STATA and using *mfx* to calculate marginal effects.

for time and controls for the increasing concentration of loan listings available which may overwhelm funding supply.

Model (3) provides the first meaningful insight into the interaction between race and credit information and its effect on funding. The sample here includes both the 3539 listings with Core Credit Controls only and the 6371 listings that include extended credit information as well as employment information. Therefore, to control for the effect this additional information may have on the decision to fund a listing, a time dummy variable, *More Info*, is added that is 0 for the earlier subsample and 1 for all observations in the later subsample; its interactions with race and gender are also included, although none are statistically significant. Additionally, I include *Borrower State Fixed Effects* to control for the various usury laws and racial compositions among different states. The coefficients on *Black* and *Hispanic* are both significant at the 1% level and concerning in their magnitude: they imply a 17.8 and 19.1 percentage point reduction in loan funding likelihood for Blacks or Hispanics based solely on race, respectively, compared to Whites when only basic credit information is available. None of the coefficients on the race and gender interactions are significant in this specification.

Model (4) provides the most controlled analysis, removing 84 degrees of freedom from the dataset for Extended Credit Controls and Employment Controls. With this new information presented to the lender, as well as the additional controls in the regression, the coefficient on *Black* retains its significance while the coefficient on *Hispanic* drops just enough to fall to the 5% level. The implied level of discrimination also decreases in either case to 12.5 and 12.4 percentage point reductions in funding likelihood for Blacks and Hispanics, respectively, suggesting some of the discriminatory behavior measured in model (3) was in response to concerns regarding unobservable characteristics that have now been revealed.[19] Furthermore, the coefficient of the interaction term *Black * Female* becomes significant at the 5% level and, with a value of 12.0%, almost entirely offsets the discrimination faced from race.[20]

5.2 Interest Rate Pricing

To determine whether observing race of the borrower has any effect on the ultimate interest rate offered, I estimate a regression specification of the general form:

$$\begin{aligned} \textit{Lender Rate}_i = {} & \beta_0 + \beta_1 \textit{Black}_i + \beta_2 \textit{Hispanic}_i + \beta_3 \textit{Asian}_i + \beta_4 \textit{Female}_i \\ & + [\textit{Interaction Effects}]_i + [\textit{Listing Controls}]_i + [\textit{Core Credit Controls}]_i \\ & + [\textit{Extended Credit Controls}]_i + [\textit{Employment Controls}]_i + \varepsilon_i \end{aligned}$$

The dependent variable investigated is the interest rate lenders were willing to offer in order to fund a listing. Regressions are linear and reported errors are the standard variance estimators for Ordinary Least Squares regression. The coefficients on the binary variables of interest can be interpreted as the additive increase or decrease in value of the dependent variable caused by a discrete change in the variable of interest. *Lender Rate* is a decimal value with an approximate range between 0.01 and 0.3575. The average rate for all funded loans is 0.178.

---

[19] Because the introduction of additional credit data restricts the dataset being analyzed to only the most recent listings, an alternative explanation of the difference in results includes the entrance of non-discriminating lenders into the marketplace, either due to their knowledge of the existing economic inefficiencies due to racial discrimination or use of investing algorithms to choose listings (which would likely be blind to race).

[20] A possible explanation for this strong effect is a lack of discrimination against Black Females at the funding stage only, as there is evidence of interest rate discrimination. See §6.2 for more information. Unless explicitly stated otherwise, ‘Blacks’ will refer to Black Males for the remainder of this analysis.

FIGURE 5.2 – MODELS OF LOAN PRICING (OLS COEFFICIENTS; STANDARD ERRORS IN PARENTHESES)

Specification: Additional Variables Included	Black	Hispanic	Asian	Female	Obs. / R^2
(1) Race and Gender Interactions	0.030139**	0.024799**	0.001819	0.017335**	5969
	(0.003340)	(0.005072)	(0.004957)	(0.001925)	0.04120
(2) Model 1 plus Listing Controls	0.022649**	0.012599**	-0.001105	0.010100**	5969
	(0.002928)	(0.004440)	(0.004338)	(0.001691)	0.27051
(3) Model 2 plus Core Credit Controls	0.009380**	0.003670	0.000529	-0.001880	5830
	(0.002811)	(0.003879)	(0.004358)	(0.001718)	0.73842
(4) Model 3 plus Extended Credit Controls and Employment Controls	0.011238**	-0.000446	-0.002127	-0.001284	3558
	(0.002351)	(0.003784)	(0.003282)	(0.001362)	0.75776

** p<0.01, * p<0.05, + p<0.1

Figure 5.2 presents measurements of racial discrimination in the pricing of loans, conditional on the loan being funded. The choice of controls for each regression parallels the previous section and allows the regressions to be evaluated similarly. Model (1) presents uncontrolled regression results which summarize the net effect on the interest groups: Blacks, Hispanics, and Females generally face higher interest rates for loans, but the race effects are mitigated for Black and Hispanic Females due to negative and statistically significant coefficients. The coefficient on *Black* represents an increase of 3 percentage points in interest rate for a Black borrower from the average for a White Male, 16.6%, which is almost one-fifth of an increase. This analysis, of course, ignores the differences among these groups in credit worthiness, and this is reflected in its weak R^2 value of 0.0412. The regression using model (2) also signals higher interest rates for the same three groups, although to a lesser degree given the addition of Listing Controls.[21] The interactions between race and gender lose their significance at this point.

[21] *Opening Rate* is no longer included in the Listing Controls. As all these loans have been funded, they have reached their market-equilibrium interest rate and the starting rate is irrelevant, provided the dummy control *Close Immediately* is included to capture the loans that do not enter the auction stage.

Model (3) once again provides the first insight into the combined subsamples of listings, including 2272 funded loans with only core credit information as well as 3558 loans funded since the introduction of extended credit profiles by Prosper.[22] Core Credit Controls are included in this regression as well as *Borrower State Fixed Effects*, raising the regression's R^2 value to 0.7384. Despite controlling for credit background, the coefficient on *Black* remains statistically significant at a 1% level and suggests a 0.93 percentage point increase in interest rates for Black borrowers; no coefficient on any other race variable is significant, suggesting targeted discrimination against Blacks rather than minority groups in general. The coefficient on *More Info* is equally significant and negative, confirming the expected result: the presence of more information lowers interest rates for all groups, as the default risk due to asymmetric information is reduced.

Model (4) contains the most controlled regression, including all Extended Credit Controls as well as Employment Controls. A negligible increase in the regression's R^2 value from 0.738 to 0.758 suggests lenders may not be utilizing this extended data significantly in pricing loans. Meanwhile, the effect of discrimination against Blacks appears larger in this regression, as the coefficient on *Black* increases to 0.011 with greater statistical significance at the 1% level, implying a 1.1 percentage point interest rate premium for being Black. No significant discrimination in loan pricing appears for Hispanics or Asians.

[22] Similarly to the funding decision analysis, a dummy variable representing this change as well as its interaction terms for the variables of interest are included in this regression.

5.3 Loan Performance

I utilize unprecedented detail regarding loan status to investigate whether relevant information regarding default risk is revealed through a borrower's race. Prosper defines a loan in default as one that is at least four months late, at which point it is considered uncollectable and sold to a debt buyer.[23] For this reason, the sample of loans analyzed has been limited to only those loans that are old enough to qualify for this state and contain a full set of controls. In order to determine whether a borrower's race has a marginal effect on his or her ability to repay loans, I estimate four separate maximum-likelihood probit regressions on the restricted dataset of positively identified funded loans, each producing an added insight. Results are reported here, although full interpretation is reserved for the Discussion section.

FIGURE 5.3 – MODELS OF LOAN DEFAULT (PROBIT DERIVATIVES; STANDARD ERRORS IN PARENTHESES)

Specification: Additional Variables Included	Black	Hispanic	Asian	Female	Obs. / Pseudo R^2
(1) None	0.078722**	0.018312	-0.025563	0.013452	2197
	(0.019330)	(0.030335)	(0.019965)	(0.010374)	0.0300
(2) Model 1 plus Listing, Core Credit, Extended	0.045942**	0.002568	-0.011990	0.009247	1751
Credit, and Employment Controls	(0.016450)	(0.017422)	(0.010778)	(0.007276)	0.373
(3) Model 1 plus Interest Rate	0.041962**	0.014110	-0.022722	0.003807	2197
	(0.014301)	(0.024066)	(0.013909)	(0.008036)	0.139
(4) Model 2 plus Interest Rate	0.038175*	0.002983	-0.011094	0.009457	1751
	(0.015086)	(0.016941)	(0.010669)	(0.007024)	0.382

** p<0.01, * p<0.05, + p<0.1

Figure 5.3 presents the results of this analysis. Model (1) uses the specification:

$$\Pr(Default_i) = \beta_0 + \beta_1 Black_i + \beta_2 Hispanic_i + \beta_3 Asian_i + \beta_4 Female_i + \varepsilon_i$$ [24]

[23] http://www.prosper.com/help/topics/lender-default_sales.aspx

[24] Race and gender interaction effects are dropped from this analysis because, in the restricted dataset, there are too few observations of Female default to allow for interpretation of regression results. Robustness checks on the full data set, either without the additional controls or with the additional controls by using the product of junk data paired with a time dummy, yielded no significant results for the interaction terms.

This regression measures only the difference in likelihood of a negative loan outcome among the groups of interest. The coefficient on *Black*, although very significant and large at 0.08, suggests only that Blacks are a higher risk group that defaults more often. In order to determine whether Blacks are more likely to default based only on their race, we must control for every other observable characteristic, an option given Prosper's online framework. Model (2) uses the following specification to achieve that goal:

$$\Pr(Default_i) = \beta_0 + \beta_1 Black_i + \beta_2 Hispanic_i + \beta_3 Asian_i + \beta_4 Female_i + [Listing\ Controls]_i + [Core\ Credit\ Controls]_i + [Extended\ Credit\ Controls]_i + [Employment\ Controls]_i + \varepsilon_i$$

Intuitively, the coefficient on *Black* measures the difference in default likelihood between two persons, one that is Black and one that is White, with all observable characteristics in Prosper's listing page equated. Model (2) yields a 0.046 coefficient on *Black* that is statistically significant at the 1% level. This implies that some bias towards default can be observed in those borrowers that are identified as *Black*.

In order to check the market efficiency of any racial discrimination that may exist, model (3) utilizes the following specification:

$$\Pr(Default_i) = \beta_0 + \beta_1 Black_i + \beta_2 Hispanic_i + \beta_3 Asian_i + \beta_4 Female_i + Borrower\ Rate_i + \varepsilon_i$$

Because the interest rate should be set taking into account all observable characteristics in an otherwise efficient market, an efficient level of statistical discrimination should yield coefficients on race variables that are not statistically different from zero. The probit regression yields a coefficient of 0.042 on *Black* with a 1% significance level. The basic interpretation suggests too little discrimination on the basis of a borrower being Black, since not all of the added default probability is reflected in a higher interest rate.

If the interest rate offered is properly accounting for default likelihood, and the market is efficient in utilizing available information to price loans, including the borrower's interest rate along with all the previous controls (from Figure 5.2) in model (4) should yield the same result as model (2) with insignificant coefficients on all controls. Therefore, I utilize the following specification:

$$\Pr(\mathit{Default}_i) = \beta_0 + \beta_1 \mathit{Black}_i + \beta_2 \mathit{Hispanic}_i + \beta_3 \mathit{Asian}_i + \beta_4 \mathit{Female}_i + \mathit{Borrower\ Rate}_i + [\mathit{Listing\ Controls}]_i + [\mathit{Core\ Credit\ Controls}]_i + [\mathit{Extended\ Credit\ Controls}]_i + [\mathit{Employment\ Controls}]_i + \varepsilon_i$$

The results from model (4) do not match those of model (2). The coefficient on *Black* drops to 0.038 at a 5% significance level. Table 3 in Appendix D notes a number of significant control variables, including the verity of the borrower's bank account information, his or her endorsement status, number of credit inquiries, revolving credit balance, bank card utilization, and length of employment status. Furthermore, joint tests of income level controls and borrower state controls exhibit significance at the 1% level.[25] The change in the coefficient on *Black*, as well as the significance of certain controls, imply that the market is not generally efficient at setting interest rate using all available information. I return to this point in my discussion.

---

[25] Regression specifications with less controls but utilizing a larger sample of the data were done to check robustness. All results were comparable to the models presented. They are omitted for the sake of simplicity.

6 Discussion

The results of my econometric analysis appear puzzling initially. Although treatment of Blacks and Hispanics is fairly uniform regarding the decision to fund a loan, the similarity ends there. In order to reconcile these outcomes, interpretations of the results for Hispanics and Blacks are analyzed separately and then brought together.

6.1 Discrimination against Hispanics

The decision to fund someone and the interest rate they receive are inexorably linked to one another. This is a benefit when analyzing the treatment of Hispanics by lenders at Prosper. The most controlled results of Figure 5.1 make it very clear that Hispanic borrowers are, on net, about 12.4 percentage points less likely to have their loan request funded than a White borrower with the same personal characteristics as well as loan terms. Figure 5.2, however, implies that although the average interest rate may be higher for Hispanics, when you control for all other observable characteristics there is no measurable interest rate premium.

What does this mean? Although Hispanics are less likely to get funded, the decision to not fund a particular Hispanic is not related to interest rates, or subsequently to their default risk. This undermines the relationship between denial and default rates that many of the critics of the Federal Bank of Boston study assumed to be true. Lenders are not using race as an indicator of default risk for Hispanics—if they were, it would be reflected in an interest rate premium. To account for this behavior, I introduce a concept of discrete heterogeneity of taste-based discrimination in lenders and utilize the listing property of *Close Immediately* and its usefulness in detecting a threshold effect. For the sake of simplicity, assume that there are generally two types of lenders on Prosper:

perfectly nondiscriminatory lenders and lenders with a prejudicial aversion to Hispanics. I postulate that discrimination occurs only in the funding stage of a Hispanic loan due to shortage of nondiscriminatory lenders.

When a nondiscriminatory person sees an attractive unfunded listing for a Hispanic borrower, he does not notice race and sees only that it is a viable listing to fund given the borrower's credit information. He then proceeds to place a proxy bid at the efficient rate for that borrower. These bids accumulate up to the total supply of available bid commitments among nondiscriminatory lenders. In the case of Hispanics, there is a shortage in the supply of these lenders. Instead of uniformly leaving all viable Hispanic listings unfunded, some subset that is not related to default risk becomes funded while other listings do not. This is because underpriced listings that are 99% funded are worthless to lenders until they reach their funding threshold. If an underpriced listing is about to hit its funding threshold, it is marginally more attractive to place a bid on that listing rather than another, all else equal, due to the higher probability that the loan will close. Once the loan becomes funded, competition among the nondiscriminatory lenders that did fund the loan drives the price down due to the marginal attractiveness of a guaranteed excess return (for an underpriced loan) compared to an unfunded listing bid. The marginal nondiscriminatory lender would still choose to bid down a funded loan to its efficient interest rate rather than place a bid on an unfunded listing, even though the starting rate on the unfunded listing might be greater, his willingness-to-pay is the same and the likelihood of getting a return is higher on the funded loan.[26]

This construction can be empirically tested. Under these assumptions, Hispanics who choose to have their loans close immediately when funded would be subject to a greater marginal interest rate premium than Whites who choose that option due to discrimination that is isolated in the first stage of loan funding. Indeed, adding race and gender interactions with the *Close Immediately* dummy variable to model (4) of Figure 5.2 reveals a coefficient of 0.013 on *Hispanic * Close Immediately* with significance at the 5% level.

Furthermore, this interpretation matches evidence from Section 5.3 regarding default data for Hispanics. Model (2) estimates no ex ante default bias for Hispanics; given the discussion above, a finding of uniform taste-based discrimination rather than statistical discrimination is able to reconcile the ex ante measurements with the results of model (3), which does not show any residual default bias after controlling for interest rate.[27] Therefore, the discrimination measured against Hispanics in the funding stage of loans on Prosper is likely to be uniform taste-based discrimination that is uncorrelated with default risk in its application; loans are accurately priced without taking race into account, although there is a shortage of lending supply for Hispanic borrowers.

---

[26] Other variations on this combination of a shortage of nondiscriminatory lending supply, two-stage loan funding, and pricing thresholds are also plausible: discriminators' ability to discern race may be heterogeneous, or the signaling quality of pictures may be heterogeneous. Then the funded loans could be those that are the weakest signal of being Hispanic, assuming this is uncorrelated with default risk.

[27] Had there been statistical bias against Hispanics with no ex ante marginally increased default risk, we would expect a reduced likelihood of default amongst Hispanic lenders when controlling for interest rate.

6.2 Discrimination against Blacks

The puzzle is more complicated when assessing the situation of Black borrowers on Prosper. Once again, the decision of who to fund and the interest rate at which to fund them are linked. Because discrimination is detected at both the funding and pricing stages, an analysis of the available results runs into concerns of simultaneity discussed by Yezer, Phillips, and Trost (1994). Although I suspect future research will divorce these effects, I instead begin with an analysis of the discrimination against Blacks within the same framework as Hispanics. I include additional explanations afterwards.

+ Two Levels of Discrimination

Similar to the case of Hispanics, the results can be interpreted in the framework of two separate stages of discrimination. Although discrimination certainly occurs at the funding stage, with a net effect of about a 12.5 percentage point reduction in likelihood of being funded, it is doubtful that this is statistical in nature due to the lack of a transformative effect it would have on the subsample of funded Black loans. If it was the case that lenders were selecting only the most qualified Black loan candidates, the funded Black candidates would, on average, have a large difference in their credit-related statistics when compared to the unfunded Black candidates. This difference in means between funded and unfunded listings for Blacks should then exceed the difference for other races; this relationship is not found. Instead, similar differences in means of characteristics are found among the different races, making it plausible that this level of discrimination is similarly taste-based in nature.

Once a loan is funded, however, its interest rate does not get bid down to be equivalent of those offered to Whites, Hispanics, or Asians, all else equal. Instead, model

(4) in Figure 5.2 shows interest rate remains a full 1.1 percentage points higher for Black borrowers. To assess the efficiency of this premium, we look to the additional default risk attributed to Black borrowers. Model (2) in Figure 5.3 tells us that, when controlling for all observable characteristics visible to the lender, Blacks have an additional default likelihood of approximately 4.6% that is not attributed to any of the present controls. This marginal likelihood of default can now be used as an ex ante measurement of additional default risk.

+ Economic Efficiency

Analysis of a simple loan pricing model based on default risk can offer insight into an appropriate interest rate adjustment for the ex ante measurement derived above. When a loan is generated, the expected payout to the lender is:

$$E[R] = D * P * (1 + r)^T$$

where R is the payout, D is the likelihood the borrower will pay back, P is the principal amount lent, r is the fixed interest rate, and T is the number of periods.[28] If groups A and B have different default rates, lenders will demand an increased interest rate from the group with a higher default rate in order to equate the two opportunities:

$$E[R_A] = D_A * P * (1 + r_A)^T = (D_A - \beta) * P * (1 + r_A + \gamma)^T = E[B]$$

where β is the additional likelihood of default and γ is the interest rate premium charged for group B. In order to compensate lenders for the marginal default risk β, the interest

[28] This analysis assumes zero proceeds from default for both Whites and Blacks. Because of the outstanding number of uncollectable loans that have not yet been sold, a definitive analysis including amount recovered from loan default is not possible for recent loans. However, historical records show the proceeds from defaults are lower for Blacks ($567.23) than they are for Whites ($597.88), which causes this analysis to understate the necessary amount of interest rate premium required.

rate premium charged to Blacks must fulfill the following conditional:

$$\gamma > \left[\left(\frac{D_A}{D_A - \beta}\right)^{\frac{1}{T}} - 1\right][r_A + 1].$$

Using average values for the repayment rate of White Males and the ex ante marginal default risk for Black Males, in the sample of loans with at least core credit information, the following inequality is evaluated:

$$\gamma > \left[\left(\frac{0.953}{0.953 - 0.046}\right)^{\frac{1}{3}} - 1\right][0.159 + 1].$$

Note that β corresponds to the ex ante marginal default risk for Black Males measured in model (2) of Figure 5.3, which is equal to 0.046. To account for this default risk, the interest rate premium must be at least:

$$\gamma > 0.019.$$

This suggested interest rate premium, which is understated according to the discussion in the previous page's footnote, is almost twice as high as that which is measured by model (4) in Figure 5.2: a value of 0.011. This suggests that the discrimination against Blacks will not only fall short of creating the reverse relationship of lower default rates hypothesized by Berkovec et al. (1994, 1996a, 1996b) under assumptions of equal ex ante default risk, it should not even absorb enough of the default risk to make lending to Blacks economically efficient. This is confirmed with model (3) of Figure 5.3 which shows that when controlling for interest rate, Blacks are still more likely to default on loans as compared to the other racial groups. This is the link missed by researchers who tried to interpret the inconsistency in the Federal Reserve Bank of Boston's data regarding denial and default: a finding of discrimination in denial rates does not require a reverse effect in default, unless it can be shown that loans to Blacks, controlling for all other variables, are not riskier than loans to Whites. The previous analysis has rejected this possibility.

+ Critics of Default Measures

Much of the aforementioned discussion is based on the assumption that in an efficiently discriminating marketplace, a regression on the likelihood of default on race variables that controls for interest rate should yield insignificant coefficients for all races. Although this appears intuitive, details regarding the Federal Bank of Boston study could call into question this analysis when applied to their data. Before moving on to other potential explanations, it is worth briefly addressing these concerns and how this analysis is an improvement on earlier work.

Ross (1996) discusses the effect of unobservable variables on default rate. He claims that, if unobservable variables affecting default correlate with race, the performance method will be biased away from finding discrimination. Let us assume for a moment that there are some unobservable variables that correlate with both default and race. A lender that owns a portfolio of loans from different races will still need to set a discriminatory higher price for the borrowers from races with a higher tendency to default, given the correlation between their race and the unobservable characteristics. Otherwise the lender would own a portfolio of equivalently priced loans, some of which have a higher likelihood of default than others. This is an impossible outcome assuming the existence of no arbitrage opportunities in an efficient marketplace.

Yinger (1996) provides an extended analysis of using loan-performance measurements to detect default. He first restates the argument Ross (1996) uses, which is addressed above. Next, he states that the validity of loan-performance measures depends on Whites not receiving more favorable treatment than minorities in foreclosure proceedings. Because Prosper treats every delinquent borrower systematically in accordance with its policies, it is unlikely that any subjective racial bias would enter into the default proceedings—this may be a unique component of home foreclosures. Additionally, Yinger states the models also rest on the assumption that the losses on minority defaults are at least equal to those for Whites. Using additional data provided by Prosper, the average loss on default for Whites is determined to be $3356.51 while the average loss for Blacks is $3403.44. It is clear from this data that Yinger's third assumption is met.

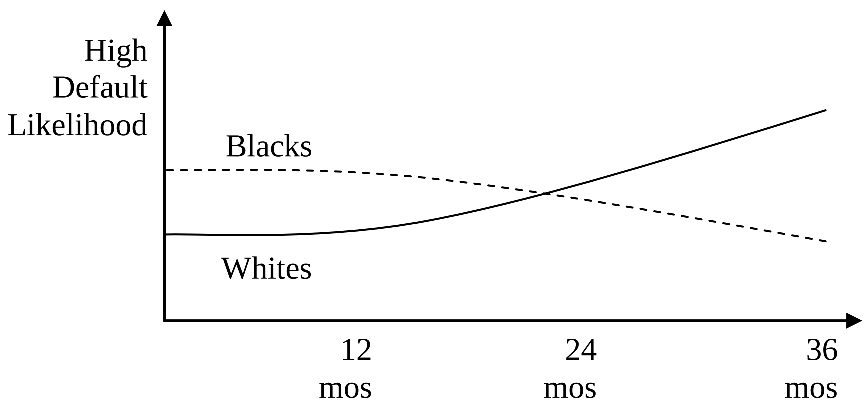
+ Market Failure

One possible explanation for the residual default measurements for Blacks is a simple inability in the market to properly price risk. Under this assumption, the effect measured for Blacks is just a consequence of general market failure. Evidence can be seen in the comparison between models (2) and (4) in Figure 5.3. Assuming that the interest rate is efficiently chosen to predict the default risk of a borrower, adding it to a regression on all its determinants should cause no difference in results. Furthermore, the coefficients on all control variables should be jointly zero. As mentioned in Section 5.3, this is not the case, as we observe changes in both the coefficients on our variables of interest as well as a number of control variables. This result signals that these control variables are not being properly priced into the interest rate of a loan, and the market is generally not efficient.

+ Loan Tenure

One possible explanation for the lingering default risk attributed to Blacks is that the measurement simply reflects heterogeneity in timing of default by the different groups. Because none of the loans in Prosper have reached their 36 month term yet (the oldest loans are 26 months old; the oldest loans with a full set of controls are only 12 months old), it is possible that the likelihood of default may look like that in Figure 6.1.

Figure 6.1: Hypothesized Marginal Default Likelihood over Time for Blacks and Whites



In this case, loans to White borrowers could be considered overpriced over time (depending on the interaction between the timing of a default and the amount recovered.) This result is marginally supported by the summary statistic on *Age* for defaulted loans when compared between Blacks (9.63) and Whites (10.01). Although a regression to show the effect of race on *Age* for defaulted loans did not produce statistically significant results at the 5% level, the effect was in the proper direction and could become significant as new observations become available.

+ Default Shock

Another possible explanation for the inefficient outcome in correcting for a Black borrower's default risk involves the existence of economic shocks. It is possible that, during the restricted time frame of our sample, some external shock affected the borrowers in a way that is correlated with race and default likelihood. Then, regardless of how many observations or control variables the sample contains, the coefficients on race variables will incorporate the effects of the shock rather than any innate differences in risk, and would not be useful for future pricing. For example, imagine the sudden increase in price of a product with high price inelasticity for Blacks and no substitutes. Blacks would have fewer resources to pay their loan obligations, and therefore may default more often. Since the shock is only temporary, prices eventually return to normal. Historical data will predict a higher default risk for Blacks based solely on race, while the value of race as a predictor of future default may be nonexistent.

7 Caveats

This paper presents the first known study of race effects in Prosper's marketplace, and is one of a small group of papers that attempt to measure the effects of discrimination in credit markets other than the mortgage market. By covering new territory, I am able to answer questions regarding discriminatory behavior and lending that may be impossible to answer in other contexts, and can provide meaningful insight into an emerging market. At the same time, however, without the guidance of previous literature, my work retains distinct weaknesses. I choose to describe these weaknesses in order to address the potential for any biases in my results and to provide a blueprint for future research to improve upon my design.

+ Measurement Error

When deciding on the best method for coding images of borrowers for race and gender, I opted to maximize consistency while minimizing cost. Therefore, the job fell on me to code over 55,000 photos. While my observations of race may be consistent, this may also create consistent errors in measurement. I do not believe this weakness to have a serious influence on my results. It is not the task of the coder to necessarily identify the actual race of the borrower, but to determine the most common subjective observation of race. Although it is unlikely that I am a perfect proxy for the common Prosper lender, this property creates some margin for error. Additionally, because I am a White Male, it is more likely that I would under sample members of other racial groups than over sample them. Since the unidentified listings are not included in my analysis, and because the same listings are more likely to be unidentifiable by other users and not treated as a member of that racial group, the net effect is most likely to be a more accurate measure of discrimination, despite the loss of observations.

Future research can determine the extent of my error and improve on the study's design by utilizing a randomly selected group of Prosper lenders to 'vote' on the race of each member. The net result would be a distribution between 0 and 1 of the likelihood that a member is identified as part of a particular race. Intuitively, an image that half the lenders believe depicts a Black man while the other half believe depicts a Hispanic man would not be a useful indicator of Black-Hispanic discrimination variance. Additionally, if actual races were desired instead of observed race, Prosper may have access to this data based on their identity verification checks, although this would require cooperation from within the organization.

+ Omitted Variable Bias

Although my analysis far exceeds previous literature in its ability to incorporate almost all observable data presented to a lender, there are minor areas where my analysis may have fallen short. Buyers are given the ability to provide some free-form information to lenders by including an optional listing description. Prosper provides a template of this description; most descriptions convey why the borrower needs a loan and his or her ability to make on time payments. It is possible that a property of these descriptions could be correlated with the dependent variables in my study as well as race. Besides conveying relevant information in the text of their listing, borrowers can also provide additional information through pictures such as the number of children they have or their marital status. Additionally, pictures could also influence lenders in more subjective ways. For example, if wearing a suit increases your likelihood of being funded, and wearing a suit is highly correlated with being Asian, it's possible that the effect of wearing a suit is captured by the race-variable on being Asian.

Again, economic factors rendered a full coding unreasonable; however, future research can certainly address these concerns. A study could randomly select among the listings included in this paper and code the descriptions and images for any additional information, as well as provide a subjective measure of 'reliability' or 'credit worthiness' to capture the effect of writing style or quality of photographs[29]. Then the correlation of these variables with race and the dependent variables can be measured. Additionally, the

[29] This would need to be done without revealing the race of the borrower to the coders. Masking names in text, and the face and body parts of persons in the images would be the most definitive way of doing this.

full sample could be coded for this additional information and the new data used in the regression models.

+ Sample Restrictions

The youth of Prosper's marketplace makes a definitive analysis of the economic efficiency of discrimination difficult. The oldest funded loans on Prosper are approximately two years old, a full year short of their three year maturity; the average age of a loan on Prosper with full details regarding credit history and employment is only 6.4 months. I employed robustness checks whenever possible to ensure the presented results generalized to other samples, however it is still possible that not enough variation is revealed due to the restricted set of funded loans with full credit information. This is especially the case for measuring the effect on very specific slices, such as Asian Females, who are generally less represented in the sample. As Prosper grows in popularity, and time passes, the dataset should become more valuable for its ability to provide significant variation across multiple interest groups.

8 Conclusion

In this paper, I utilize unique lending and demographic data to test for the presence of racial discrimination in the peer-to-peer lending marketplace Prosper.com. My dataset is a combination of Prosper's publicly available data and hand-coded observations of the race and gender of over 34,000 members. I use this data to measure the effect of racial discrimination in funding listings, pricing loans, and predicting default rates. Although my results are localized to Prosper.com's particular unsecured loan market, they may be generalized to the entire person-to-person lending market, and could shed light on discrimination in other consumer credit programs.

I report four main conclusions regarding the presence of racial discrimination and its efficiency. (1) Blacks and Hispanics experience racial discrimination during the decision to fund a listing and, relative to Whites, are 12.5 and 12.4 percentage points, respectively, less likely to be funded when requesting a loan, all else equal. (2) Blacks face additional discrimination in pricing, and are subject to a 1.1 percentage point premium on interest rates compared to other groups, all else equal. (3) Loan performance analysis indicates that Blacks are approximately 4.6 percentage points more likely to default on loans ex ante. When controlling for interest rate, Blacks are still more likely to default. (4) Therefore, Blacks face statistical discrimination due to their elevated default risk, but at too low of a level to be efficiently priced; Hispanics, however, face only taste-based discrimination unrelated to their default risk and therefore are underfunded but at an efficient price.[30]

Future research should be directed along two different paths. First, researchers should attempt more accurate measures of discrimination. This can be accomplished through additional coding of observable characteristics, addition of new data as it becomes available, and robustness checks with different subsamples of the data and functional form specifications. Although a randomized experimental design was initially planned for this study, "[t]he leadership of the company is firmly and philosophically against the idea of using phantom listings, profiles, etc."[31,32] Additionally, new research can exploit unused properties of the dataset to generate new insights into the potential

[30] It is worth mentioning that Black Females appear to fill in the third case; they do not face taste-based discrimination at the funding stage, but do face the same interest rate premium given to Black Males.
[31] Private correspondence with Tiffany Fox, Prosper's Communications Director, 10/10/2007.
[32] An experimental study may still be possible without official cooperation from within Prosper; however, the implications regarding credit fraud should be investigated before going ahead with such a plan.

causes of discrimination. Detailed row level data is available for each bid; these could be used to test hypotheses about the specific mechanisms at work within the funding decisions of Blacks and Hispanics. The level of racial discrimination could even be identified in individual lenders. Some lenders opt to include their own picture in their profile; these could be coded to check for same-sex and same-race effects. Even the control listings can be utilized to measure economic efficiency in the absence of potentially biasing data. The wealth of information contained within Prosper's dataset is only set to grow with time and should provide continued insight into the role discrimination plays in our society.

9 References

Ayres, Ian. and Peter Sigelman, "Race and Gender Discrimination in Bargaining for a New Car," *The American Economic Review*. 85:3. (1995): 304-321.

Becker, Gary, *The Economics of Discrimination*, 2nd edition. Chicago: University of Chicago Press, 1971 [1957].

Becker, Gary, "The Evidence Against Banks Doesn't Prove Bias," Business Week, April 19, 1993.

Berkovec, James A., Glenn B. Canner, Stuart A. Gabriel, and Timothy H. Hannan, "Race, Redlining, and Residential Mortgage Loan Performance," *Journal of Real Estate Finance and Economics*. 1994(9): 263-94.

Berkovec, James A., Glenn B. Canner, Stuart A. Gabriel, and Timothy H. Hannan, "Race, Redlining, and Residential Mortgage Defaults: Evidence from the FHA-Insured Single-Family Loan Program." In John Goering and Ron Wienk, eds., *Mortgage Lending Racial Discrimination and Federal Policy*. Washington, D.C.: Urban Institute Press, 1996a (251-88).

Berkovec, James A., Glenn B. Canner, Stuart A. Gabriel, and Timothy H. Hannan, "Mortgage Discrimination and FHA Loan Performance," *Cityscape: A Journal of Policy Development and Research*. 1996b(21): 9-24.

Bertrand, Marianne and Sendhil Mullainathan, "Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination," *The American Economic Review*. 94:4. (2004): 991-1013.

Blinder, Alan S., "Wage Discrimination: Reduced Form and Structural Estimates," *The Journal of Human Resources*. 8:4. (1973): 436-455.

Firth, Michael, "Racial Discrimination in the British Labor Market," *Industrial and Labor Relations Review*. 34:2. (1981): 265-272.

Gunderson, M., "Male–Female wage differentials and policy responses", *Journal of Economic Literature*. 1989(27): 46-72.

Heckman, James, and Peter Siegelman, "The Urban Institute Audit Studies: Their Methods and Findings." In M. Fix and R. Struyk, eds. *Clear and Convincing Evidence: Measurement of Discrimination in America*. Urban Institute, 1993.

Horne, David, "Evaluating the Role of Race in Mortgage Lending." *FDIC Banking Review*. (1994) 1-15.

Kenney, G. and D. Wissoker, "An analysis of the correlates of discrimination facing young Hispanic job-seekers." *American Economic Review*. 1994(84): 674-83.

King, Thomas A., "Discrimination in Mortgage Lending: A Study of Three Cities." *Office of Policy and Economic Research, Federal Home Loan Bank Board, Washington, DC*, Working Paper No. 91, 1980.

Morton, Fiona, Jorge Silva-Risso, and Florian Zettelmeyer, "Consumer Information and Discrimination: Does the Internet Affect the Pricing of New Cars to Women and Minorities?" *Quantitative Marketing and Economics*. 1:1. (2003): 65-92.

Munnell, Alicia H., Geoffrey M.B. Tootell, Lynn E. Browne, and James McEneaney, "Mortgage Lending in Boston: Interpreting HMDA Data," *American Economic Review*. 86:1. (1996): 25-53.

Oaxaca, Ronald. "Male-Female Wage Differentials in Urban Labor Markets," *International Economic Review*. 14:3. (1973): 693-709.

Ross, Stephen, "Flaws in the Use of Loan Defaults To Test for Mortgage Lending Discrimination," *Cityscape: A Journal of Policy Development and Research*. 1996(21): 41-48.

Schafer, Robert and Helen F. Ladd, *Discrimination in Mortgage Lending*. Cambridge, MA: Massachusetts Institute of Technology, 1981.

Van Order, Robert, Ann-Margaret Westin, and Peter Zorn, "Effects of Racial Composition of Neighborhoods on Defaults, and Implications for Racial Discrimination in Mortgage Markets," Paper presented at the ASSA meetings in Anaheim, California, January 1993.

Yezer, Anthony M.J., Robert F. Phillips, and Robert P. Trost, "Bias in Estimates of Discrimination and Default in Mortgage Lending: The Effects of Simultaneity and Self-Selection," *Journal of Real Estate Finance and Economics*. 9:4. (1994): 197-215.

Yinger, John, "Measuring Racial Discrimination with Fair Housing Audits: Caught in the Act," *American Economic Review*. 76:5. (1986): 881-99.

Yinger, John, "Discrimination in Mortgage Lending: A Literature Review." In John Goering and Ron Wienk, eds. *Mortgage Lending, Racial Discrimination, and Federal Policy*, Washington, D.C.: Urban Institute Press, 1996a (29-74).

Yinger, John, "Why Default Rates Cannot Shed Light on Mortgage Discrimination," *Cityscape: A Journal of Policy Development and Research*. 1996b(21): 25-32.

10 Appendices

10.1 Appendix A: Prosper Interface

Figure 1: Prosper.com Homepage

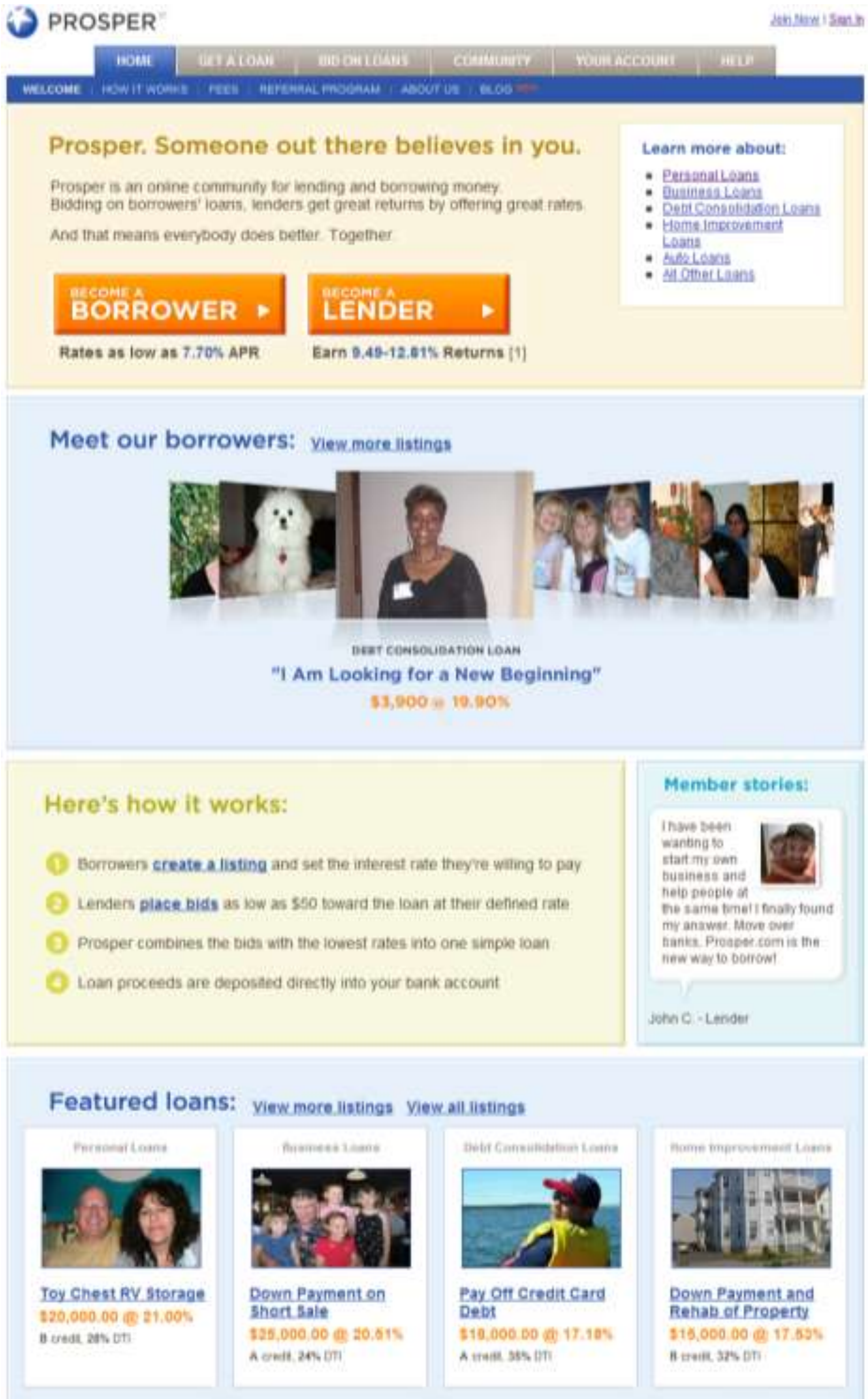


Figure 2: Prosper.com Sample Search Results

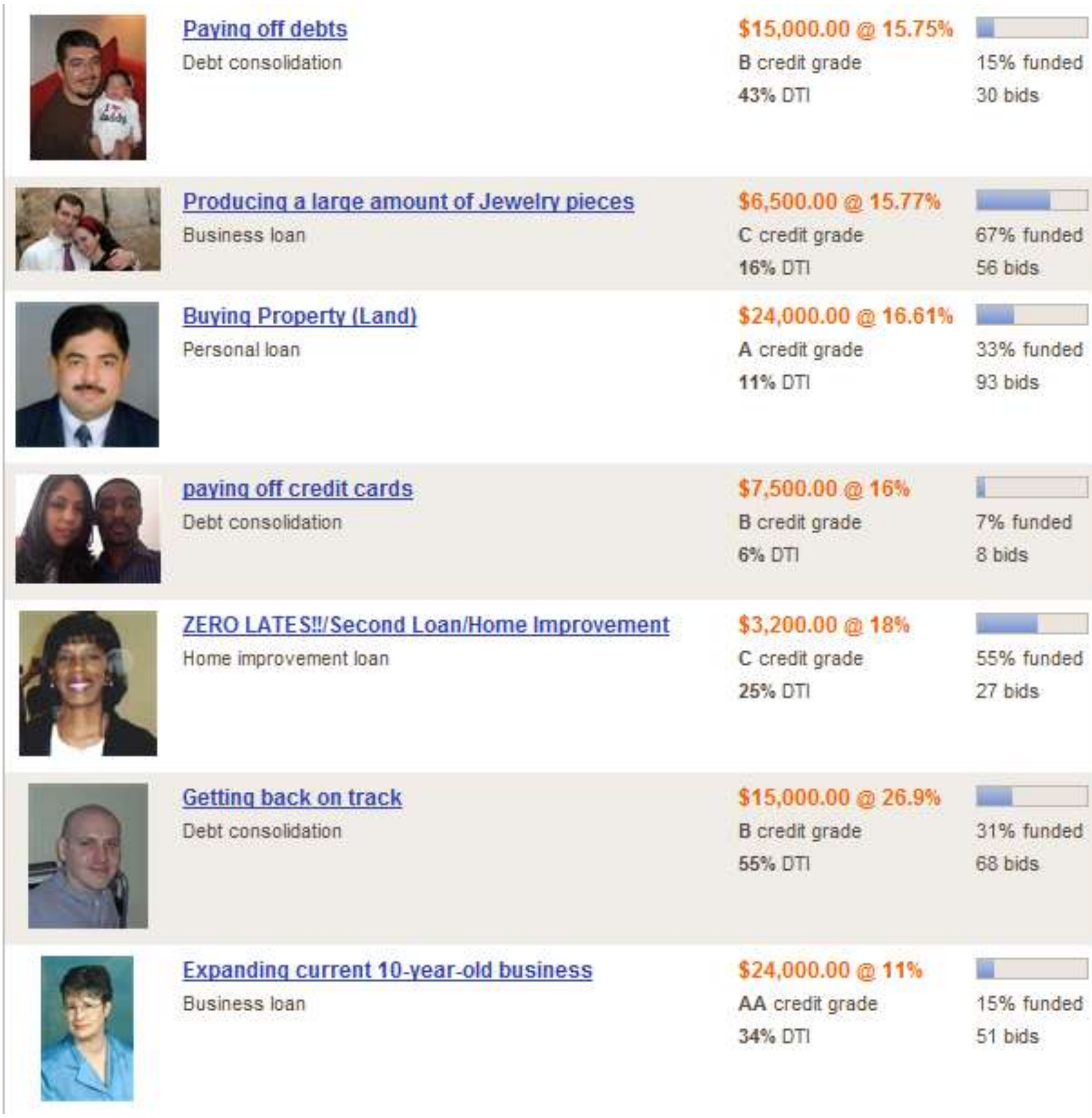


Figure 3: Prosper.com Sample Listing Page

10.2 Appendix B: Variable Descriptions

Table 1: Variables for all Listings

Variable	Description
Amount	Amount of money requested by a borrower in a listing.
Group Member	Equal to 1 if the borrower is a member of a group, 0 otherwise.
Verified Bank Account	Equal to 1 if the borrower has verified ownership of a bank account; 0 otherwise.
Home Owner	Equal to 1 if the borrower has verified ownership of a home; 0 otherwise.
Endorsed	Equal to 1 if the borrower has received an endorsement from another member.
Close Immediately	Equal to 1 if the listing closes when it is fully funded; 0 if it stays for a few days.
New Listing	Equal to 1 if this is the borrower's first listing; 0 otherwise.
Funded	Equal to 1 if the listing becomes a funded loan; 0 otherwise.
Listing Number	Sequential identifier uniquely associated with a listing.
Opening Rate	The initial rate the borrower sets when creating a listing after receiving advice from Prosper's algorithm of funding likelihood based on credit history.
Borrower State	Borrower's state of residence; loans are subject to state usury laws.

* Available for all 34,804 listings.

Table 2: Variables for Core Credit

Variable	Description
Debt to Income Ratio	Current debt of the borrower as per credit report divided by current income, as verified via W2 statement or employer's letter.
Credit Grade	Prosper groups credit scores into ranges and presents to lenders a grade rather than a number. HR (520-559), E (560-599), D (600-639), C (640-679), B (680-719), A (720-759), AA (760 and above).
Current Delinquencies	Number of current delinquencies in a borrower's credit report.
Delinquencies 7 Yrs	Number of delinquencies in the last seven years of a borrower's credit report.
Public Records 10 Yrs	Number of public records in the last ten years of a borrower's credit report.
Total Credit Lines	Number of credit lines appearing in a borrower's credit report.
Inquiries 6 Mos	Number of inquiries into a borrower's credit history contained in credit report.

* Available for 33,884 (97.4%) listings.

Table 3: Variables for Extended Credit and Employment

Variable	Description
Amount Delinquent	Total amount of borrower's debt that is currently delinquent.
Public Records 12 Mos	Number of public records in last twelve months of a borrower's credit report.
Current Credit Lines	Number of credit lines that are current (not delinquent).
Open Credit Lines	Number of credit lines that are open on a borrower's credit report.
Revolving Credit Bal	Amount of Revolving Credit Bal on a borrower's credit report.
Bankcard Utilization	Percentage of available revolving credit that is being utilized by the borrower.
Employment Status	One of five possible employment status options: not employed, part-time, full-time, self-employed, or retired.

Length of Status	Length of the above employment status in months.
Income Range	One of eight possible income ranges: Not displayed, $0 or unable to verify, $1-24,999, $25,000-49,999, $50,000-74,999, $75,000-99,999, $100,000+, or Not Employed
Borrower Occupation	One of sixty-seven possible occupational categories.

* Available for 22,227 (63.9%) listings.

Table 4: Variables for Loans

Variable	Description
Age	The current age in months of the loan in months (time since its origination.) This is held constant once a loan is paid or defaults.
Borrower Rate	The interest rate the borrower pays on a loan.
Lender Rate	The interest rate the lender receives on a loan; usually higher than *Borrower Rate* due to Prosper's fees and other surcharges.
Loan Status	One of eight different loan states: paid, current, late, 1 month late, 2 months late, 3 months late, 4+ months late, defaulted. A loan that is \`4+ months late\` is already considered uncollectable, but is renamed \`defaulted\` once it is sold to a debt buyer manually and in bulk.

* Only applicable to funded listings.

10.3 Appendix C: Summary Statistics

Table 1: All Averages for Coding Types

Averages	Control	Unknown	Identified
Observations	13109	11517	10178
Amount	6997.470	7033.675	6534.878
Group Member*	0.306	0.542	0.576
Verified Bank Account	0.491	0.729	0.723
Home Owner	0.372	0.403	0.306
Endorsed	0.041	0.123	0.156
Close Immediately	0.426	0.347	0.324
New Listing*	0.454	0.471	0.452
Funded	0.364	0.578	0.586
Opening Rate	0.180	0.193	0.196
Debt to Income Ratio	0.532	0.449	0.461
Current Delinquencies	3.805	3.052	3.037
Delinquencies 7 Yrs	11.381	8.956	9.233
Public Records 10 Yrs	0.691	0.598	0.543
Total Credit Lines	25.137	23.969	23.391
Inquiries 6 Mos	4.154	3.858	3.751
Amount Delinquent	3512.947	2541.312	2600.006
Public Records 12 Mos*	0.081	0.065	0.058
Current Credit Lines*	8.391	8.732	8.517
Open Credit Lines*	7.259	7.500	7.264
Revolving Credit Bal	11537.690	13392.694	10846.676
Bankcard Utilization	0.610	0.582	0.593
Length of Status	48.7	43.5	37.6

* Variables not included in Figure 4.1.

Table 2: All Averages for Races

Averages	White	Black	Hispanic	Asian
Observations	7068	2047	654	409
Male	0.594	0.431	0.491	0.577
Amount	6698.008	5865.134	6430.911	7234.048
Group Member*	0.570	0.606	0.534	0.584
Verified Bank Account	0.758	0.626	0.609	0.775
Home Owner	0.326	0.276	0.242	0.225
Endorsed	0.165	0.131	0.116	0.193
Close Immediately	0.304	0.382	0.381	0.271
New Listing*	0.448	0.467	0.472	0.416
Funded	0.633	0.444	0.456	0.709
Opening Rate	0.194	0.202	0.197	0.200
Debt to Income Ratio	0.464	0.474	0.397	0.463
Current Delinquencies	2.662	4.428	3.324	2.151
Delinquencies 7 Yrs	8.605	11.929	8.764	7.473
Public Records 10 Yrs	0.543	0.648	0.390	0.262
Total Credit Lines	23.949	22.180	21.903	22.156
Inquiries 6 Mos	3.421	4.473	4.637	4.448
Amount Delinquent	2253.898	3975.235	2345.994	2848.449
Public Records 12 Mos*	0.056	0.087	0.014	0.030
Current Credit Lines*	8.824	7.195	8.616	8.909
Open Credit Lines*	7.493	6.196	7.527	7.687
Revolving Credit Bal	11781.053	6936.112	10658.975	12232.925
Bankcard Utilization	0.588	0.595	0.632	0.621
Length of Status	36.4	45.8	31.4	29.1

* Variables not included in Figure 4.1.

Table 3: Averages for Race-Funded Pairs

	White		Black		Hispanic		Asian	
Averages	Unfunded	Funded	Unfunded	Funded	Unfunded	Funded	Unfunded	Funded
Observations	2596	4472	1138	909	356	298	119	290
Male	0.55	0.62	0.42	0.44	0.45	0.54	0.56	0.58
Amount	7396.30	6292.64	6162.63	5492.68	6885.17	5888.23	8304.57	6794.76
Group Member	0.475	0.626	0.511	0.725	0.447	0.638	0.403	0.659
Verified Bank Account	0.414	0.958	0.372	0.945	0.326	0.946	0.353	0.948
Home Owner	0.278	0.354	0.200	0.370	0.197	0.295	0.210	0.231
Endorsed	0.099	0.202	0.076	0.200	0.090	0.148	0.050	0.252
Close Immediately	0.350	0.278	0.423	0.331	0.388	0.372	0.353	0.238
New Listing	0.656	0.327	0.670	0.212	0.646	0.265	0.706	0.297
Opening Rate	0.182	0.201	0.182	0.225	0.184	0.213	0.190	0.205
Debt to Income Ratio	0.553	0.412	0.607	0.308	0.476	0.302	0.558	0.424
Current Delinquencies	4.154	1.798	5.614	2.978	4.236	2.221	3.559	1.570
Delinquencies 7 Yrs	12.715	6.222	14.320	9.008	10.074	7.179	12.280	5.490
Public Records 10 Yrs	0.736	0.432	0.773	0.496	0.473	0.290	0.415	0.199
Total Credit Lines	25.327	23.150	21.946	22.466	22.493	21.190	24.576	21.157
Inquiries 6 Mos	4.121	3.015	4.625	4.287	5.043	4.145	5.186	4.143
Amount Delinquent	3589.03	1342.77	5300.03	2220.21	2983.18	1456.49	4545.03	1975.92
Public Records 12 Mos	0.079	0.041	0.112	0.053	0.014	0.013	0.022	0.034
Current Credit Lines	8.087	9.328	6.309	8.368	8.264	9.107	8.989	8.869
Open Credit Lines	7.000	7.829	5.577	7.016	7.361	7.758	7.867	7.594
Revolving Credit Bal	9345.57	13443.67	4928.95	9595.09	8293.33	13961.34	10431.90	13159.16
Bankcard Utilization	0.656	0.541	0.610	0.575	0.656	0.598	0.726	0.567
Length of Status	33.9	38.1	42.6	50.1	32.1	30.5	28.4	29.6

Table 4: Groups for Race-Funded Pairs

	White		Black		Hispanic		Asian	
Credit Rating	**Unfunded**	**Funded**	**Unfunded**	**Funded**	**Unfunded**	**Funded**	**Unfunded**	**Funded**
AA	1.77%	12.09%	0.15%	3.96%	1.44%	4.70%	0.00%	8.57%
A	2.31%	11.98%	1.05%	8.12%	0.96%	8.05%	3.33%	9.71%
B	4.40%	15.20%	2.09%	12.87%	3.37%	15.44%	3.33%	24.57%
C	10.04%	19.89%	6.58%	19.60%	10.10%	24.83%	11.11%	14.86%
D	17.61%	19.71%	12.26%	23.96%	19.23%	20.81%	17.78%	17.71%
E	20.40%	9.52%	19.58%	12.28%	13.94%	14.77%	18.89%	11.43%
HR	43.48%	11.61%	58.30%	19.21%	50.96%	11.41%	45.56%	13.14%
NC	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Employment	**Unfunded**	**Funded**	**Unfunded**	**Funded**	**Unfunded**	**Funded**	**Unfunded**	**Funded**
Full-time	84.11%	85.57%	85.05%	88.12%	85.10%	87.92%	81.11%	89.71%
Not employed	1.45%	0.88%	1.20%	1.19%	0.96%	0.00%	0.00%	0.57%
Part-time	4.35%	4.14%	4.19%	4.16%	5.77%	6.71%	7.78%	5.14%
Retired	2.20%	1.68%	3.29%	1.39%	1.92%	0.00%	0.00%	0.00%
Self-employed	7.89%	7.73%	6.28%	5.15%	6.25%	5.37%	11.11%	4.57%
Income	**Unfunded**	**Funded**	**Unfunded**	**Funded**	**Unfunded**	**Funded**	**Unfunded**	**Funded**
Not Displayed	0.75%	1.21%	0.60%	0.99%	1.92%	2.01%	1.11%	0.57%
$1-24,999	18.52%	13.96%	23.92%	13.27%	24.04%	18.12%	23.33%	9.71%
$25,000-49,999	44.87%	39.85%	47.98%	43.96%	46.63%	40.94%	40.00%	43.43%
$50,000-74,999	20.83%	24.76%	18.68%	25.74%	18.75%	22.15%	23.33%	24.57%
$75,000-99,999	8.70%	10.11%	4.48%	11.68%	6.25%	11.41%	8.89%	13.71%
$100,000+	5.21%	9.45%	3.29%	3.56%	1.92%	5.37%	3.33%	7.43%
Not Employed	1.13%	0.66%	1.05%	0.79%	0.48%	0.00%	0.00%	0.57%

Table 5: Current and Defaulted Loans

	White		Black		Hispanic		Asian	
Averages	**Current**	**Default**	**Current**	**Default**	**Current**	**Default**	**Current**	**Default**
Observations	3304	403	590	182	208	35	199	31
Male	0.61	0.53	0.44	0.38	0.50	0.49	0.59	0.42
Amount	6606.38	4742.59	5651.89	4430.59	6374.04	4610.77	6934.85	4368.58
Borrower Rate	0.172	0.234	0.197	0.236	0.191	0.234	0.169	0.264
Age	8.831	10.012	9.383	9.630	9.663	10.588	8.563	10.452
Group Member	0.592	0.824	0.675	0.857	0.606	0.829	0.633	0.806
Verified Bank Account	0.997	0.720	0.995	0.808	0.986	0.743	0.995	0.774
Home Owner	0.365	0.241	0.407	0.236	0.279	0.314	0.241	0.097
Endorsed	0.194	0.176	0.198	0.176	0.144	0.171	0.281	0.097
Close Immediately	0.238	0.536	0.283	0.467	0.313	0.571	0.161	0.581
New Listing	0.332	0.203	0.229	0.148	0.288	0.200	0.296	0.194
Opening Rate	0.199	0.246	0.220	0.248	0.212	0.251	0.200	0.272
Debt to Income Ratio	0.415	0.451	0.298	0.420	0.289	0.237	0.483	0.187
Current Delinquencies	1.515	4.575	2.123	6.242	1.877	5.273	1.312	5.071
Delinquencies 7 Yrs	6.126	9.000	8.878	9.730	6.333	16.273	4.518	15.714
Public Records 10 Yrs	0.426	0.653	0.478	0.618	0.284	0.364	0.191	0.286
Total Credit Lines	23.436	20.630	23.159	20.242	20.623	22.879	21.322	20.786
Inquiries 6 Mos	2.710	4.790	3.795	5.483	3.912	4.303	3.477	5.464
Amount Delinquent	1276.10	3207.95	2299.23	2774.64	1513.03	2295.83	2068.97	3508.33
Public Records 12 Mos	0.040	0.071	0.046	0.156	0.016	0.000	0.041	0.000
Current Credit Lines	9.362	7.988	8.794	5.667	8.914	10.167	8.791	6.333
Open Credit Lines	7.870	6.894	7.369	4.889	7.594	9.667	7.541	5.667
Revolving Credit Bal	13251.27	10406.79	10141.43	2326.93	12440.09	44912.17	13776.03	7712.00
Bankcard Utilization	0.550	0.559	0.602	0.422	0.584	0.750	0.553	0.930
Length of Status	36.2	40.0	51.9	33.8	28.5	52.7	29.6	18.0
Credit Rating	**Current**	**Default**	**Current**	**Default**	**Current**	**Default**	**Current**	**Default**
AA	10.05%	0.00%	3.60%	1.70%	4.26%	0.00%	8.05%	0.00%
A	10.63%	1.50%	6.13%	0.57%	7.63%	3.23%	10.68%	0.00%
B	14.79%	2.01%	11.32%	3.27%	11.93%	3.15%	21.16%	3.41%
C	18.58%	8.46%	18.73%	7.81%	23.00%	9.16%	14.08%	3.47%
D	19.54%	13.49%	21.81%	10.49%	17.11%	23.51%	17.58%	3.47%
E	12.28%	23.50%	16.06%	18.72%	17.03%	29.87%	13.21%	35.47%
HR	13.80%	49.72%	21.98%	54.45%	18.03%	28.03%	15.23%	54.18%
NC	0.33%	1.32%	0.36%	2.98%	1.03%	3.05%	0.00%	0.00%

10.4 Appendix D: Regression Tables

Table 1: Effect of Race on Funding Decision
(Probit Derivatives)

		(1) Funded	(2) Funded	(3) Funded	(4) Funded
Variables of Interest	Black	-0.204866**	-0.203961**	-0.178234**	-0.125009**
		(0.018176)	(0.023564)	(0.047599)	(0.041311)
	Hispanic	-0.159991**	-0.154876**	-0.191535**	-0.123735*
		(0.028636)	(0.035928)	(0.059213)	(0.062294)
	Asian	0.062216+	0.091063*	0.048828	0.059362
		(0.033059)	(0.042239)	(0.088507)	(0.080406)
Interaction Effects	Female	-0.065113**	-0.041971**	0.028710	0.026310
		(0.012039)	(0.015266)	(0.032448)	(0.026908)
	Black * Female	0.045129+	0.034213	0.001046	0.120016*
		(0.024322)	(0.031714)	(0.063244)	(0.057102)
	Hispanic * Female	-0.023737	0.033185	-0.017903	0.073382
		(0.040611)	(0.050956)	(0.092445)	(0.093633)
	Asian * Female	0.045475	0.046035	0.046567	0.037346
		(0.051243)	(0.066505)	(0.140699)	(0.115762)
	More Info			0.006998	
				(0.033686)	
	Black * More Info			0.060361	
				(0.059008)	
	Hispanic* More Info			0.038868	
				(0.083185)	
	Asian* More Info			0.018310	
				(0.114450)	
	Female * More Info			-0.002378	
				(0.039644)	
	Black * Female * More Info			0.088422	
				(0.077725)	
	Hispanic * Female * More Info			0.074610	
				(0.119558)	
	Asian * Female * More Info			-0.031378	
				(0.180746)	
Listing Controls	Opening Rate		0.908953**	4.225468**	4.501979**
			(0.094628)	(0.203726)	(0.279806)
	Amount		-0.000006**	-0.000033**	-0.000037**
			(0.000001)	(0.000002)	(0.000002)
	Close Immediately		-0.077281**	0.084079**	0.064538*
			(0.014145)	(0.017469)	(0.026114)
	Group Member		-0.019761	0.067806**	0.019976
			(0.014265)	(0.018178)	(0.024662)
	Verified Bank Account		0.698726**	0.599215**	0.659751**
			(0.008390)	(0.013703)	(0.014592)

	Home Owner		0.131695**	0.015387	0.010592
			(0.013625)	(0.019486)	(0.026917)
	Endorsed		0.050234**	0.074217**	0.070913*
			(0.017608)	(0.021210)	(0.030777)
	New Listing		-0.420899**	-0.522791**	-0.543491**
			(0.011354)	(0.013242)	(0.017597)
	Listing Number		-0.000001**	-0.000002**	-0.000003**
			(0.000000)	(0.000000)	(0.000000)
Core Credit Controls	Credit Grade FEs			Y	Y
	Debt to Income Ratio			-0.039739**	-0.028870**
				(0.005634)	(0.008022)
	Current Delinquencies			-0.005276*	-0.002367
				(0.002170)	(0.004144)
	Delinquencies 7 Yrs			-0.002200**	-0.003627**
				(0.000597)	(0.000921)
	Public Records 10 Yrs			-0.036852**	-0.027045*
				(0.007803)	(0.012724)
	Total Credit Lines			-0.002920**	-0.003007*
				(0.000664)	(0.001194)
	Inquiries 6 Mos			-0.002391	-0.004067
				(0.001751)	(0.002702)
	Borrower State FEs			Y	Y
Extended Credit Controls	Amount Delinquent				-0.000003+
					(0.000002)
	Public Records 12 Mos				-0.088235*
					(0.041530)
	Current Credit Lines				0.012630*
					(0.006041)
	Open Credit Lines				-0.019127**
					(0.006524)
	Revolving Credit Bal				0.000000
					(0.000000)
	Bankcard Utilization				0.010582
					(0.031272)
Employment Controls	Employment Status FEs				Y
	Length of Status				-0.000202
					(0.000194)
	Income Range FEs				Y
	Borrower Occupation FEs				Y
	Observations	10178	10178	9910	6371
	Pseudo R-squared	0.0245	0.429	0.615	0.669

Standard errors in parentheses
** $p<0.01$, * $p<0.05$, + $p<0.1$

Table 2: Effect of Race on Interest Rate
(OLS Estimators)

		(1) Lender Rate	(2) Lender Rate	(3) Lender Rate	(4) Lender Rate
Variables of Interest	Black	0.030139**	0.022649**	0.009380**	0.011238**
		(0.003340)	(0.002928)	(0.002811)	(0.002351)
	Hispanic	0.024799**	0.012599**	0.003670	-0.000446
		(0.005072)	(0.004440)	(0.003879)	(0.003784)
	Asian	0.001819	-0.001105	0.000529	-0.002127
		(0.004957)	(0.004338)	(0.004358)	(0.003282)
Interaction Effects	Female	0.017335**	0.010100**	-0.001880	-0.001284
		(0.001925)	(0.001691)	(0.001718)	(0.001362)
	Black * Female	-0.009376*	-0.004240	-0.004572	-0.003027
		(0.004600)	(0.004020)	(0.003817)	(0.003156)
	Hispanic * Female	-0.017068*	-0.006915	-0.000736	0.002622
		(0.007522)	(0.006573)	(0.005873)	(0.005410)
	Asian * Female	0.002467	0.005918	0.010585	0.003500
		(0.007695)	(0.006721)	(0.006549)	(0.005186)
	More Info			-0.006201**	
				(0.001782)	
	Black * More Info			0.001311	
				(0.003640)	
	Hispanic* More Info			-0.003029	
				(0.005487)	
	Asian* More Info			-0.004420	
				(0.005480)	
	Female * More Info			0.002738	
				(0.002145)	
	Black * Female * More Info			0.002532	
				(0.005008)	
	Hispanic * Female * More Info			0.002977	
				(0.008094)	
	Asian * Female * More Info			-0.005654	
				(0.008450)	
Listing Controls	Amount		-0.000000+	0.000002**	0.000002**
			(0.000000)	(0.000000)	(0.000000)
	Close Immediately		0.049464**	0.030448**	0.031447**
			(0.001638)	(0.001040)	(0.001390)
	Group Member		0.004488**	-0.006145**	-0.005901**
			(0.001712)	(0.001069)	(0.001299)
	Verified Bank Account		-0.029048**	-0.004595*	-0.006723*
			(0.003428)	(0.002126)	(0.003243)
	Home Owner		-0.024669**	-0.002336*	-0.002311+
			(0.001550)	(0.001080)	(0.001363)
	Endorsed		-0.005979**	-0.005670**	-0.003955**
			(0.001826)	(0.001129)	(0.001432)
	New Listing		-0.026527**	-0.004517**	-0.006623**
			(0.001628)	(0.001048)	(0.001307)

	Listing Number		0.000000**	0.000000**	0.000000**
			(0.000000)	(0.000000)	(0.000000)
Core Credit Controls	Credit Grade FEs			Y	Y
	Debt to Income Ratio			0.001954**	0.001109*
				(0.000356)	(0.000462)
	Current Delinquencies			0.000451**	0.000467+
				(0.000146)	(0.000256)
	Delinquencies 7 Yrs			0.000162**	0.000227**
				(0.000040)	(0.000056)
	Public Records 10 Yrs			0.001614**	0.001890**
				(0.000516)	(0.000706)
	Total Credit Lines			0.000085*	0.000199**
				(0.000037)	(0.000062)
	Inquiries 6 Mos			0.000536**	0.000665**
				(0.000112)	(0.000157)
	Borrower State FEs			Y	Y
Extended Credit Controls	Amount Delinquent				0.000000
					(0.000000)
	Public Records 12 Mos				0.002109
					(0.002584)
	Current Credit Lines				-0.000528+
					(0.000309)
	Open Credit Lines				0.000683*
					(0.000332)
	Revolving Credit Bal				-0.000000*
					(0.000000)
	Bankcard Utilization				0.001080
					(0.001737)
Employment Controls	Employment Status FEs				Y
	Length of Status				-0.000010
					(0.000009)
	Income Range FEs				Y
	Borrower Occupation FEs				Y
	Constant	0.165971**	0.195122**	0.090164**	0.082434**
		(0.001190)	(0.004162)	(0.003241)	(0.007640)
	Observations	5969	5969	5830	3558
	R-squared	0.041199	0.270508	0.738425	0.757758

Standard errors in parentheses
** p<0.01, * p<0.05, + p<0.1

Table 3: Effect of Race on Default

(Probit Derivatives)

		(1) Default	(2) Default	(3) Default	(4) Default
Variables of Interest	Black	0.078722**	0.045942**	0.041962**	0.038175*
		(0.019330)	(0.016450)	(0.014301)	(0.015086)
	Hispanic	0.018312	0.002568	0.014110	0.002983
		(0.030335)	(0.017422)	(0.024066)	(0.016941)
	Asian	-0.025563	-0.011990	-0.022722	-0.011094
		(0.019965)	(0.010778)	(0.013909)	(0.010669)
Interaction Effects	Female	0.013452	0.009247	0.003807	0.009457
		(0.010374)	(0.007276)	(0.008036)	(0.007024)
Interest Rate	Borrower Rate			0.689445**	0.300593**
				(0.066213)	(0.108244)
Listing Controls	Amount		0.000002**		0.000001
			(0.000001)		(0.000001)
	Close Immediately		0.026731*		0.011521
			(0.011744)		(0.009920)
	Group Member		0.002126		0.002079
			(0.007903)		(0.007595)
	Verified Bank Account		-0.300822**		-0.280237**
			(0.062643)		(0.061890)
	Home Owner		0.002862		0.002543
			(0.008230)		(0.007849)
	Endorsed		0.016641+		0.017767+
			(0.009213)		(0.009246)
	New Listing		-0.001199		-0.000340
			(0.008390)		(0.008188)
Core Credit Controls	Credit Grade FEs		Y		Y
	Debt to Income Ratio		0.003408+		0.002980
			(0.001995)		(0.001920)
	Current Delinquencies		0.001599		0.001404
			(0.001170)		(0.001121)
	Delinquencies 7 Yrs		-0.000347		-0.000417
			(0.000300)		(0.000292)
	Public Records 10 Yrs		-0.000872		-0.001017
			(0.004068)		(0.003830)
	Total Credit Lines		0.000105		0.000040
			(0.000355)		(0.000339)
	Inquiries 6 Mos		0.001916**		0.001704*
			(0.000726)		(0.000693)
	Borrower State FEs		Y		Y
Extended Credit Controls	Amount Delinquent		-0.000001		-0.000001
			(0.000001)		(0.000001)
	Public Records 12 Mos		0.016605		0.014721
			(0.012045)		(0.011536)
	Current Credit Lines		-0.001956		-0.001967
			(0.001894)		(0.001810)

	Open Credit Lines		0.001791		0.001748
			(0.001948)		(0.001851)
	Revolving Credit Bal		0.000000+		0.000000+
			(0.000000)		(0.000000)
	Bankcard Utilization		-0.019600*		-0.017666*
			(0.009362)		(0.008955)
Employment Controls	Employment Status FEs		Y		Y
	Length of Status		-0.000189**		-0.000178**
			(0.000072)		(0.000068)
	Income Level FEs		Y		Y
	Borrower Occupation FEs		Y		Y
	Observations	2197	1751	2197	1751
	Pseudo R-squared	0.0300	0.373	0.139	0.382

Standard errors in parentheses
** p<0.01, * p<0.05, + p<0.1