Essays in Financial Economics

Assaf Aharon Shtauber

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
under the Executive Committee
of the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2013

© 2013
Assaf A. Shtauber
All rights reserved

ABSTRACT

Essays in Financial Economics

Assaf A. Shtauber

This dissertation consists of three essays in Financial Economics. The first essay concerns low-income Americans' usage of financial services. Approximately 20% of US households earning less than $30,000 a year do not have checking or savings accounts. Rather than operate within the financial mainstream, these households tend to rely on high-fee "alternative financial services" such as check-cashing and payday loans for carrying out basic financial transactions, for accessing credit, and even for saving. In this essay I examine the effects of opening a bank account, thereby gaining access to mainstream financial services, on financial behavior and "financial wellbeing". Despite significant policy efforts to "bank the poor" in recent years, there has hardly been any research about this question.

I use a unique dataset of low-income participants of financial education workshops, and exploit quasi-random variation in the likelihood of opening a bank account after the workshop to overcome selection bias and estimate causal effects. I find that opening an account reduces delinquency and raises the likelihood of credit score improvement. However, contrary to the benefits often ascribed to becoming banked, I find no effects on saving, self-reported overspending, and several measures of "financial wellbeing" such as finances-related stress. Surprisingly, I find only small aggregate effects of opening an account on actual usage of mainstream credit, as measured by credit card ownership, for example. There is heterogeneity in these effects, however, based on an individual's level of financial literacy. Those who graduated high school and who are presumably more financially literate do in fact increase their usage of mainstream credit when they open an account, but those who did not graduate high school do not. This finding is consistent with prior evidence on the links between financial literacy and usage of financial services: the less literate have been shown to be more likely to use alternative financial services, and those who do use mainstream services have been shown to make costly mistakes (e.g. pay high credit card fees). Finally, I find that access to mainstream financial services enhances the effectiveness of financial education: among the workshop participants that I study, opening an account increases self-reported financial literacy.

In the second essay I explore the theoretical effects of one of the most important consequences

of entering the mainstream financial system: facing a higher rate of return on saving and a lower cost of borrowing. I develop a simple two period life-cycle model of consumption in which voluntary default is possible and examine the effects of favorable changes in saving and borrowing rates on consumer behavior. The incorporation of default in the model is important for its applicability to the effects of entering the mainstream financial system since those who operate outside the mainstream system tend to be low-income individuals who are more prone to default. It is also novel: the effects of interest rate changes on consumer behavior ("the interest elasticity of consumption") are of great importance in economics and have been researched extensively, but the implications of incorporating default in the basic life-cycle model have never been studied.

I find that when the cost of default is not sufficiently dependent on the amount defaulted upon, the possibility of default weakens the link between first period consumption and second period utility and leads to overconsumption relative to the no-default model. It also results in a counter-intuitive *negative* marginal propensity to consume out of wealth: those who are wealthier consume *less*. It follows that the wealth effects of favorable interest rate changes imply less rather than more consumption and that such rate changes (the ultimate effects of which are determined by the combination of wealth and substitution effects) are more likely to encourage saving and to discourage borrowing than in the no-default model. Favorable rate changes decrease the probability of default and the expected defaulted-upon amount for all savers, who may default on a pre-existing obligation in the model, as well as for borrowers who initially borrow more than some threshold. I extend the model to allow for partial repayment of debt (i.e. delinquency rather than full-scale default) in both periods. I show that decreasing the borrowing rate lowers delinquency by affecting the tradeoff between delinquency and borrowing as means to finance first period consumption.

The third essay, co-authored with Andrew Ang and Paul Tetlock, examines asset pricing patterns in over-the-counter (OTC) stocks, which are stocks that trade on either the OTC Bulletin Board (OTCBB) or OTC Link (formerly Pink Sheets, or PS) interdealer quotation system. Compared to stocks that trade on the NYSE, Amex, and NASDAQ ("listed stocks"), OTC stocks are far less liquid, disclose less information, and exhibit lower institutional holdings. We exploit these different market conditions to test theories of cross-sectional return premiums. Compared to return premiums in listed markets, the OTC premium for illiquid stocks is several times higher, the OTC premiums for size, value, and volatility are similar, and the OTC premium for momentum is three times lower. The OTC premiums for illiquidity, size, value, and volatility are largest among stocks that are held almost exclusively by retail investors and those that do not disclose financial information. Theories of differences in investors' opinions and limits on short sales help to explain these return premiums. Our momentum results are most consistent with Hong and Stein's (1999) theory based on the gradual diffusion of information.

Contents

List of Figures

List of Tables

Acknowledgments

I would like to thank all of those who have contributed to the completion of this dissertation. I thank Gur Huberman for continual guidance, support, and advice. His patience and wise suggestions were essential to my progress. I am grateful to Andrew Ang and Paul Tetlock for inviting me to co-author the OTC paper with them at an early stage of the PhD program. Working with them has been a great learning experience for me. I have had the privilege of working especially closely with Paul, and I thank him for all that he has taught me. I am thankful to Steve Zeldes for taking interest in my work and for his invaluable guidance on the development of the model described in Chapter 2. I thank Kent Daniel for his kindness and encouragement and for extending very helpful advice at several key points along the way.

I am grateful to the management and staff of Neighborhood Trust Financial Partners for their generosity in sharing data with me, in collecting more data for the purpose of my research, and in answering my questions.

My classmates at Columbia have been great colleagues and friends. I am thankful to them and to all of my other friends, in Israel and in New York City, for keeping me grounded and for serving as a constant reminder that there is more to life than research.

My parents Nitza and Zvi have always motivated me and were very influential in my decision to pursue a PhD. I am grateful to them and to my sister Chen and brother Matan for being a constant source of unconditional love and support, throughout my whole life and in particular during these years in which I was far away from home.

Finally, I would like to try and express the depth of my gratitude to my wife Anat for the incredible support that she has provided me with during the past five years. Her wisdom, strength, and unwavering faith in me kept me on track and helped me through every obstacle that I encountered. I cannot imagine having completed this dissertation without her standing lovingly by my side. I dedicate it to her.

To Anat

Chapter 1

The Effects of Access to Mainstream Financial Services on the Poor: Evidence from Data on Recipients of Financial Education

"When we talk about finance we must also talk about individual access to financial services. Like money itself, the benefits that a bank account provides are easy to take for granted until you do not have one. And today, in the age of the Internet, derivatives, and embedded options, between 10 and 20 percent of American households still lack that basic passport to the broader economy. If it was an important national challenge half a century ago to ensure that essentially every American had access to electricity, to running water, and to a telephone - in [the] new economy, ensuring access to a basic bank account must be a critical challenge for today"

– Larry Summers, then Secretary of the Treasury, 2000

1.1. Introduction

Many American households operate outside of the financial mainstream: 8% of households, nearly 10 million in total, do not have checking or savings accounts. More than 80% of these households earn less than $30,000 a year.[1] These households, as well as a sizable fraction of households who do have accounts, tend to rely on high-fee "alternative financial services" such as check-cashing and payday loans for carrying out basic financial transactions, for accessing credit, and even for saving.[2] These services are not only expensive but also do not seem to represent a good substitute for mainstream financial services, and the American poor are generally considered to be credit and savings constrained (Barr (2004)). There is ample evidence to support this notion. For example, 57% of Americans earning less than $20,000 a year reported that they would not be able to come up with $2,000 in 30 days in a recent survey (Lusardi, Schneider, and Tufano (2011)). Further, only 32% of American families in the lowest income quintile reported saving over the previous year in the 2010 Survey of Consumer Finances, compared to 60%-80% in the top two quintiles (Bricker, Kennickell, Moore, and Sabelhaus (2012)). The association of figures like these with the amount of households operating outside the financial mainstream, together with the high costs and perceived predatory nature of alternative financial services, has motivated significant policy efforts to "bank the poor". These efforts have concentrated on both the supply and the demand sides, providing incentives for banks to serve the poor and in particular to offer accounts that are more suitable for their needs while encouraging the unbanked to open accounts, often through financial literacy programs.[3] Success of these efforts has been limited, however (Lyons and Scherpf (2004)). Prescott and Tatar (1999) argue that the benefits of account ownership are often overstated and that many unbanked individuals in fact cash their checks at no cost. The FDIC's 2011 National Survey of Unbanked and Underbanked Households reports that 21% of unbanked households have in fact never used alternative financial services (Federal Deposit Insurance Corporation (2012)).

In this paper I seek to estimate the causal effects of bank account ownership on the financial behavior and "financial wellbeing" of the poor. Given the amount of policy attention and the potential welfare implications, as well as the implications for our general understanding of household financial behavior, it is somewhat surprising that there has been almost no research on this

[1] The fraction of households earning less than $30,000 a year which are unbanked is 19.5%. Figures are taken from the FDIC's 2011 National Survey of Unbanked and Underbanked Households.

[2] 20.1% of American households have checking or savings accounts but regularly use alternative financial services, and are referred to as "underbanked" (Federal Deposit Insurance Corporation (2012)).

Barr (2004) notes anecdotal evidence suggesting that low-income households use money orders (as well as over-withholding of income taxes) as a saving vehicle.

[3] For example: the Community Reinvestment Act of 1977, various FDIC programs (http://economicinclusion.gov/) and "Bank On" initiatives (http://joinbankon.org/about/).

The main reasons that have been cited for poor households being unbanked include incompatibilities between mainstream banking services and their needs (e.g. fee structure, minimum balance requirements, check holding periods, lack of geographical access), cultural reasons (mistrust, dislike of dealing with banks), and past delinquencies (Barr (2004), Federal Deposit Insurance Corporation (2012), Bricker, Kennickell, Moore, and Sabelhaus (2012)).

question, particularly in a developed country setting.[4]

I divide the theoretical effects of bank account ownership in two parts. First, account ownership provides access to lower cost transactional services, raises the effective return on saving, and lowers the cost of borrowing.[5] While the demand for transactional services is likely inelastic, so that spending on these services is reduced when becoming banked,[6] economic theory does not generate a clear answer regarding the response of saving and borrowing to changes in saving and borrowing rates. In particular, although increased saving is one of the benefits most emphasized by advocates of "banking the poor" (e.g. Barr (2004)), increasing the saving rate might decrease saving if the income effect is larger than the substitution effect. In Chapter 2 I analyze the response of optimal consumption behavior to favorable changes in saving and borrowing rates, incorporating the possibility of failing to repay debts in full into the basic model. I elaborate on this and raise several related points in Section 1.3. The second type of effects is behavioral. Bank accounts have been argued to have a role as behavioral aides, helping individuals address psychological biases and cognitive constraints that result in deviations from optimal behavior, especially with respect to saving. Mullainathan and Shafir (2009) notably argue for the centrality of such a role. I elaborate on this type of effects in Section 1.3 as well.

I use a unique dataset of low-income participants of financial education workshops in New York City and exploit quasi-random variation in the likelihood of post-workshop account opening to overcome endogeneity and estimate causal effects. Workshop participants, many of whom are unbanked prior to the workshop, are encouraged through monetary and other means to open accounts at 6 specific financial institutions with a total of 8 branches around NYC. These institutions are Community Development Credit Unions which are dedicated to serving the poor. I assert that the distance between a workshop participant's home and the closest of these credit unions affects the likelihood with which the participant will open an account in one of these institutions after the workshop, and provide evidence indicating that this is indeed the case. I use distance as an instrument for post-workshop account opening at one of the collaborating credit unions. Potential endogeneity concerns such as those having to do with CDCUs' location decisions are addressed by including various control variables that are meant to isolate variation in account opening that

[4] Several studies have examined the effects of bank account ownership in developing countries (e.g. Burgess and Pande (2005) in India, Aportela (1999) in Mexico, Dupas and Robinson (2013) in Kenya). Fitzpatrick (2009) does this in the UK. In the US, Chin, Karkoviata, and Wilcox (2011) examine the effects of account ownership on savings and remittances of Mexican migrant workers, and Mills, Gale, Patterson, Engelhardt, Eriksen, and Apostolov (2008) evaluate the impact of Individual Development Account ownership. I elaborate on these studies in the next section.

[5] I discuss these points in Section 1.3. In particular, providing a safe place to store savings is likely enough to raise the effective return.

[6] As noted above, this point is debatable. See also Section 1.3.

is unrelated to outcomes. In particular, the setting allows controlling for possibly confounding unobservable location effects on outcomes, to the extent that they are time-invariant, while still retaining variation in the distance instrument. I provide evidence supporting the exogeneity of the instrument.

Throughout the paper I distinguish between those who do not have a bank account prior to the workshop (the "initially unbanked") and those who do (the "initially banked"). Interestingly, many initially banked workshop participants open accounts at the collaborating credit unions, presumably in order to enjoy services that are more suited to their needs.[7] I find no effects of opening an account on the initially banked, however, in all of the outcomes that I examine. That opening an account at a CDCU has stronger impacts for the initially unbanked than it does for the initially banked is not surprising: the latter already enjoy the potential behavioral benefits embedded in the ownership of (any type of) account, and already have access to mainstream financial services to some extent, so that they stand to gain less from opening an account. That I find no effect at all on the initially banked could point to little added value in CDCU account ownership vs. account ownership at a regular bank or to the inability of the outcomes that I examine to capture some of the effects of account ownership.

I derive most of the outcome data analyzed in this paper from credit reports, which contain detailed information on individual credit behavior with respect to mainstream (i.e. non-alternative) products such as credit cards and bank loans. Reports are available from the beginning of the workshop as well as at various time intervals after the workshop, enabling the analysis of changes in mainstream credit behavior. I begin by examining FICO scores, which aggregate the information on the credit report and which are important outcomes in their own right.[8] Less than half of initially unbanked sample individuals have a credit score at the beginning of the workshop - the remainder do not have sufficient recent credit activity for a score to be computed. I estimate a significant effect of opening an account at one of the collaborating credit unions on the likelihood of score improvement by those who start out with a score. According to my preferred estimate, the likelihood of score improvement increases by 47% at the 0.5-3 year horizon after the workshop as a result of opening an account. This finding is not attributable to account opening inducing

[7] This speaks to the evidence on the "underbanked" (see footnote 2). CDCUs are dedicated to serving the poor and as such offer products and services that are more adequate for this population in terms of fee structure (i.e. minimum balance requirements), credit access (i.e. secured loans), and savings products (i.e. lower denomination Certificates of Deposit). I provide more information about CDCUs in Section 1.4. That accounts at cooperative institutions whose mission it is to serve the poor are the type of accounts studied here should be kept in mind when deriving policy implications.

[8] Scores are used for screening and for determining conditions on many types of transactions. See Section 1.6.

increased usage of mainstream credit, as I find no evidence of an effect on credit report items which reflect usage of credit such as credit card ownership or debt balances. Account ownership is also not an input in the credit scoring algorithm, so the result is not mechanical. Rather, credit score effects seem to be driven by effects of opening an account on delinquent behavior: I estimate that opening an account reduces the recent delinquent balances[9] of initially unbanked individuals by more than $800. As I explain in Section 1.3, favorable interest rate changes (brought about by opening an account) might reduce delinquency through their effects on the amount owed as well as on the attractiveness of rolling over a due loan, and there are also potential behavioral effects.

Since, as the above quote suggests, one of the main supposed benefits of account ownership is its functioning as a gateway to the mainstream financial system, I devote special attention to mainstream credit usage on the extensive margin. Sample individuals, in particular those who are initially unbanked, both enter and exit the mainstream credit system, as measured by having a FICO score, by having any type of mainstream credit balance, and by having a credit card. I find only small effects of opening an account on the likelihood of inclusion in the mainstream credit system. I estimate that the likelihood of having a FICO score at the time of the follow-up credit pull increases by 15% as a result of opening an account, but the effects on the likelihood of having a mainstream credit balance or a credit card are indistinguishable from zero. I find no effects of opening an account on mainstream credit usage on the intensive margin. While a non-increase in overall borrowing is consistent with theory, these findings are somewhat surprising as we might expect individuals who open accounts to switch from alternative borrowing to mainstream borrowing.

I study the potential links between bank account ownership and financial literacy. Initiatives to bring unbanked individuals into the financial mainstream typically involve financial education, and financial education programs aimed at low-income individuals typically emphasize the advantages of account ownership.[10] Since the sample analyzed in this paper is composed entirely of recipients of financial education, the estimated effects of opening an account might be conditional to some extent on receiving financial education, e.g. on individuals' levels of financial literacy. Indeed, individuals with low financial literacy have been shown to make suboptimal use of financial products: Lusardi and Tufano (2009) find that the less literate pay much higher credit card fees, and that as much as one third of the fees paid by them can be attributed to lack of knowledge.

[9] These are delinquent debts which have not yet been sold to collection agencies. This data is only available for follow-up credit pulls conducted as part of this study. See Section 1.4.E.

[10] See the programs mentioned in footnote 3, for example.

Gerardi, Goette, and Meier (2010) examine subprime mortgage borrowers and find that those with low numerical ability were more likely to default and have their home foreclosed on in the recent crisis.

While I cannot directly test for interactions between the effects of opening an account and the effects of receiving financial education since there is no control group for the latter, I exploit plausible heterogeneity in individuals' levels of financial literacy in order to shed some light on this issue. In particular, I test whether initially unbanked workshop participants are differentially affected by opening an account at one of the collaborating credit unions based on whether they had graduated high-school. Formal education levels have been shown to be strong predictors of financial literacy (Lusardi and Mitchell (2013)) and there is reason to believe that those with more formal education might learn more in a classroom setting, i.e. gain more from the workshop, resulting in higher post-workshop literacy levels.[11]

I find heterogeneity based on high school graduation status in several of the effects that I examine. First, the negative effect of opening an account on recent delinquencies is limited to initially unbanked workshop participants who graduated high school. Second, while only small effects are found on the likelihood of inclusion in the mainstream credit system on aggregate, as noted above, there is significant heterogeneity in these effects: opening an account increases the likelihood that initially unbanked high school graduates have a FICO score as well as a credit card at the 0.5-3 year post-workshop horizon, while it does not predict having a FICO score among those who did not graduate high school and actually reduces the likelihood that they have a credit card. That is, opening an account does seem to function as a gateway to the mainstream financial system, but only for those with sufficient levels of financial literacy. Those with lower levels of literacy open accounts at collaborating credit unions[12] but do not seem to take advantage of the access to mainstream credit that this provides. The finding that opening an account does not reduce their delinquent balances is consistent with their effective borrowing rates remaining unchanged when they open an account. An obvious caveat to this analysis is that having graduated high school is not randomly assigned and might be correlated with other individual characteristics which have to do with the effects of account ownership. An obvious concern is that those who graduated high school are likely to be better off financially. I address this concern by examining differences

[11] Indeed, in a follow-up survey conducted as part of this study (see below), respondents who graduated high school report higher self-assessed knowledge levels than those who did not graduate high school (4.04 vs. 3.76 on a 5 point scale).

[12] While non-high school graduates open accounts at a lower rate than do high school graduates, about 20% of them open an account at one of the collaborating credit unions after the workshop.

according to employment status rather than high school graduation, and find no heterogeneity in the effects of opening an account on the likelihood of inclusion in the mainstream system.

Low literacy individuals who open accounts might be reluctant to use mainstream credit products because they realize that they are likely to make the types of costly mistakes noted above, or because they are simply uncomfortable with these products. They might also fail to realize the relative costs of alternative financial services.[13] The dependence of the demand for various types of financial services on financial literacy is well-documented: it has been found that the less literate are less likely to participate in the stock market (Van Rooij, Lusardi, and Alessie (2011)) and more likely to use alternative financial services (Lusardi and de Bassa Scheresberg (2013)). Bertrand and Morse (2011) show that providing more cost information to payday borrowers reduces the likelihood of future borrowing. Carter, Skiba, and Tobacman (2011) and Agarwal, Skiba, and Tobacman (2009) document substantial payday loan borrowing by individuals with access to cheaper sources of liquidity. My findings are consistent with this prior evidence.[14]

In order to supplement the credit report outcome data as well as the data on post-workshop account opening,[15] a telephone survey of sample individuals was conducted as part of this study. One of the most important topics was saving. Using survey data, I am unable to reject the hypothesis that opening an account does not increase reported saving among initially unbanked workshop participants on neither the extensive margin (i.e. it does not increase likelihood of saving vs. not saving) nor the intensive margin (i.e. it does not increase the amount saved). As noted above, this is consistent with the optimal response to an increase in the effective saving rate brought about by account ownership. Account ownership has also been argued to have effects on saving through addressing behavioral biases that lead to overspending, however. In order to evaluate this claim, survey respondents were asked whether they feel that they are currently spending more than they should be, and whether they feel that they are now overspending less than in the past. Here, too, I cannot reject the hypothesis that opening an account has a zero effect on self-reported overspending by those who were unbanked prior to the workshop. Unbanked sample individuals might not be subject to self-control problems or they might have alternative means of combating overspending and not gain much from opening an account, in particular after having gone through

[13]Doyle, Lopez, and Saidenberg (1998) cite a 1989 study which found that more than 30% of unbanked users of check-cashing stores either did not know whether banks or check-cashing stores were more expensive or believed that banks were more expensive.

[14]Somewhat contradictory evidence is brought by Cole, Sampson, and Zia (2011), who show that monetary incentives are much more effective than financial education in increasing the likelihood of bank account opening in a field experiment in Indonesia. Education is effective, however, for those with low initial levels of education and financial literacy.

[15]Data on account opening is partially available from administrative sources. See Section 1.4.

the workshop.[16]

If opening an account results in increased credit access, we might expect to see an effect on "financial fragility", defined as confidence in the ability to raise a relatively large amount of money in a short period of time (Lusardi, Schneider, and Tufano (2011)). This is a potentially important outcome for the welfare of low-income individuals who are likely subject to numerous income and consumption shocks (e.g. job loss, health) and who generally have less means to deal with such shocks (e.g. health insurance). Analyzing survey respondents' confidence in being able to raise $1500 in 30 days, I find no significant effect of opening an account on this level of confidence. Similarly, I do not find an effect on economic hardship, as measured by recent difficulty in paying bills (as in Melzer (2011)), or on self-reported finances-related stress. The lack of effects on these measures of "financial wellbeing" is somewhat disappointing given the high expectations for the effectiveness of banking the unbanked, as reflected in the policy efforts noted above. Again, the individuals studied here might have developed alternative ways of coping with their economic situation, and opening a bank account might not be the magic solution to their problems that it is oftentimes taken to be. As suggested above, those who open accounts might be reluctant to use the services which account ownership provides them access to because of insufficient literacy. Unfortunately, the small number of survey responses does not allow for a meaningful analysis of heterogeneous effects by literacy level. The small number of responses and low statistical power are a general caveat in interpreting the survey results: confidence intervals contain both zero and figures that are quite high.

Analyzing the survey data, I find an interesting result which speaks to the complimentarities between financial education and access to mainstream services: opening an account raises workshop participants' self-reported knowledge levels about managing their money on a daily basis. This might be because account ownership enables following one's transactions more closely or because it exposes individuals to different financial products (even if, as noted above, take-up is rather low). Another potential channel is access to financial advice: the collaborating credit unions studied in this paper offer free advice to their members, which might increase the knowledge of those who open accounts.[17] Either way, this finding points to a role of account opening in enhancing the effectiveness of financial education.[18]

[16] Mullainathan and Shafir (2009) interpret several behavioral patterns exhibited by the poor as alternative commitment savings mechanisms, for example the purchase of lottery tickets.

[17] Access to advice might play a role in some of the other findings in this paper, particularly those having to do with credit score improvements, as I note below.

[18] The effect of opening an account on knowledge might be conditional on having undergone financial education, or it might be independent of education. In any case, it seems to apply to those who receive financial education.

This chapter will proceed as follows. Section 1.2 reviews related literature. Section 1.3 outlines the theoretical effects of access to mainstream financial services. Sections 1.4 and 1.5 describe the empirical setting used in this paper and the empirical strategy. Section 1.6 describes and summarizes the credit report and survey outcome variables. Section 1.7 reports the results of reduced form regressions of outcomes on the distance instrument. Section 1.8 quantifies the causal effects of opening an account on outcomes and reports the results of the heterogeneous effects analysis. Section 1.9 concludes.

1.2. Related Literature

As noted in the introduction, there have hardly been any studies which have attempted to estimate the causal effects of bank account ownership, in particular in a developed country setting. Chin, Karkoviata, and Wilcox (2011) randomly assign assistance in obtaining an identification card (argued to be a major barrier to account opening in their setting and of little use otherwise) within a group of Mexican migrant workers in the US, generating variation in account opening. They find that total savings flows as a share of income increase by 9% for the group that received assistance 5 months after treatment (the "treatment on treated" estimate of the effect of opening an account is 23%). This finding does not seem particularly relevant for the non-migrant unbanked population in the US, however, especially given that the increase in savings comes mostly at the expense of remittances to Mexico, a finding which the authors relate to intra-household issues of control of the remitted funds.

Mills, Gale, Patterson, Engelhardt, Eriksen, and Apostolov (2008) document the results of an experiment in which access to Individual Development Accounts (IDAs) - savings vehicles which provide matching payments when balances are used for specific purposes (e.g. home purchase, education) - was randomized among eligible applicants in Oklahoma. They find modest increases in homeownership but no effects on any other investments or measures of wellbeing. As IDAs are far from being conventional bank accounts, the informativeness of this study regarding the effects of conventional accounts is limited.

A more relevant study is Fitzpatrick (2009), which exploits a policy change in the U.K that mandated the receipt of child benefits electronically to estimate the effects of account ownership on credit and consumption. The study finds a 17 percentage point increase in the likelihood of credit card ownership and little evidence of an increase in bank loans. Most of the consumption measures that are examined, including the level of weekly expenditures, are not affected by account ownership. The only measure that is affected is the number of household appliances, which increases by 3.

Several studies have examined the effects of access to mainstream financial services in developing countries. Some of these studies have exploited bank expansions as access-increasing natural experiments. Burgess and Pande (2005) study a major state-led expansion effort in India and find that branch opening in locations previously lacking banks succeeded in reducing local poverty. Aportela (1999) shows that in towns affected by rapid expansion of a government-owned bank targeted at the poor in Mexico, saving rates increased by 3-5 percentage points. Randomized experiments seeking to answer this question have also been conducted in developing countries recently. Dupas and Robinson (2013a) randomize access to non-interest bearing savings accounts among self-employed individuals in rural Kenya. Even though the accounts offered an effectively negative return (no interest and large withdrawal fees in a high-inflation environment), take-up and usage of the accounts were high among the women in their sample, who exhibited increases in saving, productive investment, and private expenditure. The authors offer two possible explanations for the success of the accounts in increasing saving and the implied highly negative returns to informal forms of saving: a tendency to overly spend money in hand due to self-control problems and demands made by friends and relatives on informal savings.

In another study conducted in Kenya, Dupas and Robinson (2013) attempt to shed light on savings mechanisms among the poor. They vary access to four different savings devices, designed towards health investment and differing in the degree of commitment they entail, and find that simply providing a safe box in which to keep savings (along with the key) increases preventive health investments by 66%-75%. Based on participants' responses to a follow-up survey, they attribute these effects to mental accounting. According to this explanation, money put into the box is labeled "for savings" and is therefore less likely to be spent on luxuries or given away to others. I elaborate on this possible mechanism below.

This paper is also related to the literature on financial education. Despite growing interest in this topic, there has not been much convincing evidence regarding the efficacy of efforts to improve financial literacy. Hastings, Madrian, and Skimmyhorn (2012) review this literature, noting the dearth of studies which involve random or quasi-random variation and concluding that "there is at best mixed evidence that financial education improves financial outcomes". Additional reviews are provided by Gale, Harris, and Levine (2012) and Collins and O'Rourke (2010).

1.3. The Theoretical Effects of Access to Mainstream Financial Services

In what ways would we expect the access to mainstream financial services which bank account ownership provides to impact the behavior and welfare of individuals, and in particular those of the poor? I divide the answer to this question in two parts. The first part consists of access to products and services "per-se" which an account provides: transactional services, savings products, and credit. The second part consists of ways in which an account can function as a "behavioral aide", helping to address biases and cognitive constraints and facilitating optimal behavior. I now elaborate on these two parts.

1. Access to products and services per-se

Account ownership provides access to financial products and services which are otherwise costlier or less available. One such type of service is transactional services, such as check cashing and bill payment, which are generally less costly for bank account owners. Assuming that the demand for these services is relatively inelastic with respect to price, spending on them is expected to decrease when becoming banked. Although this is an element that is often emphasized when discussing the benefits of account ownership (see e.g. Barr (2004)), its significance for an individual's overall financial situation is debatable. The amounts involved are usually not very large, and unbanked individuals are for the most part able to transact at relatively low costs at mainstream financial institutions even without having an account.[19] According to the FDIC's National Survey of Unbanked and Underbanked Households, only about 60% of unbanked households report using alternative financial service providers for transactional services (money orders, check cashing, and remittances) in the past year (Federal Deposit Insurance Corporation (2012)).[20]

Account ownership also provides access to a saving technology and raises the expected return on saving. This expected return is most likely negative in the absence of an account due to inflation and danger of loss or theft, as well as unwanted use by family members and demands by others.[21] Access to interest-bearing products such as money market accounts and certificates of

[19] For example, checks can usually be cashed by anyone at the issuing bank at little or no cost.

[20] In the sample studied here, 40% of unbanked individuals (and 11% of banked individuals) report cashing their checks at check-cashing stores. See Table 1.2.

[21] Unwanted use and demands on funds have been shown to be important in the developing country context, for example by Dupas and Robinson (2013). They are plausibly important in the setting studied here as well.

deposit further raises the expected return on saving of account owners.[22]

In addition to lower-cost transactional services and to a higher expected return on saving, account ownership also allows better access to credit. While credit is generally available to some extent through alternative financial services (payday loans, pawnshops, etc.), it tends to be extremely costly (e.g. payday loans usually have implied APR's of over 400% (Bertrand and Morse (2011))) and is not always fully available (e.g. payday loans are only available for the employed and only in some states).[23] Bank account ownership does not necessarily imply eligibility to receive a bank loan, of course: low-income individuals in the US are generally thought to be credit rationed at banks (Bolton and Rosenthal (2005)). However, owning a Community Development Credit Union account, the effects of which are studied in this paper, most likely does result in lower-cost credit eligibility since CDCUs are designated to serving the poor.[24] More generally, account ownership most likely improves access to credit and lowers the cost of borrowing even for those who cannot get a loan or a credit card from their own financial institution since it is taken into account by other lenders such as credit card companies. Moreover, taking a loan from a mainstream financial institution establishes formal credit history, which can result in a significant expansion of the borrowing opportunity set.[25] This "gateway to the mainstream financial system" point is often emphasized in discussions about banking the poor, for example in the Larry Summers quote brought above ("passport to the broader economy").

Economic theory does not generate a straightforward prediction for the effects of favorable changes in the expected return on saving and in the cost of borrowing on consumption, i.e. for the interest elasticity of consumption. In a standard life-cycle model, an increase in the saving rate has an ambiguous effect on the consumption of savers (since the substitution and income effects contradict each other) while a decrease in the borrowing rate unambiguously increases the consumption of borrowers (since the substitution and income effects work in the same direction) (see Romer (2011)). In Chapter 2 I show that incorporating the possibility of failing to repay debts in full (a feature that is especially relevant for low-income borrowers) in the life-cycle model

[22] Many of the low-income individuals studied in this paper have an account at a regular bank but open an additional account at a Community Development Credit Union. Their effective return on saving presumably improves as well, due to higher rates offered by CDCUs (which are cooperatively owned and have the objective of benefiting their customers-members) as well as to saving products which are more relevant for low-income individuals (i.e. lower denomination CDs).

[23] The individuals studied in this paper live in New York, which bans payday loans. These are also available on the internet, however. The FDIC's National Survey of Unbanked and Underbanked Households finds that 21% of unbanked households have never used alternative financial services (Federal Deposit Insurance Corporation (2012)).

[24] This point most likely applies to initially banked individuals who open accounts at CDCUs as well.

[25] Several of the Community Development Credit Unions studied here also offer secured "credit builder" loans, meant to establish credit history.

can alter these effects: under certain formulations for the cost of failing to repay, initial savers unambiguously decrease consumption (increase saving) in response to an increase in the saving rate and the response of initial borrowers to a decrease in the borrowing rate is ambiguous.[26] These effects are further complicated in practice by the puzzling tendency of individuals to save (at low rates) and borrow (at high rates) at the same time, contrary to the standard theoretical dichotomy between saving and borrowing.[27] Another complicating factor, potentially relevant to low-income individuals, is liquidity constraints: if account ownership not only lowers the cost of borrowing but also relaxes a binding liquidity constraint, it will increase borrowing. The extent to which low-income individuals (in particular those living in a major metropolitan area such as New York City) are liquidity constrained, i.e. cannot borrow more at any price, is debatable, however, given the proliferation of alternative financial services.

In accordance with the theoretical ambiguity, empirical studies of the interest elasticity of consumption have yielded a wide range of estimates (see e.g. Gross and Souleles (2002) and Karlan and Zinman (2008) for the elasticity of borrowing with respect to the borrowing rate and Bernheim (2002) for a survey of studies examining the elasticity of saving wrt the saving rate).

Another behavioral variable of interest for which there is no clear prediction is the amount on which the individual is delinquent. In Chapter 2 I demonstrate two channels through which favorable interest rate changes can affect delinquency. First, rate changes affect delinquency through their effect on the amount owed, which is a function of consumption and the promised rate.[28] In the model studied in Chapter 2, where repayment of debt is voluntary, the expected unpaid amount as well as the probability of not paying in full are an increasing function of the amount owed.[29] This would obviously be the case with involuntary delinquency as well. The effects of favorable interest rate changes on the amount owed are generally ambiguous, and depend on the interest elasticity of consumption.[30] A second channel has to do with the rolling over of prior debts. Lowering the cost of borrowing makes rolling over a given amount of debt (i.e. rolling over

[26]This happens when the costs of failing to repay are such that the marginal cost of consumption decreases in the amount consumed. In that case, the marginal propensity to consume out of wealth is negative, as is the income effect of a favorable interest rate change. Intuitively, a favorable interest rate change reduces the likelihood of default and strengthens the link between current consumption and future utility, "increasing the individual's stake in the future". See Chapter 2.

[27]This behavior has been documented by Gross and Souleles (2002), who characterize it as inconsistent with no-arbitrage. Zinman (2007) justifies this behavior on the grounds of household liquidity management.

[28]If we allow for simultaneous saving and borrowing, changes in the expected return on borrowing and in the cost of borrowing can both affect the amount owed.

[29]While most of that paper is devoted to full-scale bankruptcy-style default, this point applies to partial delinquency, which is explored in an extension to the basic model, as well.

[30]In particular, the amount owed decreases when the borrowing rate is lowered if the elasticity of the amount borrowed wrt the interest rate is higher than -1.

with the same lender or borrowing from another source to pay back the original loan/bill) more attractive relative to delinquency, thus lowering delinquency.

A related point has to do with the role of lowering borrowing costs and removing liquidity constraints in helping individuals face shocks to consumption and income. This has been argued to be especially important for the poor, who face numerous potential shocks to consumption and income (shocks related to health, crime, or job loss are presumably suffered disproportionately by them) and who are less able to smooth these shocks out of current income and are less likely to have access to consumption smoothing mechanisms such as health insurance. The impacts of failing to repay debts or bills as result of a shock might also be more serious for the poor: Mullainathan and Shafir (2009) emphasize the consequences of failing to repay bills, citing a study which estimates that 5% of the annual income of low-income individuals is spent on the costs of reconnection of services such as phone and gas. They note the possible dynamic consequences of failing to repay: for example, a disconnection of phone service might make it harder to find a job.

Another reason for the particular importance of credit access for the poor is the role of investment in education or entrepreneurship in emergence out of poverty, emphasized by economic theorists (e.g. Galor and Zeira (1993), Aghion and Bolton (1997), Banerjee and Newman (1993)). Better access to credit and lower borrowing costs could also be welfare reducing, however, if they cater to self-control problems (see below). This has been argued in particular with respect to alternative credit products aimed at low-income populations.[31]

2. Bank accounts as behavioral aides

In addition to providing access to financial products and services per-se, account ownership potentially acts as a "behavioral aide", helping address biases and cognitive constraints and facilitating optimal behavior. Mullainathan and Shafir (2009) are strong advocates of this approach, arguing that "financial institutions ... should not be simply viewed from a financial cost-saving point of view but instead should be understood to affect the lives of people by easing their planning, facilitating their desired actions and enabling their resistance to temptation."

Many models of consumption behavior depict individuals as suffering from behavioral biases which lead to a lack of self-control, in the sense that consumption is higher than what the individual would consider optimal ex-ante. Thaler and Shefrin (1981), Laibson (1997), Banerjee and

[31] For more on this issue and for empirical evidence in the context of access to payday lending, see Morse (2011), Melzer (2011), and Bertrand and Morse (2011).

Mullainathan (2010), and Bernheim, Ray, and Yeltekin (2013) are such models. The latter two are concerned with the special relevance of self-control problems for the poor.

Bank accounts might address lack of self-control through acting as a sort of commitment device, the demand for which has been documented by Ashraf, Karlan, and Yin (2006) and Thaler and Benartzi (2004).[32] An account might function as a commitment device through the (admittedly limited, in the age of debit cards) physical effort involved in spending money deposited in it, and, perhaps more importantly, through mental accounting (Thaler (1990)). Mental accounting might cause, as argued by Dupas and Robinson (2013) with respect to their findings, money in a savings account to be considered as non-fungible with other sources of cash and, in particular, lower the propensity to consume from it. According to Shefrin and Thaler (2004), the mental accounting model suggests a hierarchy of "money locations" arranged by how tempting it is for a household to spend the money in each, and savings accounts are at a lower level in this hierarchy than cash. If funds can be transferred to less tempting locations, Shefrin and Thaler argue, they are more likely to be saved.[33] Mullainathan and Shafir (2009) also note this potential effect of savings accounts. As mentioned above, Dupas and Robinson (2013) find that providing individuals with a safe box designated for preventive health investments (along with the key) is highly successful in increasing saving. Follow-up surveys they conduct provide support for a mental accounting story.[34] Default saving mechanisms enabled by account ownership, also mentioned in the next paragraph, may also be seen as a type of commitment device if inertia renders them costly to break. Inertia has been shown to be a powerful force in many domains, including saving (e.g. Thaler and Benartzi (2004), Madrian and Shea (2001)).

In addition to behavioral biases, limited attention might also be a cause for under-saving. Karlan, McConnell, Mullainathan, and Zinman (2012) propose a model in which future lumpy expenditures are overlooked with some probability, distorting intertemporal allocations and leading to under-saving and over-borrowing. The model predicts a role for reminders in increasing saving, which is confirmed in a randomized experiment. Account ownership itself could be seen as an element that serves as a reminder to save, increasing the salience of specific future lumpy expenditures through labeling (Karlan, McConnell, Mullainathan, and Zinman (2012)). Perhaps more importantly, account ownership facilitates default saving mechanisms like direct deposit and

[32] Alternative financial products used by the poor have also been interpreted as facilitating saving through acting as commitment devices (e.g. rent-to-own and layaway schemes) (Mullainathan and Shafir (2009)).

[33] Debit cards might increase the propensity to consume from money stored in a bank account if it is tempting to use them.

[34] However, they find that the individuals identified as time-inconsistent in their sample were not affected by this treatment, and reason that these individuals needed a stronger form of commitment.

automatic transfer of funds into savings vehicles, removing the need for active saving. Through these mechanisms, an individual who is aware of his limited attention can pre-allocate funds to unattended-to future lumpy expenditures regularly. Of course, there also has to be a mechanism to prevent the individual from spending these funds once they are in the account. This might be achieved by mental accounting. Indeed, Karlan, McConnell, Mullainathan, and Zinman (2012) propose that mental accounting is optimal precisely because of limited attention.

Related to the limited attention channel, another way in which bank accounts might function as behavioral aides is by helping cognitively constrained individuals to plan and keep track of their consumption. Individuals may be aware of the need to save but find it hard to plan their consumption accordingly, or even to keep track of how much they are actually consuming (this might also result in oversaving). An account might help in this respect, making it easier to maintain and follow a budget. Similarly, features like automatic bill payment might reduce the likelihood of missed payments and decrease delinquency. As Mullainathan and Shafir (2009) put it, bank accounts "provide the opportunity to make infrequent, carefully considered financial accounting decisions that can prove resistant to intuitive error."

1.4. Empirical Setting

A. Financial Literacy Workshops

The sample is composed entirely of graduates of financial literacy workshops run by a New York City-based nonprofit organization by the name of Neighborhood Trust Financial Partners (NTFP).[35] These workshops, titled "Getting Ahead", are targeted at the working poor and consist of five weekly meetings, each lasting two hours. In these meetings, workshop participants learn basic financial concepts (e.g. "APR", "credit limit") as well as how the credit history system works and how to read a credit report (they also receive a copy of their own report). They are encouraged to save regularly and to follow a budget, which they formulate together with the counselor during the course of a 1-on-1 counseling session that is a part of the workshop. Finally, the costs and benefits of various financial services are discussed. Participants are highly encouraged to refrain from using alternative financial services and to open checking and savings accounts. The advantages of a credit union account vs. an account at a regular bank are emphasized and participants are encouraged to open accounts at specific credit unions which NTFP collaborates

[35] http://www.neighborhoodtrust.org/. The organization was formerly called "Credit Where Credit Is Due".

with (see below).

Workshops are conducted regularly at various sites around New York City. Most of these sites belong to various partner organizations which serve the poor. These organizations offer a host of services and workshops to their clients, such as job search and computer skills, and NTFP's financial literacy workshop is one of these services.[36] Workshops are also given at NTFP's offices and at some of the credit unions with which it collaborates (see below). Participation in the workshops is voluntary.

B. Collaborations with Credit Unions

Identification of the effects of account ownership is enabled by a feature of NTFP's activity which generates plausibly exogenous variation in the likelihood of post-workshop account opening by the participants of its workshops. The organization operates a credit union (Neighborhood Trust Federal Credit Union, or NTFCU) alongside its educational activities and encourages workshop participants to open accounts there. In order to provide more convenient options to participants living far away from NTFCU, which is located in the Washington Heights neighborhood of Manhattan, the organization has initiated collaborations with five other credit unions around New York City over the past few years. These credit unions, as well as NTFCU, are "Community Development Credit Unions" (CDCU's), which are credit unions with the specific mission of serving low and moderate income communities.[37]

Collaborations with credit unions entail several elements which are meant to increase the likelihood of account opening among workshop graduates. First, costs are reduced: membership fees are waived for workshop graduates at all 6 credit unions,[38] and other fees and requirements (i.e. minimum balance requirements, monthly fees) are lowered or waived, depending on the credit union. Second, credit unions are endorsed by NTFP: workshop graduates are given a sheet of paper with the names and addresses of collaborating credit unions. These are also printed on the graduation certificate which participants receive at the end of the workshop. At the three credit unions with which collaboration is strongest, there is a NTFP advisor on-site to provide

[36] The two organizations with the most graduates in the sample are The Hope Program (14% of sample) and Common Ground (8.5%). http://www.thehopeprogram.org/, http://www.commonground.org/.

Clients of the Doe Fund (http://www.doe.org/), one of the major partners of NTFP, are not included in the sample. These are individuals living temporarily in Doe Fund facilities. They (and other workshop participants identified as living in temporary residencies or as homeless) are not included since the temporary nature of their locations does not seem to lend itself to identification using distance.

[37] See http://www.natfed.org/ for more about CDCUs.

[38] One of the credit unions has no membership fee to begin with; others have fees ranging from $5-$20.

counseling (other collaborating credit unions also provide counseling services). Finally, for the two credit unions with which collaboration is strongest (the credit union operated by the organization (NTFCU) and a Brooklyn credit union by the name of Brooklyn Co-Op, or BCFCU), remote account opening takes place at some workshop locations. In workshops given at those locations, participants are able to open accounts in these credit unions on the spot during the workshop. NTFCU and BCFCU are also the only credit unions for which I have administrative account opening data. Collaborations entail other elements in addition to those which encourage account opening. Most importantly, workshops are given at 4 of the credit unions.[39]

Figure 1.1 shows a map of New York City which displays the locations of the various credit unions with which NTFP currently collaborates and the collaboration starting dates. These dates have been inferred from Memorandum of Understandings signed between NTFP and the various credit unions and have been verified as approximate collaboration initiation dates by NTFP staff.

[Insert Figure 1.1 here]

C. Sample and Data Sources

The full sample consists of 2,165 individuals who underwent and graduated NTFP's workshop[40] in 2008-2011, for whom there is a valid address, and who are not identified as homeless or as living in temporary residencies (see also footnote 36). I also exclude individuals for whom initial banking status is not available and individuals who are members of NTFCU prior to the workshop.[41] The resulting sample includes participants from 277 workshops conducted at 50 different sites.

Comprehensive demographic and financial behavior data is collected on the first day of the workshop via a questionnaire that participants are asked to fill out. This data is summarized in Panel A of Table 1.2 and discussed in the next subsection. As part of the questionnaire, participants are asked to authorize NTFP to pull their credit reports. Reports for those who agree (almost everyone) are pulled in the first week of the workshop. Summary statistics of incoming credit report data are reported in Panel B of Table 1.2 and discussed below.

[39] A potential bias that this creates is that participants who travel further to attend the workshop (and whose "distance to the closest collaborating CU" is on average higher) may be more motivated and thus more likely to change their behavior post-workshop, regardless of account opening. This would weaken the effects of account opening estimated using distance.

[40] In order to graduate, an individual must attend at least 4 out of 5 sessions. The attrition rate is about 30%.

[41] Prior CU membership data is only available for NTFCU (see below). Clients of other credit unions might also participate in the workshops, especially given that workshops are held in 3 of the other collaborating credit unions, but I cannot identify them. Results are unchanged in a robustness check in which I filter out all those who undergo the workshop in one of these credit unions and who indicate that they have a checking or a savings account at the beginning of the workshop.

In addition to the incoming credit report pull, workshop participants are asked to authorize NTFP to pull their reports for evaluation purposes every 6 months, indefinitely into the future, and about 70% agree.[42][43] Reports for evaluation purposes have been pulled periodically for a random subset of participants since 2008. For the purpose of this study, a large number of reports was pulled in May-June 2012. In my analysis of credit report outcomes I consider workshop participants for which there are at least two credit reports available: one from the beginning of the workshop and one from at least 6 months and at most 3 years after the workshop.[44] After applying the data filters described above, there are 1,081 such participants.

In order to supplement credit report data and collect information on more outcomes, a telephone survey of NTFP's workshop participants was conducted in July 2012-March 2013. After a brief introduction in which participants were reminded of the workshop that they had taken,[45] they were asked to answer the survey in return for a chance to win a $100 American Express gift card. Those who agreed (13% of the entire sample, or 283 individuals) were told that the survey concerns their financial life and were asked to respond honestly. The outcome variables derived from survey questions are summarized in Section 1.6 and in Table 1.4, and the entire survey is in Appendix A.

In addition to comprehensive account ownership data for those who answered the survey, administrative data on account openings is available from two of the six collaborating credit unions. Data for NTFCU, the credit union operated by NTFP, is comprehensive: NTFCU's client database is periodically matched with NTFP's database and information regarding NTFCU account ownership by all NTFP workshop participants is available, regardless of when they underwent the workshop. Administrative account opening data is also available from Brooklyn Co-Op (BCFCU), one of the other collaborating credit unions. This data is obtained from reimbursement requests of waived account fees made by BCFCU, meaning that data on account ownership prior to the workshop is not available. An additional caveat is that reimbursement requests were only made starting July 2009, so account opening data is not available before that date.[46] Account opening/ownership data for the other collaborating credit unions is only available through the survey. Survey data on account openings at the other credit unions shows that the vast majority of account openings occurred in NTFCU and BCFCU, so that the missing data on openings at other credit

[42] Note that these are "soft pulls", which do not affect credit scores.

[43] Those who agree to have their reports pulled and those who do not agree are observationally similar.

[44] When there is more than one report pulled 0.5-3 years after the workshop I use the latest available report. The average time gap between pulls is 1.7 years. I consider different time intervals in a robustness check.

[45] This was done in order to increase the participation rate.

[46] Collaboration with BCFCU started in January 2009.

unions for those who did not answer the survey might not be critical (see Table 1.3). This is not surprising, since these are the credit unions with which collaboration has been in place for the longest time (see Figure 1.1) and since they are the only ones in which an account can be opened on the spot during the workshop (at some locations).

Walking distances and public transit travel times between workshop participants' addresses and collaborating credit unions, used to instrument for the likelihood of post-workshop account opening, were obtained using Google Maps.[47]

D. Demographics and Incoming Financial Behavior: Descriptions and Summary Statistics

Table 1.1 shows the distribution of initial banking status in the sample, as reported by participants on the first day of the workshop. 32% of sample individuals have neither a checking nor a savings account at the beginning of the workshop. An additional 27% have only a checking or a savings account, and the remaining 41% have both checking and savings accounts. I divide the sample into "initially banked" and "initially unbanked" workshop participants, defining the latter as those who have neither type of account. This corresponds to the FDIC's definition in its National Survey of Unbanked and Underbanked Households. To get a sense of how the 32%-68% unbanked-banked distribution in the sample compares to that of the general population, note that 34% of households in the New York metropolitan area (MSA) earning less than $15,000 a year are unbanked (26% of households earning less than $30,000). These figures are somewhat higher than the national averages, which are 28% and 19%, respectively (Federal Deposit Insurance Corporation (2012)).

[Insert Table 1.1 here]

The first two columns of Panel A of Table 1.2 show means of demographic and incoming financial behavior variables based on data collected at the beginning of the workshop. Means are computed separately for the initially banked and the initially unbanked, as defined above. Examining these figures, it can be seen that the sample is primarily composed of females. This is especially true for the initially banked, only 28% of which are male. The vast majority of sample individuals (more than 85%) are black or Hispanic, and about 40% are non-US born. Education

[47] Public transit travel times, used in a robustness check, were obtained during several weekdays, in mornings and in afternoons, and then averaged. Correlations between different runs were all higher than 98%.

levels are relatively high among the initially banked: the average banked sample individual has more than 13 years of schooling, i.e. attended college. The average unbanked individual, on the other hand, has slightly less than 12 years of schooling. Only 18% of the initially unbanked and 30% of the initially banked are married or in a domestic partnership. Average age is around 40. Unemployment rates are very high, in particular among the unbanked: 58% of initially unbanked sample individuals report being unemployed at the time of workshop (as well as 29% of the initially banked). In light of this, it is not surprising that a large fraction of the unbanked (63%) report receiving some type of welfare payments.[48] Average annual earned income, conditional on being positive, is much higher for the initially banked, though still quite low: $29,410 vs. $14,890 for the initially unbanked.[49]

The bottom three rows of the first two columns of Panel A show summary statistics for several variables which concern incoming financial behavior. While 51% of initially banked sample individuals report saving at the beginning of the workshop, only 8% of those who are initially unbanked report doing so. 40% of the initially unbanked and 11% of the initially banked report using check cashing stores to cash their checks. Finally, only 34% of initially unbanked sample individuals report having ever seen their credit report prior to attending the workshop, compared to 67% of those who are initially banked.

[Insert Table 1.2 here]

E. Credit Report Variables: Descriptions and Summary Statistics

The first two columns of Panel B of Table 1.2 report summary statistics for variables constructed using credit reports pulled at the beginning of the workshop. These reports contain information about an individual's credit history that is meant to help potential lenders assess creditworthiness. Information is collected by credit bureaus - the three major ones are Experian, Equifax, and Transunion - from lenders, collection agencies, and public records concerning legal action that has to do with debt. This study uses reports pulled from Transunion which include, in addition to credit history information, FICO scores.[50] These scores, based on an algorithm developed by the Fair Isaac Corporation and ranging between 300-850, are the most popular

[48]Defined here as public assistance, Supplemental Security Income, Disability Insurance, and food stamps.
[49]These figures are after removing the top 5% values.
[50]There are several types of scores. The scores used here are "FICO classic 04".

aggregator of credit report information and are widely used to assess creditworthiness.[51] For example, borrowers with a score below 620 are usually considered "risky" or "subprime".

Importantly, FICO scores are only computed for individuals with sufficient recent credit activity.[52] When there is insufficient recent activity, the phrase "Insufficient Credit" ("IC") appears on the report instead of a number score. This is a relevant fact for the individuals studied here, many of whom are not active users of mainstream credit. As the first row of Panel B shows, only 45% of initially unbanked sample individuals have a credit score at the beginning of the workshop. Initially banked individuals are far more likely to have a score: 78% of them have one. I come back to the score/IC issue when describing the outcome variables I examine, in Section 1.6. Among those who do have FICO scores at the beginning of the workshop, the mean score for the initially banked is 624, just above the conventional "poor-OK" cutoff of 620. This corresponds to about the 20th percentile of the US population. For the initially unbanked, the mean score is 555, corresponding to about the 7th percentile of the US population.[53]

The remainder of Panel B provides summary statistics of specific credit report items. The debt accounts reported on a credit report can be divided in two. Some accounts are "active", i.e. have not been sent to a collection agency or involved legal action. 67% of initially banked and 39% of initially unbanked sample individuals have a non-zero active credit balance. Unsurprisingly, these figures correspond quite closely to the figures for having a FICO score above. The average outstanding balance (conditional on being positive) is quite high: over \$11,000 for the initially banked and over \$9,000 for the initially unbanked.[54]

Most initially banked sample individuals (57%) have at least one open credit card versus just 17% of the initially unbanked. For those that have a positive credit card balance, the average balance is higher among the initially banked than it is among the initially unbanked.

In addition to "active" accounts, credit reports contain information about accounts at advanced stages of delinquency. Most sample individuals - 52% of the initially banked and 73% of the initially unbanked - have at least one account in collections, i.e. an account that had been sold by

[51] For more information on credit reports and FICO scores see:
http://www.myfico.com/Downloads/Files/myFICO_UYFS_Booklet.pdf, http://www.myfico.com/CreditEducation/articles/.

[52] In order for a score to be computed, a credit report must contain an account (a "tradeline") that has been open for at least 6 months and an account that has been updated in the past 6 months. See http://www.myfico.com/CreditEducation/questions/requirement-for-fico-score.aspx.

[53] For the national distribution of FICO scores see http://www.fico.com/en/Company/News/Pages/09-19-2011b.aspx.

[54] These figures are after removing mortgage balances, which are very rare (less than 5% of the sample) but relatively extremely high, and dropping the top 5% values. Still, this variable is skewed: median values for the initially banked and the initially unbanked are \$6,435 and \$3,643, respectively. The top 5% values are also removed from the credit card balance data and collections and judgments data summarized in this table.

the creditor to a collection agency. Judgments are accounts where the creditor has been awarded a legal claim on an individual's assets by the court. Around 30% of sample individuals have at least one judgment against them at the beginning of the workshop. This fraction is similar across initially banked and initially unbanked individuals. Average combined collection and judgment balances, conditional on being positive, are also similar at around $3,000.

1.5. Empirical Strategy

Since opening an account at a mainstream financial institution is an endogenous choice, selection might bias any direct non-experimental examination of the effects of opening an account on future behavior and outcomes. In the particular setting studied here, workshop graduates who decide to follow the counselor's recommendation and open an account at one of the collaborating credit unions are likely different than those who do not follow this recommendation on both observable and unobservable dimensions. For example, since opening an account is one of the main things advocated in the workshop, it may reflect a better understanding of workshop content or increased motivation to implement the lessons learned in the workshop. In turn, participants who choose to open an account might be more likely to, say, make an effort to reduce temptation spending, regardless of having opened the account. It is therefore important to find an exogenous source of variation in account opening in order to establish causal effects.

I use variation in the location of workshop participants' homes and in the time in which they underwent the workshop (relative to the existence of collaborations between NTFP and the different credit unions) as a source of exogenous variation in account opening. In particular, I use the distance between workshops participants' homes and the closest collaborating credit union at the time in which they underwent the workshop as an instrument for opening an account at one of these credit unions. In the next subsection I explain how I construct this distance variable, which I call *Proximity*. I proceed by discussing the validity of this variable as an instrument for account opening. I first provide evidence regarding *Proximity*'s strength as a predictor of account opening at collaborating credit unions and in general. I then discuss the exclusion restriction and present evidence indicating that variation in *Proximity* is, indeed, exogenous.

A. Constructing Proximity

The distance measure I use as an instrument for account opening is based on Google Maps walking distances between participants' addresses and the different collaborating credit unions. As noted above, initial collaboration dates are inferred from Memorandum of Understanding signings between NTFP and the various credit unions and have been confirmed with NTFP staff. A credit union is considered as "collaborating" at all dates following the MOU signing date.
Proximity is defined for participant i who begins the workshop on day t by:

$$Proximity_i = -log(\min_{k \in coll_t} \{distance(home_i, CU_k)\})$$

where $coll_t$ is the subset of collaborating credit unions with which collaboration was in place at time t, i.e. with which an MOU had been signed prior to participant i starting the workshop.

I use proximity, the negative of distance, so that regression coefficients have an intuitive sign. I take the natural log of distance so that *Proximity* decreases in distance at a decreasing rate. This makes intuitive sense (e.g. decreasing the distance to the closest credit union from 1.1 miles to 0.1 miles would seemingly have a larger effect on the likelihood of opening an account than decreasing the distance from 5 to 4 miles), and seems to fit the data slightly better than linear distance.

In order to control for unobservable, time-invariant place effects, I construct an additional distance variable. *Proximity_No_Coll* is identical to *Proximity* except that it does not take collaboration starting dates into account. This variable is defined as:

$$Proximity\_No\_Coll_i = -log(\min_k \{distance(home_i, CU_k)\})$$

That is, *Proximity_No_Coll* is the proximity to the closest collaborating credit union, regardless of whether that credit union already collaborates with NTFP at the time in which the individual undergoes the workshop. I elaborate on the use of this variable as a control below, where I discuss the exclusion restriction.

I winsorize both proximity variables at the 5% level on each side in order to minimize the effects of outliers. As a robustness check, I construct analogous variables using public transit travel times between participants' homes and collaborating credit unions, also from Google Maps.

The variables (and the results) are similar.[55]

B. *Proximity* as a Predictor of Account Opening

The basic premise behind using *Proximity* as an instrument for account opening is that individuals are more likely to open accounts at institutions that are closer to their homes, other things equal. This is intuitive and is also supported by evidence from the 2010 Survey of Consumer Finances, where "location" was by far the most popular reason given for choosing the financial institution where respondents had their checking accounts (Bricker, Kennickell, Moore, and Sabelhaus (2012)). Location might be important because of the costs of going to the bank and opening the account[56] as well as because the benefits of having an account at an institution might be related to distance from that institution.[57]

In any case, distance from a particular institution predicting account opening *at that institution* is not enough for proximity to the closest collaborating credit union to predict account opening in the setting studied here. If workshops are very effective in persuading participants to open accounts at collaborating credit unions and the rate of account opening in those institutions is very high, proximity to a particular credit union might predict account opening *at that credit union* (versus account opening at other credit unions), but *Proximity* will not be a good predictor of account opening in collaborating credit unions in general.

A further, similar caveat is that even if *Proximity* does predict account opening at the collaborating credit unions, those who live far away from them might be more likely to open accounts at other institutions that are closer to their homes.[58] This would complicate the analysis: in the extreme case where overall account opening rates are not affected at all by *Proximity* (i.e. *Proximity* only determines the likelihood of opening an account at a collaborating credit union versus opening an account elsewhere), the effective control group would be those who open accounts elsewhere. The exercise conducted here would then be interpreted as the estimation of the effects of opening an account at one of the collaborating credit unions vs. opening an account elsewhere. If, on the other hand, the rate of account opening in institutions other than the collaborating credit unions is not significantly affected by *Proximity*, the effective control group would

[55] Correlation between *Proximity* and the corresponding variable based on public transit travel time is 91.5%.

[56] Mullainathan and Shafir (2009) note the importance of such "channel factors" in facilitating certain behaviors and cite a study showing the importance of distance from a student health center in predicting utilization.

[57] See the discussion regarding the exclusion restriction, below, for more on this issue.

[58] That is, the value of opening an account at a nearby institution that is not a collaborating credit union might be higher than some "account opening threshold" after having gone through the workshop.

be those who do not open accounts at all.

These considerations motivate the analysis summarized in Figure 1.2 and in Table 1.3. As noted above, administrative account opening data is available from NTFCU and BCFCU, the two credit unions with which collaboration is the closest, as well as through the survey. I begin by analyzing the administrative account opening data. Panel A of Table 1.3 shows summary statistics of this data, according to which 12.2% of initially banked individuals and 9.1% of initially unbanked individuals opened accounts at NTFCU (the credit union operated by NTFP) after the workshop whereas 12.8% of initially banked and 11.7% of initially unbanked individuals for which data is available opened accounts at BCFCU.[59] I use this data to plot the likelihood of account opening in these two credit unions as a function of the distance to them in Figure 1.2. The figure is constructed by estimating the following regression:

$$Opened\_at\_NT\_BC = X'\pi + \sum_{j=1}^{10} \eta_{1j} D_j * Banked + \sum_{j=1}^{10} \eta_{2j} D_j * Unbanked + \phi Unbanked + \varepsilon$$

where $Opened\_at\_NT\_BC$ is a dummy variable indicating whether an account was opened at either NTFCU or BCFCU after the workshop and X' includes an intercept, the demographic control variables listed in footnote 93, and fixed effects for the quarter in which the workshop was taken. D_j is equal to 1 for individual i if the distance between individual i's home and the closest out of NTFCU/BCFCU is smaller than $j * 0.5$ and larger than $(j-1) * 0.5$. In other words, D_j are 0.5 mile-wide distance interval dummies, indicating whether an individual lives within a certain distance of either of these credit unions.[60]

Figure 1.2 displays the estimated η_{1j} and η_{2j} coefficients, representing distance interval effects on account opening for the initially banked and the initially unbanked, respectively. Examining the two plots in Figure 1.2, a strong distance effect whereby those who live closer are more likely to have opened accounts is evident for both the initially banked and the initially unbanked, especially for the second group. The effect seems to be decreasing in distance.

[Insert Figure 1.2 here]

[59] Administrative data for BCFCU is only available since July 2009. This is reflected in the smaller number of observations used to derive the summary statistics, as noted in Table 1.3.

[60] For individuals who participated in the workshop prior to July 2009 I consider only distance to NTFCU, since BCFCU account opening data is not available for that period.

I proceed to run regressions of the following general form:

$$Opened = X'\pi + \eta_1 Proximity * Banked + \eta_2 Proximity * Unbanked + \phi Unbanked + \varepsilon$$

where X' includes an intercept and different combinations of control variables.

Columns (1)-(4) of Panel B of Table 1.3 report the results of regressions which use the administrative account opening data. *Opened* is equal to *Opened_at_NT_BC* in these regressions and *Proximity* is constructed as explained above, except it only takes NTFCU and BCFCU into account.

[Insert Table 1.3 here]

Regression (1) controls for the demographic variables listed in footnote 93 as well as for fixed effects for the quarter in which the workshop was taken. Regression (3) adds fixed effects for the site in which the workshop was taken.[61][62] The estimated coefficients on both $Proximity * Banked$ and $Proximity * Unbanked$ are statistically significant in both specifications, indicating that distance to these two credit unions significantly affects account opening in them by both initially banked and initially unbanked workshop graduates. Although the coefficients for the initially unbanked are larger in both specifications, we cannot reject the hypothesis that they are equal to the coefficients for the initially banked, as indicated by F statistics displayed towards the bottom of Panel B. Since *Proximity* is a log transformation of distance, we can interpret the coefficients as 100*the increase in the likelihood of opening an account at NTFCU/BCFCU associated with a 1% increase in proximity (or decrease in distance) to the closest collaborating credit union of these two.

While regressions (1) and (3) establish that distance to NTFCU and BCFCU predicts account opening in these credit unions, the conclusion that minimum distance to the 6 collaborating credit unions predicts account opening *in any of the 6 collaborating credit unions* is not obvious. As discussed above, it might be the case that the overall rate of account opening in the collaborating credit unions is high and that distance to NTFCU/BCFCU is effectively a proxy for NTFCU/BCFCU being the closest out of the 6. Regressions (2) and (4) provide evidence against this story by including a dummy variable indicating whether the closest collaborating credit union

[61] Recall that workshops are given at many different sites, corresponding to different partner organizations. I discuss the importance of including these fixed effects in the context of the exclusion restriction, below.

[62] Note that it is infeasible to include a version of *Proximity_No_Coll* in these regressions because NTFCU is owned by NTFP, so there is no pre-collaboration period, and for BCFCU there is no account opening data pre-collaboration.

to the individual's home is NTFCU/BCFCU or one of the other collaborating credit unions. The results hardly change, supporting the notion that *Proximity* predicts account opening at any of the collaborating credit unions.

I now turn to account opening data derived from the telephone survey. This data is comprehensive in the sense that it includes information on account ownership at any financial institution, but it is only available for a small subset of the sample: data is available for 222 initially banked and 61 initially unbanked workshop graduates. The righthand side of Panel A of Table 1.3 summarizes account opening data derived from the survey. Figures for account opening at NTFCU and BCFCU are similar to those obtained from administrative data with the exception of initially unbanked individuals' opening rate at NTFCU. While administrative data indicates that 9.1% of the initially unbanked opened an account at NTFCU after the workshop, survey data indicates that this figure is 19.7%. This is most likely due to data errors in one or both of these sources.[63]

Regressions (5)-(7) in Panel B report the results of account opening regressions similar to regressions (1)-(4) but based on survey data. Control variables are identical to those in specification (2) (it is infeasible to include workshop site fixed effects because of the small number of observations). Regression (5) is identical to regression (2): the dependent variable is *Opened_at_NT_BC* and the proximity measure used only takes distance from NTFCU and BCFCU into account. Coefficients are, however, much larger than in regression (2), in particular for the initially unbanked: 0.266 vs. 0.079.[64] This is, again, most likely due to data errors. In Section 1.8, where I quantify the effects of account opening, I use each of these data sources and generate two sets of estimates for each outcome.

In regressions (6) and (7) I define *Proximity* considering distances from all collaborating credit unions rather than just NTFCU and BCFCU. In regression (6) the independent variable is *Opened_at_Coll_CU*, which is an indicator for having opened an account at any of the collaborating credit unions after the workshop. Summary statistics for this variable, displayed on the righthand side of Panel A, indicate that the vast majority of account openings at collaborating credit unions occurred at either NTFCU or BCFCU: only 3% of workshop graduates report opening accounts at other collaborating credit unions. Together with the relative similarity of the coefficients on the *Proximity* interactions across regressions (5) and (6), this supports the validity

[63] I add to each data source by considering an individual as having opened an account if the other source indicates so (this is effective for a small amount of individuals). Since the survey data is only available for a small subset of the sample, errors in one or both of these data sources might still cause differences in measured account opening rates.

[64] Note that both estimates are quite large, considering that the unconditional likelihood of initially unbanked workshop graduates opening accounts at NTFCU or BCFCU is 15-35% (depending on the data source used).

of using data on account openings at NTFCU and BCFCU to estimate the effect of *Proximity* on the likelihood of account opening at *any* of the collaborating credit unions.

In regression (7) the dependent variable is *Opened_Anywhere*, indicating whether an account had been opened at any mainstream financial institution following the workshop.[65] Importantly, the coefficient on $Proximity * Unbanked$ is very similar across regressions (6) and (7).[66] This provides evidence against workshop graduates living relatively far away from all collaborating credit unions being more likely to open accounts at other mainstream financial institutions (in fact, the coefficient in regression (7) is slightly *larger* than the one in regression (6), implying the opposite). We can therefore use *Proximity* to estimate the effects of opening an account at a collaborating credit union versus not opening an account at all.

C. The Exclusion Restriction

In addition to having a clear effect on the likelihood of opening an account, *Proximity* must not be related to outcomes for any other reason in order for it to be a valid instrument. *Proximity* is, however, a combination of place and time, which raises a number of potential concerns.

Time effects may be cause for concern since collaborations were signed gradually over the sample period, so that individuals undergoing the workshop in the latter part of the 2008-2011 sample were more likely to have been living close to a collaborating credit union, other things equal. Economy-wide fluctuations, in particular those of the magnitude of the last few years, may thus induce correlation between *Proximity* and outcomes that is unrelated to account opening. I include fixed effects for the quarter-year in which the workshop was taken (i.e. Q3 2009) in order to address this concern.

A potentially more serious concern is place effects: *Proximity* reflects the location of an individual's home. Individuals living in different places might exhibit different financial behavior and different propensities to change this behavior following financial education, other things equal.[67]

[65] Note that this variable is identical to *Opened_at_Coll_CU* for the initially banked. Survey respondents were asked about their current ownership of checking and savings accounts. Initially unbanked individuals who reported having an account were categorized as having opened an account after the workshop. Initially banked individuals who reported having an account at one of the collaborating credit unions were asked if they had opened the account before or after attending the workshop, and this is what the survey variables *Opened_at_NT_BC* and *Opened_at_Coll_CU* are based on. There is no data on whether initially banked individuals had opened accounts at other institutions after the workshop. This is, however, likely to be rare.

[66] Coefficients on $Proximity * Banked$ in regressions (6) and (7) are too small to be statistically significant, given the amount of noise in these regressions. In analyzing survey outcomes I concentrate on the initially unbanked.

[67] Note that most of the outcomes studied in this paper are changes relative to the beginning of the workshop (e.g. "the change in credit score from the beginning of the workshop to 0.5-3 years after the workshop"). For this type of outcomes, exclusion restriction violations would entail individuals living closer to collaborating credit unions being more or less likely to *change* their behavior following the workshop rather than being more or less likely to exhibit certain behavior at a certain point in time.

There are several specific reasons to believe that *Proximity* might be correlated with outcomes through place effects. First, the locations of the credit unions with which NTFP collaborates are not random: these credit unions, which are dedicated to serving low-income communities, presumably choose their locations based on the characteristics of residents of the surrounding areas as well as on the centrality of the specific location. Collaborations themselves are not random: NTFCU chooses to collaborate with certain credit unions and not with others. To the extent that collaborations were signed with credit unions located in places around which residents are likely to exhibit different outcomes for some reason, this will induce bias.

Second, as explained above workshops are conducted at different sites that belong to specific partner organizations. Individuals undergoing the workshop at a specific site are relatively homogenous, in terms of both geographical location and other personal characteristics that may have to do with outcomes (i.e. they are all single mothers or formerly homeless).[68] This is another channel through which place effects might induce spurious correlation between *Proximity* and outcomes.

I address this problem by controlling for several types of variables. First, I control for various observable individual characteristics which seem relevant to outcomes and which may proxy for place-related differences. I include dummy variables for whether the individual is male, black, Hispanic, born in the US, unemployed, graduated high school, is married/in a domestic partnership and is receiving some form of welfare (public assistance, SSI, SSD, or food stamps) as well as a variable equal to the individual's age. Second, I control for unobservable differences related to workshop sites by including fixed effects for the different sites.[69] Including these variables should also decrease residual variance and result in more precise estimates. Finally, I am able to directly control for any time-invariant place effects which may confound estimation - observable and unobservable - by exploiting the fact that collaborations were not all in place at the beginning of the sample period. This enables the inclusion of the control variable $Proximity\_No\_Coll$ which, as explained above, is based on the distance between workshop participants' homes and the closest collaborating credit union, regardless of whether collaboration was in place at the time in which they underwent the workshop. The remaining variation in *Proximity* is due to collaboration initiation with each credit union increasing *Proximity* for some of those who undergo the workshop from that time on, but not for those who underwent the workshop before collaboration

[68]Individuals undergoing the workshop at a specific site are also likely to receive similar educational content, as counselors are often assigned to workshops in specific sites.

[69]Recall that the sample includes participants of 277 individual workshops conducted at 50 different sites. One might also wish to control for *specific workshop* fixed effects, but this severely limits variation.

was initiated. To the extent that place effects are invariant over time, and in particular pre- and post-collaboration, they will be captured by this control variable.

A remaining potential source of bias is unobservable time-specific place effects that are related to *Proximity*. These may arise if the timing of collaborations with specific credit unions was somehow related to outcomes in their surrounding areas. For example, collaborations with specific credit unions might have been signed as part of some bigger process which the neighborhood around them had been going through at the time and which also led to better outcomes. However, NTFP staff has indicated that the timings of collaboration initiations have been practically random.

In order to provide evidence in support of the empirical approach, I examine whether *Proximity* predicts differences in incoming observable characteristics. The rightmost 6 columns of Panel A of Table 1.2 report the results of regressions of the following form:

$$Incoming = X'\beta + \rho_1 Proximity * Banked + \rho_2 Proximity * Unbanked + \phi Unbanked + \varepsilon$$

where *Incoming* includes the demographic and incoming financial behavior variables described in Section 1.4 and X' includes different combinations of the non-demographic control variables described above, as indicated in the bottom of the panel.

Three regressions are estimated for each dependent variable. The first two columns in the right-hand side of Panel A report coefficients on the *Proximity* interactions from the first regression, where no control variables are included. These coefficients indicate that workshop participants living closer to collaborating credit unions are less likely to be male, black, and born in the US and more likely to be Hispanic and married or in a domestic partnership. These coefficients are mostly smaller and less likely to be statistically significant, however, when I include control variables in regressions (2) and (3). Regression (1) indicates that unbanked workshop participants with larger values of *Proximity* are also more likely to receive welfare payments and have lower average earned income. While the former difference is not robust to the inclusion of control variables in regressions (2) and (3), *Proximity* negatively predicts the earned income of initially unbanked individuals in all regressions.[70] Education level, age, and employment status are not predicted by *Proximity*. The same is true for all of the incoming financial behavior variables examined in Panel A: saving, using check-cashing stores, and having ever seen one's credit report.

The righthand side of Panel B of Table 1.2 performs the same exercise for incoming variables

[70] This difference disappears when conditioning on other demographic variables, however.

taken from credit reports. The coefficients on the *Proximity* interactions are small and statistically insignificant for virtually all of the variables in all three specifications. Having a FICO score (vs. having "IC"), the actual score, having an active balance or an open credit card, and having accounts in collections or judgments, as well as the balances of all of these, are not significantly predicted by *Proximity*. Overall, the results presented in this table lend support to the exogeneity of *Proximity*. In particular, variables which have to do with financial/credit behavior do not seem at all related to *Proximity*, even unconditionally.

A final issue with regard to the exclusion restriction has to do with the possible dependence of the effects of opening an account on *Proximity*. For example, one might argue that, conditional on having opened an account at one of the collaborating credit unions, those living closer to their credit union are more likely to deposit cash in their account rather than keep it at home, where it is more prone to temptation spending. Conversely, funds kept in the account might be more accessible for those living closer, making temptation spending more likely.[71] Those living closer to their credit union might also be more likely to seek the financial advice given at the credit union. Strictly speaking, this would be a violation of the exclusion restriction since it is a causal effect of the instrument on outcomes.[72] In this case, however, it seems to be more an issue of defining the treatment instrumented for ("opening an account" vs. "opening an account and visiting the branch").

1.6. Outcome Variables

In this section I describe and summarize the outcome variables. This is important for being able to interpret the results in Sections 1.7 and 1.8. Summary statistics of outcome variables for sample individuals, all graduates of the financial literacy workshops, might also be taken as informative regarding the effects of these workshops. It should be kept in mind, however, that participation in these workshops is voluntary, so that selection is likely an issue.

[71] Anecdotally, an NTFP employee has told me that some of their clients prefer to live far away from their financial institution for this reason.

[72] As Angrist and Pischke (2008) (p. 117, 152-161) emphasize, the exclusion restriction has two parts when we allow for treatment effect heterogeneity. First, the instrument should be independent of potential outcomes, conditional on covariates ("independence assumption"). This is the part argued for above. Second, the instrument should not causally affect outcomes other than through the first stage ("exclusion restriction").

A. Credit Report Outcome Variables

Most of the outcome data in this paper comes from credit reports. I construct credit report outcome variables by comparing data from reports pulled between 6 months and 3 years after the workshop ("follow-up reports") to data from reports pulled at the beginning of the workshop.[73] Panel A of Table 1.4 provides summary statistics for these outcome variables, which I divide into variables having to do with credit scores or a lack thereof (see Section 1.4.E), variables having to do with mainstream credit usage, and variables having to do with delinquent credit activity. I discuss each of these types of outcomes in turn. More information on credit reports and summary statistics of incoming variables are found in Section 1.4.E.

[Insert Table 1.4 here]

1. Credit scores

As described in Section 1.4, credit scores are an aggregator of information regarding an individual's creditworthiness. They are broadly based on the extent of an individual's (mainstream) credit usage and on past delinquencies.[74] While examining the specific items that make up the score is better in order to pick up specific behavioral changes, some of these items are unavailable to me (e.g. recent credit inquiries made about the individual). The credit score provides a simple creditworthiness assessment that considers all of the information available to the credit bureau. Perhaps more importantly, credit scores are not only a reflection of credit behavior but also an extremely important outcome in their own right. They are used to determine credit conditions, from interest rates to cell-phone deposits, to screen prospective renters, and even to evaluate job applicants.[75] The importance of credit scores might reduce their informational content, however, since individuals clearly have an incentive to try to optimize their scores in various ways which do not necessarily reflect better credit behavior.[76] This is an especially relevant point here since most CDCU's offer free financial advice, where valuable tips for credit score optimization might be transmitted.

The first part of Table 1.4 summarizes variables which have to do with credit scores, separately

[73] Several variables are only available from the follow-up reports. See below.
[74] See http://www.myfico.com/CreditEducation/WhatsInYourScore.aspx.
[75] See http://www.nytimes.com/2010/04/10/business/10credit.html, http://www.nytimes.com/2011/07/17/realestate/prospective-renters-have-much-to-prove-to-landlords.html.
[76] Examples are correcting mistakes on the credit report, asking lenders to erase old delinquencies, and optimizing the mix of credit cards used. See http://money.msn.com/credit-rating/9-fast-fixes-for-your-credit-scores-weston.aspx.

for the initially banked and the initially unbanked. While 79% of initially banked individuals had a FICO score at the beginning of the workshop, only 47% of the initially unbanked had one.[77] *Improved_Score* is a dummy variable that is defined only for those who had a FICO score at the beginning of the workshop. It takes the value of 1 if the score in the follow-up pull was higher than the initial score and 0 otherwise. A similar fraction of those with an initial score were able to improve it across the initially banked and the initially unbanked: 56-59%. Those who did not improve their score either had a lower score at the time of the follow-up pull or had no score at all at that time ("Became IC" in the table). *FICO_Change* is the change in FICO score between the initial pull and the follow-up pull. It is defined only for those who have scores in both pulls.[78] The mean of *FICO_Change* is positive for both groups: it is 13 points for the initially banked and 31 points for the initially unbanked.

Established_Score is a dummy variable that is defined only for those who had "Insufficient Credit" rather than a FICO score at the beginning of the workshop. It takes the value of 1 if a score had been "established" by the time of the follow-up pull, i.e. if the follow-up pull resulted in a number score. This is the case for 33% of the initially banked and just 16% of the initially unbanked who started out "IC". I also construct the dummy variable *Has_Score* which is defined for everyone and equal to 1 if the follow-up pull resulted in a FICO score rather than in "IC". This variable captures score establishment as well as "score loss", whereby an individual who had a FICO score at the beginning of the workshop does not have one at the time of the follow-up pull. This is the case for a non-trivial fraction of individuals. In particular, 23% of initially unbanked individuals who had FICO scores at the beginning of the workshop do not have a score at the time of the follow-up pull ("Became IC").

Established_Prime_Score is defined similarly to *Established_Score*, except that it only takes the value of 1 if the follow-up pull resulted in a score higher than 620, the conventional cutoff between "poor" and "OK" credit. This was achieved by only a small fraction of those who had no score at the beginning of the workshop: 18% of the initially banked and just 7% of the initially unbanked, roughly half the fraction of those who started out "IC" and had any score at the time of the follow-up pull.

In order to make use of all available credit report observations, I construct the variable *Improved_Established*. This variable combines *Improved_Score* and *Established_Prime_Score*

[77] These figures are slightly different than the ones reported in Table 1.2 since they only take into account individuals for which there is a valid follow-up credit report pull.

[78] This might cause selection bias. See footnote 96.

to indicate whether the FICO score had improved / a 620+ score had been established.[79] *Improved_Established* can be seen as capturing a general improvement in the individual's credit situation, as reflected by his credit score.

2. Measures of mainstream credit usage

I examine several measures of mainstream credit usage which appear on the credit report, summary statistics for which are found in the second part of Panel A of Table 1.4.[80]

Has_Bal is a dummy variable equal to 1 for those who had an "active" debt balance (see Section 1.4.E) at the time of the follow-up credit pull. This is the case for 79% of the initially unbanked and 49% of the initially banked. Unsurprisingly, these figures are very similar to the ones for *Has_Score*. Comparing these two variables to the corresponding beginning-of-workshop variables, it is evident that workshop participants do not exhibit an increased likelihood of taking part in the mainstream credit system at the time of the follow-up pull, on average. As can be seen in Table 1.4, *Has_Bal* reflects both entry to and exit from the mainstream credit system (similar to *Has_Score*): some individuals begin the workshop with an active balance and have no active balance at the time of the follow-up pull, and vice-versa.

Change_Bal is the change in the outstanding active debt balance of an individual, from the beginning of the workshop to the time of the follow-up credit pull. Mortgage balances, which are very rare (less than 5% of the sample have them) but relatively extremely high, are ignored here in order to reduce noise, and the extreme top and bottom 5% are dropped. The mean change is positive at around $500 for both the initially banked and the initially unbanked.

#_Active_Accts is the number of active accounts on the individual's credit report at the time of the follow-up credit pull (this variable is not available for beginning-of-workshop pulls). After dropping the top 5% values, initially banked individuals have about 8 active accounts on average while the initially unbanked have about 3.[81]

While credit reports contain detailed information on all types of mainstream credit, I devote special attention to credit cards. 43% of the initially banked and 82% of the initially unbanked had no open credit cards at the beginning of the workshop. *Has_CC* is defined analogously to

[79] Predicting initial FICO scores for those without an initial score based on various observables and defining *Improved_Established* using improvement relative to the predicted values results in a very similar distribution of *Improved_Established*.

[80] I also consider *Established_Score* and *Has_Score*, defined above, to be measures of mainstream credit usage.

[81] Conditional on having a positive number of accounts, means are about 10 and about 4.5 for the initially banked and for the initially unbanked, respectively.

Has_Score and *Has_Bal*, measuring credit card ownership at the time of the follow-up credit pull. The overall fraction of credit card owners increases somewhat between the beginning of the workshop and the time of the follow-up pull: from 57% to 64% for the initially banked and from 18% to 23% for the initially unbanked. *Rev_Cred_Limit* is an individual's total revolving credit limit, i.e. the total credit line available on all credit cards, at the time of the follow-up pull (data is not available for beginning-of-workshop pulls). Average post-workshop revolving credit limits are much higher for the initially banked: $2,541 vs. $589 for the initially unbanked. This is mostly due to the fact that the vast majority of the initially unbanked don't own credit cards: when conditioning on a positive limit, means are much closer ($4,684 vs. $3,457).

3. Measures of delinquent credit activity

I examine the following measures of delinquent mainstream credit activity, summarized in the third part of Panel A of Table 1.4.

Amount_P_Due is the total amount that is listed as past due on active credit report accounts at the time of the follow-up pull. It is an amount that is considered delinquent but has not yet been transferred to collections, i.e. it reflects relatively recent delinquent behavior. Similar to *#_Active_Accts* and *Rev_Cred_Limit*, this variable is only available from follow-up credit reports. Around 35% of sample individuals have some amount past due at the 0.5-3 year post-workshop horizon (the figures for the initially banked and the initially unbanked are virtually identical). This may seem to be a low number, but note that in order to have amounts past due on active accounts, one needs to have an outstanding mainstream credit balance, and this is the case for 79% and only 49% of initially banked and initially unbanked individuals, respectively. The average of *Amount_P_Due*, after dropping the top 5% values, is $542 for the initially banked and $699 for the initially unbanked (when conditioning on a positive amount, averages are $1,764 and $2,065 for the initially banked and the initially unbanked, respectively).

Collections and judgments are debts at advanced stages of delinquency. *Change_Colls_Jdgs* and *%_Change_Colls_Jdgs* are, respectively, amount and percentage changes in the total amount in collections and judgments, from the beginning of the workshop to the time of the follow-up credit pull. These balances are reduced by quite a bit, on average, following the workshop: 15% for the initially banked and 22% for the initially unbanked (after dropping the top 5% values). I devote less attention to collections and judgments, however, since debts in advanced stages are likely to be impacted by negotiations and by strategic behavior (for example, in some cases it is better to wait for a statute of limitations to take effect than it is to pay off relatively old debt).

B. Survey Outcome Variables

Panel B of Table 1.4 summarizes outcome variables taken from the survey described in Section 1.4.C. 222 surveys were completed by initially banked workshop graduates and 61 surveys were completed by initially unbanked graduates. The small number of surveys completed by the initially unbanked limits statistical power for this group. I therefore only examine questions where there is reasonable variation in responses.

The first outcome variable in Panel B is *Saving*. I report the fraction of survey respondents who report that they are saving at the time of the survey,[82] conditional on whether or not they had reported saving at the beginning of the workshop ("initially saving"). The increase in the fraction of those who report saving is significant: while 56% of initially banked respondents and only 13% of initially unbanked respondents reported saving at the beginning of the workshop, 68%-69% of survey respondents report currently saving, and there is virtually no difference between the initially banked and the initially unbanked. For the initially unbanked, this represents an increase of more than 400%.

The second outcome variable summarized in Panel B, $Saving\_More$, incorporates saving changes on the intensive margin, based on self-reported information. Respondents were asked to recall whether or not they were saving 3 years ago[83] and to assess whether they were now saving more, less, or about the same as 3 years ago.[84] The answers to these questions indicate that 46% of initially banked survey respondents and 54% of initially unbanked respondents feel that they are now saving more than they did 3 years ago. Around 25% of respondents in both groups feel that they are now saving less than they did 3 years ago.

The next two variables concern self-reported overspending. The fraction of respondents who reported feeling that they were spending more than they should be[85] is about 44% and is very

[82]The question was: "Do you currently save? That is, do you put some money aside in most months?"

[83]The survey was conducted in July 2012-March 2013 among individuals who had undergone the workshop in 2008-2011. 3 years was chosen as a benchmark for self assessment of changes since around the time of the workshop, obviously applying to some respondents more than to others. The fraction of initially unbanked respondents who reported having been saving 3 years ago in the survey is much higher than the fraction that had reported saving in the questionnaire given at the beginning of the workshop (36% vs. 13%). For the initially banked, fractions are virtually identical at 56%. Correlation between these variables is somewhat low at 29%. In any case, this survey question and an analogous question on overspending are important since they represent valuable information on self-assessed changes in behavior.

[84]This question was presented only to those who indicated that they were currently saving as well as saving 3 years ago. I include those who report saving now but not 3 years ago in the "more" category and those who report saving 3 years ago but not now in the "less" category.

[85]The question was: "Do you feel that you're currently spending more than you should be?"

similar across initial banking status. Respondents were then asked to assess their current overspending relative to 3 years ago.[86] 62% of initially banked respondents and 41% of initially unbanked respondents report overspending less now than they did 3 years ago, and 19% and 34%, respectively, report overspending more now. I use this data to construct the outcome variable $Overspending\_Less$.

The next question summarized in Panel B is adapted from Lusardi, Schneider, and Tufano (2011). Respondents were asked how confident they were that they would be able to come up with $1,500 within the next 30 days.[87] Lusardi, Schneider, and Tufano (2011) take this level of confidence to be a measure of "financial fragility". 47% of initially banked and 30% of initially unbanked respondents reported being certain that they could come up with the full $1,500 in 30 days, while 29% of the initially banked and 34% of the initially unbanked reported that they could probably come up with the funds (the other options, summarized in Table 1.4, were "I could probably not come up with $1,500" and "I am certain I could not come up with $1,500"). For the purpose of the analysis I construct the dummy outcome variable $1500\_Confidence$ which takes the value of 1 if either of the first two options was chosen. The mean of this variable, 76% for the initially banked and 64% for the initially unbanked, reflects much lower "fragility" than that found by Lusardi, Schneider, and Tufano (2011). They report that only 50% of the general population (25-30% of those earning less than $30,000 a year, who are the comparable population) respond that they are certain or that they could probably come up with $2,000 within the next month. This difference is significant and likely reflects more than just the difference in the amount of money stated in the question.

The next two questions summarized in Panel B are constructed using questions in which respondents were asked to report, on a scale of 1-5, how knowledgeable they feel about managing their money on a daily basis and how stressed they feel about their personal finances and debt. These questions had also been asked in the questionnaire administered at the beginning of the workshop, which enables comparison over time.[88] Average self-reported knowledge is higher at the time of the survey than it is at the beginning of the workshop, at about 4 vs. about 3.2. Average

[86] This question, which followed the question about current overspending, was: "Many people spend money on things and later wish that they hadn't. Compared to 3 years ago, do you now spend more, less, or about the same on these types of things?"

[87] The question was: "Say that some unexpected need arose and you had to come up with $1500 within the next month. How confident are you that you would be able to do so?" I use "30 days" and "a month" interchangeably here, following Lusardi, Schneider, and Tufano (2011). The question concerns $1,500 rather than the $2,000 used in that paper to accommodate the fact that the population studied here is low-income. $1,500 is an amount which Lusardi, Schneider, and Tufano (2011) mention as being relevant for low-income households, citing Brobeck (2008).

[88] However, the questions had only been included in the beginning-of-workshop questionnaire in the latter part of the sample period (2010-2011).

self-reported stress is lower at the time of the survey, at about 3.15 vs. about 3.65 at the beginning of the workshop. Interestingly, knowledge and stress levels are similar across initial banking status at both points in time. I use the responses to these questions to construct the outcome variables $Knowledge$ and $Stress$.

The final question summarized in Panel B was adapted from Melzer (2011), who uses data from the National Survey of America's Families. Respondents were asked whether they had been unable to pay their bills on time during the past 6 months.[89] 29% of initially banked and 34% of initially unbanked survey respondents reported being unable to pay on time at some point during the past 6 months. I use these responses, which Melzer (2011) takes to be a measure of economic hardship, to construct the variable $Unable\_Bills$.

1.7. Reduced Form Proximity Effects

In this section I report the results of reduced form OLS regressions of the form:[90]

$$Outcome \quad = \quad X'\pi + \delta_1 Proximity * Banked + \delta_2 Proximity * Unbanked + \phi Banked + \varepsilon$$

where $Outcome$ is a credit report or survey outcome, as described above, and X' contains control variables and an intercept, as well as the initial value of $Outcome$, where available. In this type of specification, δ_1 and δ_2 represent proximity effects on the initially banked and on the initially unbanked, respectively.[91] I also test for differences between these coefficients, i.e. between proximity effects on the initially banked and the initially unbanked, in every regression. Note that while the effects estimated in this section are reduced form effects, they have the same sign and are indicative of the statistical significance of the causal effects of opening an account on each group. Given the partial availability of account opening data, I devote significant attention to these effects. In Section 1.8 I employ the SSIV method (Angrist and Krueger (1995)) to quantify the effects of opening an account on the outcomes for which I find statistically significant reduced

[89] The question was: "During the last 6 months, was there a time when you and your family were not able to pay your rent, mortgage, or utility bills on time?"

A similar question was asked regarding postponing medical care due to economic hardship, but there was insufficient variation to enable meaningful analysis. In particular, only 8 initially unbanked respondents answered this question with a "yes".

[90] I use a linear probability model for the many binary outcomes examined here rather than nonlinear models such as Probit or Logit. This is advocated on the grounds of simplicity, in particular in the context of instrumental variable estimation, by Angrist and Pischke (2008). The OLS coefficients I report here are similar to average marginal effects estimated using a Logit model.

[91] This is equivalent to including $Proximity$ and $Proximity * Unbanked$ on the RHS but more convenient for examining the proximity effect for each group.

form effects.

A. Proximity Effects on Credit Report Outcome Variables

Panel A of Table 1.5 reports regression results for the outcome variable *Improved_Established*. The regression reported in column (1) includes no control variables except for a dummy variable for whether the individual had a FICO score at the beginning of the workshop and the initial score.[92] These variables are included in all of the regressions reported in Panel A. Regression (2) also includes demographic variables[93] and fixed effects for the quarter in which the individual took the workshop.[94] Regression (3) includes workshop quarter fixed effects and *Proximity_No_Coll*, which effectively captures all observable and unobservable time-invariant place effects, as explained in Section 1.5.C. Regression (4) includes demographic variables, workshop quarter fixed effects, and workshop site fixed effects.

[Insert Table 1.5 here]

Across all of the specifications in Panel A, coefficients on *Proximity* interacted with *Banked* are small and statistically insignificant, whereas those on *Proximity* interacted with *Unbanked* are positive and significant. That is, proximity to a collaborating credit union significantly affects the likelihood of improving one's credit score relative to the beginning of the workshop or of establishing a non-subprime score (in case that there is no initial score), but only for initially unbanked individuals. Below I analyze the effects on each of these outcomes (improving one's score and establishing a non-subprime score) separately. Coefficients for each group are similar across specifications and range from 0.062 to 0.074 for the initially unbanked. The finding that adding control variables does not have dramatic effects on the estimated coefficients lends further support to the exogeneity of *Proximity*. In particular, the coefficient on $Proximity * Unbanked$ in regression (3), which controls for unobservable, time-invariant place effects through *Proximity_No_Coll*, is similar in magnitude to the other specifications, though far noisier since including *Proximity_No_Coll* severely limits the variation in *Proximity*. In later analyses, which often involve fewer observations, I concentrate on specifications (2) and (4), which generate more precise estimates. The

[92] Those without a FICO score are given a value of zero for the latter variable.

[93] Demographic variables include indicators for being male, black, Hispanic, US-born, employed, married, a high school graduate, and a recipient of welfare payments, as well as the individuals' age. Variables are defined in Section 1.4.D and summarized in Table 1.2. These are the variables used in all regressions which control for Demographics.

[94] The sample consists of individuals who participated in workshops in 2008-2011, so there are 15 workshop quarter fixed effect dummy variables (one is omitted).

magnitude of the estimated coefficients on the initially unbanked, which is around 0.070, should be interpreted as "living 1% closer to a collaborating credit union raises the likelihood of improving one's credit score/establishing a 620+ score by approximately 0.070%", since *Proximity* is logged.

The coefficients on the dummy variable *Unbanked*, tabulated below the coefficients on the *Proximity* interactions, are insignificant and vary widely across specifications, which makes sense given their presumed correlation to many of the control variables used. Test statistics for the equality of $Proximity * Banked$ and $Proximity * Unbanked$, tabulated towards the bottom of the table, enable us to reject the hypothesis that they are equal at conventional significance levels for all specifications. In Table 1.3 I show that while *Proximity* is a stronger predictor of account opening for the initially unbanked, it also predicts account opening by the initially banked. The null results for the initially banked, which are replicated for virtually all of the outcomes studied in this paper, should therefore not be taken as placebo effects but rather as evidence of the null effects of opening an account at a Community Development Credit Union on those who already have accounts at other institutions.

Panel B of Table 1.5 concentrates on those who had a FICO score at the beginning of the workshop and reports the results of regressions where *Improved_Score* and *FICO_Change* are the dependent variables. The coefficients on $Proximity * Unbanked$ in regressions (1)-(4), which are identical to those reported in Panel A in terms of the control variables used, are much larger than those in Panel A. It follows that the effects reported in Panel A on *Improved_Established*, which is a combination of *Improved_Score* and *Established_Prime_Score*, mostly reflect credit score improvement by those who had a score to begin with rather than score establishment by those who had no score.[95] Coefficients range around 0.110-0.130, implying that decreasing the distance from the closest collaborating credit union by 1% increases the likelihood of credit score improvement by 0.110-0.130%. In regressions (5)-(8) *FICO_Change*, the difference in FICO scores between the beginning of the workshop and the time of the follow-up pull, is the dependent variable.[96] Coefficients on $Proximity * Unbanked$ are all positive and around 10, though only the

[95] Regressions where *Established_Prime_Score* is the dependent variable yield non-significant coefficients. Panel C presents related evidence.

[96] In order to be included in regressions where *FICO_Change* is the dependent variable, an individual must have a score (rather than "IC") in both credit pulls. There might be a form of selection here, whereby *Proximity* affects the composition of those who have FICO scores at the time of the follow-up pull. In particular, since *Proximity* predicts account opening, those who live closer to collaborating credit unions might be more likely to have a score at the time of the follow-up pull. If those who are "selected in" have worse (or better) scores on average, coefficients on *Proximity* will be biased. In practice, however, *Proximity* does not strongly predict having a follow-up score (see Panel C) and results derived using the method introduced in Heckman (1976), which accounts for this potential selection problem, are very similar to OLS results.

one in regression (7) is statistically significant at conventional levels.[97]

Panels C and D of Table 1.5 provide evidence regarding the origins of credit score improvements, as well as further evidence regarding the effects of proximity to collaborating credit unions on subsequent financial behavior. Panel C reports the results of regressions where various measures of mainstream credit usage are the dependent variables. I divide these variables into those having to do with the extensive margin of mainstream credit usage ("entering the mainstream credit system") and those having to do with the intensive margin ("changes in credit usage for those already in the mainstream system"). Each column contains results for a different dependent variable, and all regressions include as controls demographics, workshop quarter fixed effects, and workshop site fixed effects (similar to specification (4) in Panel A). In regressions (1)-(3), the dependent variables have to do with the extensive margin of mainstream credit usage. In order to increase statistical power and capture both entrance into and exit out of the mainstream credit system (both of which occur to some extent, as Table 1.4 shows), I use variables that reflect inclusion in the mainstream system at the time of the follow-up pull and control for the beginning-of-workshop state.[98] Regression 1 shows a small, statistically significant effect of *Proximity* on the likelihood of having a credit score at the time of the follow-up pull (*Has_Score*) for initially unbanked workshop participants, controlling for having a score initially as well as for the initial score. Regressions (2) and (3), however, find no effect on having an active mainstream debt balance (*Has_Bal*) or on having an open credit card (*Has_CC*). All in all, *Proximity* (opening an account at a collaborating credit union) does not seem to have a substantial effect on inclusion in the mainstream credit system. This is disappointing given the "gateway to the mainstream financial system" label often given to bank account ownership. In the next section I show that initially unbanked workshop participants who graduated high school do seem to enter the mainstream system as a result of becoming banked, and propose that there is a minimal level of financial literacy that is required for this to happen.

Regressions (4)-(6) in Panel C, which include only individuals who had an active mainstream credit balance at the beginning of the workshop, show that *Proximity* does not have any statistically significant effects on the intensive margin of mainstream credit usage. In particular, the effects on the change in the total active debt balance (*Change_Bal*), the number of active accounts

[97] Other coefficients are on the verge of statistical significance. In particular, the p-value of the coefficient in regression (8) is 11.6%.

[98] Regressions which examine the effects of *Proximity* on entrance into the system for those who are out (e.g. credit card issuance for those who do not have an open credit card at the beginning of the workshop) yield similar results.

($\#\_Active\_Accts$), and the total revolving credit limit (Rev_Cred_Limit) are all insignificant. Note that the point estimate on $Proximity{*}Unbanked$ in the regression where $Change\_Bal$ is the dependent variable is negative. Given that we do not observe alternative credit usage, which most likely decreases, it is plausible that overall borrowing decreases as a result of becoming banked and facing a lower borrowing rate, consistent with the theoretical effects outlined in Section 1.3.[99] As a whole, Panel C shows that mainstream credit usage hardly changes on aggregate as a result of opening an account at a collaborating credit union, and provides evidence against this channel playing a part in the credit score improvements found in Panels A and B. Note that coefficients on $Unbanked$ (not interacted with $Proximity$) are negative and quite large in most regressions in Panel C. That is, those who are initially unbanked increase their mainstream credit usage by less after the workshop, on average, than do the initially banked, other things equal.

Panel D of Table 1.5 reports the results of regressions where measures of delinquent credit activity are the dependent variables. Regressions (1)-(4) show that $Proximity$ negatively predicts $Amount\_P\_Due$ for initially unbanked individuals. That is, relatively recent delinquent amounts, not yet sold to collection agencies, are lower at the time of the follow-up credit pull for those who live closer to collaborating credit unions and are more likely to open accounts in them. In addition to the control variables specified in the table, these regressions control for having a FICO score at the beginning of the workshop and for that initial score, similar to the regressions in Panel A. The magnitude of the estimated coefficients implies that living 1% closer to a collaborating credit union decreases amounts past due by about \$2. The results of these regressions point to reduced delinquency as a possible mechanism through which initially unbanked individuals who open accounts at collaborating credit unions are more likely to improve their credit scores. The effects are consistent with the theoretical effects outlined in Section 1.3, where several channels through which becoming banked can reduce delinquency are described. Proximity effects for the initially banked are not statistically significant, though we cannot reject the hypothesis that they are equal to the effects for the initially unbanked.

Regressions (5)-(8) in Panel D (with the exception of (7)) show that $\%\_Change\_Colls\_Jdgs$, the percentage change in the amount in collections and judgments, is not predicted by $Proximity$ for neither the initially banked nor the initially unbanked.[100] Coefficients are positive, implying,

[99] Strictly speaking, we do not observe mainstream borrowing either but rather mainstream debt balances (equal to borrowing*(1+interest)). Borrowing per-se might increase even if debt balances are unchanged. In that case, however, we might expect $\#\_Active\_Accts$ to increase (a caveat is that different accounts might be consolidated when becoming banked).

[100] Coefficients are also insignificant in all specifications but (7) in regressions where the level change $Change\_Colls\_Jdgs$ is the dependent variable.

if anything, a larger increase (or smaller decrease) in the amount in collections and judgments for those who are more likely to open accounts. As noted above, changes in collections and judgments likely reflect negotiations with collection agencies and strategic non-payment, and as such are less indicative of actual changes in individual financial behavior. I therefore do not attribute particular significance to these results.

B. Proximity Effects on Survey Outcome Variables

Panels A and B of Table 1.6 report the results of reduced form regressions where outcome variables are taken from the survey, as described in Section 1.6.B.

Recall from Table 1.3 that the effect of $Proximity$ on account opening by initially banked workshop graduates who responded to the survey is positive but too small to be statistically significant. $Proximity$ is therefore not a valid instrument for this group in the survey sample (the corresponding effect for the initially unbanked is large and statistically significant). Nevertheless, I include the initially banked in the reduced form regressions analyzed in this section in order to preserve statistical power. Due to the small number of survey observations, I control for workshop year rather than workshop quarter fixed effects. All regressions also control for the demographic variables listed in footnote 93.

[Insert Table 1.6 here]

In regressions (1) and (2) $Saving$ and $Saving\_More$ are the dependent variables. The coefficients on $Proximity * Unbanked$ are statistically indistinguishable from zero in both regressions. That is, we cannot reject the hypothesis that initially unbanked survey respondents living closer to collaborating credit unions (and more likely to open accounts in them) do not save more. This is true on both the extensive margin (as measured by $Saving$) and the intensive margin (as measured by $Saving\_More$).[101] While the theoretical effects of opening an account, as described in Section 1.3, do not necessarily imply increased saving, this result is somewhat disappointing in light of the widespread notion that account ownership encourages saving, in particular through behavioral mechanisms (Barr (2004), Mullainathan and Shafir (2009)).

Regressions (3) and (4) provide related evidence, examining self-reported overspending. If bank account ownership helps address behavioral biases which result in overspending (see Section

[101]A dummy variable indicating whether the individual had reported saving at the beginning of the workshop is included in regression (1). A variable indicating whether having been saving 3 years ago was reported in the survey is included in regression (2).

1.3), we would expect to see negative proximity effects in these regressions. The coefficients on $Proximity * Unbanked$ are, however, indistinguishable from zero. If anything, those who live closer to a collaborating credit union are slightly more likely to report that they are currently overspending and slightly less likely to report that they are overspending less than 3 years ago (as measured by $Overspending$ and $Overspending\_Less$, respectively), though the coefficients are far from being statistically significant.

In regression (5) $1500\_Confidence$ is the dependent variable. Bank account ownership might be expected to influence confidence in being able to raise a relatively large amount of money within a short time (to decrease "financial fragility") by improving credit access or by increasing savings. The coefficient on $Proximity * Unbanked$ in regression (5) is, however, negative and statistically insignificant. While regressions (1) and (2) show no evidence of an effect on saving, credit access does presumably improve when an account is opened at a CDCU, so it is surprising that no effect is found. Respondents who did not indicate that they were certain they could not come up with \$1500 within a month were also asked *how* they would go about getting the \$1,500 if they had to do so. Interestingly, only 38% of those who had an account at a collaborating credit union at the time of the survey indicated that they would take out a loan from their credit union.[102] This might point to a perceived inability to receive loans at the credit union or to an unwillingness to borrow (see heterogeneous effects analysis, below). It might also be that the initial question was phrased in a way that seemed to preclude taking out an official loan ("able to come up with \$1500" may have been understood in this way), though the relatively high rates of confidence implied by the responses would seem to contradict this interpretation.

The only statistically significant result in Table 1.6 is in regression (6), where $Knowledge$ is the dependent variable. Living 1% closer to a collaborating credit union results in a 0.0026 higher self-assessment of knowledge about managing one's money on a daily basis (recall that the scale in this question was 1-5).[103] Account ownership might increase knowledge through enabling individuals to follow their transactions more closely or through exposing them to different financial products. Another potential channel through which knowledge might increase is access to financial advice, which collaborating credit unions provide for free. Advice might also play a role in the credit report results, as credit scores are optimizable to some extent (see footnote 76). The null credit report results for the initially banked, who also gain access to free financial advice when they

[102] 69% indicated that they would draw on their own funds while 55% indicated that they would borrow from family or friends; 14% of those with accounts were certain that they could not come up with the \$1500.

[103] The initial level of self-assessed knowledge is not controlled for in this regression since that reduces the number of observations considerably. The coefficient on $Proximity * Unbanked$ more than doubles in that case.

open an account at one of the collaborating credit unions, provide evidence against this channel, however.

Regressions (7) and (8) examine the effects of *Proximity* on self-reported finances-related stress and on economic hardship (as measured by having been unable to pay bills in the past 6 months), respectively. *Stress* and *Unable_Bills* might be expected to decrease as a result of account ownership due to increased credit access and increased savings, similar to $1500\_Confidence$, as well as, perhaps, to a general improvement in one's economic situation. However, both coefficients on $Proximity * Unbanked$ are positive rather than negative and statistically insignificant.

The analysis of survey outcomes implies that opening an account at a collaborating credit union affected none of the dimensions of behavior and wellbeing examined except for knowledge. The results should, however, be taken with a grain of salt due to the small number of observations and low statistical power. Since standard errors are large, confidence intervals for some of the estimated coefficients include zero as well as figures that are quite large. For example, the 95% confidence interval for the coefficient on $Proximity * Unbanked$ in regression (1) in Table 1.6, where *Saving* is the dependent variable, is approximately $[-0.14, 0.22]$.

1.8. Additional Analyses

A. SSIV Estimates of the Effects of Opening an Account

In this section I employ the Split Sample Instrumental Variables estimator introduced by Angrist and Krueger (1995) to estimate the effects of opening an account at a collaborating credit union. The goal is to quantify the effects found in the reduced form regressions.[104] SSIV (also referred to as two-sample 2SLS) is identical to ordinary 2 Stage Least Squares estimation except that first stage and second stage regressions are estimated using different samples.[105] This is a useful method for this study because of differences in the availability of account opening data and outcome data. Recall from Section 1.4 that account opening data is available from two sources. First, administrative data is available on all account openings at NTFCU during the 2008-2011 sample period and on all openings since July 2009 at BCFCU. This data is available for many

[104]The reduced form coefficients are approximately inflated by the inverse of the first stage coefficient to generate the causal coefficient of interest. Where the reduced form effect (the proximity effect, in this case) is zero, the causal effect is zero and there is no point in performing this exercise.
[105]See Angrist and Pischke (2008), p. 149-150.

individuals for which there is no credit report outcome data.[106] Including these individuals in first stage regressions increases the precision of first stage estimates. Second, more comprehensive account opening data is available from the survey, but only for a small subset of the sample. I use both of these data sources to estimate the first stage, i.e. the effects of *Proximity* on account opening. Recall from Table 1.3 that these effects were quite different when estimated using each of the account opening data sources. I therefore generate two SSIV estimates for each outcome.

I begin by estimating the following first stage regressions for the two endogenous variables, $Opened * Banked$ and $Opened * Unbanked$, using each account opening data source:

$$Opened{*}Banked = X'\pi_1 + \eta_{11} Proximity{*}Banked + \eta_{12} Proximity{*}Unbanked + \phi_1 Unbanked + \varepsilon_1$$

$$Opened{*}Unbanked = X'\pi_2 + \eta_{21} Proximity{*}Banked + \eta_{22} Proximity{*}Unbanked + \phi_2 Unbanked + \varepsilon_2$$

The results of these four regressions are in Panel A of Table 1.7. In regressions (1) and (2), where administrative data on account openings in NTFCU and BCFCU is used, *Opened* is equal to *Opened_NT_BC* and *Proximity* only takes these two credit unions into account, similar to specifications (1)-(4) in Table 1.3. I control for either of these two credit unions being the closest to the individual's home out of all of the collaborating credit unions. I also control for demographics, workshop quarter fixed effects, whether the individual had a FICO score at the beginning of the workshop, and the value of that score.[107] In regressions (3) and (4), where account opening survey data is used, *Opened* is equal to *Opened_Coll_CU* (see Section 1.5.B) and *Proximity* is constructed considering all collaborating credit unions (rather than just NTFCU and BCFCU).

[Insert Table 1.7 here]

The coefficients on $Proximity * Banked$ ($Proximity * Unbanked$) in regressions (1) and (3) (regressions (2) and (4)) are similar to those in specifications (2) and (6) in Table 1.3. Here, too, estimated coefficients for the initially unbanked obtained using administrative data are quite different from those obtained using survey data (0.091 vs. 0.241), though both are highly statistically significant. The coefficient on $Proximity * Banked$ obtained using survey data is positive

[106] This is mostly due to many workshop graduates (about 30%) not authorizing NTFP to pull their credit reports for evaluation purposes.

[107] Including the latter two variables reduces the number of observations by about 220 (the specification is otherwise similar to specifications (2) and (5) in Table 1.3). It is necessary, however, in order to include these variables in the second stage regressions.

and larger than that obtained using administrative data (0.073 vs. 0.056), but is statistically insignificant.

Using the estimated coefficients from these first stage regressions, I generate two sets of predicted account opening probabilities for each individual in the sample: $\widehat{Opened\_Banked}$ and $\widehat{Opened\_Unbanked}$. I then proceed to run the following second stage regressions:

$$Outcome = X'\pi_3 + \rho_1 \widehat{Opened\_Banked} + \rho_2 \widehat{Opened\_Unbanked} + \phi_3 Unbanked + \varepsilon_3$$

where *Outcome* are the four outcomes for which there are significant reduced form effects in Tables 1.5 and 1.6: *Improved_Score*, *Has_Score*, *Amount_P_Due*, and *Knowledge*. Panel B of Table 1.7 displays the estimated ρ_1, ρ_2, and ϕ_3 for the different outcomes, using each of the first stage data sources. Standard errors are bootstrapped. Given the differences in first stage estimates, it is not surprising that second stage estimates differ widely depending on the account opening data source used. I take the two estimates for each outcome variable as defining a range of possible causal effect values.

Columns (1) and (2) report second stage results for *Improved_Score*, the indicator for having improved one's FICO score, defined only for those who had a score at the beginning of the workshop. The estimated coefficient on $\widehat{Opened\_Unbanked}$ using administrative (survey) account opening data implies a 120% (47.4%) increase in the likelihood of improving one's score as a result of opening an account. The former estimate is obviously implausibly large. The 47.4% increase in *Improved_Score* implied by the estimate obtained using survey account opening data is also quite large, given that the unconditional mean of *Improved_Score* for the initially unbanked, reported at the bottom of the table, is 56.4% (the predicted values for those who open and those who don't open accounts are 66.9% and 19.5%, respectively).

Columns (3) and (4) report second stage results for *Has_Score*, the indicator for having a FICO score at the time of the follow-up credit pull. Recall that a modest proximity effect on this variable was observed for the initially unbanked in Table 1.5. Accordingly, the coefficients on $\widehat{Opened\_Unbanked}$ in regressions (3) and (4) are relatively small and statistically insignificant. The more precise coefficient in regression (4), which uses survey account opening data, implies a 14.6% increase in the likelihood of having a score at the time of the follow-up pull as a result of opening an account for the initially unbanked (the p-value is 15%; the unconditional mean of *Has_Score* is 44.7%). That is, the impact of opening an account on inclusion in the mainstream credit system is not trivial but is rather small: it is about 1/3 of the impact on improving one's credit score found in regression (2).

Columns (5) and (6) report second stage results for *Amount_P_Due*, the amount that the individual is delinquent on but has not yet been sold to collection agencies. The coefficient on $\widehat{Opened\_Unbanked}$ in regression (5), which uses administrative account opening data, is large but quite noisy, and is far from being statistically significant (the p-value is 41.2%). The coefficient in regression (6), which uses survey account opening data, is smaller and much more precise, though it is still too noisy to be considered statistically significant at conventional levels (the p-value is 14.9%). It implies a decrease of \$828 in *Amount_P_Due* as a result of opening an account (the predicted values for those who open and those who don't open accounts are \$335 and \$1163, respectively)

Columns (7) and (8) report the results for *Knowledge*, the survey question in which respondents were asked to assess their level of knowledge with respect to managing their money on a daily basis on a scale of 1-5. As with the other dependent variables, the coefficient on $\widehat{Opened\_Unbanked}$ in regression (7), which uses administrative account opening data, is implausibly large. The coefficient in regression (8), which uses survey account opening data, is plausible in size at around 1 but is far from being statistically significant (the p-value is 31%). This might be expected given the small number of survey observations. Coefficients on $\widehat{Opened\_Banked}$ are statistically insignificant in almost all of the second stage regressions reported in Panel B, in accordance with the reduced form results.

B. Heterogeneous Effects by Education Level

In this section I aim to examine whether there is an interaction between bank account ownership and financial literacy, i.e. whether the effects of account ownership are different for individuals with different levels of literacy. In Table 1.8 I report the results of regressions in which triple interactions between *Proximity*, *Banked*, and *HS* are included. The latter is a dummy variable which takes the value of 1 if the individual has a high-school diploma (this is the case for 60% of the initially unbanked and 83% of the initially banked).[108] The premise behind using formal education levels to split the sample is that those with a higher level of education are more financially literate by the end of the workshop: they are likely more financially literate to begin with (Lusardi and

[108] Using an indicator for whether the individual had attended college (true for 28% of the initially unbanked and 65% of the initially banked) yields similar results but reduces statistical power for the initially unbanked, who are less in number and who are more interesting to examine given the aggregate results.

Mitchell (2013)) and are presumably able to learn more in a classroom setting, i.e. gain more from the workshop. Indeed, *Knowledge* is on average higher for respondents who graduated high school (4.04 vs. 3.76).

Regressions take the following general form:

$$\begin{aligned} Outcome &= X'\pi + \delta_{11}(Proximity * Banked) * HS + \delta_{12}(Proximity * Banked) * Not\_HS + \\ &\quad +\delta_{21}(Proximity * Unbanked) * HS + \delta_{22}(Proximity * Unbanked) * Not\_HS + \\ &\quad +\phi_1 Banked + \phi_2 HS + \varepsilon \end{aligned}$$

where Not_HS is the complement of HS, i.e. an indicator for not having graduated high school. All regressions include as controls the demographic variables listed in footnote 93 as well as workshop quarter fixed effects.

[Insert Table 1.8 here]

Regression (1) in Panel A of Table 1.8 uses account opening data from the two credit unions for which this data is widely available (NTFCU and BCFCU) to show that proximity to these two credit unions significantly predicts account opening in them (see Section 1.5) for all *Banked* / *HS* groups except for the initially banked who are not high school graduates.[109] [110] In the analysis I concentrate on the initially unbanked and examine proximity effects for high school graduates and non-graduates. The coefficients on $Proximity * Unbanked * HS$ and $Proximity * Unbanked * Not\_HS$ in column 1 of Panel A (the estimated δ_{21} and δ_{22}, in the third and fourth rows of Panel A) show that it is valid to use *Proximity* to instrument for the effects of opening an account for these two groups (assuming that the exclusion restriction is satisfied for both), though the effect for unbanked non-graduates (0.048) is smaller than the effect for unbanked graduates (0.084). The hypothesis that *Proximity* effects are equal for both subgroups of unbanked individuals cannot be rejected, however, as can be seen towards the bottom of column 1 (the F statistic is 1.87).

The remaining columns in Table 1.8 report heterogeneous proximity effects on credit report outcomes. Regression (2) in Panel A shows that *Proximity* significantly influences the likelihood of FICO score improvement (as measured by *Improved_Score*) for both groups of initially unbanked

[109] Recall that this group is relatively underrepresented in the data (83% of the initially banked graduated high school).

[110] Using survey rather than administrative account opening data in this exercise produces similar results, with larger and more noisy coefficients (in line with the results in Table 1.3).

workshop participants (high school graduates and non-high school graduates). Coefficients are similar (0.112 and 0.119), though noisier for non-graduates (as noted above, first stage coefficients, hence the implied causal effects, are different). Regression (3) shows, however, that the proximity effect on the actual change in FICO score between the beginning of the workshop and the follow-up pull (as measured by *Change_FICO*) is much larger for high school graduates and is statistically insignificant for non-graduates.

Interestingly, regression (4) shows that the negative proximity effect on *Amount_P_Due* documented in Panel D of Table 1.5 for the general initially unbanked population is in fact limited to those who graduated high school. The coefficient for this group is large at -$424 and strongly statistically significant. Unbanked individuals who did not graduate high school do not exhibit proximity effects on *Amount_P_Due*, i.e. opening an account at a collaborating credit union does not result in lowered delinquency for this group.

The regressions reported in Panel B of Table 1.8 examine heterogeneity in the effects of *Proximity* on the measures of mainstream credit usage explored in Panel C of Table 1.5. Recall that there were no substantial proximity effects on credit usage for the general initially unbanked population on neither the extensive nor the intensive margins. The results in Panel B show that there is considerable heterogeneity in these effects. First, regression (5) shows that initially unbanked workshop participants who graduated high school exhibit a positive and statistically significant proximity effect on the likelihood of having a credit score at the time of the follow-up pull, while non graduates do not exhibit a statistically significant effect (the point estimate is actually negative for the latter group). Second, regression (7) shows a positive and statistically significant proximity effect on the likelihood of having an open credit card at the time of the follow-up pull for initially unbanked high school graduates and a *negative* significant effect for initially unbanked non-graduates (coefficients are 0.065 and -0.070, respectively). On the intensive margin, there are no statistically significant coefficients for either group of initially unbanked individuals in regressions (8)-(10), although the coefficient in the regression (10), where *Rev_Cred_Limit* is the dependent variable, is large for initially unbanked high school graduates and significant at the 17% level.

That initially unbanked high school graduates increase their usage of mainstream credit on the extensive margin while initially unbanked non-graduates do not do so is consistent with financial literacy being necessary for the access to mainstream financial services provided by account ownership to be taken advantage of. As discussed in the introduction, this might be because literacy is necessary in order to avoid costly mistakes and benefit from relatively sophisticated products like credit cards, or it might be because it is necessary in order to realize the relative costs of alternative financial services.

That opening an account *lowers* the likelihood of credit card ownership for the initially unbanked who did not graduate high school is surprising. Note that while this is the only statistically significant result for this group in Panel B, all of the coefficients on $Proximity * Unbanked * Not\_HS$ in the Panel are negative. That is, $Proximity$ (opening an account) seems to predict *less* mainstream credit usage for initially unbanked non-high school graduates. A potential mechanism is financial advice given at the credit unions: if financial literacy is indeed necessary for beneficial usage of mainstream financial products, those who have low literacy might be better off not owning credit cards, and opening an account might increase the likelihood of receiving financial advice and *decreasing* credit card usage. The finding in Panel A that $Proximity$ significantly increases the likelihood of credit score improvement for this group although delinquent balances do not decrease is consistent with a role for financial advice. This is highly speculative, however.

While the analysis described in this section demonstrates heterogeneity in the effects of account ownership on eventual inclusion in the mainstream credit system, it should be kept in mind that having graduated high school is not randomly assigned and might proxy for characteristics other than financial literacy. An obvious candidate is the individual's overall financial position, likely correlated with both formal education levels and the likelihood of being eligible for mainstream credit. To evaluate this possibility, I perform a robustness check in which I replace HS with an indicator for being employed at the time of the workshop. This analysis (untabulated) yields no heterogeneity in proximity effects on variables which measure mainstream credit usage on the extensive margin.

1.9. Conclusion

This paper finds that opening an account at a Community Development Credit Union affects certain dimensions of the financial lives of low-income recipients of financial education and does not affect others. On the one hand, opening an account raises the likelihood of credit score improvement, lowers delinquent balances, and raises self-reported knowledge about managing one's money. These are important outcomes, especially for low-income individuals. On the other hand, opening an account does not seem to result in increased saving or in decreased overspending. General “financial wellbeing”, as measured by financial fragility, finances-related stress, and economic hardship, is not significantly affected either.

The lack of effects might have to with the finding that opening a bank account is, surprisingly, not synonymous with inclusion in the financial mainstream, in the sense that it does not necessarily imply actual usage of mainstream credit. A heterogeneous effects analysis which uses high school graduation to proxy for financial literacy implies that some level of literacy is needed in order to take advantage of the credit access which a bank account provides, and that even graduates of a comprehensive financial education workshop are not necessarily at that level. This highlights the importance of complementing efforts to increase account ownership with financial education.

The findings in this paper also offer lessons for the design of financial education programs aimed at the poor. The large fraction of workshop participants who opened accounts at collaborating credit unions and the strong proximity effects on the likelihood of opening an account imply that encouraging participants of financial education programs to open accounts at specific institutions can be highly successful, but critically depends on the convenience of these options. Account ownership seems to play a role in enhancing the financial literacy of program participants, and as such should be in the direct interest of financial education programs. Further research might reveal more about the complimentarities between financial education and the improvement of access to mainstream financial services. This would improve our general understanding of household financial behavior as well as help realize the potential of the financial system to increase the welfare of the most vulnerable members of society.

Figure 1.1: NTFP's Collaborations with Community Development Credit Unions

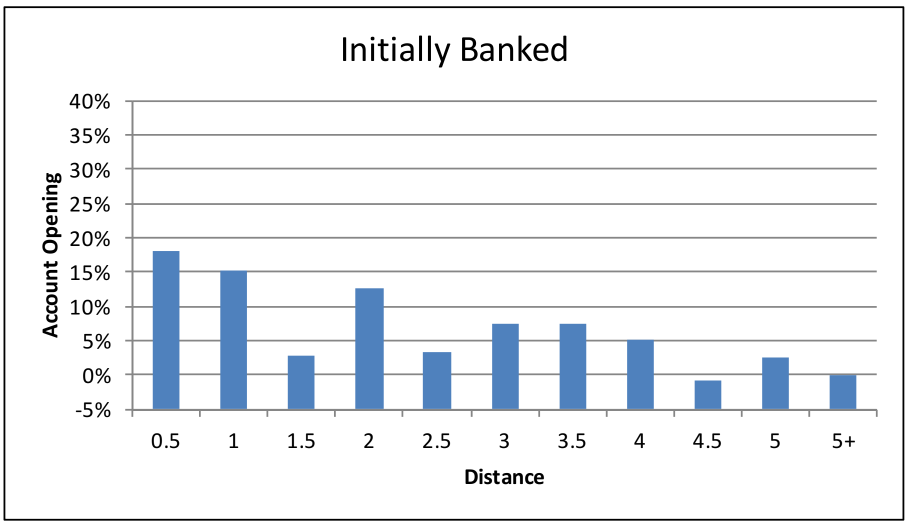
This figure shows a map of New York City with the locations of the various credit unions with which NTFP collaborates. "MOU signed" dates are dates in which Memorandum of Understandings were signed between NTFP and the various credit unions. These dates are used as collaboration starting dates in the construction of the *Proximity* instrument. Neighborhood Trust, located in the upper left corner of the map, is a credit union operated by NTFP and is considered as "always collaborating". Several of the credit unions have more than one branch.

Figure 1.2: Account Opening and Distance: Administrative Data

This figure plots coefficients estimated using the regression:

$$Opened\_at\_NT\_BC = X'\pi + \sum_{j=1}^{10} \eta_{1j} D_j Banked + \sum_{j=1}^{10} \eta_{2j} D_j Unbanked + \phi Unbanked + \varepsilon$$

where $Opened\_at\_NT\_BC$ is an indicator for whether an account was opened at either of the two credit unions for which administrative account opening data is available (NTFCU and BCFCU); X' includes an intercept, demographic control variables, and workshop quarter fixed effects; and D_j are indicators for the distance between the individual's home and the closest of these two credit unions being smaller than $j * 0.5$ and larger than $(j-1) * 0.5$. The top and the bottom plots display the estimated η_{1j} and η_{2j} coefficients, respectively.



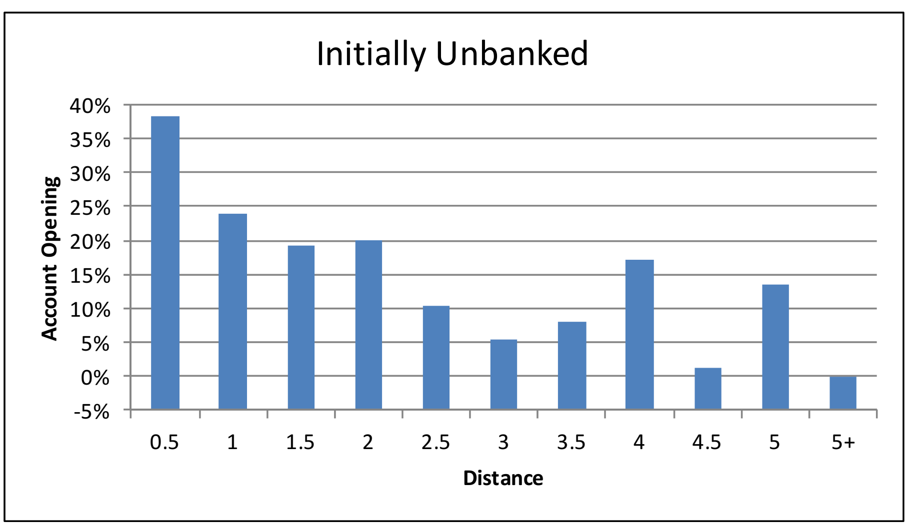


Table 1.1: Initial Banking Status of Sample Individuals

This table reports the distribution of initial banking status, as reported by sample individuals on the first day of the workshop. Individuals with no checking or savings accounts are considered “unbanked” throughout the paper.

Initial banking status	% of sample
No checking and no savings accounts ($Unbanked = 1$)	32.0%
Has just checking account ($Unbanked = 0$)	18.0%
Has just savings account ($Unbanked = 0$)	8.6%
Has both checking and savings accounts ($Unbanked = 0$)	41.4%

Table 1.2: Summary Statistics and the Effects of *Proximity* on Incoming Characteristics

The first two columns of each panel report unconditional means for various characteristics of incoming workshop participants. The next six columns of each panel contain coefficients from three regressions of the form:

$$Incoming = X'\beta + \rho_1 Proximity * Banked + \rho_2 Proximity * Unbanked + \phi Unbanked + \varepsilon$$

where *Incoming* are the incoming characteristics listed in the first column and X' contains control variables, as detailed in the bottom three rows. Estimates of ρ_1 and ρ_2 from each regression are reported in the columns labeled "Banked" and "Unbanked", respectively. Panel A reports results for demographic and financial behavior variables, and Panel B reports results for variables constructed from credit report data. Standard errors are clustered at the workshop site level. Statistical significance is denoted at the 10%, 5%, and 1% levels using *, **, and *** symbols, respectively.

Panel A: Demographics and incoming financial behavior

	Unconditional means			Coefficients on *Proximity* interactions					
			Regression:	(1)		(2)		(3)	
Initial banking status	Banked	Unbanked		Banked	Unbanked	Banked	Unbanked	Banked	Unbanked
# Observations	1472	693							
Male	28%	46%	Male	-2.4%*	-8.0%**	-4.3%**	-7.8%**	-2.3%	-4.1%*
Black	45%	53%	Black	-10.4%***	-9.3%**	-8.4%**	-9.2%*	-2.7%	-2.6%
Hispanic	43%	38%	Hispanic	11.7%***	11.4%***	10.6%**	12.4%**	2.9%**	5.0%
Born in US	56%	72%	Born in US	-4.0%	-7.7%**	-7.8%	-11.7%***	1.6%	-1.1%
Education (years)	13.6	11.7	Education (years)	-0.02	0.03	0.06	0.11	0.16**	0.14
Married/Partnership	30%	18%	Married/Partnership	2.7%*	5.0%**	3.0%	5.2%*	0.1%	2.8%
Age	41.3	38.4	Age	0.05	1.69	-0.63	1.02	0.37	0.61
Unemployed	29%	58%	Unemployed	-1.8%	-3.6%	-1.9%	-2.3%	0.8%	1.3%
On welfare	33%	63%	On welfare	1.7%	7.6%**	-0.7%	4.8%	-0.4%	3.1%
Earned Income (ann. $)	$29,410	$14,890	Earned Income (ann. $)	-1016	-3369***	-1424	-3395**	835*	-2494*
Saving	51%	8%	Saving	1.2%	-1.2%	-1.9%	-4.1%*	1.2%	-2.5%
Uses check cashing	11%	40%	Uses check cashing	-0.3%	4.2%	-1.2%	3.2%	-0.6%	3.3%
Ever seen cred. report	67%	34%	Ever seen cred. report	2.0%	3.9%	3.2%*	3.8%	0.9%	-0.8%
			Workshop Quarter FEs	-	-	Yes	Yes	Yes	Yes
			Proximity_No_Coll	-	-	Yes	Yes	-	-
			Workshop Site FEs	-	-	-	-	Yes	Yes

Panel B: Credit report variables

	Unconditional means			Coefficients on *Proximity* interactions					
			Regression:	(1)		(2)		(3)	
Initial banking status	Banked	Unbanked		Banked	Unbanked	Banked	Unbanked	Banked	Unbanked
# Observations	1290	608							
Has FICO score	78%	45%	Has FICO score	-1.4%	0.9%	2.1%	3.5%	0.1%	2.3%
Mean FICO score	624	555	Mean FICO score	4.6	3.3	0.0	-0.9	3.1	2.2
Has active balance	67%	39%	Has active balance	-2.9%*	-1.4%	0.7%	1.7%	-0.4%	0.0%
Mean active balance (>0) ($)	$11,642	$9,107	Mean active balance (>0) ($)	432	-596	123	-945	550	-1188
Has open credit card	57%	17%	Has open credit card	-2.1%	0.7%	-0.2%	1.1%	-0.7%	0.8%
Mean CC balance (>0) ($)	$4,250	$2,656	Mean CC balance (>0) ($)	93	-295	22	-401	53	-439
Has collections	52%	73%	Has collections	-1.6%	-0.2%	-0.7%	-0.7%	-1.3%	-2.4%
Has judgments	29%	34%	Has judgments	0.6%	0.3%	0.7%	0.5%	1.4%	2.2%
Mean coll+judg balance (>0) ($)	$2,880	$3,116	Mean coll+judg balance (>0) ($)	-77	200	-138	98	97	257
			Workshop Quarter FEs	-	-	Yes	Yes	Yes	Yes
			Proximity_No_Coll	-	-	Yes	Yes	-	-
			Workshop Site FEs	-	-	-	-	Yes	Yes

Table 1.3: Account Openings

Panel A reports summary statistics of administrative and survey account opening data. *Opened_at_NT_BC* is an indicator for whether an account was opened at NTFCU or BCFCU, the two credit unions for which administrative data is available; *Opened_at_Coll_CU* is an indicator for whether an account was opened at any of the collaborating credit unions; and *Opened_Anywhere* is an indicator for whether an account was opened at any financial institution following the workshop.

Panel B reports the results of regressions of the form:

$$Opened = X'\pi + \eta_1 Proximity * Banked + \eta_2 Proximity * Unbanked + \phi Unbanked + \varepsilon$$

where *Opened* are the various account opening indicators, as indicated in the second row. Standard errors, reported in parentheses, are clustered at the workshop site level. Statistical significance is denoted at the 10%, 5%, and 1% levels using *, **, and *** symbols, respectively.

Panel A: Summary statistics

Administrative data			Survey data (N=283)		
	Initially banked	Initially Unbanked		Initially banked	Initially Unbanked
Opened at NTFCU (N=2165)	12.2%	9.1%	Opened at NTFCU	14.4%	19.7%
Opened at BCFCU (N=1493)	12.8%	11.7%	Opened at BCFCU	11.7%	16.4%
Opened_at_NT_BC	21.3%	16.5%	*Opened_at_NT_BC*	26.1%	36.1%
			Opened_at_Coll_CU	29.3%	39.3%
			Opened_Anywhere	29.3%	65.6%

Panel B: Regressions

	Administrative data				Survey data		
Dependent Variable:	*Opened_at_NT_BC*				*Opened_at_NT_BC*	*Opened_at_Coll_CU*	*Opened_Anywhere*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$Proximity * Banked$	0.058***	0.052***	0.038***	0.034**	0.078**	0.046	0.047
	(0.01)	(0.01)	(0.01)	(0.02)	(0.03)	(0.05)	(0.05)
$Proximity * Unbanked$	0.084***	0.079***	0.055**	0.050**	0.266***	0.188**	0.220**
	(0.03)	(0.03)	(0.02)	(0.02)	(0.09)	(0.08)	(0.10)
$Unbanked$	0.034	0.036	0.049	0.050	0.279**	0.194	0.506***
	(0.04)	(0.04)	(0.03)	(0.03)	(0.13)	(0.12)	(0.11)
Obs.	1,937	1,937	1,937	1,937	271	271	271
R^2	0.083	0.085	0.179	0.180	0.205	0.152	0.229
F stat for eq. of coeffs	0.89	0.97	0.45	0.46	4.28**	2.38	2.31
Dummy for NT/BC closest		Yes		Yes	Yes		
Demographics	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Workshop Quarter FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Workshop Site FEs			Yes	Yes			

Table 1.4: Summary Statistics of Outcome Variables

This table presents summary statistics for the outcome variables studied in this paper, computed separately for initially banked and initially unbanked workshop participants. Panel A presents statistics for outcome variables based on credit report data, defined in Section 1.6.A. Panel B presents statistics for outcome variables based on survey data, defined in Section 1.6.B.

Panel A: Credit report outcomes

	Initially Banked	Initially Unbanked
# Observations	748	333
I. Credit Scores		
Had FICO score	79%	47%
Improved_Score	59%	56%
Became *IC*	6%	23%
Score decreased	35%	21%
Had "Insufficient Credit"	21%	53%
Established_Score	33%	16%
Established_Prime_Score	18%	7%
Has_Score (post-workshop)	81%	45%
Improved_Established	50%	30%
FICO_Change (mean)	13	31
II. Mainstream Credit Usage		
Had active debt balance	75%	46%
Has active balance (post-workshop)	94%	85%
Had no active debt balance	25%	54%
Has active balance (post-workshop)	32%	18%
Has_Bal (post-workshop)	79%	49%
Change_Bal (mean) ($)	$510	$512
#_Active_Accts (post-workshop)	7.8	2.9
Had no open credit cards	43%	82%
Has open card (post-workshop)	28%	13%
Had an open credit card	57%	18%
Has open card (post-workshop)	92%	69%
Has_CC (post-workshop)	64%	23%
Rev_Cred_Limit (post-workshop mean) ($)	$2,541	$589
Rev_Cred_Limit (post-workshop mean) (>0) ($)	$4,684	$3,457
III. Delinquencies		
Has past due (post-workshop)	33%	33%
Amount_P_Due (post-workshop mean) ($)	$542	$699
Amount_P_Due (post-workshop mean) (>0) ($)	$1,764	$2,065
Had Collections/judgments	61%	80%
Change_Colls_Jdgs (mean) ($)	-$108	-$293
*%_Change_Colls_Jdg*s (mean)	-15%	-22%

Panel B: Survey outcomes

	Initially Banked	Initially Unbanked
# Observations	222	61
I. Saving		
Initially Saving	56%	13%
Saving at time of survey	71%	88%
Initially not saving	44%	87%
Saving at time of survey	64%	65%
Saving (at time of survey)	68%	69%
Saving_More (relative to 3 years ago)	46%	54%
Saving about the same as 3 years ago	27%	21%
Saving less than 3 years ago	27%	25%
II. Overspending		
Overspending	43%	44%
Overspending more than 3 years ago	19%	34%
Overspending about the same as 3 years ago	19%	23%
Overspending_Less	62%	41%
III. Financial fragility		
Certain that could come up with $1500 in 30 days	47%	30%
Could probably come up with $1500 in 30 days	29%	34%
Could probably not come up with $1500 in 30 days	10%	7%
Certain that could not come up with $1500 in 30 days	14%	30%
IV. Self-assessed knowledge (*Knowledge*)		
Beginning of workshop	3.20	3.25
Survey	4.10	3.94
V. Finances-related stress (*Stress*)		
Beginning of workshop	3.60	3.68
Survey	3.10	3.19
VI. Economic hardship		
Unable to pay bills in past 6 months? (*Unable_Bills*)		
Yes	29%	34%
No	71%	66%

Table 1.5: Reduced Form Results: Credit Report Outcomes

This table reports the results of reduced form regressions of the form:

$$Outcome \quad = \quad X'\pi + \delta_1 Proximity * Banked + \delta_2 Proximity * Unbanked + \phi Banked + \varepsilon$$

where *Outcome* are various outcome variables constructed using credit report data, specified in the top of each panel, and X' contains control variables, as detailed in the bottom of each panel. All outcome variables are defined in Section 1.6.A. Standard errors, reported in parentheses, are clustered at the workshop site level. Statistical significance is denoted at the 10%, 5%, and 1% levels using *, **, and *** symbols, respectively.

Panel A: Credit score improvement/Establishment of a 620+ score

Dependent Variable:	*Improved_Established*			
	(1)	(2)	(3)	(4)
$Proximity * Banked$	0.009	0.008	0.013	0.026
	(0.02)	(0.02)	(0.04)	(0.02)
$Proximity * Unbanked$	0.062***	0.070***	0.072*	0.074***
	(0.02)	(0.02)	(0.04)	(0.02)
$Unbanked$	-0.045	0.022	-0.036	0.010
	(0.03)	(0.04)	(0.03)	(0.04)
Obs.	1,081	1,036	1,081	1,036
R^2	0.212	0.241	0.223	0.277
F stat for eq. of coeffs	4.88**	7.27**	5.74**	3.56*
Demographics		Yes		Yes
Workshop Quarter FEs		Yes	Yes	Yes
Proximity_No_Coll			Yes	
Workshop Site FEs				Yes

Panel B: Credit score improvement

Dependent Variable:	*Improved_Score*				*FICO_Change*			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$Proximity * Banked$	0.005	0.003	0.021	0.019	-0.03	-1.02	5.10	1.37
	(0.03)	(0.03)	(0.04)	(0.03)	(3.03)	(3.23)	(4.92)	(4.10)
$Proximity * Unbanked$	0.105***	0.114***	0.136***	0.130***	8.02	8.61	13.21**	8.63
	(0.04)	(0.03)	(0.05)	(0.04)	(6.17)	(5.69)	(5.65)	(5.30)
$Unbanked$	0.005	0.081	0.032	0.062	10.35	21.63**	11.33	12.86
	(0.06)	(0.06)	(0.05)	(0.08)	(9.79)	(10.32)	(9.29)	(10.88)
Obs.	744	711	744	711	675	645	675	645
R^2	0.026	0.073	0.049	0.122	0.129	0.183	0.155	0.245
F stat for eq. of coeffs	4.51**	6.45**	6.98**	3.72*	1.27	1.99	1.50	1.31
Demographics		Yes		Yes		Yes		Yes
Workshop Quarter FEs		Yes	Yes	Yes		Yes	Yes	Yes
Proximity_No_Coll			Yes				Yes	
Workshop Site FEs				Yes				Yes

Panel C: Measures of mainstream credit usage

	Extensive Margin			Intensive Margin		
Dependent Variable:	*Has_Score*	*Has_Bal*	*Has_CC*	*Change_Bal*	*#_Active_Accts*	*Rev_Cred_Limit*
	(1)	(2)	(3)	(4)	(5)	(6)
$Proximity * Banked$	-0.004	0.002	0.014	130.27	-0.071	-314.74
	(0.01)	(0.02)	(0.02)	(381.93)	(0.39)	(587.38)
$Proximity * Unbanked$	0.037*	-0.020	0.020	-382.60	-0.786	912.64
	(0.02)	(0.02)	(0.03)	(831.71)	(0.72)	(878.77)
$Unbanked$	-0.072**	-0.088**	-0.113**	-1,325.17*	-4.081***	-282.74
	(0.03)	(0.03)	(0.04)	(664.08)	(0.99)	(922.17)
Obs.	1,036	1,036	1,036	617	568	567
R^2	0.514	0.531	0.518	0.125	0.267	0.257
F stat for eq. of coeffs	3.61*	0.81	0.03	0.45	1.00	2.26
Only those with initial balance				Yes	Yes	Yes
Demographics	Yes	Yes	Yes	Yes	Yes	Yes
Workshop Quarter FEs	Yes	Yes	Yes	Yes	Yes	Yes
Workshop Site FEs	Yes	Yes	Yes	Yes	Yes	Yes

Panel D: Measures of delinquent credit activity

Dependent Variable:	*Amount_P_Due* ($)				*%_Change_Colls_Jdgs*			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$Proximity * Banked$	-48.35	-43.11	-38.47	-83.70	1.499	6.322	13.744**	5.527
	(58.87)	(62.82)	(105.52)	(76.74)	(4.38)	(4.68)	(6.56)	(4.39)
$Proximity * Unbanked$	-212.43*	-204.99**	-203.55	-223.90*	3.351	8.166	14.809**	5.950
	(110.06)	(94.42)	(127.22)	(114.72)	(6.15)	(6.76)	(7.10)	(7.25)
Unbanked	-49.52	-94.05	-31.50	-157.65	-5.429	-9.621	-9.667	-11.515
	(172.11)	(168.63)	(175.65)	(184.71)	(8.35)	(10.28)	(8.05)	(9.57)
Obs.	866	857	866	857	653	629	653	629
R^2	0.117	0.144	0.124	0.186	0.004	0.039	0.033	0.093
F stat for eq. of coeffs	1.62	1.84	1.61	0.96	0.07	0.06	0.03	0.00
Demographics		Yes		Yes		Yes		Yes
Workshop Quarter FEs		Yes	Yes	Yes		Yes	Yes	Yes
Proximity_No_Coll			Yes				Yes	
Workshop Site FEs				Yes				Yes

Table 1.6: Reduced Form Results: Survey Outcomes

This table reports the results of reduced form regressions of the form:

$$Outcome \quad = \quad X'\pi + \delta_1 Proximity * Banked + \delta_2 Proximity * Unbanked + \phi Banked + \varepsilon$$

where *Outcome* are various outcome variables constructed using survey data, specified in the top of each panel, and X' contains control variables, as detailed in the bottom of each panel. All outcome variables are defined in Section 1.6.B. Standard errors, reported in parentheses, are clustered at the workshop site level. Statistical significance is denoted at the 10%, 5%, and 1% levels using *, **, and *** symbols, respectively.

Panel A

Dependent Variable:	$Saving$ (1)	$Saving\_More$ (2)	$Overspending$ (3)	$Overspending\_Less$ (4)
$Proximity * Banked$	0.034	0.007	-0.027	0.019
	(0.03)	(0.04)	(0.05)	(0.05)
$Proximity * Unbanked$	0.043	0.002	0.080	-0.047
	(0.09)	(0.09)	(0.09)	(0.09)
$Unbanked$	0.050	-0.008	0.120	-0.190**
	(0.09)	(0.12)	(0.14)	(0.09)
Obs.	263	271	269	270
R^2	0.104	0.218	0.064	0.081
F stat for eq. of coeffs	0.01	0.00	1.15	0.56
Demographics	Yes	Yes	Yes	Yes
Workshop Year FEs	Yes	Yes	Yes	Yes

Panel B

Dependent Variable:	$1500\_Confidence$ (5)	$Knowledge$ (6)	$Stress$ (7)	$Unable\_Bills$ (8)
$Proximity * Banked$	0.020	0.117	0.005	-0.055
	(0.04)	(0.08)	(0.09)	(0.04)
$Proximity * Unbanked$	-0.057	0.259*	0.125	0.013
	(0.07)	(0.15)	(0.18)	(0.08)
$Unbanked$	-0.158	-0.057	-0.067	0.036
	(0.10)	(0.23)	(0.25)	(0.12)
Obs.	270	270	268	269
R^2	0.098	0.087	0.064	0.069
F stat for eq. of coeffs	0.82	0.8	0.33	0.58
Demographics	Yes	Yes	Yes	Yes
Workshop Year FEs	Yes	Yes	Yes	Yes

Table 1.7: SSIV Estimates of the Effects of Opening an Account

This table reports the results of applying the SSIV method (Angrist and Krueger (1995)) to estimate the effects of opening an account at a collaborating credit union. Panel A reports the results of first stage regressions of the form:

$$Opened * Banked = X'\pi_1 + \eta_{11} Proximity * Banked + \eta_{12} Proximity * Unbanked + \phi_1 Unbanked + \varepsilon_1$$

$$Opened * Unbanked = X'\pi_2 + \eta_{21} Proximity * Banked + \eta_{22} Proximity * Unbanked + \phi_2 Unbanked + \varepsilon_2$$

where *Opened* are the account opening indicators specified in the top of each column (defined in Section 1.5.B), constructed using the data source specified in the top row.
Panel B reports the results of regressions of the form:

$$Outcome = X'\pi_3 + \rho_1 \widehat{Opened\_Banked} + \rho_2 \widehat{Opened\_Unbanked} + \phi Unbanked + \varepsilon_3$$

where *Outcome* are the outcome variables specified in the top of the panel and $\widehat{Opened\_Banked}$ and $\widehat{Opened\_Unbanked}$ are predicted account opening probabilities formed using the first stage regressions reported in Panel A. Regressions labeled “Administrative” use predicted probabilities from regressions (1) and (2) in Panel A, and regressions labeled “Survey” use predicted probabilities from regressions (3) and (4) in Panel A.
X' contains various control variables, as detailed in the bottom of each panel. Standard errors in both panels, reported in parentheses, are clustered at the workshop site level. Standard errors in Panel B are bootstrapped. Statistical significance is denoted at the 10%, 5%, and 1% levels using *, **, and *** symbols, respectively.

Panel A: First stage regressions

	Administrative data		Survey data	
Dependent Variable:	*Opened_NT_BC* $*Banked$	*Opened_NT_BC* $*Unbanked$	*Opened_Coll_CU* $*Banked$	*Opened_Coll_CU* $*Unbanked$
	(1)	(2)	(3)	(4)
$Proximity * Banked$	0.056***	0.002	0.073	0.005
	(0.02)	(0.003)	(0.05)	(0.01)
$Proximity * Unbanked$	-0.008	0.091***	0.010	0.241***
	(0.01)	(0.03)	(0.06)	(0.07)
$Unbanked$	-0.263***	0.295***	-0.340***	0.564***
	(0.04)	(0.04)	(0.07)	(0.10)
Obs.	1,719	1,719	252	252
R^2	0.138	0.200	0.217	0.489
Demographics	Yes	Yes	Yes	Yes
Workshop Quarter FEs	Yes	Yes	Yes	Yes
Dummy for NT/BC closest	Yes	Yes		

Panel B: Second stage

Dependent Variable:	*Improved_Score*		*Has_Score*		*Amount_P_Due* ($)		*Knowledge*	
Data Source:	Administrative	Survey	Administrative	Survey	Administrative	Survey	Administrative	Survey
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\widehat{Opened\_Banked}$	0.443	0.012	-0.054	-0.192	467.47	-536.71	1.284	1.692*
	(0.41)	(0.34)	(0.23)	(0.19)	(913.14)	(708.80)	(1.26)	(1.00)
$\widehat{Opened\_Unbanked}$	1.201*	0.474**	0.314	0.146	-1,508.71	-828.24	3.863*	0.990
	(0.64)	(0.23)	(0.35)	(0.10)	(1,839.21)	(573.65)	(2.28)	(0.98)
Unbanked	-0.176	-0.183	-0.195**	-0.230***	444.75	190.91	-0.594	0.027
	(0.16)	(0.15)	(0.09)	(0.08)	(454.33)	(348.50)	(0.55)	(0.55)
Obs.	711	711	1,036	1,036	857	857	252	252
R^2	0.075	0.073	0.491	0.491	0.143	0.144	0.152	0.143
Demographics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Workshop Quarter FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
NT/BC closest dummy	Yes		Yes		Yes		Yes	
Unconditional mean unbanked	0.564		0.447		$699		3.94	

Table 1.8: Heterogeneous Effects by Education Level

This table reports the results of reduced form regressions of the form:

$$\begin{aligned} Outcome &= X'\pi + \delta_{11}(Proximity * Banked) * HS + \delta_{12}(Proximity * Banked) * Not\_HS + \\ &+\delta_{21}(Proximity * Unbanked) * HS + \delta_{22}(Proximity * Unbanked) * Not\_HS + \\ &+\phi_1 Banked + \phi_2 HS + \varepsilon \end{aligned}$$

where *Outcome* are various outcome variables (with the exception of regression (1) where *Outcome* is equal to *Opened_at_NT_BC*), as specified in the top of each column. X' contains various control variables, as detailed in the bottom of each panel. All outcome variables are defined in Section 1.6.A and *Opened_at_NT_BC* is defined in Section 1.5.B. Standard errors, reported in parentheses, are clustered at the workshop site level. Statistical significance is denoted at the 10%, 5%, and 1% levels using *, **, and *** symbols, respectively.

Panel A: First stage and reduced form

	First stage	Reduced form		
Dependent Variable:	*Opened_at_NT_BC*	*Improved_Score*	*Change_FICO*	*Amount_P_Due* ($)
	(1)	(2)	(3)	(4)
$Proximity * Banked*$				
$*HS$	0.055***	0.012	-1.06	-47.82
	(0.01)	(0.03)	(3.26)	(69.40)
$*Not\_HS$	-0.006	-0.074	-0.59	-44.85
	(0.03)	(0.06)	(6.97)	(85.47)
$Proximity * Unbanked*$				
$*HS$	0.084***	0.112***	10.68*	-423.84***
	(0.03)	(0.04)	(6.29)	(138.14)
$*Not\_HS$	0.048*	0.119*	3.80	68.10
	(0.03)	(0.06)	(8.32)	(102.11)
$Unbanked$	0.043	0.091	22.13**	-150.27
	(0.04)	(0.06)	(10.77)	(151.29)
HS	0.144***	-0.001	-5.59	53.74
	(0.04)	(0.06)	(8.15)	(130.68)
Obs.	1,937	711	645	857
R^2	0.084	0.075	0.184	0.152
F stat for equality of				
$Proximity * Unbanked$ interactions	1.87	0.01	0.56	7.32***
Dummy for NT/BC closest	Yes			
Demographics	Yes	Yes	Yes	Yes
Workshop Quarter FEs	Yes	Yes	Yes	Yes

Panel B: Measures of mainstream credit usage: reduced form

	Extensive Margin			Intensive Margin		
Dependent Variable:	*Has_Score*	*Has_Bal*	*Has_CC*	*Change_Bal*	*#_Active_Accts*	*Rev_Cred_Limit*
	(5)	(6)	(7)	(8)	(9)	(10)
$Proximity * Banked*$						
$*HS$	-0.003	0.003	0.003	-167.02	0.268	-126.56
	(0.01)	(0.01)	(0.01)	(256.46)	(0.37)	(426.24)
$*Not\_HS$	-0.069**	-0.052**	0.008	342.07	1.384*	1492.61
	(0.03)	(0.02)	(0.04)	(481.15)	(0.81)	(958.06)
$Proximity * Unbanked*$						
$*HS$	0.058**	-0.036	0.065*	-474.63	0.115	859.36
	(0.02)	(0.03)	(0.03)	(916.81)	(0.78)	(616.41)
$*Not\_HS$	-0.006	-0.027	-0.070**	-791.71	-1.107	-195.49
	(0.03)	(0.03)	(0.03)	(844.65)	(1.21)	(949.33)
$Unbanked$	-0.072**	-0.106***	-0.111**	-784.03	-4.262***	-763.81
	(0.03)	(0.03)	(0.04)	(519.51)	(0.65)	(887.50)
HS	0.116***	0.059	0.142***	-233.54	2.732***	803.16
	(0.03)	(0.04)	(0.04)	(474.33)	(0.67)	(992.15)
Obs.	1,036	1,036	1,036	586	557	553
R^2	0.494	0.513	0.509	0.044	0.206	0.244
F test for equality of $Proximity * Unbanked$ interactions	4.30**	0.09	28.30***	0.09	0.96	1.39
Only those with initial bal.				Yes	Yes	Yes
Demographics	Yes	Yes	Yes	Yes	Yes	Yes
Workshop Quarter FEs	Yes	Yes	Yes	Yes	Yes	Yes

1.10. References

Agarwal, Sumit, Paige M. Skiba, and Jeremy Tobacman, 2009, Payday loans and credit cards: New liquidity and credit scoring puzzles?, *American Economic Review: Papers & Proceedings* 99, 412–417.

Aghion, Philippe, and Patrick Bolton, 1997, A theory of trickle-down growth and development, *The Review of Economic Studies* 64, 151–172.

Angrist, Joshua D., and Alan B. Krueger, 1995, Split-sample instrumental variables estimates of the return to schooling, *Journal of Business & Economic Statistics* 13, 225–235.

Angrist, Joshua D., and Jörn-Steffen Pischke, 2008, *Mostly harmless econometrics: An empiricist's companion* (Princeton University Press).

Aportela, Fernando, 1999, Effects of financial access on savings by low-income people, Unpublished.

Ashraf, Nava, Dean Karlan, and Wesley Yin, 2006, Tying odysseus to the mast: Evidence from a commitment savings product in the philippines, *The Quarterly Journal of Economics* 121, 635–672.

Banerjee, Abhijit, and Sendhil Mullainathan, 2010, The shape of temptation: Implications for the economic lives of the poor, NBER Working Paper #15973.

Banerjee, Abhijit V., and Andrew F. Newman, 1993, Occupational choice and the process of development, *Journal of Political Economy* 101, 274–298.

Barr, Michael S., 2004, Banking the poor, The John M. Olin Center for Law and Economics Working Paper Series.

Bernheim, B. Douglas, 2002, Taxation and saving, in Alan J. Auerbach, and Martin S. Feldstein, ed.: *Handbook of Public Economics,*vol. 3 . pp. 1173–1249 (Elsevier, Amsterdam).

——— , Debraj Ray, and Sevin Yeltekin, 2013, Poverty and self-control, NBER Working Paper #18742.

Bertrand, Marianne, and Adair Morse, 2011, Information disclosure, cognitive biases, and payday borrowing, *The Journal of Finance* 66, 1865–1893.

Bolton, Patrick, and Howard Rosenthal, 2005, *Credit Markets for the Poor* . chap. 1 (Russell Sage Foundation, New York).

Bricker, Jesse, Arthur B. Kennickell, Kevin B. Moore, and John Sabelhaus, 2012, Changes in US family finances from 2007 to 2010: Evidence from the survey of consumer finances, *Federal Reserve Bulletin* 98, 1–80.

Brobeck, Stephen, 2008, Understanding the emergency savings needs of low- and moderate- income households: A survey-based analysis of impacts, causes, and remedies, Consumer Federation of America.

Burgess, Robin, and Rohini Pande, 2005, Do rural banks matter? evidence from the indian social banking experiment, *The American Economic Review* 95, 780–795.

Carter, Susan P., Paige M. Skiba, and Jeremy Tobacman, 2011, Pecuniary mistakes? payday borrowing by credit union members, in Olivia S. Mitchell, and Annamaria Lusardi, ed.: *Financial literacy: implications for retirement security and the financial marketplace* (Oxford University Press, Oxford).

Chin, Aimee, Leonie Karkoviata, and Nathaniel Wilcox, 2011, Impact of bank accounts on migrant savings and remittances: evidence from a field experiment, Unpublished.

Cole, Shawn, Thomas Sampson, and Bilal Zia, 2011, Prices or knowledge? what drives demand for financial services in emerging markets?, *The Journal of Finance* 66, 1933–1967.

Collins, J. Michael, and Collin M. O'Rourke, 2010, Financial education and counseling - still holding promise, *Journal of Consumer Affairs* 44, 483–498.

Doyle, Joseph J., Jose A. Lopez, and Marc R. Saidenberg, 1998, How effective is lifeline banking in assisting the 'unbanked'?, *Current Issues in Economics and Finance* 4, 1–6.

Dupas, Pascaline, and Jonathan Robinson, 2013a, Savings constraints and microenterprise development: Evidence from a field experiment in Kenya, *American Economic Journal: Applied Economics* 5, 163–192.

——— , 2013b, Why don't the poor save more? evidence from health savings experiments, *The American Economic Review* 103, 1138–1171.

Federal Deposit Insurance Corporation, 2012, 2011 FDIC National Survey of Unbanked and Underbanked Households, Available at http://www.fdic.gov/householdsurvey/2012_unbankedreport.pdf.

Fitzpatrick, Katie, 2009, The effect of bank account ownership on credit and consumption: Evidence from the United Kingdom, Unpublished.

Gale, William G., Benjamin H. Harris, and Ruth Levine, 2012, Raising household saving: Does financial education work?, *Social Security Bulletin* 72, 39–48.

Galor, Oded, and Joseph Zeira, 1993, Income distribution and macroeconomics, *The Review of Economic Studies* 60, 35–52.

Gerardi, Kristopher, Lorenz Goette, and Stephan Meier, 2010, Financial literacy and subprime mortgage delinquency: Evidence from a survey matched to administrative data, Federal Reserve Bank of Atlanta Working Paper Series.

Gross, David B., and Nicholas S. Souleles, 2002, Do liquidity constraints and interest rates matter for consumer behavior? evidence from credit card data, *Quarterly Journal of Economics* 117, 149–185.

Hastings, Justine S., Brigitte C. Madrian, and William L. Skimmyhorn, 2012, Financial literacy, financial education and economic outcomes, NBER Working Paper #18412.

Heckman, James J., 1976, The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models, 5, 475–492.

Karlan, Dean, Margaret McConnell, Sendhil Mullainathan, and Jonathan Zinman, 2012, Getting to the top of mind: how reminders increase saving, Unpublished.

Karlan, Dean S., and Jonathan Zinman, 2008, Credit elasticities in less-developed economies: Implications for microfinance, *The American Economic Review* 98, 1040–1068.

Laibson, David, 1997, Golden eggs and hyperbolic discounting, *The Quarterly Journal of Economics* 112, 443–478.

Lusardi, Annamaria, and Carlo de Bassa Scheresberg, 2013, Financial literacy and high-cost borrowing in the United States, NBER Working Paper #18969.

Lusardi, Annamaria, and Olivia S. Mitchell, 2013, The economic importance of financial literacy: theory and evidence, Pension Research Council Working Paper.

Lusardi, Annamaria, Daniel Schneider, and Peter Tufano, 2011, Financially fragile households: Evidence and implications, *Brookings Papers on Economic Activity* Spring 2011, 83–134.

Lusardi, Annamaria, and Peter Tufano, 2009, Debt literacy, financial experiences, and overindebtedness, NBER Working Paper #14808.

Lyons, Angela C., and Erik Scherpf, 2004, Moving from unbanked to banked: Evidence from the Money Smart program, *Financial Services Review* 13, 215–231.

Madrian, Brigitte C., and Dennis F. Shea, 2001, The power of suggestion: Inertia in 401(k) participation and savings behavior, *The Quarterly Journal of Economics* 116, 1149–1187.

Melzer, Brian T., 2011, The real costs of credit access: Evidence from the payday lending market, *The Quarterly Journal of Economics* 126, 517–555.

Mills, Gregory, William G. Gale, Rhiannon Patterson, Gary V. Engelhardt, Michael D. Eriksen, and Emil Apostolov, 2008, Effects of individual development accounts on asset purchases and saving behavior: Evidence from a controlled experiment, *Journal of Public Economics* 92, 1509 – 1530.

Morse, Adair, 2011, Payday lenders: Heroes or villains?, *Journal of Financial Economics* 102, 28–44.

Mullainathan, Sendhil, and Eldar Shafir, 2009, Savings policy and decision making in low-income households, in Rebecca M. Blank, and Michael S. Barr, ed.: *Insufficient Funds: Savings, Assets, Credit, and Banking Among Low-Income Households* (Russell Sage Foundation, New York).

Prescott, Edward S., and Daniel D. Tatar, 1999, Means of payment, the unbanked, and EFT '99, *Federal Reserve Bank of Richmond Economic Quarterly* 85, 49–70.

Romer, David, 2011, *Advanced Macroeconomics* (McGraw-Hill, New York).

Shefrin, Hersh M., and Richard H. Thaler, 2004, Mental accounting, saving, and self-control, in Colin F. Camerer, George Loewenstein, and Matthew Rabin, ed.: *Advances in Behavioral Economics* (Russell Sage Foundation, New York and Princeton University Press, Princeton).

Thaler, Richard H., 1990, Anomalies: Saving, fungibility, and mental accounts, *The Journal of Economic Perspectives* 4, 193–205.

——— , and Shlomo Benartzi, 2004, Save More Tomorrow™: Using behavioral economics to increase employee saving, *Journal of Political Economy* 112, S164–S187.

Thaler, Richard H., and Hersh M. Shefrin, 1981, An economic theory of self-control, *Journal of Political Economy* 89, 392–406.

Van Rooij, Maarten, Annamaria Lusardi, and Rob Alessie, 2011, Financial literacy and stock market participation, *Journal of Financial Economics* 101, 449–472.

Zinman, Jonathan, 2007, Household borrowing high and lending low under no-arbitrage, Unpublished.

Chapter 2

The Theoretical Effects of Favorable Interest Rate Changes on Consumer Behavior in the Presence of Default

2.1. Introduction

The effects of changes in the return to saving and in the cost of borrowing on consumer behavior have long been of significant interest to economists and policy makers. The main motivations behind studying these effects have been interest in how capital income taxation can stimulate or discourage individual saving and in how the response of households to changes in monetary or fiscal policy could impact the effectiveness of such policy changes.[1] Another piece of motivation, which has not been highlighted in the literature so far, is the implications for the effects of bank account ownership. There have been significant efforts in recent years to promote account ownership among the 8% of American households (20% of households earning less than $30,000 a year) that are unbanked (Federal Deposit Insurance Corporation (2012)).[2] Individuals who become banked face a higher expected return on saving as well as a lower cost of borrowing.[3]

In this paper I study the effects of interest rate changes on consumer behavior when defaulting on debt is possible. The analysis is relevant for the general understanding of these effects in light of the high prevalence of bankruptcy. It is particularly relevant, however, for thinking about the value of becoming banked, since the unbanked tend to be lower-income individuals who are relatively more likely to default.[4] With the banking context in mind, I concentrate on the effects of *favorable* interest rate changes - an increase in the saving rate and a decrease in the borrowing rate - on consumer behavior.

I study these effects using a simple two period life-cycle model with uncertain second period income. Individuals start the first period with some initial wealth and can borrow or save. In the second period, income realizes and debts are due. Individuals choose between paying in full and defaulting, in which case debts are discharged in full and a cost of default is incurred. I derive optimal consumption and default behavior and the responses of these to favorable changes in saving and borrowing rates, and show that incorporating the possibility of default in the standard model has potentially important implications.

---

[1] See Boskin (1978) for several other issues for which these effects are central.

[2] There have also been efforts elsewhere to promote account ownership and general access to lower cost financial services. In particular, micro-finance initiatives lower the cost of borrowing for low-income households in developing countries. The model explored here is relevant for thinking about the effects of such initiatives as well.

[3] The expected return on saving in the absence of a bank account is most likely negative, due to inflation, danger of theft and unwanted use by family members. Alternative lenders such as payday lenders, pawnshops and loan sharks offer loans at notoriously high rates.

In addition to improving saving and borrowing rates, account ownership lowers the cost of transactional services (e.g. check cashing) and is believed to help individuals overcome behavioral biases and cognitive constraints (see e.g. Mullainathan and Shafir (2009)).

[4] Median annual income for the approximately 1.3 million households who filed for bankruptcy in 2011 was $33,372. See http://www.uscourts.gov/uscourts/Statistics/BankruptcyStatistics/BAPCPA/2011/BAPCPA-report.pdf

I begin by showing that when the cost of default is not sufficiently dependent on the amount defaulted upon,[5] first period consumption of individuals who know they might default in the second period[6] is higher than it would have been absent the possibility of default. The intuition is simple: the intertemporal tradeoff between first period and second period consumption is altered when default is introduced since second period consumption is not affected by the amount consumed in the first period in states in which the individual defaults. If the cost of default is not sufficiently dependent on the amount defaulted upon, first period consumption is less costly in terms of its marginal effect on second period utility than it would have otherwise been and is thus higher.

Important for the effects studied in this paper, introducing the possibility of default also affects the way in which the intertemporal tradeoff changes with first period consumption: when the cost of default is not sufficiently dependent on the defaulted-upon amount, the marginal cost associated with first period consumption ("the opportunity cost of first period consumption") decreases in first period consumption for those who might default. This has important implications: not only is optimal consumption higher than it would have otherwise been, but the effects of changes in variables that influence the consumption choice are also different. First, marginal propensity to consume out of initial wealth is negative for individuals who might default. That is, higher initial wealth results in *lower* first period consumption for these individuals. This happens because additional wealth strengthens the link between first period consumption and second period utility by decreasing the probability of default, i.e. by increasing the number of states in which reducing current consumption increases own future consumption as opposed to decreasing the amount defaulted upon. In other words, it gives individuals "more stake in the future" and encourages lower consumption, moving them closer to the amount they would have consumed absent the possibility of default.

Second, and related to the MPC result, the effects of favorable interest rate changes on the optimal consumption of individuals who know they might default are more likely to be negative, i.e. to decrease consumption, than they would have been absent default. Favorable rate changes make both savers and borrowers better off, but where MPC is negative this works to decrease consumption rather than to increase it as standard intuition would predict. This negative wealth effect of

[5] In particular, throughout the paper I assume a fixed cost of default, i.e. a cost that is independent of the amount defaulted upon. Modeling the cost of default as independent of this amount is the standard practice in the literature (e.g. Athreya (2008), Livshits, MacGee, and Tertilt (2007),Chatterjee, Corbae, Nakajima, and Ríos-Rull (2007)).

[6] I show that there is an initial wealth threshold such that relatively wealthy individuals never find it optimal to default in the second period. Relatively poor individuals default in relatively bad second period states and repay their debts in relatively good states.

rate changes, which is combined with a substitution effect to determine the ultimate response of consumption, again reflects the strengthening of the link between first period consumption and second period utility brought about by a reduction in the probability of default. The substitution effect of favorable rate changes works to increase consumption for those who initially borrow (i.e. those who borrow prior to the rate changes) and to decrease it for those who initially save, regardless of the possibility of default.[7] For initial borrowers who might default, the income effect contradicts the substitution effect and the result is that consumption increases in response to a decrease in the borrowing rate for some parameter values and levels of initial borrowing and decreases for others. For initial savers who might default,[8] both effects go in the same direction and consumption always decreases in response to an increase in the saving rate. This stands in contrast to the standard effects of favorable interest rate changes on consumption (i.e. the effects absent the possibility of default), whereby initial borrowers always increase their consumption and the response of initial savers could be positive or negative, depending on the parameters used and the level of initial saving.

I also examine the effects of favorable interest rate changes on expected default: the probability of default and the expected defaulted-upon amount. I show that both of these quantities move together with *Debt*, the amount owed in the second period, in response to rate changes. While all initial savers who might default decrease *Debt* (and expected default) in response to an increase in the saving rate, some initial borrowers decrease *Debt* in response to a decrease in the borrowing rate and some increase it. The condition for *Debt* (and expected default) of borrowers to decrease is that the fraction of consumption that is financed by borrowing is higher than the elasticity of intertemporal substitution.[9]

In an extension to the model I allow for partial repayment of debt in both periods. I focus on the first period, which the individual starts with some pre-existing debt. Here the individual chooses how much of this debt he wishes to repay, where the remainder is considered delinquent. Simultaneously, he chooses how much he wishes to save or borrow. First period consumption can be viewed as being financed by delinquency and borrowing, and there is an optimal mix of these for different levels of consumption. Decreasing the borrowing rate makes borrowing relatively more attractive and reduces delinquency. This analysis highlights another channel through which

[7] The substitution effect is stronger for a given initial consumption level when default is possible.

[8] Default is possible when saving because there is a pre-existing obligation. See below.

[9] With time-separable utility, the EIS is the inverse of the Arrow-Pratt coefficient of relative risk aversion, so that with CRRA preferences with a coefficient of 3, for example, *Debt* decreases for all those who initially borrow more than one third of how much they consume.

favorable interest rate changes can affect optimal consumption and debt repayment when it is possible not to repay in full.

This paper contributes to a wide literature on the interest elasticity of consumption in life-cycle type models of intertemporal utility maximization.[10] The factors that impact this elasticity in the basic life-cycle model (e.g. the substitution and income effects) are well-understood and appear in many economics textbooks (e.g. Romer (2011)). Feldstein and Tsiang (1968) is an early paper which derives the elasticity formally for the basic model. Later studies have refined the basic derivation (Summers (1981) is an early example) and have explored the implications for this elasticity of some of the many extensions to the basic life-cycle model that have been proposed over the years. Bernheim (2002) surveys studies which have examined the implications of incorporating various types of bequest motives, uncertain earnings, and liquidity constraints as well as bounded rationality and self-control problems in the model. Attanasio and Wakefield (2010) note several types of extensions to the basic model for which implications for the interest elasticity of consumption have not been studied. These include endogenous labor supply, intertemporal nonseparabilities (e.g. durable goods and habit formation type models) and housing.[11]

This paper explores the unstudied implications of another type of extension to the life-cycle model, namely the incorporation of default (in the basic model, debt is assumed to be paid with certainty). Several recent papers have studied elaborate general equilibrium models where individuals can take on unsecured debt and subsequently default (see e.g. Athreya, Tam, and Young (2012), Athreya (2008), Livshits, MacGee, and Tertilt (2007), Livshits, MacGee, and Tertilt (2010)). These papers have, however, focused on explaining aggregate quantities such as the amount of debt outstanding, the rate of default and the distribution of interest rates as well as on analyzing different default arrangements and proposed reforms to bankruptcy laws. None of these papers examines how the incorporation of the possibility of default in the model impacts the effects of changes in variables

[10] Studies of this elasticity have typically framed it as measuring the effect of capital income taxation on individual saving. Bernheim (2002) and Attanasio and Wakefield (2010) provide comprehensive surveys of the theoretical and empirical literature.

[11] In addition to the theoretical literature, many papers have attempted to empirically estimate the the elasticity of intertemporal substitution, i.e. the sensitivity of expected consumption *growth* to the interest rate (e.g. Hall (1998), Vissing-Jorgensen (2002)), or to directly estimate the interest elasticity of consumption. Studies of the latter type have estimated the elasticity of borrowing with respect to the borrowing rate (e.g. Gross and Souleles (2002), Karlan and Zinman (2008)) or the elasticity of saving with respect to the saving rate (see Bernheim (2002) for a survey). Both types of studies have yielded a wide range of estimates. A common finding is that elasticity is lower for those who are closer to their debt limits, have lower incomes or have lower credit scores (see e.g. Gross and Souleles (2002), Attanasio, Goldberg, and Kyriazidou (2008)). This is normally taken as evidence of the importance of liquidity constraints but is also consistent with negative wealth effects for low-income, default-prone individuals, as shown in this paper.

Some papers have also looked at the effect of lowering the borrowing rate on consumer default. Karlan and Zinman (2008) and Ausubel (1999) find that default is reduced.

such as wealth and the interest rate on optimal consumption. This is not surprising since the effects studied here, which depend critically on the marginal cost of first period consumption (the "opportunity cost") decreasing in first period consumption for those who might default, most likely do not exist in the models examined in those papers. In those models, interest rates are determined on a loan-by-loan basis. Higher first period consumption (and borrowing) implies a higher probability of default, other things equal, and thus a higher break-even interest rate. The opportunity cost, which depends on the interest rate, might therefore be increasing rather than decreasing in consumption when interest rates vary in the amount borrowed, in which case the effects established in this paper would not hold.[12]

Here I assume that there is a single borrowing rate for loans of all sizes, corresponding to the lender's zero profit condition holding on aggregate rather than on a loan-by-loan basis. In Appendix B I show that if there is an information asymmetry whereby expected second period income varies across different "types" and is unknown to the lender, no screening contract based on the borrowed amount can be incentive compatible and there must be a single borrowing rate for loans of all sizes. Asymmetric information about borrowers' risk of default is one common justification for a lack of dispersion in borrowing rates,[13] the other being the fixed costs of creating differentiated credit contracts. Athreya, Tam, and Young (2012) and Livshits, MacGee, and Tertilt (2011) argue that improvements in available information and technological advances which have reduced these fixed costs have caused the large increase in the dispersion of interest rates reported by credit card holders in the past few decades. The increased dispersion in rates might, however, reflect an increase in the number of lenders (who also compete on dimensions other than price) rather than non-flat rate-amount curves offered by specific lenders. Moreover, while information available to lenders has certainly improved, it is still relatively poor for some populations, for example for unbanked individuals who usually have little credit history. This is another sense in which this paper is especially relevant for thinking about the effects of banking the unbanked.

This chapter will proceed as follows. After a brief description of the model in Section 2.2, I examine the individual's optimization problem in Section 2.3. I find that there is a second period income

[12]Indeed, in Athreya (2008) the incorporation of default significantly *lowers* the borrowing and consumption of young households (corresponding to the first period here) rather than increasing borrowing, as would be implied if incorporating default meant switching from an increasing to a decreasing opportunity cost. This despite of the cost of default being independent of the amount defaulted upon which, I show below, leads to a decreasing opportunity cost when interest rates do not vary in the amount borrowed. Modeling the marginal cost of default as decreasing in the amount defaulted upon might produce a decreasing opportunity cost with loan-by-loan pricing.

[13]For example, Athreya (2002) justifies his use of a single rate with the difficulty in assessing the default probability for any single lender given poor information from credit bureaus, in particular the absence of good labor income data.

threshold below which it is optimal to default and above which it is optimal to repay. I then turn to the first period problem, distinguishing consumption levels that imply zero probability of default (i.e. imply that it will never be optimal to default in the second period) from levels that imply positive probability of default and characterizing the first order condition in each region. I examine the slope of the opportunity cost of first period consumption, which plays an important role in the paper, as explained above, and show that it is negative where default is possible when the cost of default is fixed. In Section 2.4 I characterize optimal consumption. I establish the existence of initial wealth thresholds which determine whether an individual's optimal consumption leads to positive probability of second period default, whether it implies saving, borrowing or neither and whether it implies borrowing up to the borrowing constraint. In this section I also derive the marginal propensity to consume out of wealth and show that it is negative for individuals who have positive probability of default. In Section 2.5 I examine the effects of favorable interest rate changes on optimal behavior. I first examine the effects on the initial wealth thresholds established in the previous section and show that as a result of favorable rate changes more individuals consume an amount such that they have zero probability of default, less individuals borrow up to the borrowing constraint, and more individuals save and more borrow (as opposed to doing neither). I then turn to examining the effects of favorable rate changes on optimal consumption and on expected default, which are at the heart of the paper. I extend the model to allow for voluntary extent of delinquency in both periods, as described above, in Section 2.6. Section 2.7 concludes. In Appendix B I demonstrate that an information asymmetry can lead to a single interest rate for all borrowed amounts being the only incentive compatible contract. In Appendix C I provide proofs omitted from the main text.

2.2. Model

I model an individual consumer who lives for 2 periods and derives utility from consumption in both, maximizing a time-separable expected utility function with discount rate β:

$$U = U_1 + \beta E(U_2)$$

The individual starts the first period with initial wealth Y_1[14] and chooses first period consump-

[14]More precisely, Y_1 is what the individual is left with after repayment of all previous debts, termed "cash on hand" in the literature.

tion C_1. The difference, $(C_1 - Y_1)$, which I denote by B_1, represents borrowing if positive and saving if negative.

Borrowing is allowed up to a borrowing constraint $\overline{B_1}$.[15] The borrowing (saving) rate for any borrowed (saved) amount is R_B (R_S).[16] There is a positive spread $R_B - R_S$.

In the second period, the individual receives stochastic income $\tilde{Y_2}$. It is distributed uniformly over some interval: $\tilde{Y_2} \sim U[\underline{Y_2}, \overline{Y_2}]$ where $(\overline{Y_2} - \underline{Y_2})$ is normalized to 1. $\tilde{Y_2}$ is meant to capture any shocks to income and to unavoidable expenses (such as medical expenses) that the individual is exposed to in the second period.

After $\tilde{Y_2}$ realizes, debts are due. If the individual had saved in period 1, i.e. if $B_1 = C_1 - Y_1 < 0$, he receives $B_1 R_S$. If the individual had borrowed, i.e. if $B_1 > 0$, he owes $B_1 R_B$. Where it is not important whether the individual is saving or borrowing, I use R to denote the applicable interest rate.

There is also a pre-existing fixed obligation $\bar{D_0}$ due in the second period, regardless of the first period consumption decision. In particular, it is due even if the individual saves in the first period. $\bar{D_0}$ can be thought of as representing utility bills or older installment loans that cannot be re-financed.[17]

If the individual owes on net, i.e. if $Debt = B_1 R + \bar{D_0} > 0$, he can choose to default after having observed $\tilde{Y_2}$.[18] Default here is similar to chapter 7 bankruptcy: in case of default, all debts are discharged and second period consumption is equal to second period income $\tilde{Y_2}$ (in case of no-default, it is equal to $\tilde{Y_2} - Debt$).[19] Default involves a cost, generally denoted as $\Lambda(Debt)$, representing the pecuniary (i.e. legal fees, seizure of non-exempt assets, exclusion from formal credit markets) as well as non-pecuniary (i.e. stigma) costs of default. Throughout most of the paper I assume that the cost of default is fixed, i.e. independent of $Debt$, as is standard in the literature.

[15] I provide more information regarding the borrowing constraint in the next section

[16] In Appendix B I show that an information asymmetry results in a single borrowing rate for all loan sizes being the only incentive compatible contract.

[17] Y_1 is defined net of this fixed obligation. Gross and Souleles (2002) show evidence of simultaneous saving and borrowing.

[18] Note that when $\tilde{Y_2} < Debt$, the individual is unable to cover all of his debts and is forced to default.

[19] In Section 2.6 I model partial delinquency, where the individual can choose how much of his debt he wishes to repay.

2.3. The Individual's Optimization Problem

A. The Second Period Default Decision

The individual comes into the second period owing $Debt$ (or due to receive $-Debt$, if $Debt < 0$). The period starts with the realization of $\tilde{Y_2}$, followed by the default decision. As noted above, the individual is forced to default when $\tilde{Y_2} < Debt$.

When $\tilde{Y_2} \geq Debt$, paying in full and consuming $C_2 = \tilde{Y_2} - Debt$ is possible. The individual might also choose to default (this is only possible if $Debt > 0$), consuming $\tilde{Y_2}$ and incurring a cost $\Lambda(Debt)$. Default is optimal when $\tilde{Y_2} \geq Debt$ if and only if:

$$u(\tilde{Y_2}) - \Lambda(Debt) > u(\tilde{Y_2} - Debt) \tag{2.3.1}$$

Proposition 1 shows that there exists a second period income threshold $\widehat{Y_2}$ that determines second period behavior. In relatively good states, where $\tilde{Y_2} \geq \widehat{Y_2}$, it is optimal to pay back debts. In relatively bad states $\tilde{Y_2} < \widehat{Y_2}$, default is optimal.[20]

Proposition 1. *There exists a second period income threshold* $\widehat{Y_2}$ *such that:*

- If $\tilde{Y_2} < \widehat{Y_2}$, the individual defaults, consuming $\tilde{Y_2}$ and incurring the cost of default $\Lambda(Debt)$
- If $\tilde{Y_2} \geq \widehat{Y_2}$, the individual pays $Debt$ in full, consuming $\tilde{Y_2} - Debt$

$\widehat{Y_2}$ is a function of $Debt$, defined by:

$$u(\widehat{Y_2}) - \Lambda(Debt) = u(\widehat{Y_2} - Debt) \tag{2.3.2}$$

Proof. See Appendix C. □

B. The First Period Problem

In the first period the individual chooses C_1, implicitly choosing $Debt = (C_1 - Y_1)R + \bar{D_0}$. Since $\widehat{Y_2}$ is increasing in C_1,[21] there is, for any initial wealth level Y_1, some first period consumption level $\widehat{C_1}$ s.t.:

[20] I assume that $\widehat{Y_2} > Debt$ for all possible levels of $Debt$, i.e. there is always some range of $\tilde{Y_2}$ where the individual is able to pay in full ($\tilde{Y_2} > Debt$) but chooses not to do so. This amounts to assuming $u(Debt) > \Lambda(Debt)$ for all possible levels of $Debt$.

[21] This can be seen easily by applying the implicit function theorem on (2.3.2).

$$\widehat{Y_2}(\widehat{C_1}) = \underline{Y_2} \tag{2.3.3}$$

For first period consumption levels $C_1 \leq \widehat{C_1}$, we have $\widehat{Y_2} \leq \underline{Y_2}$, i.e. the individual chooses to repay in full even in the worst possible second income realization and the ex-ante probability of default in the second period is zero. These are "default impossible" C_1 levels, and I refer to individuals who choose such C_1 levels as "never defaulters".

For $C_1 > \widehat{C_1}$, we have $\widehat{Y_2} > \underline{Y_2}$ and the ex-ante probability of default in the second period is positive (it is known ex-ante that if a low income state realizes in the second period, the individual will choose to default). These are "default possible" C_1 levels, and I refer to individuals who choose them as "potential defaulters".

Using this useful distinction, expected utility can be written as:

$$\begin{gathered} U(C_1) = u(C_1)+ \\ +\beta \int_{\underline{Y_2}}^{\overline{Y_2}} u(\tilde{Y_2} - Debt)d\tilde{Y_2} * 1(C_1 \leq \widehat{C_1})+ \\ +\beta(\int_{\widehat{Y_2}}^{\overline{Y_2}} u(\tilde{Y_2} - Debt)d\tilde{Y_2} + \int_{\underline{Y_2}}^{\widehat{Y_2}}(u(\tilde{Y_2}) - \Lambda(Debt))d\tilde{Y_2}) * 1(C_1 > \widehat{C_1}) \end{gathered} \tag{2.3.4}$$

Note that expected second period utility, in the second and third rows, takes on a different functional form depending on whether there is ex-ante potential for default.

In the first period, the individual solves:

$$\begin{array}{cc} max & U(C_1) \\ C_1 & \\ s.t. & C_1 - Y_1 \leq \overline{B_1} \end{array}$$

where $\overline{B_1}$ is a borrowing constraint, limiting the amount that the individual can consume in the first period to $\overline{C_1} = Y_1 + \overline{B_1}$. Although there is a single borrowing rate here, implicitly set on aggregate rather than loan-by-loan,[22] the lender should still be reluctant to make individual loans on which it is certain to lose, and should thus limit the amount it is willing to lend for any given interest rate. Moreover, I assume below that the cost of default is fixed, and show that in that case the value function increases without bound for high C_1 values. The existence of a solution to the individual's problem therefore requires that consumption and borrowing are limited at some point.

[22]See the introduction and Appendix B for discussions of this point.

An intuitive rule for setting this limit is that the individual should not be allowed to borrow an amount such that he will always default (recall that the lender gets nothing in case of default).[23] According to this rule, the borrowing constraint $\overline{B_1}$ is set s.t.:

$$\widehat{Y_2}(\overline{B_1}) = \overline{Y_2}$$

That is, an individual who borrows $\overline{B_1}$ is indifferent between defaulting and repaying in full if the best second period income state realizes, and prefers to default if any other state realizes.

C. The FOC and the Slope of the Opportunity Cost Curve

The interior solution to the problem is determined by the following first order condition:

$$\begin{aligned} u'(C_1) = \quad & \beta R[\int_{\underline{Y_2}}^{\overline{Y_2}} u'(\tilde{Y_2} - Debt) d\tilde{Y_2}] * 1(C_1 \leq \widehat{C_1}) + \\ & + \beta R[(\int_{\widehat{Y_2}}^{\overline{Y_2}} u'(\tilde{Y_2} - Debt) d\tilde{Y_2} + \int_{\underline{Y_2}}^{\widehat{Y_2}} \Lambda'(Debt)) d\tilde{Y_2})] * 1(C_1 > \widehat{C_1}) \end{aligned} \tag{2.3.5}$$

The LHS of the first order condition is the marginal benefit of first period consumption and the RHS is the marginal cost, equal to the expected loss of second period utility associated with consuming another unit in the first period. The latter quantity, which I refer to as "the opportunity cost of first period consumption" or "opportunity cost" and denote *opp_cost*, plays a central role in this paper. In particular, the sign of its derivative with respect to C_1 will be shown to have important implications.

Let us examine this derivative in the two regions defined by $\widehat{C_1}$. At $C_1 \leq \widehat{C_1}$ values, where default is impossible (i.e. it will not be chosen even in the worst second period state), the opportunity cost is increasing in C_1, as is standard:

$$\frac{d}{dC_1} opp\_cost \mid C_1 \leq \widehat{C_1} = -\beta R^2 \int_{\underline{Y_2}}^{\overline{Y_2}} u''(\tilde{Y_2} - Debt) d\tilde{Y_2} > 0$$

The region $C_1 > \widehat{C_1}$, where default is possible (i.e. where default will be chosen in relatively

[23] Note that the particular value of the borrowing constraint does not generally influence the results derived in this paper, as long as it is set at a level of borrowing which implies positive probability of default. See footnote 7 in Appendix C for the only caveat to this.

bad states), is more interesting to consider. The derivative of the opportunity cost wrt C_1 in this region is:

$$\frac{d}{dC_1} opp\_cost \mid C_1 > \hat{C_1} = \tag{2.3.6}$$
$$= \beta R^2 [- \int_{\widehat{Y_2}}^{\overline{Y_2}} u''(\tilde{Y_2} - Debt) d\tilde{Y_2} + \int_{\underline{Y_2}}^{\widehat{Y_2}} \Lambda''(Debt) d\tilde{Y_2} + \frac{d\widehat{Y_2}}{dDebt}(-u'(\widehat{Y_2} - Debt) + \Lambda'(Debt))]$$

The main insight of this paper is that when (2.3.6) is negative ("when the opportunity cost is decreasing"), which is the case for certain formulations of $\Lambda(Debt)$, the optimal consumption behavior of potential defaulters and its response to changes in variables such as wealth and the interest rate are different than they would have been absent the possibility of default. In particular, optimal consumption is higher for any given initial wealth level, the marginal propensity to consume out of wealth is negative, and the response of consumption to favorable interest rate changes is different. I explore these implications in detail in Section 2.4.

The intuition behind the effects of the sign of (2.3.6) is simple: when it is negative, the marginal cost associated with C_1 (which is balanced against the marginal benefit of C_1 in the first order condition) decreases in C_1. This alters the intertemporal tradeoff as well as how this tradeoff changes with respect to C_1. First, for any level of C_1 the opportunity cost (the marginal cost of C_1) is lower than it would have been if the opportunity cost were increasing, so that the optimal C_1 is higher than it would have been.[24] Second, the effects of changes in variables that impact the intertemporal tradeoff, such as initial wealth and the interest rate, are different since they depend on how the opportunity cost and the marginal benefit of consumption change in C_1 (how the intertemporal tradeoff changes).

D. The Determinants of the Sign of the Slope of the Opportunity Cost Curve

Given the importance of the sign of $\frac{d}{dC_1} opp\_cost$ for the results, I examine the elements of (2.3.6) one by one. The first element is similar to the derivative of the opportunity cost in the "default impossible" region $C_1 < \widehat{C_1}$ and represents the effect of increasing C_1 on marginal utility in second period states in which debts are paid in full. Increasing C_1 decreases consumption in those states, *increasing* overall opportunity cost (since marginal utility is decreasing). Note,

[24]Hubbard, Skinner, and Zeldes (1995) show that social insurance programs with means tests based on assets discourage wealth accumulation through a similar mechanism.

however, that the lower limit of the integral is $\widehat{Y_2}$ rather than $\underline{Y_2}$: consumption in $\tilde{Y_2} < \widehat{Y_2}$ states is equal to second period income and is not affected by increasing C_1. This is a crucial point.

The second element is non-zero if and only if the marginal cost of default changes in C_1. If it decreases in C_1, as seems plausible, increasing C_1 decreases the marginal cost of default and works to *decrease* overall opportunity cost. Note that this occurs only in second period states in which the individual defaults, i.e. those in which $\tilde{Y_2} < \widehat{Y_2}$.

The third element reflects the effect of increasing C_1 on $\widehat{Y_2}$, the default threshold. Increasing C_1 increases $Debt$, which increases $\widehat{Y_2}$, i.e. the individual defaults in more second period states.[25] This has the effect of decreasing the number of states in which the opportunity cost is $u'(\tilde{Y_2} - Debt)$ and increasing the number of states in which it is $\Lambda'(Debt)$. Since $u'(\widehat{Y_2} - Debt) > \Lambda'(Debt)$ for any level of $Debt$,[26] C_1 has the effect of *decreasing* overall opportunity cost through this channel.

As is apparent from (2.3.6), the way in which we model the cost of default $\Lambda(Debt)$ determines how the opportunity cost changes wrt C_1. In particular, if the marginal cost of default is not sufficiently high (for example, if it is zero, as with a fixed cost of default), the opportunity cost decreases in C_1. To see this, rewrite (2.3.6):

$$\frac{d}{dC_1} opp\_cost \mid C_1 > \hat{C_1} = \tag{2.3.7}$$
$$= \beta R^2[(1 - \frac{d\hat{Y_2}}{dDebt})u'(\widehat{Y_2} - Debt) - u'(\overline{Y_2} - Debt) + \frac{d\widehat{Y_2}}{dDebt}\Lambda'(Debt) + (\widehat{Y_2} - \underline{Y_2})\Lambda''(Debt)]$$

The sign of (2.3.7) depends on $\Lambda'(Debt)$.[27] Proposition 2 shows that when the cost of default is fixed, s.t. $\Lambda(Debt) = \Lambda$, the opportunity cost decreases in C_1. This makes intuitive sense: when the cost of default is not dependent on the extent of default, C_1 has no effect on second period utility in states in which the individual already defaults ($\tilde{Y_2} < \widehat{Y_2}$), i.e. the opportunity cost in these states is zero, regardless of the level of C_1. Increasing C_1 therefore increases the opportunity cost through its effect on second period consumption, but only in non-default ($\tilde{Y_2} \geq \widehat{Y_2}$) states. Importantly, increasing C_1 increases $\widehat{Y_2}$, i.e. decreases the number of states in which C_1 matters, thereby decreasing the opportunity cost. The net effect is negative: increasing C_1 decreases the

[25] I show that $\frac{d\widehat{Y_2}}{dDebt} > 0$ in the proof of Proposition 2.

[26] This is always the case for $\widehat{Y_2}(Debt)$ if $u'(.)$ is decreasing and $\Lambda'(.)$ is non-increasing

[27] The sign of the first element depends on the sign of $(1 - \frac{d\widehat{Y_2}}{dDebt})$. As I show in the proof of Proposition 2, $\frac{d\widehat{Y_2}}{dDebt} \geq 1$ if and only if $u'(\widehat{Y_2}) \geq \Lambda'(Debt)$. The second element is always negative. The third element is $\frac{d\widehat{Y_2}}{dDebt}\Lambda'(Debt)$. As I show in the proof of Proposition 2, $\frac{d\widehat{Y_2}}{dDebt} > 0$, so that this element is non-negative and depends on $\Lambda'(Debt)$. The fourth element is non-positive if we assume $\Lambda''(.) \leq 0$, which seems reasonable.

opportunity cost.

It is important to note that modeling the cost of default as non-dependent on the extent of default is the common practice in models which feature consumer default (see e.g. Athreya (2008),Livshits, MacGee, and Tertilt (2007), Chatterjee, Corbae, Nakajima, and Ríos-Rull (2007)). However, as I explain in the introduction, the opportunity cost most likely does not decrease in C_1 in those models since the interest rate is determined on a loan-by-loan basis rather than on aggregate.

Proposition 2. *When the cost of default is fixed, s.t.* $\Lambda(Debt) = \Lambda$*, the opportunity cost increases in* C_1 *where default is impossible (at* $C_1 \leq \widehat{C_1}$*) and decreases in* C_1 *where default is possible (at* $C_1 > \widehat{C_1}$*).*

Proof. See Appendix C. □

Figure 2.1 plots the marginal benefit and the opportunity cost of C_1 with a fixed cost of default for two initial wealth levels. The point where the opportunity cost curve is kinked in each plot is $\widehat{C_1}$, the first period consumption threshold that separates "default impossible" and "default possible" C_1 levels (i.e. C_1 levels where default is never chosen in the second period and C_1 levels where it is sometimes chosen). As can be seen in the figure, and as I show in Proposition 2, the opportunity cost curve is increasing at $C_1 < \widehat{C_1}$, where default is impossible, and decreasing at $C_1 > \widehat{C_1}$, where it is possible. Note that the kink point $\widehat{C_1}$ represents a higher C_1 level in the top plot, where initial wealth is higher. This is because $\widehat{Y_2}$, the second period income threshold above which it is optimal to repay and below which it is optimal to default, is a function of $Debt = (C_1 - Y_1)R + \bar{D}_0$. $\widehat{C_1}$, the C_1 level for which $\widehat{Y_2}(C_1) = \underline{Y_2}$, is thus increasing in initial wealth Y_1.

In the top plot of Figure 2.1, where wealth is relatively high, the two curves cross in the upwards sloping part of the opportunity cost curve. That is, optimal consumption is lower than $\widehat{C_1}$, and default is never chosen in the second period. In the bottom plot, where wealth is relatively low, the two curves cross in the downwards sloping part and optimal consumption is higher than $\widehat{C_1}$, so that the ex-ante probability of default is positive. In Proposition 3 I formally show that there is an initial wealth threshold which separates "never defaulters", i.e. those who choose $C_1 < \widehat{C_1}(Y_1)$ levels, from "potential defaulters", i.e. those who choose $C_1 > \widehat{C_1}(Y_1)$ levels. Note that in the bottom plot, where optimal consumption is such that default is possible, optimal consumption is lower than it would have been absent the possibility of default (the opportunity cost curve would have intersected the marginal benefit curve at a lower level of C_1 had there not been a kink).

[Insert Figure 2.1 here]

2.4. Characterizing Optimal Consumption

In this section I characterize optimal consumption behavior. I operate under the assumption of a fixed cost of default,[28] but derive general results where possible. I first establish the existence of initial wealth thresholds which determine consumption behavior. I then discuss the marginal propensity to consume out of initial wealth, showing that it is negative for potential defaulters.

A. Initial Wealth Thresholds

The next several propositions establish that optimal consumption behavior is determined to a large extent by initial wealth and describe behavior in different initial wealth regions. Proposition 3 establishes the existence of a threshold $\widehat{Y_1}$ which separates "never defaulters" from potential defaulters: those with initial wealth $Y_1 \geq \widehat{Y_1}$ choose C_1 such that their ex-ante probability of default is zero (they choose $C_1 \leq \widehat{C_1}(Y_1)$ and know ex-ante that they will not find it optimal to default in any $\tilde{Y_2}$ state), while those with initial wealth $Y_1 < \widehat{Y_1}$ choose C_1 such that they have positive ex-ante probability of default (they choose $C_1 > \widehat{C_1}(Y_1)$). Proposition 4 establishes the existence of a threshold $\check{Y_1}$ which separates those who borrow up to the borrowing constraint from those who borrow less than the borrowing constraint or save. Proposition 5 establishes the existence of thresholds $\grave{Y_1}(R_B)$ and $\grave{Y_1}(R_S)$ which separate savers from borrowers. Proposition 5 also shows that those with initial wealth between these thresholds ($\grave{Y_1}(R_B) < Y_1 < \grave{Y_1}(R_S)$) neither borrow nor save. The propositions establish the ordering of thresholds shown in Figure 2.4.

Proposition 3. *There exists an initial wealth threshold $\widehat{Y_1}$ which separates "never defaulters" from "potential defaulters":*

- Individuals with initial wealth $Y_1 \geq \widehat{Y_1}$ are "never defaulters": they choose $C_1 \leq \widehat{C_1}(Y_1)$ and have zero ex-ante probability of default.

[28]Note that all of the results would hold with any other form of $\Lambda(.)$ which leads to a decreasing opportunity cost in the "default possible" region.

- Individuals with initial wealth $Y_1 < \widehat{Y_1}$ are "potential defaulters": they choose $C_1 > \widehat{C_1}(Y_1)$ and have positive ex-ante probability of default. At low second income realizations $\tilde{Y_2} < \widehat{Y_2}$, they default on their debts, consume their entire second period income and incur the cost of default. At high second period income realizations $\tilde{Y_2} \geq \widehat{Y_2}$, they pay back in full.

Proof. See Appendix C. $\square$

Proposition 4. *For any borrowing constraint* $\overline{B_1}$[29] *there exists an initial wealth threshold* $\check{Y}_1$*, where* $\check{Y}_1 \leq \widehat{Y_1}$*, such that individuals with initial wealth* $Y_1 \leq \check{Y}_1$ *borrow up to the constraint and individuals with initial wealth* $Y_1 > \check{Y}_1$ *borrow less than the constraint or save.*

Proof. See Appendix C. $\square$

Proposition 5. *For saving and borrowing rates* $R_S < R_B$*, there exist first period income thresholds* $\grave{Y}_1(R_B)$ *and* $\grave{Y}_1(R_S)$*, such that* $\check{Y}_1 < \grave{Y}_1(R_B) < \grave{Y}_1(R_S) < \widehat{Y_1}$*, for which:*

- Individuals with initial wealth $Y_1 < \grave{Y}_1(R_B)$ borrow; Borrowing decreases in Y_1 over $\check{Y}_1 < Y_1 < \grave{Y}_1(R_B)$.
- Individuals with initial wealth $\grave{Y}_1(R_S) < Y_1$ save; Saving increases in Y_1 over $\grave{Y}_1(R_S) < Y_1$.
- Individuals with initial wealth $\grave{Y}_1(R_B) < Y_1 < \grave{Y}_1(R_S)$ neither borrow nor save.

Proof. See Appendix C. $\square$

B. Marginal Propensity to Consume

Proposition 6 shows that when the cost of default is fixed, the marginal propensity to consume out of initial wealth is negative for potential defaulters. That is, rather than the standard splitting of an additional $1 of wealth between the present (through increased current consumption) and the future (through increased saving or decreased borrowing), potential defaulters increase their saving/decrease their borrowing by *more than $1* when their initial wealth increases by $1, thereby *decreasing* their current consumption.

[29] I assume that any borrowing constraint implies maximum first period consumption $\overline{C_1}$ s.t. $\overline{C_1} > \widehat{C_1}$, i.e. s.t. there is positive ex-ante probability of default.

This result, which is directly related to the opportunity cost of C_1 decreasing at "default possible" C_1 levels,[30] is central to the effects of favorable interest rate changes on optimal consumption behavior, as I discuss in the next section. A similar result is obtained by Hubbard, Skinner, and Zeldes (1995) who show that means-tested social insurance could lead to a decreasing opportunity cost and to negative MPC over some wealth ranges.[31]

To better understand the result and its relation to the sign of the slope of the opportunity cost curve, consider an individual who is initially optimizing, and moves an entire additional $1 of initial wealth to the second period, leaving C_1 unchanged. Let us analyze this consumption choice relative to the new, higher wealth, optimum, for the standard case of an opportunity cost that is increasing in C_1 and for the case of an opportunity cost that is decreasing in C_1.

A necessary condition for an optimum is $u'(C_1) = \mathit{opp\_cost}$. In both cases, $u'(C_1)$ is unchanged. Since $\frac{\partial}{\partial Y_1}\mathit{opp\_cost} = -\frac{d}{dC_1}\mathit{opp\_cost}$,[32] opportunity cost will be lower than it was in the old optimum in the case of an opportunity cost that is increasing in C_1, and we will have $u'(C_1) > \mathit{opp\_cost}$. The new optimum can be reached by moving some of the extra $1 back to the first period, i.e. by increasing C_1. This will lower $u'(C_1)$ and increase *opp_cost*, eventually equating them. The implied MPC is between 0 and 1.

In the case of an opportunity cost that is decreasing in C_1, moving the entire $1 to the second period will *increase* the opportunity cost, so that $u'(C_1) < \mathit{opp\_cost}$. The new optimum will therefore involve *decreasing* C_1, which will increase both $u'(C_1)$ and *opp_cost*. A new optimum will indeed be reached since $u'(C_1)$ increases faster than *opp_cost* around an optimum,[33] and they will be equated at a *lower* C_1 level than the old optimal choice, implying $MPC < 0$.

The intuition behind this exercise can be understood by recalling the discussion in Section 2.3.C and noting that receiving an extra $1 in initial wealth and keeping C_1 unchanged is the same as lowering C_1 by $1 in terms of the effect on the opportunity cost. The opportunity cost of potential defaulters is therefore higher when the extra $1 is moved to the second period, and the current level

[30] "Never defaulters", for whom the opportunity cost is increasing in C_1 at the optimum, do split additional wealth between the present and future (e.g. have $0 < MPC < 1$). Assuming a cost of default that generates an increasing opportunity cost at "default possible" C_1 levels (see Section 2.3.C) would generate $0 < MPC < 1$ for potential defaulters as well.

[31] However, they describe negative MPC as the result of the effect of extra wealth on the relative value of several local maxima, representing different probabilities of receiving insurance payments (in the example they give, second period income has a discrete distribution and the opportunity cost increases in between several downward jumps, corresponding to the different states, rather than decreasing continuously, leading to multiple maxima (see p. 395)). Here there is at most a single "default possible" maximum for a given initial wealth level. Negative MPC corresponds to the single optimal C_1 decreasing as initial wealth increases rather than to switching between several maxima (see Figure 2.2).

[32] Increasing Y_1 while holding C_1 constant has the same effect on *Debt* (and on the opportunity cost) as decreasing C_1 since $\mathit{Debt} = (C_1 - Y_1)R + \bar{D}_0$.

[33] This follows from the concavity of the value function around an optimum.

of consumption seems too costly in terms of its effect on second period utility. The opportunity cost is higher because moving the extra \$1 to the second period increases the number of no-default states (decreases $\widehat{Y_2}$). That is, it increases the number of states in which reducing current consumption increases own future consumption as opposed to decreasing the amount defaulted upon (decreasing the lender's losses).[34] In this sense, extra wealth gives potential defaulters "more stake in the future": they expect to get more of their own marginal saving, so they are encouraged to save more (borrow less).[35]

Proposition 6. *Marginal propensity to consume out of initial wealth Y_1 is negative for potential defaulters (those with initial wealth $Y_1 < \widehat{Y_1}$, see Proposition 3). For "never defaulters" ($Y_1 > \widehat{Y_1}$), it is positive and smaller than 1.*

Proof. See Appendix C. □

To illustrate the negative MPC result as well as some other results derived earlier, Figure 2.2 plots the value function, i.e. expected utility as a function of first period consumption, for several initial wealth levels. At the initial wealth levels plotted in the top row ($Y_1 = 23$ and $Y_1 = 25$), the value function reaches its peak, marked by $*$, at the C_1 level implied by the borrowing constraint, where the value function drops to zero. That is, both wealth levels are lower than the borrowing constrained threshold $\check{Y}_1$ with the parameters used in this exercise, and individuals with these initial wealth levels choose to borrow up to the constraint. At the initial wealth levels plotted in middle row ($Y_1 = 27$ and $Y_1 = 29$), the value function reaches its peak at an interior C_1 level, i.e. at a value lower than that implied by the borrowing constraint. The peaks of both value functions are attained at higher C_1 values than the respective $\widehat{C_1}$ values (marked by the dashed lines), i.e. at "default possible" C_1 values. Individuals with initial wealth levels $Y_1 = 27$ and $Y_1 = 29$ therefore choose C_1 levels which imply that they will find it optimal to default in low second period income states. Note that the amount consumed at $Y_1 = 29$ is lower than the amount consumed at $Y_1 = 27$ (27.28 vs. 29.93), e.g. MPC is negative (the maximum of the value function moves to the left as we increase initial wealth).

At the initial wealth levels plotted in the bottom row ($Y_1 = 31$ and $Y_1 = 33$), the value function reaches its peak at lower C_1 values then the respective $\widehat{C_1}$ values (note that $\widehat{C_1}$ increases in Y_1),

[34] Recall that in default states the individual consumes his second period income, regardless of how much he consumed in the first period.

[35] As Hubbard, Skinner, and Zeldes (1995) note in the context of means-tested social insurance, there is an implicit tax of 100% on saving in the states in which the individual defaults.

i.e. at "default impossible" C_1 values. Individuals with initial wealth levels $Y_1 = 31$ and $Y_1 = 33$ therefore choose C_1 levels which imply that they will never find it optimal to default in the second period (both initial wealth levels are lower than $\widehat{Y_1}$, the "never default" threshold). The amount consumed at $Y_1 = 33$ is higher than the amount consumed at $Y_1 = 31$ (26.52 vs. 25.44), e.g. MPC is positive (the maximum of the value function moves to the right as we increase initial wealth in this region).

[Insert Figure 2.2 here]

Figure 2.3 plots the consumption function - optimal consumption as a function of initial wealth - for a range of initial wealth values, with the same parameters used in the previous figures. The slope of the curve is the marginal propensity to consume. At low initial wealth levels $Y_1 < \check{Y}_1$, such as those in the top row of Figure 2.2, this slope is 1 since the individual borrows up to the borrowing constraint so that $C_1 = Y_1 + \overline{B_1}$. At $\check{Y}_1$ there is a sharp drop in consumption as the individual moves from a regime of borrowing up to the constraint to one of choosing an interior solution. At initial wealth levels $\check{Y}_1 < Y_1 < \widehat{Y_1}$, such as those in the middle row of Figure 2.2, interior solutions which imply positive ex-ante probability of default are chosen, and consumption decreases with initial wealth (MPC is negative). At initial wealth levels $Y_1 > \widehat{Y_1}$, such as those in the bottom row of Figure 2.2, interior solutions which imply zero probability of default are chosen, and consumption increases with initial wealth (MPC is positive).

[Insert Figure 2.3 here]

2.5. The Effects of Favorable Interest Rate Changes

In this section I analyze the effects of favorable changes in interest rates, i.e. an increase in the saving rate R_S and a decrease in the borrowing rate R_B, on optimal consumption behavior. I first examine the effects of such changes on the initial wealth thresholds established in Propositions 3, 4 and 5. I then examine the effects on optimal consumption. Finally, I examine the effects of interest rate changes on expected default. As in Section 2.4, I operate under the assumption of a fixed cost of default.[36]

[36]See footnote 28

A. Effects on Initial Wealth Thresholds

Proposition 7 shows that favorable interest rate changes affect the initial wealth thresholds established in previous propositions in the following way. $\widehat{Y_1}$, the threshold which separates potential defaulters from "never defaulters", decreases, increasing the fraction of "never defaulters". The borrowing constrained threshold $\check{Y_1}$ increases, decreasing the fraction of individuals who hit the borrowing constraint (which increases when R_B decreases). The saving and borrowing thresholds $\grave{Y}_1(R_B)$ and $\grave{Y}_1(R_S)$ move closer to each other so that more individuals borrow and more save. There is no transition between saving and borrowing or vice-versa. The changes are summarized in Figure 2.4.

Proposition 7. *Favorable interest rate changes, i.e. an increase in the saving rate and a decrease in the borrowing rate, have the following effects on the thresholds established above:*

1. $\widehat{Y_1}$ decreases
2. Under reasonable assumptions, $\check{Y_1}$ decreases
3. $\grave{Y}_1(R_B)$ increases and $\grave{Y}_1(R_S)$ decreases

Proof. See Appendix C. □

[Insert Figure 2.4 here]

B. Effects on First Period Optimal Consumption

In Proposition 8 I derive the effects of favorable interest rate changes on optimal consumption. As is well-known, changing the interest rate generally has the following effects on optimal first period consumption:[37]

- An increase in the interest rate raises the price of current consumption relative to future consumption, while a decrease in the interest rate lowers this relative price ("substitution effect")
- An increase in the interest rate allows for higher second period consumption, all else equal, making the individual better off, while a decrease in the interest rate makes him worse off ("income effect")

[37] See Deaton (1992), p. 3

- An increase in the interest rate decreases the present discounted value of second period income and makes the individual worse off, while a decrease makes him better off ("human capital effect")

I refer to the combination of the latter 2 effects as "the wealth effect". As might be obvious, this combined effect is always positive for favorable interest rate changes. That is, favorable interest rate changes always make the individual better off.[38]

In Proposition 8 I derive the following expression for the elasticity of first period consumption wrt the interest rate:

$$\frac{dC_1}{dR}\frac{R}{C_1} = -(1-MPC)EIS + \frac{Y_1 - C_1}{C_1}MPC \tag{2.5.1}$$

where $EIS(C_1)$ is the elasticity of intertemporal substitution at C_1 and MPC is the marginal propensity to consume out of initial wealth at C_1.

The first term represents the substitution effect, and is always negative. That is, an increase in the relative price of C_1 works to decrease C_1. It is more negative for a given C_1 level when default is possible, since the term multiplying EIS is $(1-MPC)$, which is larger than 1 when default is possible and smaller than 1 when it is not. While the EIS determines how expected consumption *growth* responds to a change in R, how much of that response is manifested in changes in C_1 versus changes in expected C_2 (that is, changes in expected C_2 in non-default states) depends on how the intertemporal tradeoff between C_1 and expected C_2 changes with C_1. $1 - MPC = \frac{u''(C_1)}{u''(C_1) - \frac{d}{dC_1}opp\_cost}$ measures how the tradeoff changes,[39] and is larger where default is possible and the opportunity cost is decreasing (where $\frac{d}{dC_1}opp\_cost < 0$).[40]

The second term represents the wealth effect. The amount saved/borrowed (scaled by C_1), representing the impact of a change in R on the individual ("by how much does a change in R make the individual better off"), is multiplied by MPC to determine the effect on C_1. It follows that where MPC is positive, i.e. where default is impossible, the wealth effect of favorable interest

[38] An increase in the interest rate has a positive wealth effect on savers and a negative effect on borrowers. A decrease has the opposite effects. Favorable rate changes are defined as an increase in the saving rate and a decrease in the borrowing rate, and always have positive wealth effects.

[39] $u''(C_1)$ measures how the marginal benefit of C_1 changes while $\frac{d}{dC_1}opp\_cost$ measures how the opportunity cost, or marginal cost of C_1, changes with C_1.

[40] To illustrate this, consider the certainty case. In that case, $EIS = \frac{dC_2/C_1}{dR}\frac{R}{C_2/C_1} = \frac{dC_2}{dR}\frac{R}{C_2} - \frac{dC_1}{dR}\frac{R}{C_1}$. The substitution effect dictates that a 1% increase in the interest rate decrease C_1 by $(1-MPC)EIS$% and increase C_2 by $MPC * EIS$%, so that the growth rate C_2/C_1 increase by EIS%. When $MPC < 0$, the substitution effect of a 1% increase in the interest rate on C_1 is a decrease of more than EIS%, and the effect on C_2 is negative rather than positive.

rate changes implies increased consumption, as is standard. Where MPC is negative, however, i.e. where default is possible,[41] the wealth effect of favorable rate changes works to *decrease* consumption.

The elasticity is ultimately determined by the combination of wealth and substitution effects. Let us examine this combination where default is possible and where it is impossible, keeping in mind that the "default impossible" case corresponds to the standard textbook case (with uncertain second period income). For savers, an increase in R_S yields a negative substitution effect, implying lower C_1, regardless of the potential for default. Where default is impossible, MPC is positive and the wealth effect implies higher C_1, contradicting the substitution effect and overpowering it for some parameters and levels of initial saving. Where default is possible, however, MPC is negative and the wealth effect *reinforces* the substitution effect. It follows that while savers might increase or decrease their consumption in response to an increase in R_S where default is impossible, depending on the parameters used and the level of initial saving,[42] savers who are potential defaulters *always* decrease their consumption in response to this change.

For borrowers, it is a mirror image.[43] A decrease in R_B yields a positive substitution effect, implying higher C_1, regardless of the potential for default. Where default is impossible,[44] MPC is positive and the wealth effect reinforces the substitution effect, while where default is possible, it is negative and *contradicts* the substitution effect, overpowering it for some parameters and levels of initial borrowing. It follows that while borrowers always increase their consumption in response to a decrease in R_B where default is impossible, borrowers who are potential defaulters might increase or decrease their consumption in response to this change, depending on the parameters used and the level of initial borrowing.[45]

That is, the potential for default shifts the effects of favorable interest rate changes in a more conservative direction in the sense that all savers decrease their consumption in response to an increase in R_S (where only some would have otherwise) and some borrowers decrease their consumption in response to a decrease in R_B (where all would have increased it otherwise). This is brought about by the negative sign of the slope of the opportunity cost curve where default

[41] As shown in Section 2.4.B, the MPC is negative where the opportunity cost is decreasing in C_1. As shown in Section 2.3.D, this is the case for potential defaulters if the marginal cost of default is not high enough and in particular in the case of a fixed cost of default.

[42] For certain parameters there is a threshold level of initial saving (corresponding to a threshold level of initial wealth) above which the wealth effect is stronger than the substitution effect.

[43] The effect of decreasing R_B is the negative of (2.5.1).

[44] Although there are no borrowers with zero probability of default here by assumption (see footnote 4 in Appendix C), this is useful as a benchmark for evaluating the effects of incorporating the possibility of default.

[45] For certain parameters there is a threshold level of initial borrowing (corresponding to a threshold level of initial wealth) above which the wealth effect is stronger than the substitution effect.

is possible, and stands in contrast to the standard effects of favorable rate changes on optimal consumption.

Let us now compare the actual elasticities (rather than just signs) where default is possible and where it is not possible, for given levels of initial consumption. It should be noted that this comparison ignores the higher initial consumption brought about by the possibility of default (as does the above, which compares savers and borrowers from both groups). Still, it is instructive to consider the differences in elasticities holding C_1 fixed.

For given levels of C_1 which imply saving, the substitution effect is stronger where default is possible (implying a stronger decrease in consumption) and the wealth effect implies decreasing rather than increasing consumption. It follows that the interest elasticity of consumption is always more negative (implying a stronger decrease in consumption in response to an increase in R_S) where default is possible. For given levels of C_1 which imply borrowing, the stronger substitution effect where default is possible implies a stronger *increase* in consumption while the wealth effect implies decreasing rather than increasing consumption. It follows that, although the elasticity is positive for some parameters where default is possible and always negative where it is not, which elasticity is larger for borrowers depends on the parameters used.

Proposition 8 derives the effects discussed in this section formally.

Proposition 8. *Favorable interest rate changes, i.e. an increase in the saving rate and a decrease in the borrowing rate, have the following effects on optimal consumption:*

1. Where default is impossible (among “never defaulters”):

   (a) Initial savers decrease consumption (increase saving) if and only if the wealth effect is smaller (in absolute value) than the substitution effect:

   $$\frac{Y_1 - C_1}{C_1} MPC < (1 - MPC)EIS$$

   (b) All initial borrowers increase consumption (increase borrowing)[46]

2. Where default is possible (among “potential defaulters”):

   (a) All initial savers decrease consumption (increase saving)[47]

[46] See footnote 44

[47] Moreover, some potential defaulters who are savers increase their saving enough to become “never defaulters” (see Proposition 7).

(b) Initial borrowers increase consumption (increase borrowing) if and only if the wealth effect is smaller (in absolute value) than the substitution effect:[48]

$$-\frac{C_1 - Y_1}{C_1} MPC < (1 - MPC) EIS$$

3. For given levels of C_1 which imply saving, the interest elasticity of C_1 is more negative where default is possible than where it is not possible. For C_1 levels which imply borrowing, this depends on the parameters used.

Proof. See Appendix C. □

C. Effects on Expected Default

In this section I examine the effects of favorable interest rate changes on expected default: the probability of default as well as the closely related expected defaulted-upon amount.[49]

Recall that a potential defaulter defaults on his entire debt in relatively low income states $\tilde{Y_2} < \widehat{Y_2}$. The expected defaulted-upon amount is therefore:[50]

$$E(Def) = Pr.(Def) * Debt = (\widehat{Y_2} - \underline{Y_2}) Debt$$

The effect of changing R on this amount is:

$$\frac{dE(Def)}{dR} = \frac{d\widehat{Y_2}}{dR} Debt + (\widehat{Y_2} - \underline{Y_2}) \frac{dDebt}{dR}$$

where the first term represents the effect of changing R on the default threshold $\widehat{Y_2}$, i.e. on the probability of default, and the second term represents the effect of changing R on $Debt$, the amount owed (and potentially defaulted upon).

[48] This applies to those who transition between interior solutions in response to interest rate changes. Note that there will also be a small initial wealth interval where borrowers decrease their consumption sharply in response to a decrease in R_B as they transition from a corner solution (hitting the borrowing constraint) to an interior one (see Proposition 7).

[49] As shown in Proposition 7, favorable rate changes cause some potential defaulters to become "never defaulters" (but not "never defaulters" to become potential defaulters). Here I examine the effects for potential defaulters who still have positive probability of default following the rate changes.

[50] Recall that $\tilde{Y_2} \sim U[\underline{Y_2}, \overline{Y_2}]$ and that we normalize $\overline{Y_2} - \underline{Y_2} = 1$.

Proposition 9 shows that the directions of the effects of favorable interest rate changes on the probability of default and on the expected defaulted-upon amount are the same, and that they depend on the sign of $\frac{dDebt}{dR}$. As the proposition shows, increasing R_S always decreases *Debt* for initial savers, therefore always decreases both the probability of default and the expected defaulted-upon amount.[51]

Decreasing R_B increases *Debt* for some initial borrowers and decreases it for others. Proposition 9 shows that *Debt* decreases, i.e. decreasing R_B decreases expected default, if and only if:

$$\frac{C_1 - Y_1}{C_1} > -\frac{dC_1}{dR}\frac{R}{C_1} \tag{2.5.2}$$

That is, if and only if $\frac{C_1 - Y_1}{C_1}$, the fraction of first period consumption that is financed using borrowing, is larger than minus the elasticity of C_1 with respect to R.[52] Where this elasticity is positive, i.e. where decreasing R_B lowers the amount borrowed, the condition always holds and decreasing R_B always lowers expected default among borrowers. As Proposition 8 shows, the elasticity is positive for certain parameters where default is possible, and always negative when it is impossible (note that the condition might hold where the elasticity is negative as well).[53] Proposition 9 shows that with CRRA preferences, an equivalent condition to (2.5.2) is:

$$\frac{C_1 - Y_1}{C_1} > \frac{1}{\rho} \tag{2.5.3}$$

where ρ is the Arrow-Pratt coefficient of relative risk aversion, also equal to the inverse of the elasticity of intertemporal substitution. That is, probability of default and expected defaulted-upon amount decrease for borrowers who finance at least $\frac{1}{\rho}$ of their consumption through borrowing.

The fraction of consumption financed through borrowing decreases in initial wealth and increases in borrowing, so that the condition corresponds to an initial wealth and an initial borrowing threshold, s.t. those who are relatively poor and initially borrow relatively large amounts decrease expected default when R_B decreases, while those who are relatively wealthy and initially borrow relatively small amounts increase expected default.

[51] Recall that some savers have $Debt > 0$ because of the pre-existing obligation $\bar{D}_0$, which they can default on. The assumption that second period consumption in case of default is equal to second period income implies that any savings are seized in default and that $\bar{D}_0$ is partially paid back rather than defaulted upon in full when an individual who had saved in the first period defaults.

[52] This corresponds to the elasticity of the amount borrowed $(C_1 - Y_1)$ wrt R being higher than -1.

[53] In other words, if borrowing (B_1) decreases when R_B decreases, $Debt = B_1 R_B$ also decreases. If borrowing increases, the product might still decrease.

Note that (2.5.3) is a general condition for decreasing R_B to decrease $Debt$ and increase second period consumption in no-default states.[54] That is, it is a condition for the elasticity of second period consumption with respect to R_B to be positive, reflecting the combination of substitution and wealth effects on second period consumption.

Although the slope of the opportunity cost does not appear in (2.5.3), the possibility of default and its implications for optimal consumption behavior impact the effects of favorable rate changes on $Debt$. First, where default is possible C_1 and $\frac{C_1-Y_1}{C_1}$ are higher for given initial wealth. There is therefore an initial wealth range for which (2.5.3) holds where default is possible but not where it is not possible. Second, the possibility of default impacts the magnitude of the effect on $Debt$. Proposition 9 shows that with CRRA preferences, the derivative of $Debt$ wrt R is:

$$\frac{dDebt}{dR} = C_1(1 - MPC)\left(\frac{C_1 - Y_1}{C_1} - \frac{1}{\rho}\right) \tag{2.5.4}$$

For given C_1 levels, the term multiplying $(\frac{C_1-Y_1}{C_1} - \frac{1}{\rho})$ is larger where default is possible, since MPC is smaller. It follows that for C_1 levels for which $\frac{dDebt}{dR}$ is positive (for which decreasing R_B decreases expected default for borrowers), the effect is larger where default is possible.[55] For C_1 levels for which $\frac{dDebt}{dR}$ is negative,[56] whether the effect is smaller or larger where default is possible depends on the parameters.

Proposition 9 derives the effects discussed in this section formally.

Proposition 9. *Favorable interest rate changes, i.e. an increase in the saving rate and a decrease in the borrowing rate, have the following effects on the probability of default and the expected defaulted-upon amount of potential defaulters:*

1. Savers who are potential defaulters (who might default on $\bar{D_0}$, the pre-existing obligation) decrease their probability of default as well as their expected defaulted-upon amount.

2. Borrowers decrease their probability of default as well as their expected defaulted-upon amount if and only if:
$$\frac{C_1 - Y_1}{C_1} > -\frac{dC_1}{dR}\frac{R}{C_1}$$

[54]Second period consumption in no-default states is $\tilde{Y_2} - Debt$, so $\frac{d}{dR}Debt = -\frac{d}{dR}C_2$ in those states.
[55]Note that a given C_1 implies a higher level of $\frac{C_1-Y_1}{C_1}$ where default is possible relative to where it is not possible, which contributes to the larger effect.
[56]This includes all C_1 levels which imply saving (for which increasing R_S decreases expected default), as well as C_1 levels which imply borrowing for which decreasing R_B increases expected default.

and increase these quantities otherwise. With $CRRA$ preferences, an equivalent condition is:

$$\frac{C_1 - Y_1}{C_1} > \frac{1}{\rho}$$

where ρ is the Arrow-Pratt coefficient of relative risk aversion. $\frac{C_1 - Y_1}{C_1}$ corresponds to an initial wealth threshold and an initial borrowing threshold.

3. As shown in Proposition 7, favorable rate changes cause some of those who were initially neither borrowing nor saving to start saving and some to start borrowing. Those who start saving lower their probability of default and expected defaulted-upon amount while those who start borrowing increase these quantities.

Proof. See Appendix C. □

2.6. Extension: Voluntary Extent of Delinquency in Both Periods

In this section I extend the basic model in two ways. First, I allow the individual to choose *how much* of his debt he wishes to repay. That is, the choice is not between defaulting on all obligations and repaying in full: there is flexibility to choose any point between these two extremes. This might be thought of as corresponding to delinquency, i.e. failing to repay some portion of the debt on time, rather than to bankruptcy and full discharge of debts.[57]

More importantly, I allow for *first period* delinquency. In the main version of the model, failing to repay in full is possible only in the second period: $\bar{D_0}$, the pre-existing fixed obligation that is due in every period, is assumed to always be paid back in full in the first period (recall that Y_1 is defined as net of $\bar{D_0}$). Here I relax this assumption and allow the individual to be delinquent on $\bar{D_0}$ in the first period. The goal is to examine how favorable interest rate changes might affect consumption and delinquency when borrowing is possible (in the main version of the model, it is not possible to be delinquent and borrow at the same time).

The analysis here, which shows that lowering the borrowing rate reduces the likelihood and extent of first period delinquency, illustrates another channel through which favorable changes

[57] Athreya, Sánchez, Tam, and Young (2012) make the distinction between bankruptcy (corresponding to the all-or-nothing default setup used above) and delinquency in their model and examine the tradeoff between them. Dawsey and Ausubel (2004) show that delinquency, which they term "informal bankruptcy", is quite prevalent, even among relatively creditworthy borrowers, and study it empirically using data on credit card accounts.

in interest rates can affect consumer behavior, in particular the repayment of debts. Individuals essentially trade off borrowing and delinquency in the first period to finance consumption: every dollar of $\bar{D_0}$ that is repaid has to be borrowed in order to maintain the same level of consumption. Lowering the cost of borrowing lowers delinquency by making borrowing and repaying (i.e. rolling over) more attractive relative to delinquency.

I first solve the second period problem in which the individual chooses the extent to which he wishes to be delinquent on $Debt$. Assuming a constant marginal cost of delinquency,[58] the solution is characterized by a floor level of consumption $\underline{C_2}$: the individual is delinquent if and only if second period income is insufficient to cover $Debt$ as well as consume $\underline{C_2}$, and the extent of delinquency is as much as is needed in order to consume this level. I then move on to the first period problem and show that there is a tradeoff between using borrowing and being delinquent on $\bar{D_0}$ as means of financing first period consumption. I characterize the optimal delinquency-borrowing mix for different levels of C_1 and then, given this optimal mix for the different levels of C_1, characterize the optimal choice of C_1 for different first period income levels. As in the second period, there is a floor level of consumption $\underline{C_1}$ that delinquency is used to ensure. Finally, I derive the effects of favorable interest rate changes on optimal first period delinquency.

A. The Second Period Problem

In this version of the model, the individual enters the second period with $Debt = B_1 R + \bar{D_0}$, receives $\tilde{Y_2}$ and then chooses Del_2, the delinquent amount. Second period consumption is:

$$C_2 = \tilde{Y_2} - Debt + Del_2$$

The cost of delinquency is $\Lambda_2(Del_2)$. In this case, this cost can be thought of as reflecting damage to the individual's credit score (resulting in higher future borrowing costs) as well as any future expected payments as a result of delinquency (i.e. at least some of the delinquent amount might be expected to be paid eventually, possibly in addition to interest and fees).[59]

If $\tilde{Y_2}$ is insufficient to cover $Debt$, the individual must be delinquent to some extent (i.e. C_2 cannot be negative). Moreover, Del_2 cannot be negative and cannot exceed $Debt$. The resulting second period optimization problem is:

[58] Note that in this case, the opportunity cost curve increases in C_1 everywhere. See footnote 9 in Appendix C.
[59] This can be modeled explicitly in a model with more than 2 periods (see conclusion).

$$
\begin{array}{ll}
max & u(\tilde{Y_2} - Debt + Del_2) - \Lambda_2(Del_2) \\
Del_2 & \\
s.t. & Del_2 \geq max\{0, Debt - \tilde{Y_2}\} \\
 & Del_2 \leq Debt
\end{array}
$$

The interior solution to this problem is defined by the first order condition:

$$u'(\tilde{Y_2} - Debt + Del_2) = \Lambda_2'(Del_2) \tag{2.6.1}$$

For simplicity, I assume a constant marginal cost of delinquency: $\Lambda_2(Del_2) = \Lambda_2 * Del_2$.[60] Denoting the C_2 level at which marginal utility is equal to the marginal cost of delinquency by $\underline{C_2}$:

$$(u')^{-1}(\Lambda_2) \equiv \underline{C_2}$$

The interior solution (which satisfies (2.6.1)) is:

$$Del_2 = \underline{C_2} - (\tilde{Y_2} - Debt)$$

Incorporating the constraints and assuming $\underline{Y_2} > \underline{C_2}$,[61] we get the following second period behavior:

$$Del_2 = \begin{cases} 0 & \text{if } \tilde{Y_2} \geq \underline{C_2} + Debt \\ \underline{C_2} - (\tilde{Y_2} - Debt) & \text{if } \underline{C_2} \leq \tilde{Y_2} < \underline{C_2} + Debt \end{cases}$$

$$C_2 = \begin{cases} \tilde{Y_2} - Debt & \text{if } \tilde{Y_2} \geq \underline{C_2} + Debt \\ \underline{C_2} & \text{if } \underline{C_2} \leq \tilde{Y_2} < \underline{C_2} + Debt \end{cases}$$

That is, in the second period the individual is delinquent if and only if $\tilde{Y_2}$ is not sufficiently high to ensure some floor level of consumption $\underline{C_2}$. The extent of delinquency, in that case, will be as much as is needed in order to consume $\underline{C_2}$. This pattern is not surprising, given the assumption

[60] A fixed cost of delinquency as in the basic model would effectively reduce this to the "all or nothing default" case analyzed above.

[61] This assumption ensures that the constraint $Del_2 \leq Debt$ is never hit, i.e. the individual never wishes to be delinquent on more than he owes. In particular, when the individual saves enough s.t. $Debt < 0$, i.e. $-B_1 R_S > \bar{D_0}$, he pays $\bar{D_0}$ in full ($Del_2 = 0$).

that the marginal cost of delinquency is constant and given that marginal utility is monotonically decreasing.

B. The First Period Problem

The individual starts the first period owing $\bar{D_0}$.[62] After receiving Y_1 (termed "first period income" rather than "initial wealth" here to emphasize that it is not defined net of prior debt, as it is in the main version of the model), he chooses C_1 and Del_1. Assume that the cost of delinquency is fixed in this period as well: $\Lambda_1(Del_1) = \Lambda_1$.[63]

Going back to the second period problem, define a second period income threshold:

$$\widehat{Y_2} = max\{\underline{Y_2}, \underline{C_2} + Debt\}$$

and note that similarly to the original version of the model, the individual optimally chooses to pay his second period obligations in full if and only if $\tilde{Y_2} \geq \widehat{Y_2}$. If $\tilde{Y_2} < \widehat{Y_2}$, delinquency occurs to some extent.

Assuming $Y_1 > \bar{D_0}$, i.e. first period income is always sufficient to cover the pre-existing obligation, the individual solves the following problem in the first period:

$$\begin{aligned} \underset{C_1, Del_1}{max} \quad & u(C_1) - \Lambda_1 Del_1 + \beta\left(\int_{\widehat{Y_2}}^{\overline{Y_2}} u(\tilde{Y_2} - Debt) d\tilde{Y_2} + \int_{\underline{Y_2}}^{\widehat{Y_2}} (u(\underline{C_2}) - \Lambda_2 Del_2) d\tilde{Y_2}\right) \qquad (2.6.2) \\ s.t. \quad & Debt = B_1 R + \bar{D_0} \\ & B_1 = (C_1 - Y_1) + (\bar{D_0} - Del_1) \\ & 0 \leq Del_1 \leq \bar{D_0} \\ & Del_2 = max\{\underline{C_2} - (\tilde{Y_2} - Debt), 0\} \end{aligned}$$

C. Characterizing the Delinquency-Borrowing Tradeoff

(2.6.2) can be viewed as a two stage problem, where there is an optimal mix of delinquency and borrowing that is used to finance any C_1 level in excess of Y_1. This point is clear when re-writing

[62] Note that $\bar{D_0}$ is owed in each period.

[63] In principle, the delinquent amount in the first period should be due in the second period, but I ignore that here for simplicity.

the first period budget constraint:

$$C_1 = Y_1 + B_1 + (Del_1 - \bar{D_0})$$

The optimal mix of delinquency and borrowing is determined by the constraint $0 \leq Del_1 \leq \bar{D_0}$ combined with the first order condition wrt Del_1:

$$\Lambda_1 = \beta R(\int_{\underline{Y_2}}^{\widehat{Y_2}} \Lambda_2 d\tilde{Y_2} + \int_{\widehat{Y_2}}^{\overline{Y_2}} u'(\tilde{Y_2} - Debt) d\tilde{Y_2}) \tag{2.6.3}$$

where the LHS is the marginal cost of delinquency and the RHS is the marginal cost of borrowing.

Define $\breve{B}_1$ as the borrowing level at which (2.6.3) holds:

$$\Lambda_1 = \beta R(\int_{\underline{Y_2}}^{\widehat{Y_2}(\breve{B}_1 R)} \Lambda_2 d\tilde{Y_2} + \int_{\widehat{Y_2}(\breve{B}_1 R)}^{\overline{Y_2}} u'(\tilde{Y_2} - \breve{B}_1 R - \bar{D_0}) d\tilde{Y_2}) \tag{2.6.4}$$

Proposition 10 shows that a constant marginal cost of delinquency leads to the delinquency-borrowing tradeoff having three distinct first period consumption regions. Relatively low levels of consumption are completely financed by borrowing (or, at $C_1 < Y_1$, the individual is saving) and $\bar{D_0}$ is paid in full (the constraint $Del_1 \geq 0$ is binding in this region). At intermediate levels of consumption, $\breve{B}_1$ is borrowed and the individual is delinquent on $\bar{D_0}$ to the extent needed to finance consumption in excess of $Y_1 - \bar{D_0} + \breve{B}_1$ (this is the only region in which there is an interior solution, i.e. in which (2.6.3) holds). At high levels of consumption, the individual is delinquent on the entire $\bar{D_0}$ and borrows as much as is needed (the constraint $Del_1 \leq \bar{D_0}$ is binding in this region).

Proposition 10. *When the marginal cost of delinquency is constant, the delinquency-borrowing tradeoff is characterized by three C_1 regions:*

1. $C_1 < Y_1 - \bar{D_0} + \breve{B}_1$: the individual borrows $B_1 \leq \breve{B}_1$ and pays $\bar{D_0}$ in full

2. $Y_1 - \bar{D_0} + \breve{B}_1 \leq C_1 < Y_1 + \breve{B}_1$: the individual borrows $B_1 = \breve{B}_1$ and is delinquent on $(C_1 - (Y_1 - \bar{D_0} + \breve{B}_1))$

3. $Y_1 + \breve{B}_1 \leq C_1$: the individual borrows $B_1 > \breve{B}_1$ and is delinquent on the entire fixed obligation $\bar{D_0}$

where $\breve{B}_1$ is defined by (2.6.4).

Proof. See Appendix C. $\square$

D. Characterizing Optimal First Period Consumption

Now that we have established the optimal delinquency-borrowing mix for any C_1 level, let us examine the optimal choice of C_1. The relevant first order condition from (2.6.2) is the one wrt C_1:

$$u'(C_1) = \beta R(\int_{\underline{Y_2}}^{\widehat{Y_2}} \Lambda_2 d\tilde{Y_2} + \int_{\widehat{Y_2}}^{\overline{Y_2}} u'(\tilde{Y_2} - Debt)d\tilde{Y_2}) \tag{2.6.5}$$

Incorporating the findings of Proposition 10, Proposition 11 characterizes optimal first period consumption and delinquency for different levels of first period income, establishing regions that correspond to the consumption regions established in Proposition 10.

As in the second period, there is a floor level of consumption $\underline{C_1}$ where the marginal utility of consumption is equal to the marginal cost of delinquency. At high levels of first period income, consumption is higher than this level and is financed completely by borrowing ($\bar{D}_0$ is paid in full; the constraint $Del_1 \geq 0$ is binding). At intermediate levels of first period income, $\breve{B}_1$ is borrowed and the individual is delinquent on as much of $\bar{D}_0$ as is needed in order to finance the floor level of consumption $\underline{C_1}$. At low levels of first period income, the individual is delinquent on the entire $\bar{D}_0$ (the constraint $Del_1 \leq Debt$ is binding), borrows more than $\breve{B}_1$ and consumes less than $\underline{C_1}$.

Proposition 11. *First period consumption and delinquency are characterized by the following first period income (Y_1) regions:*

1. $Y_1 > \underline{C_1} + \bar{D}_0 - \breve{B}_1$: the individual borrows $B_1 < \breve{B}_1$ and consumes $C_1 > \underline{C_1}$. There is no delinquency: $\bar{D}_0$ is paid in full.

2. $\underline{C_1} - \breve{B}_1 < Y_1 \leq \underline{C_1} + \bar{D}_0 - \breve{B}_1$: the individual borrows $\breve{B}_1$ and consumes $C_1 = \underline{C_1}$. There is partial delinquency: $Y_1 - (\underline{C_1} - \breve{B}_1)$ out of $\bar{D}_0$ is paid.

3. $Y_1 \leq \underline{C_1} - \breve{B}_1$: the individual borrows $B_1 > \breve{B}_1$ and consumes $C_1 < \underline{C_1}$. The individual is delinquent on the entire $\bar{D}_0$.

Proof. See Appendix C. $\square$

E. The Effects of Favorable Interest Rate Changes

Favorable interest rate changes, in particular a decrease in the borrowing rate, increase the attractiveness of borrowing relative to delinquency in the first period. Delinquency thus decreases, in terms of the fraction of individuals who are delinquent as well as the average delinquent amount.

This can be seen by observing that changing R affects the borrowing-delinquency tradeoff expressed in (2.6.3) by lowering the marginal cost of borrowing for any given borrowing level. The marginal cost of delinquency and the marginal cost of borrowing are then equated at a higher borrowing level. That is, $\breve{B}_1$, as defined in (2.6.4), increases when R decreases.[64] In light of the thresholds and quantities established in Proposition 11, a higher $\breve{B}_1$ has the following effects on first period delinquency:

- The fraction of individuals who pay $\bar{D}_0$ in full increases
- The fraction of individuals who are delinquent on the entire $\bar{D}_0$ decreases
- Individuals who are partially delinquent on $\bar{D}_0$ decrease their delinquent amount (some, as noted above, to zero)

2.7. Conclusion

This paper shows that incorporating the possibility of default in the basic life-cycle model can have significant implications for consumer behavior, as well as for the response of this behavior to changes in other variables such as wealth and the interest rate. When the cost of default is not sufficiently dependent on the amount defaulted upon, consumption is higher than it would have otherwise been for those who might default, and marginal propensity to consume out of wealth is negative. This has implications for the response to any price changes that make the consumer better off, such as favorable interest rate changes. The wealth effects of such changes work to decrease consumption through strengthening the link between first period consumption and second period utility, a link that is initially weakened by the possibility of default.

As noted in the introduction, the results are relevant for multiple economic issues linked to the effects of the return to saving and the cost of borrowing on consumer behavior. They are particularly relevant for thinking about the effects of bank account ownership, and imply that

[64] To see this, note that the partial derivative of the RHS of (2.6.4) wrt R is positive (the marginal cost of borrowing increases in R), while the derivative wrt $\breve{B}_1$ is negative. Since the LHS of (2.6.4) (the marginal cost of delinquency) is constant, decreasing R increases $\breve{B}_1$, the borrowing level at which the marginal costs are equal.

models without the possibility of default might underestimate the role of account ownership in encouraging saving and discouraging excessive borrowing of previously unbanked individuals - a relatively default-prone population for whom there might not be enough information to enable the accurate pricing of individual loans. A more comprehensive model of the effects of bank account ownership might also include transactional costs such as check cashing and, perhaps more importantly, self-control problems and limited attention to the need to save which result in over-consumption. Account ownership has been argued to help address self-control problems through mental accounting (Dupas and Robinson (2013)) and might also address limited attention through default mechanisms such as direct deposit of income into savings vehicles.[65] However, a more general point is that if individuals are subject to self-control problems that result in over-consumption, lowering the borrowing rate might be welfare-reducing because it would cater to these problems, similar to lowering the price of any good that is subject to self-control problems.[66]

This paper is mainly concerned with the implications of allowing for bankruptcy-style default where debts are discharged in full. Delinquency, the failure to repay some portion of due debts on time, is another important phenomenon in unsecured credit markets. Extending the model to several periods where both default and delinquency are possible, as Athreya, Sánchez, Tam, and Young (2012) have recently done, is therefore an obvious way to enrich it. A step in that direction is taken in Section 2.6, which allows choice of the amount repaid but does not specify separate implications for delinquency and default.[67] As this paper demonstrates, the way in which the costs of failing to repay in full are modeled, in particular their dependence on the amount unpaid, will determine the implications for the effects of interest rate changes on consumer behavior.

[65]As Mullainathan and Shafir (2009) put it, "financial institutions ... should not be simply viewed from a financial cost-saving point of view but instead should be understood to affect the lives of people by easing their planning, facilitating their desired actions and enabling their resistance to temptation."

[66]Morse (2011) makes this point in the context of access to payday loans.

[67]In Athreya, Sánchez, Tam, and Young (2012) both delinquency and default carry a fixed cost and exclude the household from credit markets for one period. Default relieves the household of all obligations, however, while delinquency allows lenders to garnish some portion of income and involves a revaluation of the amount outstanding.

Figure 2.1: The Marginal Benefit and Opportunity Cost Curves

I plot the marginal benefit and the opportunity cost of first period consumption for a range of first period consumption (C_1) levels. The top plot includes the two curves for an individual with initial wealth (Y_1) equal to 38, while the bottom plot includes the curves for an individual with initial wealth (Y_1) equal to 28. Both plots use *CRRA* preferences with a coefficient of relative risk aversion equal to 3, as well as the following parameters: $\beta = 1$, $R_S = R_B = 1$, $\underline{Y_2} = 20, \overline{Y_2} = 50$, $\bar{D}_0 = 10$. The cost of default is fixed and equal to $u(20) - u(15)$.

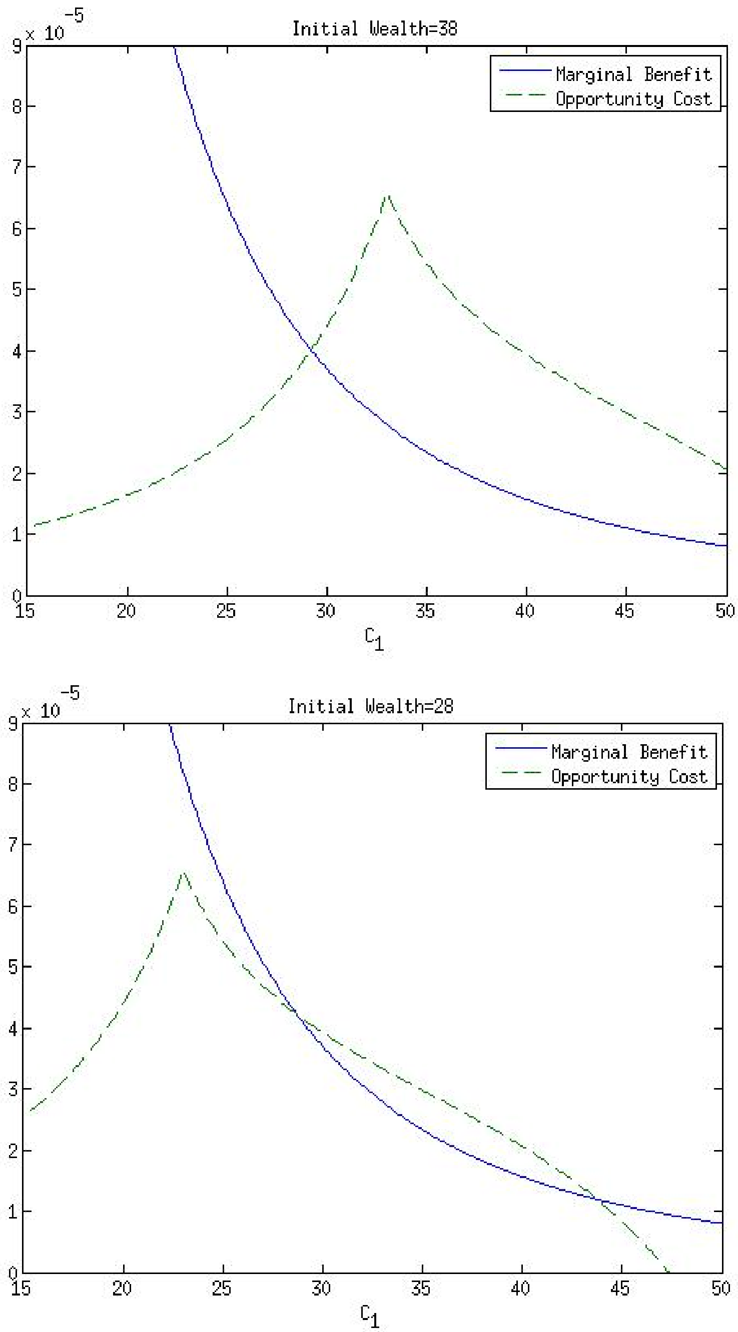


Figure 2.2: The Value Function for Various Levels of Initial Wealth

I plot the value function for several initial wealth levels, specified at the top of each plot. The dashed lines mark $\widehat{C_1}(Y_1)$, as defined in Section 2.3.B, and the maxima of the value functions are marked by $*$. I use $CRRA$ preferences with a coefficient of relative risk aversion equal to 3, as well as: $\beta = 1$, $R_S = R_B = 1$, $\underline{Y_2} = 20, \overline{Y_2} = 50$, $\bar{D}_0 = 10$. The cost of default is fixed and equal to $u(20) - u(15)$.

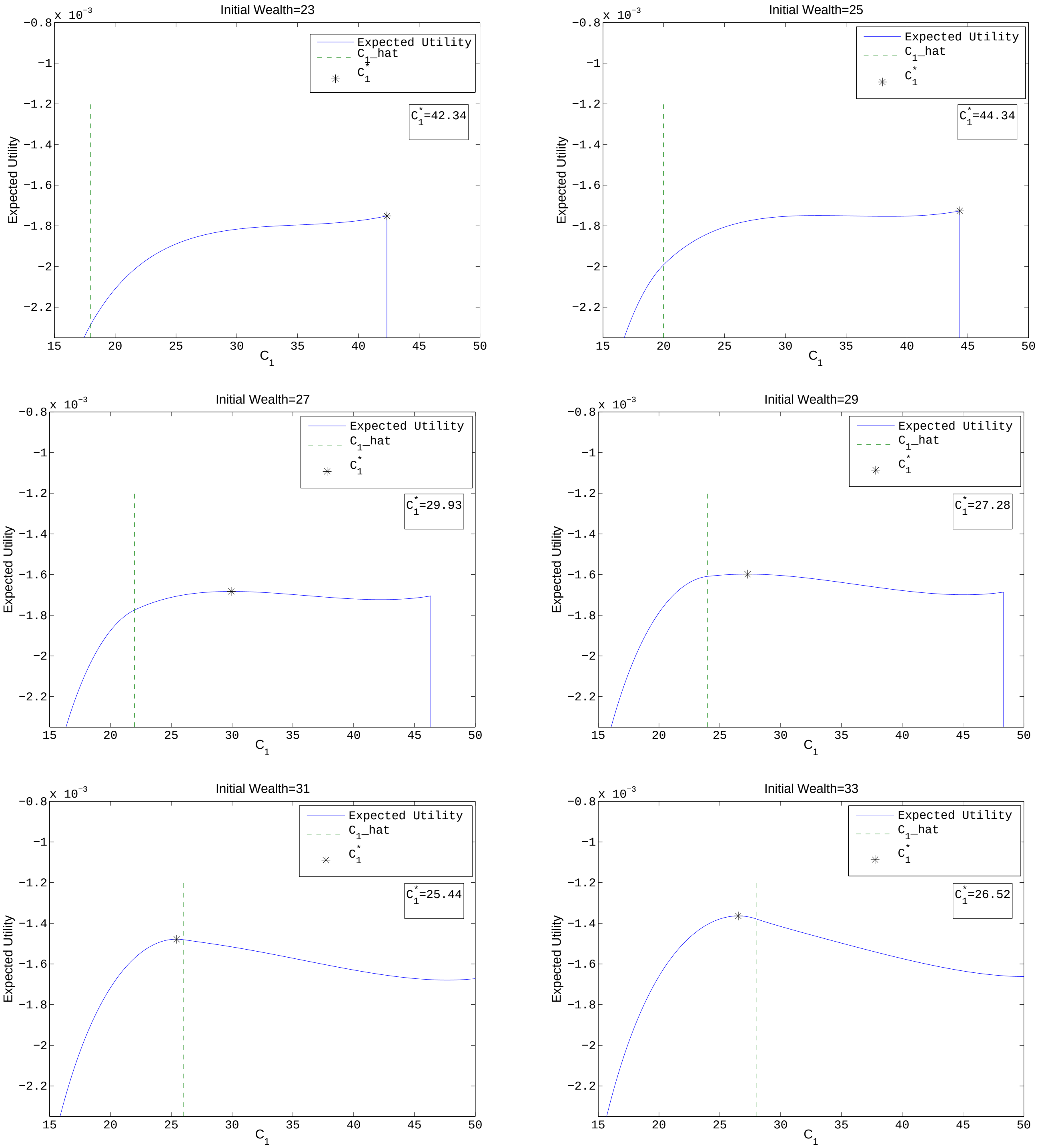


Figure 2.3: The Consumption Function

I plot the consumption function for a range of initial wealth (Y_1) values. $\widehat{Y_1}$ and $\check{Y}_1$, as defined in Propositions 3 and 4, are marked on the figure. I use $CRRA$ preferences with a coefficient of relative risk aversion equal to 3, as well as the following parameters: $\beta = 1$, $R_S = R_B = 1$, $\underline{Y_2} = 20$,$\overline{Y_2} = 50$, $\bar{D}_0 = 10$. The cost of default is fixed and equal to $u(20) - u(15)$.

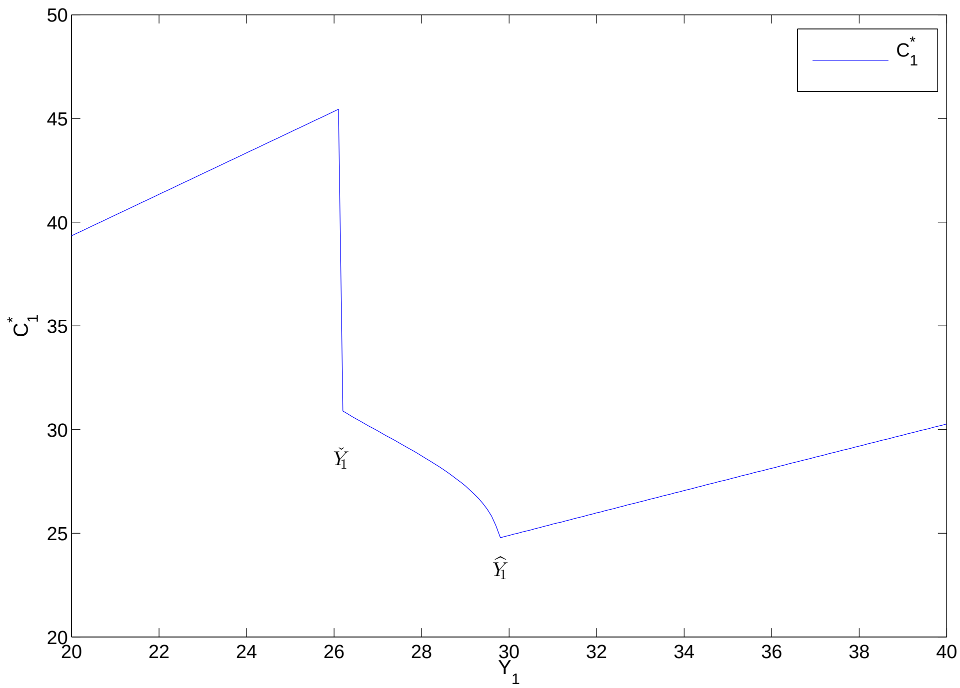
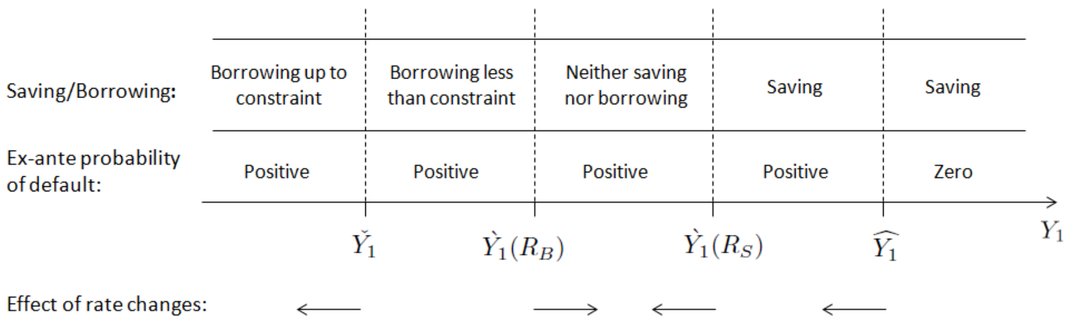


Figure 2.4: Initial Wealth Thresholds

I divide initial wealth (Y_1) into the regions established in Propositions 3, 4 and 5. The top two rows contain information on saving and borrowing in the various regions and on the ex-ante probability of default. The bottom row shows the directions of the effects of favorable interest rate changes on the thresholds separating the regions, as established in Proposition 7.



2.8. References

Athreya, Kartik, Juan M. Sánchez, Xuan S. Tam, and Eric R. Young, 2012, Bankruptcy and delinquency in a model of unsecured debt, Federal Reserve Bank of St. Louis Working Paper Series.

Athreya, Kartik, Xuan S. Tam, and Eric R. Young, 2012, A quantitative theory of information and unsecured credit, *American Economic Journal: Macroeconomics* 4, 153–183.

Athreya, Kartik B., 2002, Welfare implications of the Bankruptcy Reform Act of 1999, *Journal of Monetary Economics* 49, 1567–1595.

——— , 2008, Default, insurance, and debt over the life-cycle, *Journal of Monetary Economics* 55, 752–774.

Attanasio, Orazio P., Pinelopi K. Goldberg, and Ekaterini Kyriazidou, 2008, Credit constraints in the market for consumer durables: Evidence from micro data on car loans, *International Economic Review* 49, 401–436.

Attanasio, Orazio P., and Matthew Wakefield, 2010, The effects on consumption and saving of taxing asset returns, in Stuart Adam, Timothy Besley, Richard Blundell, Stephen Bond, Robert Chote, Malcolm Gammie, Paul Johnson, Gareth Myles, and James Poterba, ed.: *Dimensions of Tax Design: The Mirrlees Review* (Oxford University Press, Oxford).

Ausubel, Lawrence M., 1999, Adverse selection in the credit card market, Unpublished.

Bernheim, B. Douglas, 2002, Taxation and saving, in Alan J. Auerbach, and Martin S. Feldstein, ed.: *Handbook of Public Economics,*vol. 3 . pp. 1173–1249 (Elsevier, Amsterdam).

Boskin, Michael J., 1978, Taxation, saving, and the rate of interest, *Journal of Political Economy* 86, S3–S27.

Chatterjee, Satyajit, Dean Corbae, Makoto Nakajima, and José-Víctor Ríos-Rull, 2007, A quantitative theory of unsecured consumer credit with risk of default, *Econometrica* 75, 1525–1589.

Dawsey, Amanda E., and Lawrence M. Ausubel, 2004, Informal bankruptcy, Unpublished.

Deaton, Angus, 1992, *Understanding Consumption* (Clarendon Press, Oxford).

Dupas, Pascaline, and Jonathan Robinson, 2013, Why don't the poor save more? evidence from health savings experiments, *The American Economic Review* 103, 1138–1171.

Federal Deposit Insurance Corporation, 2012, 2011 FDIC National Survey of Unbanked and Underbanked Households, Available at http://www.fdic.gov/householdsurvey/2012_unbankedreport.pdf.

Feldstein, Martin S., and Sho-Chieh Tsiang, 1968, The interest rate, taxation, and the personal savings incentive, *Quarterly Journal of Economics* 82, 419–434.

Gross, David B., and Nicholas S. Souleles, 2002, Do liquidity constraints and interest rates matter for consumer behavior? evidence from credit card data, *Quarterly Journal of Economics* 117, 149–185.

Hall, Robert E., 1998, Intertemporal substitution in consumption, *Journal of Political Economy* 96, 339–357.

Hubbard, R. Glenn, Jonathan Skinner, and Stephen P. Zeldes, 1995, Precautionary saving and social insurance, *Journal of Political Economy* 103, 360–399.

Karlan, Dean S., and Jonathan Zinman, 2008, Credit elasticities in less-developed economies: Implications for microfinance, *The American Economic Review* 98, 1040–1068.

Livshits, Igor, James MacGee, and Michèle Tertilt, 2007, Consumer bankruptcy: A fresh start, *The American Economic Review* 97, 402–418.

———, 2010, Accounting for the rise in consumer bankruptcies, *American Economic Journal: Macroeconomics* 2, 165–193.

———, 2011, Costly contracts and consumer credit, NBER Working Paper #17448.

Morse, Adair, 2011, Payday lenders: Heroes or villains?, *Journal of Financial Economics* 102, 28–44.

Mullainathan, Sendhil, and Eldar Shafir, 2009, Savings policy and decision making in low-income households, in Rebecca M. Blank, and Michael S. Barr, ed.: *Insufficient Funds: Savings, Assets, Credit, and Banking Among Low-Income Households* (Russell Sage Foundation, New York).

Romer, David, 2011, *Advanced Macroeconomics* (McGraw-Hill, New York).

Summers, Lawrence H., 1981, Capital taxation and accumulation in a life cycle growth model, *The American Economic Review* 71, 533–544.

Vissing-Jorgensen, Annette, 2002, Limited asset market participation and the elasticity of intertemporal substitution, *Journal of Political Economy* 110, 825–853.

Chapter 3

Asset Pricing in the Dark: The Cross Section of OTC Stocks[1]

3.1. Introduction

While hundreds of studies have investigated expected return patterns in listed stocks that trade on the NYSE, Amex, and NASDAQ, many U.S. stocks—roughly one-fifth of the number of stocks listed on the major exchanges—trade in OTC markets. The definition of an OTC stock is one that trades on either the OTC Bulletin Board (OTCBB) or OTC Link (formerly Pink Sheets, or PS) interdealer quotation system, where at least one licensed broker-dealer agrees to make a market in the stock. We examine market data for 6,668 OTC firms from 1977 through 2008. To our knowledge, this is the largest dataset of U.S. stock prices to be introduced to research since the Center for Research on Security Prices (CRSP) added data on NASDAQ stocks in 1984.

The OTC and listed stock markets consist of many similar firms and market participants. More than 80% of OTC firms with market capitalizations above $1 million are traded in listed markets either before, concurrently, or after their OTC trading activity. Most broker-dealers who act as market makers in OTC stocks are also market makers in listed markets. Moreover, many investors, including retail investors and hedge funds, actively trade both groups of stocks.

There are, however, three important differences between OTC and listed stocks. First, there is far lower liquidity in OTC markets than on the major exchanges. Second, whereas firms in listed

[1]This chapter is co-authored with Andrew Ang and Paul Tetlock of Columbia University.

stock markets must file regular financial disclosures, disclosure requirements for firms traded in OTC markets are minimal, if non-existent, for most of our sample period.[2] Third, non-institutional (i.e., retail) investors are the primary owners of most OTC stocks, whereas institutional investors hold significant stakes in nearly all stocks on listed exchanges, including small stocks. Possibly as a consequence of low ownership by institutions, the main lenders of shares, short selling of OTC stocks is difficult, expensive, and rare.

We exploit these features of OTC and listed stock markets to distinguish among myriad theories of return premiums. Differentiating theories whose predictions depend on stocks' information environments and investor clientele using only the listed markets is challenging because all listed stocks are subject to the same reporting requirements and nearly all are held by institutions.[3] We estimate return premiums both within and across OTC markets and listed markets, sorting stocks by the characteristics that distinguish the two markets. This combined cross-market and within-market identification strategy allows for powerful tests of competing theories because the data exhibit enormous heterogeneity along both dimensions.

In light of the large cross-market differences in liquidity, we devote special attention to measuring illiquidity premiums. We find that the return premium for illiquid stocks is much higher in OTC markets than in listed markets. One of our key liquidity measures is the proportion of non-trading days (PNT), where higher PNT indicates higher illiquidity, and we sort OTC stocks into PNT quintiles. When constructing listed return factors, we focus on "comparable" listed stocks with market capitalizations similar to the typical OTC stock to control for differences in firm size. We first evaluate factors' pre-cost returns. We find that an OTC illiquidity factor has an annual Sharpe ratio of 0.91, whereas the comparable listed illiquidity factor has a Sharpe ratio of just 0.14.

Asset pricing theories based on transaction costs, such Amihud and Mendelson (1986) and Constantinides (1986), do not explain the OTC illiquidity premium. These theories predict that stocks exhibit *positive* pre-cost risk-adjusted returns that increase with bid-ask spreads to compensate rational investors for their expected trading costs. Empirically, the most liquid OTC stocks exhibit risk-adjusted monthly pre-cost returns of –4.0%, implying that their post-cost returns are even more negative. In addition, the typical OTC investor incurs trading costs of less than 50

[2]After June 2000, firms listed on the OTCBB but not the PS must have at least 100 shareholders, file annual reports, hold annual shareholder meetings, and meet other governance requirements (see Bushee and Leuz, 2005).

[3]Researchers can also use international data, like Bekaert, Harvey, and Lundblad (2007) who estimate illiquidity premiums, or different asset classes like Karolyi and Sanders (1998), to study determinants of return premiums. International studies are hampered by different treatments of creditor rights and securities not having the same claims to cash flows across countries.

basis points per month, suggesting that the magnitudes of trading costs are too small to explain our findings. Data errors or microstructure biases in OTC stocks also do not explain the OTC illiquidity premium. Such errors and biases should be smaller in the most liquid stocks and would bias the returns of OTC stocks *upward*, implying their returns after adjusting for illiquidity effects and data errors should be even more negative.

The strongly negative returns of liquid OTC stocks are consistent with the idea that limits to arbitrage allow the OTC illiquidity premium to remain so high during our 32-year sample. Given the difficulty in short selling even liquid OTC stocks, an arbitrageur could be unable to attain the high Sharpe ratio of the OTC illiquidity premium. We also provide evidence that trading costs, while relatively insignificant for the typical OTC investor who trades very infrequently, could severely limit the effectiveness of short-horizon arbitrage in OTC stocks.

Next we test whether the well-known return premiums for stocks with low market capitalizations ("size"), high ratios of book equity to market equity ("value" or B/M), low idiosyncratic volatility ("volatility"), and high past returns ("momentum") generalize to OTC markets.[4] Interestingly, the return premiums for size, value, and volatility are similarly large in OTC markets and comparable listed markets. In contrast, the return premium for momentum is considerably smaller and less robust in OTC markets than in listed markets.[5] Most of the OTC return premiums above are driven by the negative returns on the short legs of the long-short portfolios, again consistent with theories in which limits to short selling affect prices.

We find that traditional factor models—using factors constructed from listed returns—do not account for the large illiquidity, size, value, and volatility return premiums in OTC markets. We also show that the correlations between OTC return factors and their listed counterparts are typically well below 0.5. The correlation between the OTC illiquidity factor and the listed Pastor and Stambaugh's (2003) illiquidity factor is close to zero. These facts show that the OTC factor structure differs significantly from the factor structure of listed stocks, presenting a challenge for explanations of return premiums based on economy-wide risk factors.

Our final tests examine whether theories based on behavioral biases and limits to arbitrage can explain OTC and listed return premiums. Models analyzing the impact of differences in opinions and limits on short sales could apply to both OTC and listed markets. In Appendix D, we present

[4]Studies of listed stocks by Banz (1981), Fama and French (1992), Ang, Hodrick, Xing, and Zhang (2006), and Jegadeesh and Titman (1993) provide early evidence of the size, value, volatility, and momentum premiums, respectively.

[5]Momentum is often thought to be pervasive in that it occurs in many different countries and asset classes (see, for example, Asness, Moskowitz, and Pedersen (2013)).

a model of OTC stock pricing inspired by the theories of Miller (1977), Duffie, Garleanu, and Pedersen (2002), and Scheinkman and Xiong (2003). The key mechanism is that, when investors' opinions diverge, costs of short selling discourage the participation of investors with the most pessimistic views of a stock. This causes overpricing followed by negative risk-adjusted returns. In the model, investors' overconfidence in their preferred valuation signals causes disagreement. Disclosure of financial information reduces differences in opinion by resolving uncertainty over which investors can disagree.

The model predicts that differences in opinion and overpricing are associated with high values of four firm characteristics: trading volume, return volatility, market capitalization, and market-to-book equity ratio (M/B). These relations are stronger for stocks with higher investor overconfidence and those with fewer disclosures. The model's first four predictions are consistent with the evidence that OTC stocks with high volume, volatility, size, and M/B exhibit negative abnormal returns. Importantly, we also find evidence consistent with both sets of the model's predicted interaction effects. Motivated by Barber and Odean (2000) evidence that retail investors are overconfident, we use a stock's institutional ownership as an inverse measure of investor overconfidence. We show that the return premiums for PNT, volume, volatility, value, and size are 1.0% to 4.4% per month larger in OTC stocks that are not held by institutions. We then measure OTC firms' disclosure of book equity data, which is basic financial information relevant for valuation. Empirically, OTC return premiums based on three proxies for disagreement—PNT, volume, and volatility—are 1.4% to 1.6% per month larger among stocks with undisclosed book equity.

Our cross-market findings are also consistent with the idea that our model of overpricing applies more to OTC markets than listed markets. Our evidence indicates that short selling is more difficult in OTC markets; and the lower disclosure and higher proportion of retail clientele in OTC markets suggest investor disagreement could be greater. The fact that the OTC illiquidity premium exceeds the listed premium is consistent with this notion. Moreover, we find that the return on the entire OTC market is actually significantly negative at –0.8% per month, implying widespread overpricing of OTC stocks. This negative return is driven by the OTC stocks with the most trading activity, which likely exhibit the highest investor disagreement.

Although our model of overpricing provides a plausible account of many return premiums, it does not make clear predictions for the momentum premium. We investigate momentum further and find evidence that is most consistent with Hong and Stein's (1999) model based on the gradual diffusion of information across investors. The lack of momentum for most OTC stocks is consistent with the idea that investors do not attend closely to most OTC firms' fundamentals, perhaps because these firms lack credibility. We also find that momentum is strongest among OTC stocks that disclose basic financial information and the largest OTC firms, which presumably have more credibility. Furthermore, momentum among large OTC firms does not exhibit any reversal over five years, which is consistent with Hong and Stein's (1999) model but is hard to reconcile with some alternative models of momentum.

3.2. Related Studies of OTC Stocks

Only a few studies investigate stock pricing in OTC markets.[6] Studies by Luft, Levine, and Larson (2001) and Eraker and Ready (2011) find that the average OTC market return is negative during their sample periods spanning 1995 to 2008. Although we use the OTC market return as a factor in some of our tests, we focus on the cross section of OTC returns.[7] In many cases, the differences among OTC stocks' returns are much larger than the (negative) OTC market premium and are not explained by exposures to the OTC market factor.

Studies of OTC firms' liquidity and disclosure are also relevant. Three studies examine how liquidity changes for stocks moving from listed markets to the OTC markets. Sanger and Peterson (1990) show that quoted bid-ask spreads triple for 57 firms that are delisted and then trade in OTC markets from 1971 to 1985. Harris, Panchapagesan, and Werner (2008) show that volume falls by two-thirds, quoted bid-ask spreads double, and effective spreads triple for 1,098 firms that are delisted from NASDAQ in 1999 to 2002 and subsequently trade on OTC markets. Macey, O'Hara, and Pompilio (2008) also find higher spreads for most of the 58 NYSE stocks moving to OTC markets in 2002. These studies suggest that the shift in trading to OTC venues actually causes stocks to become less liquid.

Leuz, Triantis, and Yue Wang (2008) investigate a firm's decision to "go dark," which means a firm ceases to report to the SEC while continuing to trade publicly in OTC markets. They find that the 480 firms going dark between 1998 and 2004 experience negative average abnormal returns of −10% upon announcement. Our study analyzes the returns of all OTC firms, including those that have chosen to go dark (a minority), those that have never reported to the SEC, and those

[6] Bollen and Christie (2009) examine various aspects of OTC stock microstructure, but do not investigate cross-sectional return premiums.

[7] Luft and Levine (2004) also explore the how OTC stocks' returns are related to their size and liquidity, but they do not perform formal statistical tests presumably because their sample spans only the five years from 1996 to 2000.

that currently report to the SEC. All OTC firms' past disclosure policies and financial reports are available to investors and thus should be reflected in stock prices insofar as they affect investors' valuations.

3.3. OTC Market Data

A. Institutional Details

Our data consist of US common stocks traded in the OTCBB and PS markets from 1977 through 2008. We obtain these data through MarketQA, which is a Thomson Reuters data analytics platform. The OTC markets are regulated by the Financial Industry Regulatory Authority (FINRA), formerly the National Association of Securities Dealers (NASD), and the SEC to enhance market transparency, fairness, and integrity. For most of our sample, the defining requirement of an OTC stock is that at least one FINRA (formerly NASD) member must be willing to act as a market maker for the stock.

As of June 2010, over 211 FINRA firms were market makers in OTC stocks, facilitating daily trading activity of $395 million ($100 billion annualized). The most active firms, such as Archipelago Trading Services and Knight Equity Markets, are also market makers in stocks listed on the NASDAQ and are SEC-registered broker-dealers. FINRA requires market makers to trade at their publicly displayed quotations.

Prior to 2000, the key formal disclosure requirement for firms traded on the OTCBB and PS was Section 12(g) of the Exchange Act. This provision applies only to OTC firms with more than 500 shareholders of record and $10 million in assets. Yet the vast majority of beneficial owners of OTC firms are not shareholders of record as their shares are held in "street name" through their brokers. So even large OTC firms can circumvent this disclosure requirement.

FINRA and SEC regulation of OTC markets, however, has increased substantially since 2000. After June 2000, firms quoted on the OTCBB must have at least 100 shareholders, file annual reports, hold annual shareholder meetings, and meet other governance requirements (Bushee and Leuz, 2005). However, these disclosure requirements do not apply to PS firms, and they did not apply to OTCBB firms for most of our sample.

We later provide evidence suggesting that the majority of investors in the firms traded exclusively on OTC markets are individuals rather than institutions. Individual investors can buy and sell OTCBB and PS stocks through most full service and discount brokers, such as E-Trade, Fidelity, and Schwab. However, short selling OTC stocks is difficult for investors, especially individuals. We collect short selling data for a sample of 50 OTC stocks and 50 similarly-sized listed stocks in June 2012.[8] A retail customer of Fidelity could buy all 100 of these stocks, but the broker would allow short selling in only one of the OTC stocks and eight of the listed stocks. Despite the constraints on individuals, for the 50 listed stocks, short interest as a percentage of floating shares averages 4.1% and exceeds 0.1% for all 50. In contrast, for the 50 OTC stocks, short interest averages just 0.5% and is lower than 0.1% for 28 of the stocks—though it is positive for all but seven stocks. We infer that it is hard for individual investors to short most small stocks; and nearly all investors have difficulty shorting OTC stocks. Thus, the OTC market is a natural place to test theories of limits on short sales.

B. OTCBB and PS Data

We examine the universe of firms incorporated in the US with common stocks that are traded in OTC markets from 1977 through 2008. Our analysis uses only OTC firms without stocks that have been listed on the NYSE, NASDAQ, or Amex exchanges within the last three months. We purposely exclude listed firms to ensure that we are analyzing a set of firms that is as orthogonal as possible to those listed on the traditional venues. MarketQA provides daily trading volume, market capitalization, and closing, bid, and ask prices for these firms.

To ensure adequate data quality, we further restrict the sample to firms meeting the following requirements in the previous month:

- Non-missing data on stock price, market capitalization, and returns
- Stock price exceeds $1
- Market capitalization exceeds $1 million in 2008 dollars
- At least one non-zero daily return
- Positive trading volume—imposed only after 1995 when volume data are reliable.[9]

The price restriction above follows Ince and Porter (2006), who find that errors in computed returns are more likely for firms with prices of less than $1.[10] The market capitalization restriction

[8] These data are available upon request.

[9] Prior to 1995, some OTC firms' volume data is recorded as missing when it is actually zero and vice versa. We set all missing volume to zero prior to 1995 because we find that such firms have low volume when volume data become available. Our results are virtually unchanged if we treat these firms' volume data as missing instead.

[10] In untabulated results, we find that using a minimum price of $0.10 results in similar OTC return premiums.

is designed to eliminate thinly traded and economically unimportant firms that would otherwise dominate equal-weighted portfolios. The non-zero return and positive volume restrictions exclude thinly traded firms that suffer from bid-ask bounce and nonsynchronous trading issues.[11] Our final OTC sample includes an average of 486 firms per month.

C. Comparison to Listed Stocks

We compare our sample of OTC stocks to common stocks listed on the NYSE, NASDAQ, or Amex exchanges using CRSP data. We define three groups of stocks: active, eligible, and comparable. *Active* stocks have at least one non-zero daily return in the past year. *Eligible* stocks meet our data requirements in Section 3.3.B. *Comparable* stocks in the listed sample consist of the $2N$ eligible listed firms with the lowest market capitalizations, where N is the number of listed firms with a market capitalization below the median market capitalization in OTC markets in each month. These listed firms are comparable to OTC firms in terms of size.

Table 3.1 provides a snapshot of summary statistics for the OTC, comparable listed, and eligible listed samples in July of 1997—a typical month of OTC market activity. In this month, the median market capitalization of an OTC stock is $12.9 million, as compared to $36 million for the eligible listed sample. The difference in total market capitalization is much larger ($21.3 billion versus $9.59 trillion) because the largest listed firms are enormous and because there are 12 times fewer OTC stocks (600 OTC stocks versus 7,127 listed stocks). The annualized median OTC trading volume is only 2.2% of the median eligible listed trading volume ($2.3 million versus $101 million, respectively).[12] The aggregate annualized transactions in OTC stocks exceed $8.2 billion, whereas trades in eligible listed stocks exceed $11.4 trillion.

[Insert Table 3.1 here]

By design, the OTC sample is more similar to the comparable listed sample described in the second column of Table 3.1. In particular, the median size is identical in the two samples ($12.9 million). Although median sizes match perfectly, the mean size in the OTC markets is larger ($35.5 million) than that of the comparable listed sample ($12.7 million) because some OTC firms are quite large, as discussed below.[13] In July 1997, the mean of OTC trading volume at $13.7

[11] These filters also minimize the impact of market manipulation on our results. Aggarwal and Wu (2006), Böhme and Holz (2006), and Frieder and Zittrain (2007) show that market manipulation can affect OTC stocks.

[12] Listed trading volume statistics do not adjust for possible double-counting of NASDAQ interdealer trades.

[13] The average fraction of shares floating is reasonably similar for the smaller samples of 50 OTC firms (53% floating) and 50 similarly-sized listed firms (35% floating) in June of 2012.

million is very similar to that of the comparable listed sample at $12.8 million. Although mean volumes match well, the median OTC volume is smaller than that of the comparable listed sample ($2.3 million vs. $6.1 million, respectively), which is not surprising given the thinner OTC market. In summary, the comparable listed sample is a benchmark group that is close in terms of size and trading characteristics to the OTC firms.

Averaging across all months in our sample, the number of firms is 5,228 in the listed sample and is 5,708 in the active listed universe. The averages are 486 in our OTC sample and 3,357 in the active OTC universe. The OTC sample contains fewer firms than the active OTC universe partly because 30% of OTC firms have a stock concurrently listed on the NASDAQ, making them ineligible for the sample.[14] When imposed individually, our sample filters for a non-zero daily return, minimum price of $1, non-missing price, minimum market capitalization of $1 million, and non-missing market capitalization eliminate 28%, 28%, 21%, 19%, and 16% of active OTC firms, respectively. Notably, none of these sample requirements has much impact on the listed sample, which contains 92% of the active firms in CRSP in an average month.

We now compare the size, volume, and number of firms in the OTC and eligible listed samples over time. For this comparison, we transform the size and volume data to minimize the influence of outliers which sometimes reflect data errors. In each month, we compute the difference in the cross-sectional average of the logarithms of size and ($1 plus) volume in the two samples. After taking the difference, we invert the log transform to obtain a ratio that can be interpreted as the OTC characteristic divided by the listed characteristic.

Figure 3.1 summarizes the size, trading volume, and number of firms in the OTC sample as a percentage of the corresponding amounts in the eligible listed sample. The number of firms in the OTC sample averages 10% of the number in the listed sample, though this percentage increased to 24% by the end of 2008. The average firm size and trading volume in the OTC sample are an order of magnitude smaller than they are in the listed sample. The average OTC stock is 11% of the size of the average listed stock. The average OTC stock's volume is just 6% of that of the average listed stock. The relative size of OTC stocks has almost always been higher than their relative volume, consistent with lower liquidity in OTC markets. This gap between relative size and volume widens after 2000, as more illiquid firms are now traded in OTC markets relative to listed markets.[15] The

[14] In untabulated tests, we find that cross-listed OTC and NASDAQ stocks exhibit return premiums much like other listed stocks. The impact of NYSE versus NASDAQ listing choice has been studied in Baruch and Saar (2009) and others. International cross-listing effects have been studied by Baruch, Karolyi, and Lemmon (2007) and others.

[15] As explained in footnote 9, a structural break in OTC volume reporting causes the gap to appear to widen in July 1995. Average OTC volume would be lower prior to July 1995 if volume data on all OTC firms were available.

increase in the number of OTC firms in the late 1990s outpaces the concurrent rise in the number of listed firms. The relative increase in OTC firms after 2003 coincides with the Sarbanes-Oxley Act when many listed firms to chose to "go dark."

[Insert Figure 3.1 here]

Although the typical OTC firm is smaller than most listed firms, there are several large OTC firms that have market capitalizations similar to large listed firms. Table 3.2 lists the firm size and month in which the 10 largest firms in our sample attain their peak size. These firms have market capitalizations measured in billions. The largest firm, Publix Supermarkets, reaches a market capitalization of $88 billion at the end of our sample in December 2008. It would rank 18th in size in the listed sample in that month, which exceeds the median of the top percentile. Several large companies, such as Delphi Corp., trade on PS after delisting from NYSE, NASDAQ, or Amex. We inspect the entire time series of data for all 77 OTC firms with peak sizes exceeding $1 billion. We correct 19 errors arising from an incorrect number of shares outstanding. Such errors apply mainly to the largest of these 77 firms and do not affect their returns. Still, these data errors suggest one should be careful when interpreting OTC size data and value-weighted portfolio returns.

[Insert Table 3.2 here]

In summary, the typical OTC stock is smaller, less liquid, and harder to short than the typical listed stock. However, the largest 10% of OTC stocks are comparable in size to the median-sized listed stock. The number of firms in our OTC sample is substantial, averaging almost 10% of all listed stocks and increasing dramatically after 2000. Thus, although the OTC market is much smaller than the market for listed stocks, the OTC universe is a powerful new venue to test the determinants of return premiums.

3.4. Variable Definitions

This section summarizes the key variables used in our analyses. Our return predictability tests require estimates of stocks' monthly returns and betas. We also measure several firm characteristics known to predict returns in listed stocks, such as size, book-to-market equity, past returns, idiosyncratic volatility, and illiquidity.

We compute a stock's return as the monthly percentage change in MarketQA's "total return index" variable, which is a cumulative stock price that accounts for dividends and splits.[16] We assign a monthly index value based on the last available daily index value. Our sample filters ensure that this value is available within the last month. Our tests use two past return variables: past one-month returns (Ret[-1]) which capture short-term serial correlation and past 12-month returns (Ret[-12,-2]), not including the past month, which capture stock price momentum.

Idiosyncratic volatility is defined relative to the Fama-French (1993) three-factor model, as in Ang, Hodrick, Xing, and Zhang (2006). To estimate a stock's volatility in month t, we use a time-series regression from month t–2 to t–1 of the stock's daily return on the daily market (MKT), size (SMB) and value (HML) factors, as defined in Fama and French (1993). The stock's idiosyncratic volatility ($Volatility$) in month t is the log of the standard deviation of the residuals from its time series regression. We use the same regression procedure as described in Appendix E, except that we apply this to daily rather than monthly observations.

Our analyses use three measures of individual stock liquidity. The main illiquidity measure is the proportion of days with no trading volume (PNT) in each month. The PNT variable measures an investor's ability to trade a stock at all, which is highly relevant in illiquid markets such as the OTC market. This measure more directly measures a lack of trading than Lesmond, Ogden, and Trzcinka (1999) proportion of days with zero returns. The variable $Volume$ is the log of one plus a stock's monthly dollar volume. The variable $Spread$ is the difference between a stock's ask and bid quotes divided by the bid-ask midpoint from the last day when both quotes are available. These other two illiquidity measures capture the amount of trading and the cost of trading in a stock, respectively.

Our return predictability tests use data on firm disclosure, institutional holdings, size, and book-to-market ratios. Firm disclosure ($Disclose$) is a dummy variable that is one if a firm's book equity data is available from either Compustat, Reuters Fundamentals, or Audit Analytics. We define book equity data as available if it appears in a firm's annual report dated between 7 and 19 months ago. Institutional holdings ($InstHold$) is a dummy variable indicating whether a firm's stock appears as a holding of at least one institutional manager or mutual fund that filed Form 13F, N-CSR, or N-Q with the SEC in the past three months, as recorded by Thomson Reuters. Firm $Size$ is the log of the most recently available market capitalization, as computed by

[16] Much like Ince and Porter (2006), we correct firms' returns in cases in which extremely improbable return reversals occur—e.g., a firm's stock price changes from \$57.00 to \$5.70 and back to \$57.00. None of the main results depend on our correction procedure, which is available upon request.

MarketQA. The book-to-market variable (B/M) is the log of the ratio of book-to-market equity. We Winsorize all independent variables at the 5% level to minimize the influence of outliers.

[Insert Table 3.3 here]

Table 3.3 reports summary statistics of returns and variables for OTC stocks and comparable listed stocks in Panels A and B, respectively. The mean monthly return of OTC stocks is slightly negative at –0.04% compared to 0.66% for comparable listed stocks, which is consistent with Luft, Levine, and Larson (2001) and Eraker and Ready (2011). The cross section of monthly OTC returns is also significantly more disperse than listed stocks, with cross-sectional standard deviations of 28.08% and 19.46%, respectively. OTC stocks are substantially more volatile than comparable listed stocks, with average monthly average volatilities of 6.56% and 4.29% for the OTC and listed samples, respectively. The size and book-to-market distributions of firms in the OTC and comparable listed samples are similar.

However, the OTC and listed samples exhibit very different levels of disclosure, institutional ownership, and liquidity. The mean of the *Disclose* dummy for book equity data is 0.60 in the OTC sample and 0.83 in the comparable listed sample, suggesting that 40% of OTC firms choose not to disclose accounting data whereas only 17% of small listed firms omit this information.[17] Table 3.3 shows that an average of 26% of OTC stocks are held by institutions ($InstHold$), as compared to 71% of comparable listed stocks. This suggests that the investor clientele in OTC markets is mainly retail, while institutions play a bigger role in listed markets.

The average of log volume ($Volume$) is much smaller for OTC stocks (8.25) than for listed stocks (10.77). OTC stocks also trade much less frequently: the mean fraction of days with no trading in a month, PNT, is 0.55 for OTC stocks compared to 0.20 for listed stocks. The 95th percentile PNT value is 0.94, implying the least frequently traded OTC stocks trade just one day per month. Average OTC *Spreads* are quite high at 0.15 versus 0.08 for comparable listed stocks. We explicitly account for the impact of the bid-ask bounce bias in OTC stocks' average returns using the Asparouhova, Bessembinder, and Kalcheva (2010) method described below.

Panel C in Table 3.3 shows average cross-sectional correlations among OTC firms' characteristics and their betas on listed return factors. Nearly all of the pairwise correlations are much less

[17] Some of the lack of book equity data reflects incomplete coverage in our data sources. In unreported analyses, we find that our three data sources have significantly overlapping coverage, but no single source subsumes the others.

than 0.5. The exception is the large negative correlation of –0.84 between PNT and $Volume$, which indicates that these two variables reflect a common source of OTC illiquidity.

3.5. Comparing the Cross Sections of OTC and Listed Returns

Following researchers studying listed stocks, we construct calendar-time portfolios of OTC stocks ranked by characteristics to estimate the expected returns of OTC factors. We compare OTC factor returns to those in the comparable listed and eligible listed samples. Forming factors has the advantage that the means of the portfolios have economic interpretations as return premiums. These portfolio tests also do not require linearity assumptions imposed by regressions. The disadvantages of portfolios are that confounding effects can obfuscate return premiums based on univariate sorts and they lead to less powerful tests. Accordingly, we also present cross-sectional regressions below in which we jointly estimate return premiums. Our analysis focuses on portfolios ranked by two illiquidity measures, PNT and $Volume$. We also estimate the returns of factor portfolios ranked by size, value, volatility, and momentum.

To construct portfolios, we sort firms into quintiles at the end of each month based on the firm characteristic of interest, such as a firm's PNT value in that month. A long-only quintile portfolio return in month t is the weighted average of returns in month t of firms in the quintile, as ranked by their characteristics in month t–1 among sample firms. A long-short factor portfolio return is the difference between the returns of the top and bottom quintile portfolios. The portfolios use three sets of weights: equal-weighted (EW), value-weighted (VW), and weighted by the prior month's gross return (gross-return weighted or GRW). Asparouhova, Bessembinder, and Kalcheva (2010) show that the expected return of a GRW portfolio is the same as that of an EW portfolio, except that it corrects for the bid-ask bounce bias noted by Blume and Stambaugh (1983).[18] A long-only portfolio's excess return is its monthly return minus the monthly risk-free rate prevailing at the end of the prior month. Each factor portfolio's alpha is the intercept from a time-series regression of its monthly returns on various monthly factor returns. All standard errors are based on the robust estimator in Newey and West (1987).[19]

[18] In unreported tests, we simulate OTC stock returns in the presence of empirically realistic bid-ask bounce and non-trading, as well as persistent 50% errors in recorded prices that occur with 5% probability. For portfolios sorted by PNT values, we find that the bias in observed monthly GRW portfolio returns is always less than 0.85%, and adjusting for the bias would only strengthen our main results.

[19] We follow Newey and West (1994) recommendation to set the number of lags equal to the highest integer less

To measure factor loadings in portfolios that may be infrequently traded, we include six monthly lags of each factor and report the sum of the contemporaneous and six lagged coefficients as the factor loading.[20] We analyze five factors based on listed returns, including the MKT, SMB, HML, momentum (UMD), and illiquidity (ILQ) factors. We define UMD using Carhart's 12-month momentum measure (1997) and ILQ using Pastor and Stambaugh's (2003) volume-induced reversal measure. We create a sixth factor equal to the value-weighted OTC market return minus the standard (30-day Treasury Bill) risk-free rate, which we refer to as "OTC MktVW." Our three return benchmarks are the OTC CAPM, Listed CAPM, and the Listed Five-Factor models. The OTC CAPM and Listed CAPM models include only the OTC market and listed market factors, respectively. The Listed Five-Factor model consists of the MKT, SMB, HML, UMD, and ILQ factors.

We summarize the return premiums for each OTC factor in Table 3.4. Panel A shows the Sharpe ratios of each OTC and listed factor and their information ratios (alphas divided by idiosyncratic volatilities) relative to the factor model benchmarks. Panel B displays the average monthly returns and alphas of each OTC factor relative to the factor model benchmarks. Panel C shows the listed factor loadings of OTC factors. Panels D and E report the analyses of Panels B and C for comparable listed stocks. The returns in Table 3.4 do not include trading costs, and we use them to test theories' predictions of pre-cost returns.

[Insert Table 3.4 here]

Table 3.4 shows three interesting comparisons between factor premiums in OTC markets and those in comparable listed markets: (1) the illiquidity return premium is much larger in OTC markets; (2) the size, value, and volatility premiums are similar in OTC and listed markets;[21] and (3) the momentum premium is much smaller in OTC markets.

A. Liquidity Premiums

The first four rows of Table 3.4, Panel A report the illiquidity premiums. The raw Sharpe Ratios of the OTC illiquidity factors based on PNT and $Volume$ are both large at 0.91 and –0.90, respectively. Both PNT, which captures whether investors trade, and $Volume$, which quantifies

than $4 * (T/100)^{(2/9)}$, where T is the number of periods in the sample. For our sample of 383 months, applying this formula results in a lag length of 5 months.

[20] Our method is the monthly analog to the one proposed by Dimson (1979), who analyzes stocks that are infrequently traded at the daily frequency.

[21] All OTC and listed value portfolios exclude firms with negative book equity.

how much they trade, appear to be relevant aspects of liquidity for OTC stocks. The average returns of the value-weighted PNT factor (PNT_{VW}) are also highly positive and significant. They are lower than the GRW returns partly because size-based weightings place the lowest weights on the least liquid stocks, which have the highest returns.[22]

In contrast to the large OTC premiums based on the PNT and $Volume$ measures of illiquidity, the listed premiums based on these measures are tiny and insignificant. For comparable and eligible listed stocks, the Sharpe ratios and information ratios based on either liquidity measure are 0.30 or lower and are statistically insignificant. Our analysis of illiquidity premiums complements the results from numerous studies of listed US and international stocks, including Amihud and Mendelson (1986), Lee and Swaminathan (2000), Pástor and Stambaugh (2003), Bekaert, Harvey, and Lundblad (2007), and Hasbrouck (2009). These studies show that the least liquid listed stocks have higher returns than the most liquid listed stocks, though the magnitude of the listed illiquidity premium depends on the liquidity measure and time horizon. In particular, listed illiquidity premiums constructed by sorting on price impact rather than volume measures could differ from those examined here.

Neither the Listed CAPM nor the Listed Five-Factor model, which includes the illiquidity (ILQ) factor of Pástor and Stambaugh (2003), can explain the OTC PNT and $Volume$ illiquidity premiums. In fact, the OTC PNT factor's information ratio of 1.34 with respect to the Listed Five-Factor model is larger than its Sharpe ratio of 0.91. The OTC illiquidity premiums become larger after controlling for listed risk factors mainly because the OTC illiquidity factors are negatively correlated with the listed market and SMB factors. Panel C of Table 3.4 shows that the OTC PNT factor has negative market and SMB betas of −1.24 and −1.02, respectively, and an insignificant ILQ beta. The very negative beta on the market and SMB factors and the insignificant ILQ beta pose a serious challenge for theories in which the OTC illiquidity premium represents compensation for bearing systematic risk as measured by listed factors.

Next we test whether asset pricing theories that emphasize transaction costs, such as Amihud and Mendelson (1986) and Constantinides (1986), can account for the OTC illiquidity premium. In such theories, prices adjust until investors' *post*-cost risk-adjusted expected returns are equal across assets and equal to the risk-free rate, assuming one can costlessly trade the risk-free asset. This implies that all risky portfolios' *pre*-cost alphas should be positive by an amount reflecting

[22] In general, we do not focus on the value-weighted returns of OTC portfolios because these results are sensitive to interactions between the large OTC size premium and the other factor premiums. Panel A of Table 3.5 in the following section reports how each return premium varies with firm size.

the cost of trading risky assets, where cost is equal to bid-ask spread times the average investor's turnover. We test this hypothesis in Table 3.5 for OTC and listed portfolios sorted by illiquidity measures. In each month, we either sort stocks into PNT deciles (Panel A), or into 10 bid-ask spread ranges (Panel B), using increments of 2.5% from 0% to 25%. Because these finely partitioned sorts result in portfolios with fewer than 10 firms in the early years when liquidity data are limited, Table 3.5 only includes data from August 1995 through December 2008.

[Insert Table 3.5 here]

The results in Table 3.5 are inconsistent with several implications of trading cost theories. First and foremost, the pre-cost CAPM alphas of the OTC stocks in all but one of the bottom four (eight) deciles of PNT (*Spread*) are significantly negative, implying that their post-cost alphas must be even more negative. The OTC stocks with the lowest PNT values have especially negative pre-cost alphas of –3.98% per month, whereas the comparable listed stocks with the lowest PNT values have roughly zero pre-cost alphas of –0.06%. Both groups of low PNT stocks have similar turnover and the OTC stocks actually have higher bid-ask spreads (6.3% versus 4.6%). Thus, a transaction cost theory would predict that the OTC stocks should have higher returns, rather than returns that are 3.92% lower; and it would not predict negative risk-adjusted returns for any group of stocks.

Moreover, the magnitudes of trading costs incurred by OTC investors are small relative to the pre-cost return premiums in Table 3.4. In Constantinides' (1986) model, an asset's illiquidity premium is equal to the representative investor's one-way trading cost, which is the asset's turnover multiplied by half of its bid-ask spread. The last two columns in Table 3.5 report *twice* this amount and show that the round-trip costs range from 0.14% for the highest PNT stocks to 1.30% for the lowest PNT stocks. These magnitudes are much smaller than the top minus bottom decile PNT premium of 5.34% (3.98 – (–1.36)). Furthermore, because equilibrium trading costs *decrease* with PNT, subtracting trading costs from returns would increase the magnitude of the PNT premium. Unreported tests show the same point applies to the $Volume$ premium and five of the other six premiums reported in Table 3.4. OTC investors incur higher trading costs in low PNT and high $Volume$ OTC stocks because they trade these stocks more by definition, which more than offsets the lower average spreads associated with these stocks. This is an important difference between liquidity measures based on volume versus price impact, such as bid-ask spread. Although OTC investors trade low *Spread* stocks more often, they incur lower costs in such stocks (see Panel B) because of their low spreads.

We also test the unique predictions of Amihud and Mendelson's (1986) model, which assumes heterogeneous investors with exogenously specified horizons. This theory predicts that the risk-adjusted returns of portfolios sorted by bid-ask spreads will be *increasing* and weakly *concave*. Intuitively, the marginal compensation for illiquidity diminishes with bid-ask spreads because investors with longer horizons choose to hold illiquid stocks in equilibrium, and they require less additional compensation per unit increase in spread than short-horizon investors. We formally test for monotonicity and concavity by constructing long-short portfolios based on the 10 spread-sorted portfolios in Panel B. The monotonicity portfolio puts increasing weights of (–5, –4, –3, –2, –1, 1, 2, 3, 4, 5) / 15 on the 10 spread portfolios, while the concavity portfolio applies initially increasing and then decreasing weights of (–2, –1, 0, 1, 2, 2, 1, 0, –1, –2) / 3. The concavity portfolio represents the difference between two long-short illiquidity factors formed within spread ranges of [0%, 12.5%] and [12.5%, 25%]. Its expected return is zero if the return-spread relation is linear, positive if it is concave, and negative if it is convex.

The results from the monotonicity and concavity tests are ostensibly inconsistent with the implications of trading cost theories. The monthly alpha of the monotonicity portfolio based on spread sorts is only slightly positive (0.54%) and is statistically insignificant. The monthly alpha of a monotonicity portfolio formed from PNT sorts in Panel A is significantly higher at 3.75%. Furthermore, the concavity portfolio based on spread sorts exhibits a significantly *negative* alpha of 2.63% per month, meaning that the spread-return relation is actually convex, not concave.

The results in Table 3.5 are also inconsistent with the hypothesis that data errors and microstructure biases, such as bid-ask bounce, explain the OTC illiquidity premium. Both panels demonstrate that the negative alphas of liquid OTC stocks are the primary driving force behind the observed illiquidity premium. These negative alphas are unlikely to be spurious because errors and microstructure biases are smaller among liquid stocks and typically produce an *upward* bias, implying that the liquid OTC stocks' true alphas may be even more negative.

In unreported tests, we investigate whether the OTC illiquidity premium is driven by survivorship bias. As we show in Table 3.7 below, the annual return of a PNT factor portfolio with a 12-month holding period is 32.9% (12 * 2.74%). For the top and bottom PNT decile portfolios, 12-month returns are missing for 15.5% and 16.5% of firms during the post-formation period. The similarity in these 12-month disappearance rates suggests survivorship bias does not explain the OTC illiquidity premium. Furthermore, the annual return of the 12-month PNT factor portfolio of 32.9% is twice as high as the 16% disappearance rates above. Thus, even an enormous return differential of –50% between the disappearing high and low PNT firms would explain only one quarter (–50% * 16% / 32.9% = 24.3%) of the OTC illiquidity premium.

B. Size and Value Premiums

Table 3.4 shows that the size, value, and volatility premiums found in listed markets also exist in OTC markets and have similar magnitudes. Panel A indicates that the annual Sharpe ratios of the GRW size and value factors in the OTC market are –1.02 and 0.82, respectively, as compared to –0.98 and 1.19 in the comparable listed sample. This evidence demonstrates that the size and value premiums are robust to the differences across OTC and listed markets.

While the magnitudes of these premiums are similar, neither the listed size nor the listed value factor explains much of the variation in the OTC size and value factors. In Panel B, the monthly alpha of the OTC size factor is –2.81% after controlling for its loading on the listed size factor and the other four listed factors. These listed factors explain just 8.1% of the variance in the OTC size factor, as reported in the R^2 columns in Panel C. Even after controlling for the five listed factors, the monthly alpha of the OTC value factor is still 2.29%. Although the loading on the listed value (HML) factor is positive, all five listed factors explain just 25.3% of the variance in the OTC value factor. Hence there are independent size and value factors in the OTC market that are not captured by listed factors.

C. Volatility Premium

Panel A in Table 3.4 shows that OTC stocks with high volatility have lower average returns than those with low volatility. The Sharpe ratio of the OTC volatility factor at –0.55 is close to the corresponding listed Sharpe ratios at –0.75 and –0.64. Panel B shows that the alpha of the OTC volatility factor with respect to the listed CAPM is significantly negative at –2.63% per month. At first glance, OTC stocks with high idiosyncratic volatility seem to exhibit low returns just like listed stocks with high idiosyncratic volatility.

Interestingly, the OTC volatility factor's negative alpha is much smaller in the OTC CAPM regression. The OTC market itself has an overall negative return: Panel A of Table 3.4 reports that the Sharpe ratio of the OTC market is –0.52. The fact that there is no idiosyncratic volatility effect in OTC markets after controlling for the OTC market factor implies that a single root cause could explain both the low return of the OTC market and the low returns of highly volatile OTC stocks. Panel C shows that the OTC market beta of the long-short OTC volatility factor is 1.07 and that exposure to the OTC market explains 15.5% of the variance in the volatility factor. Panel C of Table 3.4 also indicates that the OTC volatility factor has a negative loading of –1.38 on the listed illiquidity factor, implying that the volatility effect in OTC stocks is related to the modest illiquidity premium in listed stocks.

D. Momentum

The third key result is that the return premium for momentum in OTC markets is surprisingly small. Whereas the Sharpe ratio of 1.56 for listed momentum is the largest among all the comparable listed premiums in Table 3.4, Panel A, the Sharpe ratio of 0.41 for OTC momentum is the smallest of the OTC premiums. Panel E in Table 3.4 shows that the OTC and listed momentum factors are significantly positively correlated.[23] This explains why the information ratio of the OTC momentum factor against the Listed Five-Factor model, which includes listed momentum, is close to zero at 0.09.

The OTC momentum premium shown in Table 3.4 is much smaller than the momentum premium in listed stocks reported in Jegadeesh and Titman (1993) and the high Sharpe ratio of 1.30 for momentum in the eligible listed universe. The average OTC momentum premium has the same sign as the listed premium, but the magnitude of the OTC premium is at least three times smaller, depending on the exact specification. This evidence contrasts with the robust evidence that illiquidity, size, value, and volatility premiums exist in the OTC markets. Only the OTC illiquidity premium is significantly larger than its listed counterpart.

E. OTC Market Returns

The last rows in Panels A to C of Table 3.4 report time-series regressions using the excess return on the value-weighted OTC market as the dependent variable. The alpha of the OTC market is negative, regardless of which listed factor model is used (also see Eraker and Ready (2011)). In addition, the listed CAPM explains only 43.5% of the variation in the OTC market, while the five-factor model explains 57.3% and leaves 42.7% unexplained. This is broadly consistent with the inability of the other systematic listed factors to explain much of the variation in the OTC

[23] Like the listed momentum factor, the OTC momentum factor exhibits statistically and economically significantly lower returns in January: its January Sharpe ratio is –0.89 versus a non-January Sharpe ratio of 0.54.

size, value, momentum, illiquidity, and volatility factors.

Motivated by the differences in volatility and liquidity between OTC and listed stocks in Table 3.3, we explore the empirical relationship between the OTC market premium and the OTC volatility and illiquidity premiums. In an untabulated regression, we find that the OTC market has highly significant loadings on the OTC volatility and PNT factors with t-statistics of 3.85 and –5.98, respectively. Moreover, after controlling for these two factors, the OTC market's alpha changes from –0.74% to 0.01% (i.e., near zero). This regression establishes strong links between the OTC volatility and illiquidity premiums and the negative OTC market premium.

F. Multivariate Cross-sectional Regressions

We also estimate return premiums using monthly multivariate linear regressions that simultaneously control for firms' betas and characteristics. Table 3.6 reports Fama and MacBeth (1973) return predictability coefficients, along with Newey and West (1987) standard errors in parentheses. The point estimate is the weighted-average of monthly coefficients, where each coefficient's weight is the inverse of its squared monthly standard error as in Ferson and Harvey (1999). As before, we use GRW returns to correct for bid-ask bounce bias. We group regressors into firms' betas on the MKT, SMB, HML, and UMD factors and firms' characteristics based on size, book-to-market equity, volatility, past returns, and illiquidity.[24] Regressions I, II, and III include only betas, only characteristics, and both betas and characteristics, respectively. In Appendix E, we explain how we estimate firms' betas and adjust them to account for nonsynchronous trading. The three sets of columns in Table 3.6 represent estimates of return premiums in the OTC, comparable listed, and eligible listed samples.

[Insert Table 3.6 here]

There are two main findings from Table 3.6. First, firms' betas do not strongly predict returns in any of the three samples, especially in Regression III which includes both firms' betas and characteristics. This echoes Daniel and Titman (1997) findings in listed stock markets. Although using estimated betas as regressors induces an attenuation bias in the coefficients on betas, this bias cannot explain why half of the beta coefficients are negative and statistically significant in Regression I. Furthermore, controlling for firms' betas has virtually no impact on the coefficients on firms'

[24]Regression specifications I and II also include an unreported dummy variable for firms with missing or negative book equity variable to keep these firms in the sample without affecting the coefficient on book-to-market equity.

characteristics, which are nearly identical in Regressions II and III. The weak predictability from betas indicates that most of the predictive power in the cross section comes from characteristics, and supports our use of characteristics in constructing portfolios.

Second, with few exceptions, jointly estimating return premiums on firms' betas and characteristics results in premiums that are quite similar to those using portfolio methods. For example, the PNT coefficient in the OTC sample in Regression III is 4.053, which implies a 3.36% per month ($4.053 \cdot (0.08 - 0.91)$) difference in returns between firms ranked at the medians of the top and bottom quintiles of PNT (0.08 and 0.91, respectively). This magnitude closely matches the top-to-bottom quintile difference in the GRW returns of PNT portfolios of 2.92% per month in Table 3.4.B. The same qualitative result applies to the other return premiums. These findings in Table 3.6 show that none of the return premiums estimated using univariate portfolio sorts in Table 3.4 is due to the correlations among firm characteristics. This makes sense in light of the low cross-correlations among the variables reported in Table 3.3.C. Consequently, we focus on portfolio tests in the rest of the paper.

3.6. Testing Theories of Limited Arbitrage and Behavioral Biases

We exploit the differences between the OTC and listed markets as well as within-market heterogeneity on several dimensions to test asset pricing theories based on limits to arbitrage and behavioral biases. Our main strategy is to contrast return premiums in subsamples of OTC and listed stocks, and we use additional tests to shed further light on the momentum premium.

A. Trading Costs as a Limit to Arbitrage

We first test whether trading costs limit the extent to which arbitrageurs can exploit the pre-cost returns of OTC factors in Table 3.4. We estimate the post-cost returns of an arbitrageur who takes positions in each of the OTC factors, assuming that the investor pays each stock's bid-ask spread in every round-trip trade. Studies such as Frazzini, Israel, and Moskowitz (2012) show that spread data overstate the trading costs incurred by arbitrageurs who use sophisticated strategies to minimize costs. Our post-cost return calculation is more relevant for the average investor in OTC markets.

We compute post-cost returns at rebalancing frequencies between 1 and 24 months to evaluate how arbitrageurs' profitability depends on their portfolio turnover. We rebalance portfolios at n-month frequencies using the Jegadeesh and Titman (1993) method in which $1/n$ of the firms in each portfolio can change in each month based on rankings of firms' characteristics in the prior month. As before, we focus on portfolios with GRW weights, which remain gross-return weighted in the absence of rebalancing. We also analyze VW and liquidity-weighted (LW) portfolios to assess whether arbitrageurs lower their trading costs by concentrating on large and liquid stocks. The LW weights are inversely proportional to stocks' bid-ask spreads.[25]

Table 3.7 reports pre-cost and post-cost returns of GRW portfolios and breakeven rebalancing frequencies and spreads for the post-cost factor portfolios. The breakeven frequency (spread) is the rebalancing frequency (bid-ask spread) at which the post-cost return of the factor portfolio is closest to 0%. Table 3.7 reports the pre-cost returns, post-cost returns, and breakeven spreads of the GRW OTC factors with rebalancing frequencies of 1 and 12 months. We complement the table with Figures 3.3A and 3.3B, which show the GRW OTC factors' pre-cost returns and post-cost returns at rebalancing frequencies ranging from 1 to 24 months.

[Insert Table 3.7 here]

[Insert Figures 3.3A and 3.3B here]

The main finding in Table 3.7 is that the post-cost returns for arbitrageurs trying to exploit the OTC factors are much lower than the factors' pre-cost returns. Even at the annual rebalancing frequency, the post-cost GRW returns of all six OTC factors are less than 1% per month and are not statistically significantly greater than 0%—in contrast to the pre-cost returns that are as high as 2.74% per month and almost always statistically significant. Only the PNT, $Volume$, and $Value$ factors exhibit positive post-cost GRW returns at the annual frequency, which is why the GRW breakeven horizons of these factors are less than one year. If an arbitrageur uses VW or LW strategies, the breakeven horizons decline for these three strategies and the breakeven horizon for the $Size$ factor decreases to less than one year. However, one cannot profitably exploit the OTC $Momentum$ and $Volatility$ factors with a one-year rebalancing frequency, regardless of which weighting scheme one uses.

[25] Because limited spread data are available, we compute post-cost returns only in the second half of the sample (1993 to 2008) and estimate costs based on average portfolio turnover multiplied by average bid-ask spreads.

Figures 3.3A and 3.3B show that pre-cost GRW factor returns monotonically decrease with frequency presumably because the information used to form the portfolios gradually becomes outdated at longer frequencies. Despite this effect, the post-cost factor returns steadily increase with frequency because the longer frequency portfolios have much lower trading costs. At the 24-month frequency, the post-cost returns of the PNT and $Value$ factors exceed 1% per month, but only the $Value$ factor return is statistically significant at the 5% level.

The breakeven spread columns in Table 3.7 indicate that effective bid-ask spreads must be quite high—the average across the six factors is 12.3%—in order to deter arbitrage at the one-year rebalancing frequency. However, because the median OTC spread in Table 3.3.B is 10%, it seems that OTC trading costs are indeed high enough to limit the effectiveness of arbitrage, especially when one also considers the limits on short selling in OTC markets noted earlier. Such limits help explain why these large OTC return premiums persist, but one needs a model of investor behavior - such as the one in Appendix D - to understand why premiums arise in the first place. We now turn to tests that allow us to distinguish among theories of limited arbitrage.

B. Evidence from Double Sorts

We measure return premiums within each market in subsamples of stocks sorted by characteristics that distinguish OTC and listed markets: institutional holdings, disclosure, and size. We select these three characteristics to construct powerful tests of competing theories of return premiums. We form double-sorted portfolios by first ranking stocks based on a distinguishing characteristic in month t–1 and sorting them into portfolios with sufficiently many stocks. In these initial sorts, we use two portfolios when sorting on the two binary variables ($InstHold$ and $Disclose$), and three portfolios when sorting on size. Within each of these portfolios, such as stocks not held by institutions, we sort stocks into terciles based on the characteristics, such as liquidity, used in constructing factors. Holding each distinguishing characteristic (e.g., institutional holdings) constant, we measure return premiums (e.g., illiquidity) as the difference between returns in month t of stocks in the top and bottom terciles from the second sort. Our method also allows us to test whether the distinguishing characteristic is priced within each tercile from the second sort.

Table 3.8 shows the excess returns from these double-sorted portfolios. Panel A shows that the return premiums for illiquidity (both PNT and $Volume$) and size are much larger within OTC stocks that are not held by institutions. Panel B indicates that both illiquidity premiums and the volatility premium are roughly twice as large among OTC stocks that do not disclose book equity. Panel C shows the OTC illiquidity premium is larger among small stocks, while the OTC momentum premium is four times larger among big stocks. Twelve of the 13 statistically significant differences in return premiums in Table 3.8 exhibit the same signs in the OTC and comparable listed samples, though the magnitudes are often smaller in the listed sample. We now discuss the implications of these results and others for theories of return premiums.

[Insert Table 3.8 here]

C. Testing Theories of Investor Disagreement and Limits on Short Sales

We test Miller's (1977) hypothesis that investor disagreement combined with limits on short sales leads to overpricing and negative abnormal returns. As we show in Appendix D, this theory can help explain the illiquidity, size, volatility, and value premiums in OTC and listed markets because these characteristics are natural proxies for investor disagreement. In particular, both of our OTC illiquidity measures are based on trading volume, which is directly linked to investor disagreement as formalized in Propositions 1 and 2 in Appendix D.

There are several additional testable implications of this theory. If retail (institutional) investors are more (less) likely to disagree, stocks not held by institutions should exhibit higher return premiums based on proxies for disagreement. A complementary story is that a lack of institutional ownership could be a proxy for limits on short sales, as suggested by Nagel (2005), which are associated with larger overpricing in Miller's (1977) theory. Consistent with both interpretations, Panel A in Table 3.8 shows that the return premiums for illiquidity (both PNT and $Volume$ measures), volatility, value, and size are 0.96% to 4.39% per month larger in OTC stocks that are not held by institutions. The differences in the illiquidity and size premiums are especially large and highly statistically significant. Hinting at a role for limits on short sales, the premiums among non-held stocks arise mainly from the extremely negative returns of stocks with high liquidity, size, volatility, and valuation. There are also significant differences in the illiquidity (PNT and $Volume$) premiums between stocks held and not held by institutions in the comparable listed sample, suggesting similar mechanisms could operate in listed markets.

In the model in Appendix D, the impact of differences in opinion is especially strong among OTC stocks that do not disclose basic financial information. Investors are likely to hold widely divergent views about the financial condition of firms without disclosures, implying overpricing of such firms' stocks will be more severe. Consistent with this idea, Panel B in Table 3.8 shows that the return premiums based on four proxies for disagreement—PNT, volume, volatility, and size—are 1.38% to 1.64% per month larger among OTC stocks that do not disclose book equity. The differences in all premiums except for size are significant at the 5% level. The difference in size premiums is significant at the 10% level.

We further test disagreement theories by analyzing whether disclosure itself can predict returns. If the disclosure of financial information helps to resolve investor disagreement, as predicted by the model in Appendix D, disclosing firms will earn higher returns than non-disclosing firms.[26] We look for a disclosure premium within firms in the top terciles of liquidity and volatility, where disagreement could significantly affect investors' valuations. Panel B of Table 3.8 shows that disclosing firms do exhibit higher returns than non-disclosing firms, especially among liquid and volatile firms. The disclosure premium is 1.52% [= –1.04 – (–2.56)], 1.78%, and 1.37% per month, respectively, when evaluated within the PNT, volume, and volatility terciles representing the most liquid and volatile firms. All three premiums are statistically significant, economically large, and in line with the theory in Appendix D.

Furthermore, the negative market returns on OTC stocks are consistent with the overpricing argument. Investor disagreement can cause overpricing of the entire market when there are market-wide limits on short sales (e.g., Jarrow (1980)). Because few OTC stocks can be shorted and there is no tradable index of OTC stocks that can be shorted, limits on short sales plausibly apply to the OTC market as a whole. Thus, disagreement combined with limits on short sales could explain the negative returns of the OTC market. It could also help explain the strong empirical links between the OTC market premium and the OTC premiums for illiquidity and volatility, which could all stem from the same underlying investor disagreement.

Lastly, Miller's (1977) theory could help explain why the coefficients on market beta are negative and statistically significant in predicting returns in Table 3.6. He argues that "the riskiest stocks are also those about which there is the greatest divergence of opinion." If so, in the presence of limits on short sales, stocks with the highest systematic risk (i.e., beta) could become so overpriced that they exhibit lower future returns than stocks with low risk.

[26] Hirshleifer and Teoh (2003) develop a theory of attention that makes a similar prediction. Firms can choose whether to disclose financial information to investors with limited attention. In equilibrium, firms do not disclose if they have negative news, knowing that investors fail to take this self-selection into account. This theory predicts that investors overprice firms that do not disclose, implying that these firms have lower returns than disclosing firms.

D. Testing Theories of Momentum

Firms traded in OTC markets disclose much less information than those in listed markets, and retail investors dominate in OTC markets. This suggests that theories emphasizing how investors react to information and the role of institutions could explain the relatively small OTC momentum premium. This section presents evidence that is most consistent with Hong and Stein's (1999) model of momentum based on the gradual diffusion of information.

Two elements in Hong and Stein's (1999) model are necessary for momentum. First, there must be a group of "newswatcher" investors who only attend to firms' fundamentals and disregard firms' stock price movements. Such newswatchers may not follow many OTC firms. Greenstone, Oyer, and Vissing-Jorgensen (2006) argue that investors view financial information disclosed by most OTC firms as less credible than information from listed firms. In contrast, OTC firms' stock prices are reliable, verifiable, and widely available. If OTC stocks lack newswatchers, they would not exhibit momentum. This argument is consistent with the evidence in Tables 3.4 and 3.5 showing that OTC momentum is on average lower than listed momentum.

The second key element in Hong and Stein's (1999) model is the gradual transmission of information across newswatchers. The model predicts that momentum is stronger and longer-lasting when information transmission is slower. Because fewer investors hold and discuss OTC stocks, information transmission is likely to be slower in OTC stocks than in listed stocks. Under this reasoning, momentum should be strong and long-lasting among OTC stocks that newswatchers might follow, such as large OTC firms and those that disclose key financial information. Consistent with this prediction, Panels B and C of Table 3.8 shows that momentum is two to four times higher among OTC stocks that newswatchers might follow. Specifically, momentum is 1.78% and 1.55% per month, and highly statistically significant, among the largest OTC firms and those that disclose book equity, respectively, while it is only 0.41% and 0.61%, and insignificant, among the smallest OTC firms and those that do not disclose book equity.

Next we examine the time horizon of momentum in OTC markets. We construct long-short momentum portfolios at various horizons using the Jegadeesh and Titman (1993) method, similar to the rebalanced portfolios examined in Table 3.7.[27] Table 3.9 reports the momentum portfolios' GRW and VW returns at horizons up to five years. There is no momentum (–0.08% per month)

[27] This procedure entails two steps. First, we form top and bottom quintile portfolios based on stocks' Ret[12,-2] as of month t–k. Second, to measure returns n years after portfolio formation in each month t, we apply GRW weights to the 12 monthly returns of the extreme quintile portfolios formed in months $t – n * 12$ to $t – n * 12 – 11$. The average difference in the extreme quintile portfolios' returns is the momentum premium at the n-year horizon.

at the one-year horizon in OTC markets using the GRW method. There is, however, significant one-year momentum (1.57% per month) in the VW OTC portfolios, but this places extremely large weights on a few big OTC firms.

[Insert Table 3.9 here]

Analyzing the long-term returns of momentum portfolios in OTC and listed markets helps us differentiate theories of momentum. In the models of Hong and Stein (1999) and Barberis, Shleifer, and Vishny (1998), momentum originates from investors' underreaction to tangible firm-specific information, such as news about firm earnings, and thus momentum need not reverse.[28] In contrast, in Daniel, Hirshleifer, and Subrahmanyam (1998) theory, momentum arises from "continuing overreaction" to intangible information, implying that momentum eventually reverses. Table 3.9 shows that VW momentum portfolios in OTC markets exhibit *positive* but statistically insignificant returns of 0.45% per month in years two through five after portfolio formation. In addition, momentum in listed markets exhibits limited reversal in the eligible sample and no reversal in the comparable-size sample in years two through five.[29] The observed lack of reversal lends support to the two underreaction theories of momentum: Hong and Stein (1999) and Barberis, Shleifer, and Vishny (1998).

An alternative explanation for the weak GRW momentum premium in OTC markets is the small role of institutional investors in OTC markets. In listed stock markets, institutions herd (e.g., Nofsinger and Sias, 1999; Sias, 2004) and institutions follow momentum strategies (e.g., Badrinath and Wahal (2002); Griffin, Harris, and Topaloglu (2003)). Gutierrez and Prinsky (2007) and Vayanos and Woolley (2013) argue that momentum in listed markets partly arises because of agency issues in these delegated institutional money managers. Our cross-market evidence is broadly consistent with this view. Table 3.4 shows that momentum is three times higher among comparable listed stocks, which are far more likely to be held by institutions (see Table 3.3).

However, the evidence within the OTC market is ostensibly inconsistent with the theory that institutions per se cause momentum. Panel A in Table 3.8 shows that OTC stocks experience nearly identical momentum (1.97% versus 2.18% per month) whether or not they are held by institutions. Nevertheless, the types of institutions likely differ across OTC and listed markets.

[28] Because we lack earnings data for OTC firms, we cannot test several predictions of the Barberis, Shleifer, and Vishny (1998) model, which is based on a representative investor's underreaction and overreaction to sequences of news. However, Loh and Warachka (2012) argue that listed stock price reactions to sequences of earnings surprises are inconsistent with this model.

[29] Lee and Swaminathan (2000) and Jegadeesh and Titman (2001) show that momentum in listed stocks partially reverses in their samples.

Large asset managers that are subject to the delegated agency problems described by Vayanos and Woolley (2013) play important roles in listed markets. Table 3.3 shows that few large institutions invest in OTC stocks. However, small hedge funds without reporting obligations could significantly affect OTC market prices. These smaller institutions may not be subject to the same agency issues as the largest institutions. Future theories on institutional investors and momentum should account for the different roles played by these various types of investors.

3.7. Concluding Discussion

While many cross-sectional return premiums in listed markets, such as size, value, and volatility, generalize to OTC markets, other return premiums are strikingly different. The premium for illiquidity in OTC markets is several times larger than in listed markets. The pronounced momentum effect in listed markets is economically small in OTC markets. Listed return factors cannot explain the majority of the variation in OTC return factors.

Variation in the illiquidity, size, value, and volatility premiums within OTC markets is consistent with theories in which disagreement and limits on short sales cause temporary overpricing. Variation in the momentum premium within OTC markets is most consistent with Hong and Stein's (1999) theory based on the gradual diffusion of information. We test and find only limited support for several alternative explanations of these premiums, including theories based on exposures to systematic factor risk and those based on transaction costs.

The return premiums in OTC markets offer insights into the future of listed markets. For example, the finding that size, value, and volatility premiums exist in OTC markets provides new evidence that these premiums are robust to differences in market structure and liquidity, and therefore could persist in the future. The finding that the most actively traded OTC stocks appear to be overpriced could also have an important counterpart in listed markets: Ofek and Richardson (2003), Baker and Stein (2004), and others show that apparent speculative bubbles are often associated with high trading volume. Our evidence suggests that such bubbles are magnified when investors must price assets in the dark, and thus improved financial disclosures could mitigate future bubbles.

Figure 3.1: OTC Sample Characteristics as a Percentage of Listed Sample Characteristics

For each month, we plot the average size, average trading volume, and number of stocks in the OTC sample as a percentage of the corresponding statistics in the eligible listed sample. To minimize the influence of outliers and possible data errors, we transform the size and volume data for this comparison. In each month, we compute the difference in the cross-sectional average of the logarithms of size and ($1 plus) volume in the two samples. Then we invert the log transform to obtain a ratio that can be interpreted as the OTC characteristic divided by the listed characteristic. We exclude volume data from firms with zero monthly volume prior to July 1995, which is the date when volume data become reliable. The eligible listed sample consists of the CRSP stocks satisfying the same data requirements as the OTC sample described in Section 3.3.B.

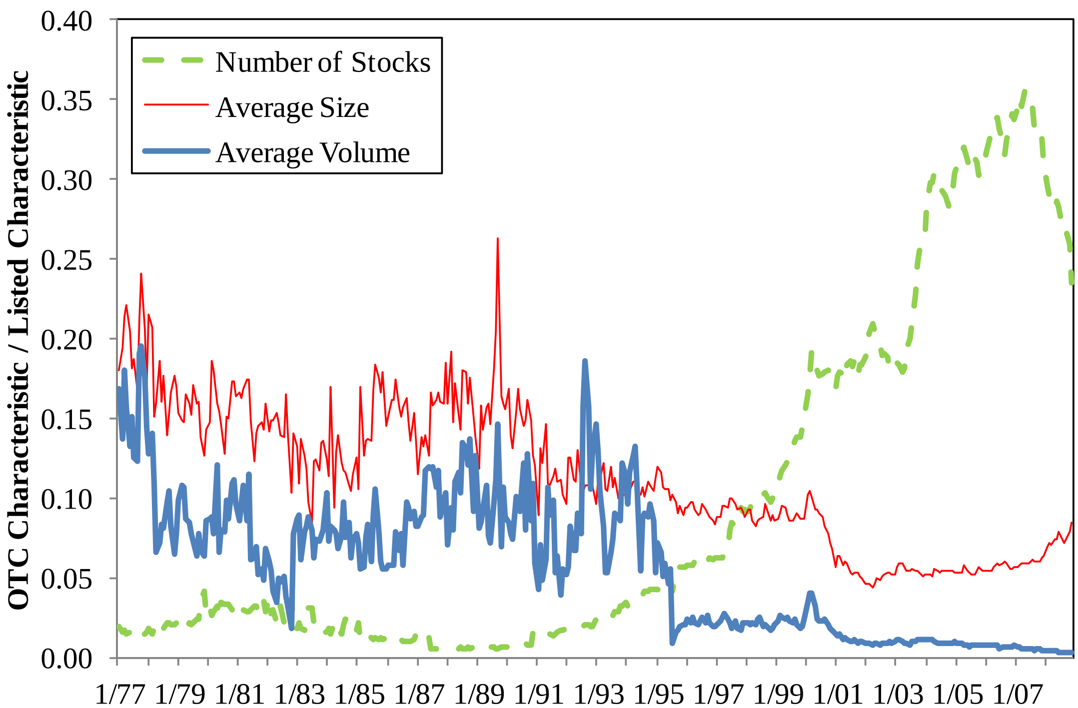


Figure 3.2: The Value of $1 Invested in Illiquidity Factors

We graph the cumulative returns for illiquidity factors in the OTC, comparable listed, and eligible listed samples. We use a logarithmic scale to represent the evolution of the value of a $1 investment from December 1976 to December 2008 for the illiquidity factors from each market. In all three markets, we sort stocks into quintiles according to their monthly PNT values, where PNT is the fraction of non-trading days in a month. Each PNT factor return is the difference between the gross-return-weighted returns of firms in the top and bottom PNT quintiles. We also plot the cumulative return of the value-weighted Pastor-Stambaugh illiquidity factor from the eligible listed sample. We assume that an investor begins with $1 long and $1 short and faces no margin or other funding requirements. To facilitate comparison, we scale the long-short portfolio positions in the OTC and eligible listed factors so that the volatility of these portfolios is equal to the volatility of the long-short portfolio based on the comparable listed factor. The comparable listed sample consists of stocks that are comparable to stocks in the OTC sample in terms of size, as described in Section 3.3.C. The eligible listed sample consists of all listed stocks that satisfy the same data requirements as the OTC sample described in Section 3.3.B.

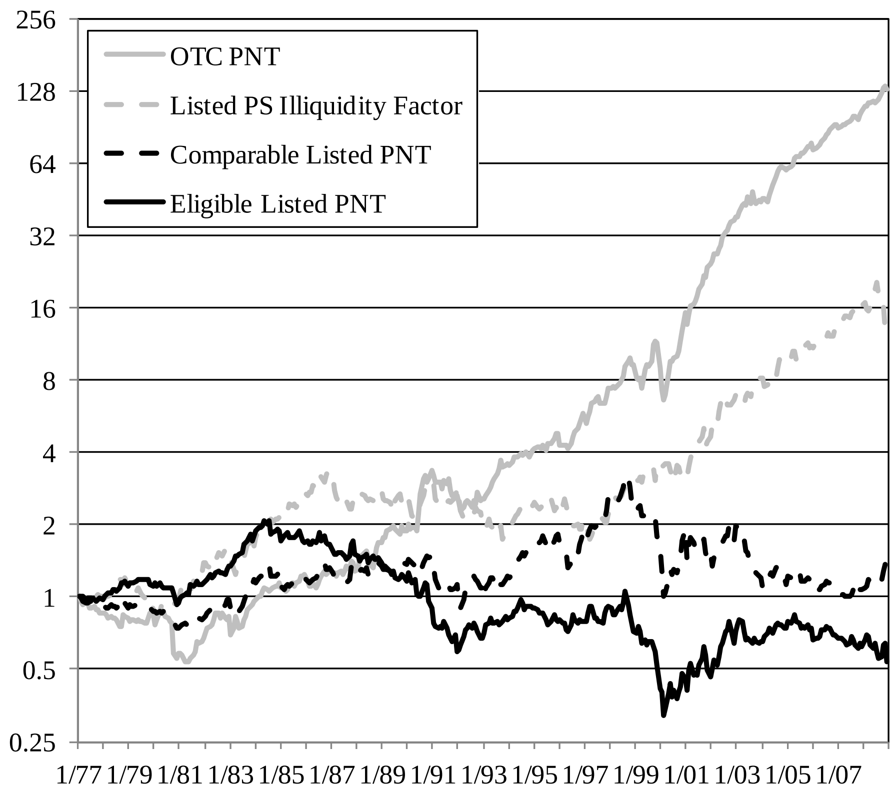
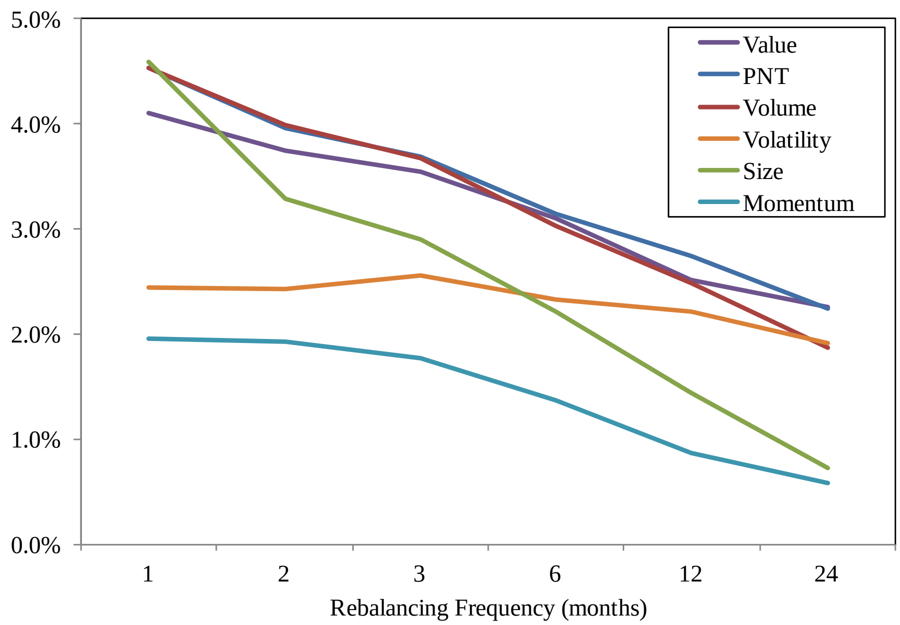


Figure 3.3: Impact of Trading Costs and Rebalancing Frequency on Arbitrageur Returns

We plot the average monthly returns of long-short OTC factor portfolios that are rebalanced at various frequencies using the method in Jegadeesh and Titman (1993) in which up to $1/n$ of the firms in each portfolio change in each month, based on rankings of OTC firms' values of the characteristics listed in the first column in the prior month. In both figures, rebalancing frequencies are indicated on the x-axis and stocks' returns are weighted by their prior month's gross return (GRW). In Panel A, we plot average pre-trading cost returns. In Panel B, we plot average post-trading cost returns for an arbitrageur who pays stocks' bid-ask spreads on each round-trip trade. Estimated monthly costs are equal to average portfolio turnover multiplied by average bid-ask spreads. Figures are based on 192 months of data from January 1993 through December 2008.

Panel A: Pre-Cost Monthly Returns of OTC Factor Portfolios



Panel B: Post-Cost Monthly Returns of OTC Factor Portfolios

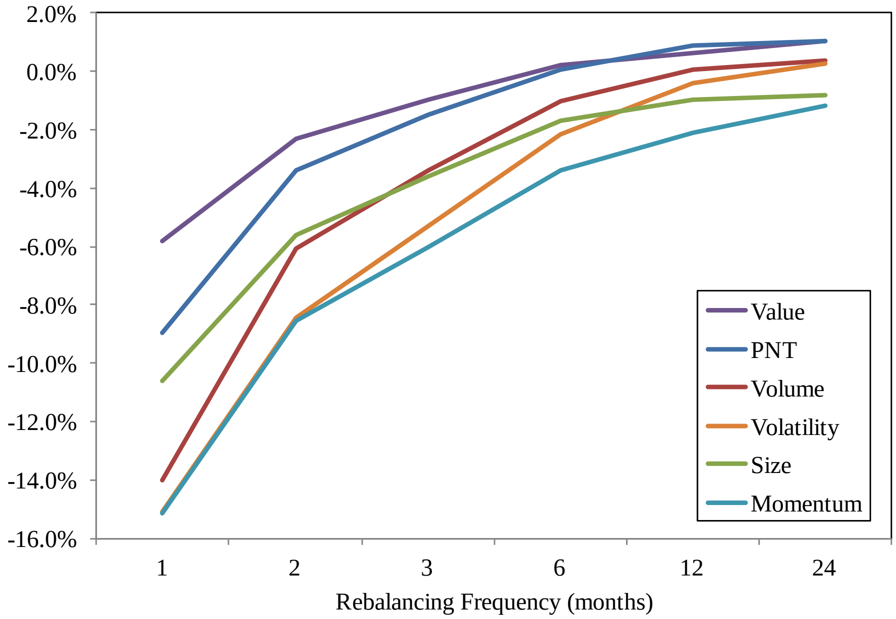


Table 3.1: Summary Statistics for the OTC and Listed Samples in July 1997

We report statistics for size, volume, and the number of firms in the OTC, comparable listed, and eligible listed samples in July of 1997, a typical month in terms of our OTC sample size. We construct the comparable listed sample to have the same median size as the OTC sample. The eligible listed sample consists of all listed stocks that satisfy the same data requirements as the OTC stocks in our sample, as described in Section 3.3.B.

	OTC	Comparable Listed	Eligible Listed
Total Market Capitalization (Billions)	21.3	15.1	9,592
Median Market Capitalization (Millions)	12.9	12.9	36
Mean Market Capitalization (Millions)	35.5	12.7	1,346
Trading Volume (Annualized Billions)	8.2	15.2	11,472
Median Trading Volume (Annualized Millions)	2.3	6.1	101
Mean Trading Volume (Annualized Millions)	13.7	12.8	1,608
Number of Firms	600	1,190	7,127

Table 3.2: The Peak Sizes of the Largest 10 OTC Firms

This table describes the ten largest OTC firms in our sample from 1977 to 2008. The first column shows the month in which each firm attains its peak size. The third column shows its size in that month. The two rightmost columns show each OTC firm's size rank and percentile within the eligible listed sample. The eligible listed sample consists of all listed stocks that satisfy the same data requirements as the OTC stocks in our sample, as described in Section 3.3.C.

Company Name	Peak Month	Trading Venue	Peak Size in Billions	Size Rank in Listed Sample	Size Percentile in Listed Sample
PUBLIX SUPER MARKETS INC	Dec-08	OTCBB	88.5	18th	99.5%
DELPHI CORP	Mar-08	Pink Sheets	13.0	225th	94.8%
MCI INC	Jan-04	Pink Sheets	7.7	292th	93.9%
MAXIM INTEGRATED PRODS INC	May-08	Pink Sheets	7.1	381th	91.2%
LEVEL 3 COMMUNICATIONS INC	Feb-98	OTCBB	6.6	297th	95.8%
NAVISTAR INTL CORP NEW	May-08	Pink Sheets	5.3	464th	89.3%
COMVERSE TECHNOLOGY INC	May-07	Pink Sheets	4.7	567th	87.6%
MERCURY INTERACTIVE CORP	Oct-06	Pink Sheets	4.6	515th	88.8%
ACTERNA CORP	Oct-00	OTCBB	3.0	623th	89.8%
HEALTHSOUTH CORP	Dec-04	Pink Sheets	2.5	734th	84.4%

Table 3.3: Cross-Sectional Summary Statistics for Key Variables

We summarize the distributions of monthly returns and the main firm characteristics for the OTC and comparable listed samples in Panels A and B, respectively. We construct the comparable listed sample to have the same median size as the OTC sample. Panel C contains average cross-sectional correlations between betas and characteristics among OTC sample firms. We compute all statistics below separately for the cross section of stocks in each month and then average across months. We measure all firm characteristics other than PNT using logarithms. We Winsorize all firm characteristics at the 5% level, but we do not Winsorize returns. The first seven columns report monthly averages of means, standard deviations, and various percentiles. The second to last column presents the average number of firms with non-missing values of each variable in each month. The last column presents the total number of months in which there is any data for each variable.

Panel A: OTC Stocks

	Monthly Averages								Total
Variable	Mean	SD	P5	P25	P50	P75	P95	Firms	Months
$Return$ (%)	-0.04	28.08	-34.73	-9.95	-1.30	4.86	39.23	486	383
$Disclosure$	0.60	0.46	0.00	0.29	0.65	1.00	1.00	486	383
$Size$	2.35	1.30	0.19	1.36	2.32	3.28	4.72	486	383
B/M	1.09	2.17	0.06	0.30	0.69	1.28	3.28	231	383
$Volatility$	6.56	5.52	0.79	2.33	4.95	8.97	20.57	476	383
$Volume$	8.25	3.57	4.43	5.67	7.01	10.96	14.62	486	383
PNT	0.55	0.34	0.01	0.28	0.63	0.82	0.94	486	383
$Spread$	0.15	0.14	0.02	0.05	0.10	0.20	0.51	391	192
$InstHold$	0.26	0.41	0.00	0.00	0.00	0.47	1.00	477	344

Panel B: Comparable Listed Sample

	Monthly Averages								Total
Variable	Mean	SD	P5	P25	P50	P75	P95	Firms	Months
$Return$ (%)	0.66	19.46	-24.45	-8.99	-1.22	7.28	32.16	1018	383
$Disclosure$	0.83	0.33	0.28	0.65	1.00	1.00	1.00	1018	383
$Size$	2.21	0.53	1.08	1.85	2.32	2.66	2.89	1018	383
B/M	1.29	1.64	0.18	0.54	0.96	1.57	3.26	789	383
$Volatility$	4.29	2.13	1.22	2.65	3.97	5.61	8.99	1005	383
$Volume$	10.77	1.98	8.11	9.48	10.27	12.35	14.26	1018	383
PNT	0.20	0.21	0.00	0.03	0.13	0.33	0.67	1018	383
$Spread$	0.08	0.04	0.02	0.04	0.07	0.10	0.18	538	303
$InstHold$	0.71	0.39	0.08	0.51	0.82	0.99	1.00	890	344

Panel C: Cross-sectional Correlations among OTC Stocks

	β_{MKT}	β_{SMB}	β_{HML}	β_{UMD}	$Size$	B/M	$Volatility$	$Ret[-1]$	$Ret[-12,-2]$	PNT	$Volume$	$Disclosure$	$InstHold$
β_{MKT}	1.00	-0.08	0.42	0.02	0.03	-0.09	0.05	-0.01	-0.02	-0.15	0.12	0.06	0.02
β_{SMB}	-0.08	1.00	0.13	-0.01	-0.04	-0.05	0.10	-0.01	0.01	-0.14	0.11	0.03	-0.03
β_{HML}	0.42	0.13	1.00	0.03	0.03	-0.04	-0.03	0.01	-0.04	-0.03	0.03	0.03	0.04
β_{UMD}	0.02	-0.01	0.03	1.00	0.06	-0.02	-0.03	-0.01	0.02	0.00	0.02	0.01	0.02
$Size$	0.03	-0.04	0.03	0.06	1.00	-0.19	-0.36	0.05	0.15	-0.17	0.36	0.06	0.27
B/M	-0.09	-0.05	-0.04	-0.02	-0.19	1.00	-0.03	-0.03	-0.13	0.22	-0.19	-0.22	-0.02
$Volatility$	0.05	0.10	-0.03	-0.03	-0.36	-0.03	1.00	0.02	-0.01	-0.06	-0.11	0.01	-0.19
$Ret[-1]$	-0.01	-0.01	0.01	-0.01	0.05	-0.03	0.02	1.00	-0.01	0.04	0.01	0.02	-0.01
$Ret[-12,-2]$	-0.02	0.01	-0.04	0.02	0.15	-0.13	-0.01	-0.01	1.00	0.00	0.05	0.04	0.00
PNT	-0.15	-0.14	-0.03	0.00	-0.17	0.22	-0.06	0.04	0.00	1.00	-0.84	-0.12	-0.06
$Volume$	0.12	0.11	0.03	0.02	0.36	-0.19	-0.11	0.01	0.05	-0.84	1.00	0.10	0.17
$Disclosure$	0.06	0.03	0.03	0.01	0.06	-0.22	0.01	0.02	0.04	-0.12	0.10	1.00	0.17
$InstHold$	0.02	-0.03	0.04	0.02	0.27	-0.02	-0.19	-0.01	0.00	-0.06	0.17	0.17	1.00

Table 3.4: Time Series Analysis of OTC and Comparable Listed Factor Portfolios

This table summarizes the returns and risk of long-short factor portfolios constructed using data on OTC stocks and comparable listed stocks from 1977 through 2008. We construct the comparable listed sample to have the same median size as the OTC sample. To construct each factor, we sort firms in each sample into quintiles at the end of each month based on the firm characteristics in the Factor column. Each factor's return for month t is the difference between the weighted returns of firms in the top and bottom quintiles, as ranked in month t–1. We use either equal weights (EW), a firm's prior month gross returns (GRW), or its prior month size (VW) when computing quintile portfolio returns. The PNT_{VW} and OTC Mkt_{VW} portfolios are marked with ǂ to indicate that they are always value-weighted while all other returns are weighted as indicated in the table. We estimate time series regressions of the monthly factor returns on various contemporaneous listed return factors and six lags of these factors to account for non-synchronous trading. Each factor loading is the sum of the estimated coefficients on the contemporaneous factor and its six lags. The regressors in these time series regressions are either the OTC market (OTC CAPM model), the listed MKT (Listed CAPM model), or the listed MKT, SMB, HML, UMD, and ILQ (Listed 5-Factor model) return factors. The last three columns in Panel B report the intercepts from these three regressions for each factor, while the first two columns show the average factor returns. Panel C shows the factor loadings from each regression, along with the R^2 statistics. Panels D and E report the analogous statistics for the comparable listed sample. Panel A shows the ratio of the intercepts in Panels B and D to the volatilities of the factors, where all ratios have been annualized by multiplying by the square root of 12. See the text for further details and definitions. We denote statistical significance at the 5% and 1% levels using * and ** symbols, respectively. These statistical tests employ Newey and West (1987) standard errors with the number of lags based on the formula from Newey and West (1994).

Panel A: Evaluating OTC and Comparable Listed Factor Returns

	Annualized Sharpe Ratios (GRW returns)			Annualized Information Ratios (GRW returns)			
				Listed CAPM			5-Factor Model
Factor	OTC	Comparable Listed	Eligible Listed	OTC	Comparable Listed	Eligible Listed	OTC
PNT	0.91**	0.14	-0.01	1.24**	0.29	0.08	1.34**
	(0.20)	(0.19)	(0.17)	(0.19)	(0.19)	(0.24)	(0.32)
PNT_{VW}‡	0.66**	0.04	0.13	1.00**	0.21	0.32	1.06**
	(0.21)	(0.20)	(0.20)	(0.23)	(0.19)	(0.27)	(0.32)
$Volume$	-0.90**	0.07	0.15	-1.14**	0.16	0.30	-1.23**
	(0.20)	(0.18)	(0.18)	(0.20)	(0.19)	(0.24)	(0.35)
$Size$	-1.02**	-0.98**	0.04	-0.98**	-0.81**	0.20	-0.92**
	(0.21)	(0.20)	(0.19)	(0.19)	(0.19)	(0.21)	(0.28)
$Value$	0.82**	1.19**	0.53*	1.19**	1.22**	0.68**	1.00**
	(0.24)	(0.20)	(0.21)	(0.22)	(0.22)	(0.25)	(0.33)
$Momentum$	0.41**	1.56**	1.30**	0.54**	1.71**	1.35**	0.09
	(0.15)	(0.15)	(0.16)	(0.14)	(0.15)	(0.17)	(0.20)
$Volatility$	-0.55**	-0.75**	-0.64**	-0.79**	-1.08**	-1.01**	-0.50
	(0.21)	(0.20)	(0.21)	(0.19)	(0.19)	(0.20)	(0.28)
$OTCMkt_{VW}$‡	-0.52*			-1.21**			-1.52**
	(0.23)			(0.19)			(0.26)

Panel B: Evaluating OTC Factor Returns

	Monthly Returns		Alphas by Model (GRW returns)		
	EW	GRW	OTC	Listed	Listed
Factor	Returns	Returns	CAPM	CAPM	5-Factor
PNT	2.94**	2.92**	2.22**	3.70**	3.67**
	(0.58)	(0.63)	(0.54)	(0.57)	(0.86)
PNT_{VW}‡	1.68**	N/A	1.01*	2.19**	2.19**
	(0.53)		(0.42)	(0.49)	(0.66)
$Volume$	-3.16**	-2.77**	-2.22**	-3.36**	-3.44**
	(0.56)	(0.63)	(0.59)	(0.57)	(0.99)
$Size$	-3.45**	-3.07**	-3.14**	-2.95**	-2.81**
	(0.56)	(0.63)	(0.76)	(0.57)	(0.85)
$Value$	1.99**	2.08**	1.77**	2.88**	2.29**
	(0.54)	(0.60)	(0.55)	(0.52)	(0.76)
$Momentum$	0.49	1.39**	1.28*	1.84**	0.30
	(0.43)	(0.53)	(0.60)	(0.49)	(0.69)
$Volatility$	-0.85	-1.87**	-1.00	-2.63**	-1.59
	(0.62)	(0.72)	(0.71)	(0.62)	(0.90)
$OTCMkt_{VW}$‡	-0.74*	N/A	N/A	-1.32**	-1.5**
	(0.33)			(0.21)	(0.26)

Panel C: Systematic Variation in OTC Return Factors

	Factor Loadings							R^2 by Model		
Factor	β_{OMKT}	β_{MKT_CAPM}	β_{MKT_5F}	β_{SMB}	β_{HML}	β_{UMD}	β_{LIQ}	OTC CAPM	Listed CAPM	Listed 5-Factor
PNT	-1.05**	-1.41**	-1.24**	-1.02*	0.89	-0.16	0.13	24.3%	15.3%	34.1%
	(0.25)	(0.36)	(0.36)	(0.43)	(0.57)	(0.42)	(0.39)			
PNT_{VW}‡	-0.90**	-1.06**	-0.88**	-0.91*	0.70	-0.03	-0.14	36.1%	27.1%	40.1%
	(0.20)	(0.25)	(0.30)	(0.40)	(0.41)	(0.31)	(0.36)			
$Volume$	0.86**	1.04**	0.97*	0.82	-0.75	0.16	-0.01	17.7%	11.5%	26.5%
	(0.25)	(0.36)	(0.41)	(0.47)	(0.66)	(0.45)	(0.41)			
$Size$	0.02	-0.36	-0.01	-1.01	0.16	-0.39	0.33	2.4%	2.6%	8.1%
	(0.31)	(0.40)	(0.50)	(0.61)	(0.67)	(0.56)	(0.51)			
$Value$	-0.71**	-1.19**	-0.85**	0.15	0.67	-0.54	1.00*	11.3%	9.6%	25.3%
	(0.22)	(0.28)	(0.30)	(0.39)	(0.41)	(0.43)	(0.47)			
$Momentum$	-0.34	-0.62	-0.22	-0.72	0.74	1.09**	0.47	3.0%	2.2%	12.0%
	(0.26)	(0.40)	(0.39)	(0.51)	(0.47)	(0.41)	(0.44)			
$Volatility$	1.07**	1.63**	0.87*	1.06*	-1.11	0.31	-1.38*	15.5%	8.6%	21.8%
	(0.27)	(0.40)	(0.37)	(0.42)	(0.65)	(0.50)	(0.56)			
$OTCMkt_{VW}$‡	N/A	1.17**	1.15**	0.59**	0.00	-0.02	0.11	N/A	43.5%	57.3%
		(0.11)	(0.13)	(0.17)	(0.17)	(0.14)	(0.18)			

Panel D: Evaluating Comparable Listed Factor Returns

	Monthly Returns		Alphas by Model (GRW returns)		
	EW	GRW	OTC	Listed	Listed
Factor	Returns	Returns	CAPM	CAPM	5-Factor
PNT	0.11	0.22	-0.01	0.40	0.07
	(0.30)	(0.30)	(0.29)	(0.26)	(0.28)
PNT_{VW}‡	0.06	N/A	-0.22	0.28	-0.14
	(0.31)		(0.29)	(0.25)	(0.28)
$Volume$	0.16	0.10	0.17	0.22	0.21
	(0.27)	(0.27)	(0.27)	(0.26)	(0.30)
$Size$	-1.01**	-0.98**	-1.21**	-0.79**	-0.43
	(0.19)	(0.20)	(0.24)	(0.19)	(0.25)
$Value$	1.39**	1.36**	1.36**	1.40**	1.40**
	(0.23)	(0.23)	(0.24)	(0.25)	(0.24)
$Momentum$	1.77**	2.10**	1.95**	2.23**	2.06**
	(0.21)	(0.21)	(0.20)	(0.19)	(0.28)
$Volatility$	-0.91*	-1.35**	-0.81*	-1.76**	-1.87**
	(0.36)	(0.36)	(0.37)	(0.30)	(0.28)

Panel E: Systematic Variation in Comparable Listed Return Factors

	Factor Loadings							R^2 by Model		
Factor	β_{OMKT}	β_{MKT_CAPM}	β_{MKT_5F}	β_{SMB}	β_{HML}	β_{UMD}	β_{LIQ}	OTC CAPM	Listed CAPM	Listed 5-Factor
PNT	-0.28*	-0.41**	-0.20	-0.66**	0.76**	0.17	-0.05	32.9%	26.5%	56.7%
	(0.14)	(0.14)	(0.16)	(0.20)	(0.19)	(0.18)	(0.14)			
PNT_{VW}‡	-0.32**	-0.51**	-0.31	-0.57**	0.72**	0.29	-0.09	37.4%	31.7%	60.2%
	(0.12)	(0.14)	(0.16)	(0.21)	(0.19)	(0.17)	(0.15)			
$Volume$	0.01	-0.10	-0.18	0.39	-0.46*	0.10	0.01	32.6%	26.6%	58.0%
	(0.13)	(0.14)	(0.15)	(0.21)	(0.19)	(0.16)	(0.14)			
$Size$	-0.32**	-0.36**	-0.31*	-0.35	0.04	-0.19	-0.28	7.9%	8.0%	21.0%
	(0.10)	(0.11)	(0.14)	(0.26)	(0.22)	(0.22)	(0.15)			
$Value$	0.01	-0.05	0.14	-0.37**	0.49**	-0.38**	0.29	5.9%	3.4%	40.2%
	(0.09)	(0.14)	(0.12)	(0.13)	(0.15)	(0.13)	(0.15)			
$Momentum$	-0.20*	-0.29*	-0.29	-0.23	-0.07	0.34*	-0.10	6.4%	9.1%	35.0%
	(0.09)	(0.12)	(0.16)	(0.17)	(0.17)	(0.16)	(0.14)			
$Volatility$	0.69**	0.87**	0.63**	1.21**	-0.44	0.12	-0.03	34.6%	22.2%	54.9%
	(0.17)	(0.16)	(0.19)	(0.29)	(0.28)	(0.25)	(0.21)			

Table 3.5: Testing Transaction Cost Theories of the Illiquidity Premium

This table reports the risk-adjusted returns and summary statistics for portfolios sorted by two illiquidity measures, PNT in Panel A and *Spread* in Panel B. In Panel A, we rank firms based on their PNT values in each month and divide them into decile portfolios. In Panel B, we divide firms into portfolios containing firms with the *Spread* ranges noted in the first column of Panel B. We require at least 5 firms in all portfolios in each month. We include data from August 1995 through December 2008 when volume and bid-ask data are widely available. A decile portfolio return for month t is based on month t–1 sorting. We compute returns corrected for bid-ask bounce by weighing each firm's return by its prior month's gross return. The first two columns in both panels report CAPM alphas for portfolios composed of OTC stocks and of stocks included in the comparable-size listed sample, as described in Section 3.3.C. These alphas are the intercepts from time series regressions of monthly portfolio returns on the listed MKT factor, including six lags to account for non-synchronous trading. Columns 8 and 9 in both panels report mean $Turnover$ values for each portfolio, while columns 10 and 11 report mean monthly trading costs. $Turnover$ is defined as monthly volume divided by end-of-month market capitalization. Trading costs are defined as $Spread * Turnover$. We denote statistical significance at the 5% and 1% levels using * and ** symbols, respectively. These statistical tests employ Newey and West (1987) standard errors with the number of lags based on the formula from Newey and West (1994).

Panel A: Sorts by PNT

	CAPM Alphas (GRW)			Mean *PNT*		Mean *Spread*		Mean *Turnover*		Trading Costs	
PNT Decile	OTC	Comp. Listed	Difference	OTC	Comp. Listed	OTC	Comp. Listed	OTC	Comp. Listed	OTC	Comp. Listed
1 Liquid	-3.98**	-0.06	-3.92**	0.000	0.000	6.3%	4.6%	20.7%	18.7%	1.30%	0.85%
	(0.95)	(0.55)	(0.67)								
2	-3.40**	-0.02	-3.39**	0.051	0.048	9.8%	5.6%	9.5%	8.2%	0.93%	0.46%
	(0.86)	(0.48)	(0.89)								
3	-2.12	0.11	-2.23	0.113	0.092	11.2%	5.8%	7.5%	5.8%	0.84%	0.34%
	(1.09)	(0.57)	(1.23)								
4	-1.93**	-0.19	-1.74**	0.198	0.137	12.7%	6.3%	5.6%	4.5%	0.71%	0.29%
	(0.56)	(0.44)	(0.59)								
5	-1.24	0.27	-1.52	0.301	0.183	14.2%	6.5%	3.5%	3.6%	0.50%	0.24%
	(0.79)	(0.43)	(0.84)								
6	-0.55	0.13	-0.68	0.410	0.231	15.4%	6.6%	2.8%	3.1%	0.43%	0.21%
	(0.58)	(0.44)	(0.66)								
7	0.22	0.74	-0.52	0.519	0.285	15.9%	7.0%	1.8%	2.7%	0.29%	0.19%
	(0.69)	(0.56)	(0.90)								
8	0.88	0.31	0.57	0.629	0.352	18.5%	7.3%	1.4%	2.5%	0.26%	0.18%
	(1.28)	(0.42)	(1.30)								
9	0.47	0.18	0.29	0.757	0.464	22.2%	7.9%	0.9%	1.9%	0.19%	0.15%
	(0.62)	(0.32)	(0.67)								
10 Illiquid	1.36	-0.17	1.52**	0.898	0.661	30.9%	8.8%	0.5%	1.0%	0.14%	0.09%
	(0.70)	(0.34)	(0.58)								
Monotonicity	3.75**	0.20	3.55**								
	(0.76)	(0.38)	(0.76)								

Panel B: Sorts into Bid-Ask Spread Ranges

Bid-Ask Spread Range	CAPM Alphas (GRW) OTC	Comp. Listed	Difference	Mean *PNT* OTC	Comp. Listed	Mean *Spread* OTC	Comp. Listed	Mean *Turnover* OTC	Comp. Listed	Trading Costs OTC	Comp. Listed
(0.000,0.025]	-1.25	0.48	-1.73*	0.215	0.137	1.5%	1.5%	14.7%	18.2%	0.21%	0.28%
	(0.68)	(0.39)	(0.68)								
(0.025,0.050]	-1.52**	0.59	-2.12**	0.297	0.178	3.7%	3.6%	10.5%	8.5%	0.39%	0.31%
	(0.52)	(0.46)	(0.50)								
(0.050,0.075]	-1.62*	0.14	-1.76**	0.336	0.214	6.2%	6.1%	7.8%	5.8%	0.48%	0.36%
	(0.75)	(0.43)	(0.66)								
(0.075,0.100]	-2.30**	-0.88	-1.43**	0.353	0.242	8.7%	8.6%	6.7%	5.1%	0.58%	0.44%
	(0.51)	(0.54)	(0.52)								
(0.100,0.125]	-2.27**	-0.15	-2.11**	0.369	0.278	11.2%	11.1%	6.3%	3.9%	0.71%	0.44%
	(0.64)	(0.61)	(0.73)								
(0.125,0.150]	-2.21**	-0.64	-1.58	0.388	0.297	13.7%	13.6%	5.3%	3.6%	0.72%	0.50%
	(0.77)	(0.76)	(0.96)								
(0.150,0.175]	-1.57*	0.25	-1.82	0.417	0.311	16.2%	16.1%	4.5%	4.0%	0.73%	0.65%
	(0.77)	(0.93)	(1.19)								
(0.175,0.200]	-2.47**	-0.68	-1.79*	0.434	0.333	18.6%	18.6%	4.7%	3.4%	0.88%	0.63%
	(0.75)	(0.73)	(0.90)								
(0.200,0.225]	-0.36	-1.93	1.57	0.456	0.387	21.4%	21.2%	3.4%	3.1%	0.73%	0.65%
	(1.23)	(1.15)	(2.29)								
(0.225,0.250]	-0.28	-1.51	1.22	0.483	0.398	24.0%	23.8%	2.6%	2.9%	0.62%	0.69%
	(1.10)	(1.31)	(2.23)								
Monotonicity	0.54	-1.73*	2.27**								
	(0.54)	(0.66)	(1.00)								
Concavity	-2.63**	-0.38**	-2.25**								
	(0.98)	(0.93)	(1.61)								

Table 3.6: Cross-Sectional Regressions of Monthly Returns on Firm Characteristics

This table displays corrected estimates of cross-sectional regressions of monthly stock returns on several firm characteristics and factor loadings. We estimate monthly cross-sectional weighted least squares regressions as in Asparouhova, Bessembinder, and Kalcheva (2010), using each stock's gross return in the previous month as the weighting. The table reports average coefficients that weight each monthly coefficient by the inverse of its squared standard errors as in Ferson and Harvey (1999). We compute Newey and West (1987) standard errors with five lags based on the formula from Newey and West (1994). The R^2 in the bottom row is the average from the monthly regressions. We denote statistical significance at the 5% and 1% levels using * and ** symbols, respectively.

	OTC Sample			Comparable Listed Sample			Eligible Listed Sample		
	I	II	III	I	II	III	I	II	III
β_{MKT}	-0.228**		-0.140*	-0.233**		-0.057	-0.282**		-0.069
	(0.063)		(0.054)	(0.072)		(0.059)	(0.086)		(0.059)
β_{SMB}	-0.160**		-0.063*	-0.128**		-0.014	-0.199**		-0.047
	(0.034)		(0.031)	(0.038)		(0.032)	(0.052)		(0.031)
β_{HML}	0.141**		0.091*	0.061		0.012	0.198**		0.054
	(0.044)		(0.042)	(0.039)		(0.028)	(0.062)		(0.034)
β_{UMD}	-0.065		-0.060	0.007		-0.005	0.047		0.028
	(0.044)		(0.041)	(0.027)		(0.026)	(0.029)		(0.023)
$Size$		-0.692**	-0.688**		-0.607**	-0.625**		-0.134**	-0.142**
		(0.141)	(0.124)		(0.097)	(0.095)		(0.038)	(0.038)
B/M		0.380**	0.316**		0.659**	0.631**		0.522**	0.475**
		(0.119)	(0.117)		(0.104)	(0.102)		(0.083)	(0.074)
$Volatility$		-0.247**	-0.245**		-0.356**	-0.347**		-0.436**	-0.414**
		(0.034)	(0.033)		(0.043)	(0.038)		(0.060)	(0.046)
$Ret[-1]$		-0.038**	-0.038**		-0.064**	-0.065**		-0.043**	-0.046**
		(0.007)	(0.007)		(0.006)	(0.006)		(0.005)	(0.005)
$Ret[-12,-2]$		0.008**	0.008**		0.018**	0.019**		0.013**	0.014**
		(0.001)	(0.001)		(0.001)	(0.001)		(0.001)	(0.001)
PNT		4.302**	4.053**		-0.364	-0.475		0.050	-0.086
		(0.642)	(0.639)		(0.334)	(0.301)		(0.373)	(0.306)
Average R^2	6.8%	10.6%	15.0%	1.6%	3.7%	4.7%	2.6%	4.8%	5.8%
Avg. Stocks	454	441	439	919	905	905	4,809	4,762	4,762

Table 3.7: Impact of Trading Costs and Rebalancing Frequency on Arbitrageur Returns

This table evaluates the returns for an arbitrageur trying to implement the OTC factor returns who pays stocks' bid-ask spreads on each round-trip trade. We compute summary statistics for long-short factor portfolios that are rebalanced at frequencies of 1 and 12 months using the method in Jegadeesh and Titman (1993) in which up to $1/n$ of the firms in each portfolio change in each month, based on rankings of OTC firms' values of the characteristics listed in the first column in the prior month. The first two columns report factor portfolios' average pre-cost returns for 1- and 12-month rebalancing frequencies. Columns 3 and 4 report factor portfolios' average post-cost returns at these frequencies. Estimated monthly costs are equal to average portfolio turnover multiplied by average bid-ask spreads. Columns 5 and 6 show the bid-ask spreads such that average post-cost returns would be zero for the two rebalancing frequencies. In Columns 1 to 6, all stocks' returns are weighted by their prior month's gross return (GRW). Columns 7, 8 and 9 report rebalancing frequencies at which, using actual bid-ask spreads, average post-cost returns would be closest to zero for three portfolio weighing methods: GRW (as used in columns 1-6), value-weighted (VW) returns, and liquidity-weighted (LW) returns, which are weighted by the inverse of stocks' bid-ask spreads. These statistics are based on 192 months of data from January 1993 through December 2008. We denote statistical significance at the 5% and 1% levels using * and ** symbols, respectively. These statistical tests employ Newey and West (1987) standard errors with five lags based on the formula from Newey and West (1994).

	Pre-cost Returns		Post-cost Returns		Breakeven Spread		Breakeven Frequency		
OTC factor	1 Months	12 Months	1 Months	12 Months	1 Months	12 Months	GRW	VW	LW
PNT	4.53%**	2.74%**	-8.94%**	0.87%	5.41%	17.04%	6	4	4
$Volume$	4.53%**	2.48%**	-14.02%**	0.05%	4.73%	14.12%	12	9	6
$Size$	4.59%**	1.44%*	-10.59%**	-0.96%	6.42%	9.25%	24+	9	10
$Value$	4.1%**	2.51%**	-5.81%**	0.64%	6.33%	16.19%	6	3	3
$Momentum$	1.96%**	0.87%	-15.17%**	-2.11%**	2.19%	4.41%	24+	24+	24+
$Volatility$	2.44%*	2.22%**	-15.11%**	-0.43%	2.69%	12.87%	17	24+	24+

Table 3.8: Double Sorted Portfolios

This table contains average monthly returns for double sorted portfolios within OTC stocks and within stocks included in the comparable listed sample, which consists of stocks that are comparable to stocks in the OTC sample in terms of size, as described in Section 3.3.C. We first rank stocks according to one characteristic of interest and sort them into portfolios. We then rank stocks within these portfolios according to other characteristics and again sort into portfolios. We sort stocks into terciles rather than quintiles to ensure that we have a sufficient number of stocks in each portfolio, and require at least 10 stocks in each tercile. Within each double-sorted tercile, we compute returns corrected for bid-ask bounce by weighing each stock's return by its prior month's gross returns. We display returns for the top and bottom terciles (i.e., the extreme terciles) according to the second sort within the first-sort extreme terciles. For binary variables ($InstHold$ and $Disclose$), we sort stocks into two portfolios based on their values. Panel A reports the returns of double-sorted portfolios where stocks are first sorted according to $InstHold$. Panel B reports returns where stocks are first sorted according to $Disclose$. Panel C reports returns where stocks are first sorted according to $Size$. We denote statistical significance at the 5% and 1% levels using * and ** symbols, respectively. These statistical tests employ Newey and West (1987) standard errors with five lags based on the formula from Newey and West (1994).

Panel A: Double Sorted Portfolios: Initial Sort Based on Institutional Holdings

	Held stocks monthly returns			Non-held stocks monthly returns			Premium
	Top tercile	Bottom tercile	Premium	Top tercile	Bottom tercile	Premium	Difference (%)
OTC Stocks							
PNT	0.21	-1.44	1.65	1.11	-4.12	5.23**	-3.58**
Size	-0.31	0.40	-0.71	-2.13	1.74	-3.87**	3.16**
Volume	-0.80	0.51	-1.30	-3.97	1.72	-5.70**	4.39**
Value	1.18	-1.36	2.54**	1.10	-2.56	3.66**	-1.12
Momentum	0.77	-1.20	1.97**	-0.28	-2.46	2.18**	-0.21
Volatility	-0.76	0.52	-1.28	-2.01	0.23	-2.24**	0.96
Comparable Listed Stocks							
PNT	0.46	0.35	0.11	0.54	-0.28	0.82*	-0.71*
Size	0.17	0.90	-0.73**	-0.05	0.70	-0.75*	0.02
Volume	0.57	0.28	0.29	-0.18	0.53	-0.71	1.00**
Value	0.89	-0.03	0.92**	1.08	-0.76	1.84**	-0.92*
Momentum	1.23	-0.34	1.56**	1.08	-0.93	2.01**	-0.44
Volatility	-0.22	0.88	-1.10**	-0.67	0.98	-1.65**	0.55

Panel B: Double Sorted Portfolios: Initial Sort Based on Disclosure

	Disclosing stocks monthly returns			Non-disclosing monthly returns			Premium
	Top tercile	Bottom tercile	Premium	Top tercile	Bottom tercile	Premium	Difference (%)
OTC Stocks							
PNT	0.89	-1.04	1.94**	0.75	-2.56	3.31**	-1.38*
Size	-0.22	1.16	-1.38**	-1.47	1.42	-2.89**	1.51
Volume	-0.62	1.02	-1.64**	-2.40	0.89	-3.28**	1.64*
Momentum	0.89	-0.66	1.55**	-0.04	-0.65	0.61	0.94
Volatility	-0.24	0.70	-0.94	-1.61	0.93	-2.54**	1.60*
Comparable Listed Stocks							
PNT	0.69	0.36	0.33	0.41	-0.35	0.76	-0.43
Size	0.25	1.08	-0.83**	-0.03	0.41	-0.45	-0.38
Volume	0.40	0.55	-0.15	-0.15	0.35	-0.50	0.35
Momentum	1.45	-0.14	1.59**	1.27	-0.73	2.00**	-0.41
Volatility	-0.12	1.04	-1.16**	-0.90	0.89	-1.79**	0.63*

Panel C: Double Sorted Portfolios: Initial Sort Based on Size

	Big stocks monthly returns			Small stocks monthly returns			Premium
	Top tercile	Bottom tercile	Premium	Top tercile	Bottom tercile	Premium	Difference (%)
OTC Stocks							
PNT	0.12	-2.00	2.12*	2.31	-1.32	3.62**	-1.50
$Volume$	-1.47	-0.33	-1.14	-1.59	3.21	-4.80**	3.65**
$Value$	0.33	-2.72	3.05**	2.03	0.19	1.84	1.20
$Momentum$	-0.09	-1.86	1.78**	1.26	0.84	0.41	1.37
$Volatility$	-2.12	0.44	-2.55**	0.95	1.53	-0.58	-1.97
Comparable Listed Stocks							
PNT	0.31	0.10	0.21	0.77	0.83	-0.06	0.27
$Volume$	0.41	0.11	0.29	1.02	0.61	0.42	-0.12
$Value$	0.50	-0.19	0.70**	1.46	0.54	0.92**	-0.23
$Momentum$	1.08	-0.73	1.81**	1.54	0.24	1.29**	0.51*
$Volatility$	-0.72	0.78	-1.50**	0.47	1.19	-0.72*	-0.78**

Table 3.9: Long–term Returns of Momentum Portfolios

This table contains average returns for long-short momentum portfolios constructed at various time horizons using the method described in Jegadeesh and Titman (1993). We first form top and bottom quintile portfolios for each month $t-1$ based on stocks' momentum, defined as the return from month $t-12$ to month $t-2$. Returns within each extreme quintile portfolio are either weighted by the prior month's gross returns ("GRW returns") or value weighted ("VW returns"). Then, to measure momentum returns n years after portfolio formation in each month t, we equally weight the 12 monthly returns of the extreme quintile portfolios formed in months $t-n*12$ to $t-n*12$–11. The top minus bottom quintile portfolio return is the momentum premium at the n-year horizon. We compute returns for portfolios within our 3 samples: OTC stocks, stocks included in the comparable listed sample, which consists of stocks that are comparable to stocks in the OTC sample in terms of size, as described in Section 3.3.C, and stocks included in the eligible listed sample, which consists of all listed stocks that satisfy the same data requirements as the OTC stocks in our sample, as described in Section 3.3.B. We denote statistical significance at the 5% and 1% levels using * and ** symbols, respectively. These statistical tests employ Newey and West (1987) standard errors with five lags based on the formula from Newey and West (1994).

	OTC Stocks		Comparable Listed Stocks		Eligible Listed Stocks	
Horizon in Months	GRW Returns	VW Returns	GRW Returns	VW Returns	GRW Returns	VW Returns
[1,1]	1.39**	3.15**	2.10**	1.97**	1.68**	1.29**
[1,12]	-0.08	1.57**	0.58**	0.75**	0.44*	0.47
[13,24]	-0.75	0.71	-0.12	-0.03	-0.21	-0.23
[25,36]	-0.07	0.37	0.13	0.24	-0.17	-0.11
[37,48]	-0.66	0.37	0.05	0.05	0.10	0.08
[49,60]	-0.99	0.42	-0.08	0.18	-0.29**	-0.20
[13,60]	-0.56	0.45	0.02	0.12	-0.13	-0.10

3.8. References

Aggarwal, Rajesh K., and Guojun Wu, 2006, Stock market manipulations, *The Journal of Business* 79, 1915–1953.

Amihud, Yakov, and Haim Mendelson, 1986, Asset pricing and the bid-ask spread, *Journal of Financial Economics* 17, 223–249.

Ang, Andrew, Robert J. Hodrick, Yuhang Xing, and Xiaoyan Zhang, 2006, The cross-section of volatility and expected returns, *The Journal of Finance* 61, 259–299.

Asness, Clifford S., Tobias J. Moskowitz, and Lasse Heje Pedersen, 2013, Value and momentum everywhere, *The Journal of Finance* 68, 929–985.

Asparouhova, Elena, Hendrik Bessembinder, and Ivalina Kalcheva, 2010, Liquidity biases in asset pricing tests, *Journal of Financial Economics* 96, 215–237.

Badrinath, Swaminathan G., and Sunil Wahal, 2002, Momentum trading by institutions, *The Journal of Finance* 57, 2449–2478.

Baker, Malcolm, and Jeremy C. Stein, 2004, Market liquidity as a sentiment indicator, *Journal of Financial Markets* 7, 271–299.

Banz, Rolf W., 1981, The relationship between return and market value of common stocks, *Journal of Financial Economics* 9, 3–18.

Barber, Brad M., and Terrance Odean, 2000, Trading is hazardous to your wealth: The common stock investment performance of individual investors, *The Journal of Finance* 55, 773–806.

Barberis, Nicholas, Andrei Shleifer, and Robert Vishny, 1998, A model of investor sentiment, *Journal of Financial Economics* 49, 307–343.

Baruch, Shmuel, Andrew G. Karolyi, and Michael L. Lemmon, 2007, Multimarket trading and liquidity: theory and evidence, *The Journal of Finance* 62, 2169–2200.

Baruch, Shmuel, and Gideon Saar, 2009, Asset returns and the listing choice of firms, *Review of Financial Studies* 22, 2239–2274.

Bekaert, Geert, Campbell R. Harvey, and Christian Lundblad, 2007, Liquidity and expected returns: Lessons from emerging markets, *Review of Financial Studies* 20, 1783–1831.

Blume, Marshall E., and Robert F. Stambaugh, 1983, Biases in computed returns: An application to the size effect, *Journal of Financial Economics* 12, 387–404.

Böhme, Rainer, and Thorsten Holz, 2006, The effect of stock spam on financial markets, Technische Universität Dresden Working Paper.

Bollen, Nicolas P.B., and William G. Christie, 2009, Market microstructure of the pink sheets, *Journal of Banking & Finance* 33, 1326–1339.

Carhart, Mark M., 1997, On persistence in mutual fund performance, *The Journal of Finance* 52, 57–82.

Constantinides, George M., 1986, Capital market equilibrium with transaction costs, *Journal of Political Economy* 94, 842–862.

Daniel, Kent, David Hirshleifer, and Avanidhar Subrahmanyam, 1998, Investor psychology and security market under- and overreactions, *The Journal of Finance* 53, 1839–1885.

Daniel, Kent, and Sheridan Titman, 1997, Evidence on the characteristics of cross sectional variation in stock returns, *The Journal of Finance* 52, 1–33.

D'avolio, Gene, 2002, The market for borrowing stock, *Journal of Financial Economics* 66, 271–306.

Dimson, Elroy, 1979, Risk measurement when shares are subject to infrequent trading, *Journal of Financial Economics* 7, 197–226.

Duffie, Darrell, Nicolae Garleanu, and Lasse Heje Pedersen, 2002, Securities lending, shorting, and pricing, *Journal of Financial Economics* 66, 307–339.

Eraker, Bjørn, and Mark J. Ready, 2011, Over the counter stock markets, University of Wisconsin, Madison Working Paper.

Fama, Eugene F., and Kenneth R. French, 1992, The cross-section of expected stock returns, *The Journal of Finance* 47, 427–465.

——— , 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3–56.

Fama, Eugene F., and James D. MacBeth, 1973, Risk, return, and equilibrium: Empirical tests, *Journal of Political Economy* 81, 607–636.

Ferson, Wayne E., and Campbell R. Harvey, 1999, Conditioning variables and the cross section of stock returns, *The Journal of Finance* 54, 1325–1360.

Frazzini, Andrea, Ronen Israel, and Tobias J. Moskowitz, 2012, Trading costs of asset pricing anomalies, University of Chicago Working Paper.

Frieder, Laura, and Jonathan Zittrain, 2007, Spam works: Evidence from stock touts and corresponding market activity, *Hastings Communications and Entertainment Law Journal* 30, 479–520.

Greenstone, Michael, Paul Oyer, and Annette Vissing-Jorgensen, 2006, Mandated disclosure, stock returns, and the 1964 securities acts amendments, *The Quarterly Journal of Economics* 121, 399–460.

Griffin, John M., Jeffrey H. Harris, and Selim Topaloglu, 2003, The dynamics of institutional and individual trading, *The Journal of Finance* 58, 2285–2320.

Gutierrez, Roberto C. Jr., and Christo A. Prinsky, 2007, Momentum, reversal, and the trading behaviors of institutions, *Journal of Financial Markets* 10, 48–75.

Harris, Jeffrey H., Venkatesh Panchapagesan, and Ingrid M. Werner, 2008, Off but not gone: a study of NASDAQ delistings, Fisher College of Business Working Paper.

Hasbrouck, Joel, 2009, Trading costs and returns for US equities: Estimating effective costs from daily data, *The Journal of Finance* 64, 1445–1477.

Hirshleifer, David, and Siew Hong Teoh, 2003, Limited attention, information disclosure, and financial reporting, *Journal of Accounting and Economics* 36, 337–386.

Hong, Harrison, and Jeremy C. Stein, 1999, A unified theory of underreaction, momentum trading, and overreaction in asset markets, *The Journal of Finance* 54, 2143–2184.

Ince, Ozgur S., and R. Burt Porter, 2006, Individual equity return data from thomson datastream: Handle with care!, *Journal of Financial Research* 29, 463–479.

Jarrow, Robert, 1980, Heterogeneous expectations, restrictions on short sales, and equilibrium asset prices, *The Journal of Finance* 35, 1105–1113.

Jegadeesh, Narasimhan, and Sheridan Titman, 1993, Returns to buying winners and selling losers: Implications for stock market efficiency, *The Journal of Finance* 48, 65–91.

——— , 2001, Profitability of momentum strategies: An evaluation of alternative explanations, *The Journal of Finance* 56, 699–720.

Karolyi, G. Andrew, and Anthony B. Sanders, 1998, The variation of economic risk premiums in real estate returns, *The Journal of Real Estate Finance and Economics* 17, 245–262.

Lee, Charles, and Bhaskaran Swaminathan, 2000, Price momentum and trading volume, *The Journal of Finance* 55, 2017–2069.

Lesmond, David A., Joseph P. Ogden, and Charles A. Trzcinka, 1999, A new estimate of transaction costs, *Review of Financial Studies* 12, 1113–1141.

Leuz, Christian, Alexander Triantis, and Tracy Yue Wang, 2008, Why do firms go dark? causes and economic consequences of voluntary SEC deregistrations, *Journal of Accounting and Economics* 45, 181–208.

Lo, Andrew W., and Craig A. MacKinlay, 1990, An econometric analysis of nonsynchronous trading, *Journal of Econometrics* 45, 181–211.

Loh, Roger K., and Mitch Warachka, 2012, Streaks in earnings surprises and the cross-section of stock returns, *Management Science* 58, 1305–1321.

Luft, Carl, and Lawrence M. Levine, 2004, Over the counter bulletin board exchange: The impact of liquidity and size to return, volatility, and bid/ask spread, *The Journal of Alternative Investments* 7, 95–106.

——— , and Scott Larson, 2001, Over the counter bulletin board exchange: Market structure, risk, and return, *The Journal of Alternative Investments* 4, 33–42.

Macey, Jonathan, Maureen O'Hara, and David Pompilio, 2008, Down and out in the stock market: the law and economics of the delisting process, *Journal of Law and Economics* 51, 683–713.

Miller, Edward M., 1977, Risk, uncertainty, and divergence of opinion, *The Journal of Finance* 32, 1151–1168.

Nagel, Stefan, 2005, Short sales, institutional investors and the cross-section of stock returns, *Journal of Financial Economics* 78, 277–309.

Newey, Whitney K., and Kenneth D. West, 1987, A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica* 55, 703–708.

——— , 1994, Automatic lag selection in covariance matrix estimation, *The Review of Economic Studies* 61, 631–653.

Ofek, Eli, and Matthew Richardson, 2003, Dotcom mania: The rise and fall of internet stock prices, *The Journal of Finance* 58, 1113–1138.

Pástor, L'uboš, and Robert F. Stambaugh, 2003, Liquidity risk and expected stock returns, *Journal of Political Economy* 111, 642–685.

Sanger, Gary C., and James D. Peterson, 1990, An empirical analysis of common stock delistings, *Journal of Financial and Quantitative Analysis* 25, 261–272.

Scheinkman, Jose A., and Wei Xiong, 2003, Overconfidence and speculative bubbles, *Journal of Political Economy* 111, 1183–1220.

Vayanos, Dimitri, and Paul Woolley, 2013, An institutional theory of momentum and reversal, *Review of Financial Studies* 26, 1087–1145.

Appendix A

Survey Questions

I. Account ownership

1. Do you currently have any checking accounts?

- Yes
- No

** If answered *Yes*: Do you have a checking account at a bank, a credit union, or both? Or maybe someplace else?

- At a bank
- At a credit union
- Someplace else

** If answered *At a Credit Union*: Which credit union do you have your checking account at? (Please remember that your answers will remain completely confidential)

[Neighborhood Trust; Brooklyn Cooperative; Lower East Side People's FCU; Bethex; Union Settlement; ERDA; other credit union]

2. Do you currently have any savings accounts? This includes money market, CD and Christmas club accounts.

- Yes

- No

** If answered *Yes*: Do you have a savings account at a bank, a credit union, or both? Or maybe someplace else?

- At a bank
- At a credit union
- Someplace else

** If answered *At a Credit Union*: Which credit union do you have your savings account at? (Please remember that your answers will remain completely confidential)

[Neighborhood Trust; Brooklyn Cooperative; Lower East Side People's FCU; Bethex; Union Settlement; ERDA; other credit union]

** If has any credit union account:

3. Did you become a credit union member before or after attending the "Getting Ahead" workshop?

- I became a member before attending the workshop
- I became a member after attending the workshop
- I don't remember

** If answered *No* to both 1 and 2 (if has no type of account):

4. Have you ever had any type of checking or savings account?

- Yes
- No

II. Saving

5. Do you currently save? That is, do you put some money aside in most months?

- Yes

- No

** If has savings account and is saving:

6. Where do you save? Your savings account? Or someplace else (like at home)? Or both?

- My savings account

- Someplace else (like home)

7. Thinking back to 3 years ago, were you saving at that time?

- Yes

- No

** If answered *Yes* to both 4 and 6 (if saving both now and 3 years ago):

8. Compared to 3 years ago, do you now save more, less, or about the same? That is, do you put aside more, less or about the same?

- I save more now than I did 3 years ago

- I save less now than I did 3 years ago

- I save about the same now as I did 3 years ago

** If currently saving:

9. The next question is a bit sensitive and you can refuse to answer if it makes you uncomfortable. Do keep in mind, though, that your answers will be kept totally confidential. The question is: thinking about all of your savings, about how much do you have saved up right now (in $)?

** If currently not saving:

9. The next question is a bit sensitive and you can refuse to answer if it makes you uncomfortable. Do keep in mind, though, that your answers will be kept totally confidential. The question is: do you have any money saved up, and if so, about how much (in $)?

III. Financial Services Usage

** If answered *Yes* to either 1 or 2 (if has any type of account):

10. Do you have direct deposit set up on your account?

- Yes
- No
- I don't know

** If answers to 1 and 2 indicate any type of credit union account:

11. Have you ever taken out a loan from your credit union?

- Yes
- No

** If answered *Yes*: Which type of loan did you take from your credit union?

(read out the options. respondent can give more than one answer)

- A credit-builder/secured loan (also known as goal-saver, socio-credito and sociedad)
- A debt consolidation loan
- Another type of loan (such as a personal loan)

** If answers to 1 and 2 indicate any type of credit union account:

12. Have you ever used the 1-on-1 financial advice or credit counseling services offered at your credit union?

- Yes
- No

** Everyone:

13. Do you sometimes cash your checks at check cashing places?

(if the person deposits checks in his account using check-cashing places, this counts as "No")

- Yes, I do so sometimes
- No, I never do so

14. In the past year, have you taken out a loan anywhere that is not a bank or a credit union? I'm talking about things like payday loans, pawnshops, loan sharks or loans offered on the internet.

- Yes
- No

15. Thinking back to 3 years ago, did you take out any loans of this type around that period?

- Yes
- No
- I don't know

IV. "Financial Fragility"

16. Do you have an emergency fund to help you deal with situations such as an expensive health issue or a car breaking down?

- Yes
- No

** If answered *Yes*: This again is a bit of a sensitive question, but if you can, I'd appreciate it if you could tell me: About how much money do you have in your emergency fund (in $)?

17. Say that some unexpected need arose and you had to come up with $1500 within the next month. How confident are you that you would be able to do so? The options are:

(please read all of the options out loud)

- I am certain I could come up with the full $1,500
- I could probably come up with $1,500
- I could probably not come up with $1,500

- I am certain I could not come up with $1,500

** If answered anything other than "certain that could not": And how would you go about getting these $1500 if you had to? I'll read you some options. Please tell me if you would use that method (you can say as many as you would like).

- draw from emergency fund/ savings
- borrow from family or friends
- use my credit card
- take out a loan from bank/credit union
- take out a loan from a pawnshop/payday loan/loan shark/ internet
- some other method

V. Spending Behavior

18. Are you currently using a budget to plan and follow your monthly income and expenses?

- Yes
- No

** If answered *Yes*: Are you using the budget that you received at the "Getting Ahead" workshop?

- Yes
- No

19. Do you feel that you have a good sense of how much money you normally spend on various categories, like food, transportation, child care or entertainment?

- Yes, I feel that I have a good sense of how much I spend on various categories
- No, I do not feel that I have a good sense of how much I spend on various categories

20. Do you feel that you're currently spending more than you should be?

- Yes, I feel that I'm currently spending more than I should be
- No, I do not feel that I'm currently spending more than I should be

21. Many people spend money on things and later wish that they hadn't. Compared to 3 years ago, do you now spend more, less or about the same on these types of things?

- I spend more now on these types of things than I did 3 years ago
- I spend less now on these types of things than I did 3years ago
- I spend about the same now on these types of things as I did 3 years ago

VI. Miscellaneous

22. On a scale of 1-5, how stressed do you feel about your personal finances and debt in general? 1 being the least stressed and 5 being the most stressed. 1, 2, 3, 4 or 5?

23. On a scale of 1-5, how knowledgeable do you feel about managing your money on a daily basis? 1 being the least knowledgeable and 5 being the most knowledgeable. 1, 2, 3, 4 or 5?

24. During the last 6 months, was there a time when you and your family were not able to pay your rent, mortgage, or utility bills on time?

- Yes
- No

25. During the last 6 months, did you not get or postpone getting any type of medical care because you couldn't afford it? This includes filling a prescription for drugs.

- Yes
- No

26. What is your current employment status? Are you

- Employed full-time?
- Employed part-time?

- Self-employed?
- Unemployed and looking for work?
- Disabled/retired/student?

Appendix B

An Information Asymmetry Leads to a Flat Rate-Amount Schedule

In this section I show that when the borrower's expected second period income is unknown to the lender, a flat borrowing rate-borrowed amount schedule is the only incentive compatible contract.

Consider the setting described in Section 2.2 except that rather than a single $\tilde{Y_2}$ distribution, there is a continuum of second period income types $j \in [\underline{J}, \overline{J}]$.[1]

$\tilde{Y_2^j}$, the second period income of type j, is uniformly distributed. The distributions of all types have the same variance but different means, such that types range from $\underline{J}$ ("worst type") to $\overline{J}$ ("best type"):

$$\tilde{Y_2^j} \sim U[\underline{Y_2^j}, \overline{Y_2^j}]$$

$$[\underline{Y_2^j}, \overline{Y_2^j}] = [j - a, j + a]$$

where $2a$ is normalized to 1. Type is uncorrelated with initial wealth Y_1 and unknown to the lender.

In a full-information competitive setting, each type would be offered a rate-amount schedule that would satisfy a lender zero-profit condition. Since the lender gets nothing in case of default (when $\tilde{Y_2^j} < \widehat{Y_2}(B_1, R)$),[2] this schedule would satisfy, for each type j:

[1] The above analysis can be thought of as applying to each individual $\tilde{Y_2^j}$ type.

[2] This ignores any non-exempt assets that the lender might be able to seize.

$$R(B_1)^j = \frac{R^f + \psi}{\overline{Y_2^j} - \widehat{Y_2}(B_1, R)}$$

where R^f is the risk-free rate and ψ is the cost of intermediation, proportional to the amount borrowed.

When types are unknown, borrowers might be screened through their choice of the borrowed amount B_1. I now show that in the setting described in this model (with a fixed cost of default Λ) there is no rate-amount schedule which successfully screens borrowers, and the only incentive compatible contract is one where there is a single borrowing rate for loans of all sizes. For simplicity I assume that all types have the same initial wealth.

A rate-amount schedule is a contract which specifies B_1^j and R^j for each type j. I first show that a rate-amount schedule cannot be incentive compatible for any two types G and B s.t. $E(\tilde{Y_2^G}) > E(\tilde{Y_2^B})$ if it implies $Debt^B > Debt^G$, where both types have zero probability of default for their own allocation as well as for the other type's allocation.[3] Recalling that $\widehat{Y_2}$ is an increasing function of $Debt$, we get the following ordering:

$$\widehat{Y_2^G} < \widehat{Y_2^B} < \underline{Y_2^B} < \underline{Y_2^G} < \overline{Y_2^B} < \overline{Y_2^G}$$

Incentive compatibility requires $U^G(B_1^G, R^G) > U^G(B_1^B, R^B)$ and $U^B(B_1^B, R^B) > U^B(B_1^G, R^G)$. We have $C_1^j = Y_1 + B_1^j$ and $C_2^j = \tilde{Y_2^j} - Debt^j$ in all states here, so the IC constraints are:

$$u(Y_1 + B_1^G) + \beta \int_{\underline{Y_2^G}}^{\overline{Y_2^G}} u(\tilde{Y_2} - Debt^G) d\tilde{Y_2} \geq u(Y_1 + B_1^B) + \beta \int_{\underline{Y_2^G}}^{\overline{Y_2^G}} u(\tilde{Y_2} - Debt^B) d\tilde{Y_2}$$

$$u(Y_1 + B_1^B) + \beta \int_{\underline{Y_2^B}}^{\overline{Y_2^B}} u(\tilde{Y_2} - Debt^B) d\tilde{Y_2} \geq u(Y_1 + B_1^G) + \beta \int_{\underline{Y_2^B}}^{\overline{Y_2^B}} u(\tilde{Y_2} - Debt^G) d\tilde{Y_2}$$

Combining the two IC constraints, we get:

[3] Recall that $\widehat{Y_2} < \underline{Y_2}$ implies that it is never optimal to default in the second period. In the original formulation of the model it was assumed that $\underline{Y_2}$ is s.t. all borrowers might find it optimal to default if a bad state realizes, but with many types it seems reasonable to assume that those with relatively high expected incomes never default for some (low) levels of borrowing.

$$\int_{\underline{Y_2^G}}^{\overline{Y_2^G}} (u(\tilde{Y_2} - Debt^G) - u(\tilde{Y_2} - Debt^B))d\tilde{Y_2} \geq \int_{\underline{Y_2^B}}^{\overline{Y_2^B}} (u(\tilde{Y_2} - Debt^G) - u(\tilde{Y_2} - Debt^B))d\tilde{Y_2}$$

$$\int_{\underline{Y_2^B}}^{\overline{Y_2^B}} (u(\tilde{Y_2} + X - Debt^G) - u(\tilde{Y_2} + X - Debt^B))d\tilde{Y_2} \geq \int_{\underline{Y_2^B}}^{\overline{Y_2^B}} (u(\tilde{Y_2} - Debt^G) - u(\tilde{Y_2} - Debt^B))d\tilde{Y_2}$$

where $X = E(\tilde{Y_2^G}) - E(\tilde{Y_2^B}) > 0$. This inequality does not hold since marginal utility is decreasing. That is, such a contract cannot be incentive compatible for both types.

I now show that a rate-amount schedule cannot be incentive compatible if it implies $Debt^B < Debt^G$ for any types G and B s.t. $E(\tilde{Y_2^G}) > E(\tilde{Y_2^B})$, where both types have *positive* probability of default for their own allocation as well as for the other type's allocation:

$$\underline{Y_2^B} < \underline{Y_2^G} < \widehat{Y_2^B} < \widehat{Y_2^G} < \overline{Y_2^B} < \overline{Y_2^G}$$

Any type j choosing $Debt^k$ will find it optimal to default in the second period if and only if $\tilde{Y_2^j} < \widehat{Y_2}(Debt^k) = \widehat{Y_2^k}$. It follows that the IC constraints for types G and B in this case are:

$$u(Y_1+B_1^G)+\beta[\int_{\underline{Y_2^G}}^{\widehat{Y_2^G}} (u(\tilde{Y_2})-\Lambda)d\tilde{Y_2}+\int_{\widehat{Y_2^G}}^{\overline{Y_2^G}} u(\tilde{Y_2}-Debt^G)d\tilde{Y_2}] \geq u(Y_1+B_1^B)+\beta[\int_{\underline{Y_2^G}}^{\widehat{Y_2^B}} (u(\tilde{Y_2})-\Lambda)d\tilde{Y_2}+\int_{\widehat{Y_2^B}}^{\overline{Y_2^G}} u(\tilde{Y_2}-Debt^B)d\tilde{Y_2}]$$

$$u(Y_1+B_1^B)+\beta[\int_{\underline{Y_2^B}}^{\widehat{Y_2^B}} (u(\tilde{Y_2})-\Lambda)d\tilde{Y_2}+\int_{\widehat{Y_2^B}}^{\overline{Y_2^B}} u(\tilde{Y_2}-Debt^B)d\tilde{Y_2}] \geq u(Y_1+B_1^G)+\beta[\int_{\underline{Y_2^B}}^{\widehat{Y_2^G}} (u(\tilde{Y_2})-\Lambda)d\tilde{Y_2}+\int_{\widehat{Y_2^G}}^{\overline{Y_2^B}} u(\tilde{Y_2}-Debt^G)d\tilde{Y_2}]$$

Similar to above, we can combine the IC constraints to get:

$$\int_{\overline{Y_2^B}}^{\overline{Y_2^G}} u(\tilde{Y_2} - Debt^B)d\tilde{Y_2} \leq \int_{\overline{Y_2^B}}^{\overline{Y_2^G}} u(\tilde{Y_2} - Debt^G)d\tilde{Y_2}$$

Since $Debt^B < Debt^G$, this does not hold.

It follows that the only possible type of contract is one where *Debt* is constant across types. A rate-amount schedule where R varies with the amount borrowed B_1 while *Debt* is held constant (i.e. where R decreases in B_1 and their product is held constant) cannot be incentive compatible since second period consumption is determined by *Debt*. No type would ever choose to borrow a low amount at a high rate if he can keep second period consumption constant and consume more in the first period by borrowing a higher amount at a lower rate.

It follows that a screening contract based on the amount borrowed cannot be incentive compatible in the setting examined here. The borrowing rate will therefore be flat in the amount borrowed. It will be determined by an *aggregate* zero-profit condition rather than on a loan-by loan basis:

$$R = \frac{(R^f + \psi)\int_{\underline{J}}^{\overline{J}} B_1^j(R)dj}{\int_{\underline{J}}^{\overline{J}}(\overline{Y_2^j} - \widehat{Y_2^j})B_1^j(R)dj}$$

where $\widehat{Y_2^j} = \widehat{Y_2}(B_1^j(R), R)$.

Appendix C

Proofs of Propositions in Chapter 2

Proof of Proposition 1

Proof. Assuming decreasing marginal utility of consumption ($u''(.) < 0$), the RHS of (2.3.1) increases faster in $\tilde{Y_2}$ than does the LHS when $Debt > 0$. The existence of a threshold $\widehat{Y_2}$ directly follows. □

Proof of Proposition 2

Proof. When $\Lambda(Debt) = \Lambda$, (2.3.7) is reduced to:

$$\beta R^2[(1 - \frac{d\widehat{Y_2}}{dDebt})u'(\widehat{Y_2} - Debt) - u'(\overline{Y_2} - Debt)]$$

As marginal utility is always positive, it is sufficient that $\frac{d\widehat{Y_2}}{dDebt} \geq 1$ for this quantity to be negative. I now show that this is always the case with a fixed cost of default $\Lambda(Debt) = \Lambda$.

Applying the implicit function theorem on (2.3.2) with a general cost of default $\Lambda(Debt)$, the derivative of $\widehat{Y_2}$ wrt $Debt$ is:

$$\frac{d\widehat{Y_2}}{dDebt} = \frac{u'(\widehat{Y_2} - Debt) - \Lambda'(Debt)}{u'(\widehat{Y_2} - Debt) - u'(\widehat{Y_2})} \tag{C.0.1}$$

Assume that the sign of $\frac{d\widehat{Y_2}}{dDebt}$ is the same for all debt levels. The denominator is positive since marginal utility is decreasing. The numerator is positive for any default cost function where $\Lambda'(.)$ is non-increasing, since, re-writing (2.3.2) (and assuming, as in (2.3.2), that $\Lambda(0) = 0$), $\widehat{Y_2}$ is defined by:

$$\int_0^{Debt} u'(\widehat{Y_2} - x)dx = \int_0^{Debt} \Lambda'(x)dx$$

A non-positive numerator implies $u'(\widehat{Y_2} - Debt) \leq \Lambda'(Debt)$. Assuming $u''(.) < 0$ and $\Lambda''(.) \leq 0$, this implies:

$$\int_0^{Debt} u'(\widehat{Y_2} - x)dx < \int_0^{Debt} u'(\widehat{Y_2} - Debt)dx \leq \int_0^{Debt} \Lambda'(Debt)dx \leq \int_0^{Debt} \Lambda'(x)dx$$

which contradicts (2.3.2). It follows that the numerator of (C.0.1) is positive as well.

A necessary and sufficient condition for $\frac{d\widehat{Y_2}}{dDebt} \geq 1$ is therefore $u'(\widehat{Y_2}) \geq \Lambda'(Debt)$. For a fixed cost of default $\Lambda(Debt) = \Lambda$, this holds with a strong inequality. □

Proof of Proposition 3

Proof. I first show that there is some threshold $\dddot{Y_1}$ above which a unique "default impossible" solution exists and is preferred to any "default possible" interior solution.

In the C_1 region where default is impossible, the derivative of the value function wrt C_1 is:

$$u'(C_1) - \beta RE(u'(\tilde{Y_2} - Debt))$$

The second derivative is $u''(C_1) + \beta R^2 E(u''(\tilde{Y_2} - Debt)) < 0$. It follows that there is at most one "default impossible" solution $C_1^* \leq \widehat{C_1}$. For such a solution to exist, the value function must decrease at $\widehat{C_1}$ (it must be past its maximum), i.e.

$$u'(\widehat{C_1}) \leq \beta RE(u'(\tilde{Y_2} - Debt(\widehat{C_1}))) = \beta RE(u'(\tilde{Y_2} - \bar{D_0} - (\widehat{C_1} - Y_1)R))$$

Since $\frac{\partial \widehat{C_1}}{\partial Y_1} = 1$,[1] the LHS is decreasing in Y_1 and the RHS is constant in Y_1. It follows that there is some threshold Y_1 above which a unique "default impossible" solution exists and below which no such solution exists.

[1]This is the case since $Debt(\widehat{C_1})$ is constant (see (2.3.3)).

When a "default impossible" solution exists, there might also exist an interior "default possible" solution. I now show that when this is the case, the value of the former relative to the latter increases in Y_1. It follows that there must be a $\dddot{Y_1}$ threshold above which the "default impossible" solution is chosen over any other interior solution.

Consider two initial wealth values $\dot{Y_1} < \ddot{Y_1}$ for which both types of interior solutions exist: "default impossible" solutions $\dot{C_1^*}$ and $\ddot{C_1^*}$ and "default possible" solutions $\dot{C_1^{**}}$ and $\ddot{C_1^{**}}$. The value of the "default possible" relative to the "default impossible" solution for the wealthier individual is:

$$U(\ddot{C_1^{**}}) - U(\ddot{C_1^*}) = \int\limits_{\ddot{C_1^*}}^{\ddot{C_1^{**}}} (u'(C_1) - opp\_cost(C_1, \ddot{Y_1}))dC_1$$

Examining the RHS of (2.3.5), note that the opportunity cost of first period consumption depends only on the value of $Debt = (C_1 - Y_1)R + \bar{D_0}$. It follows that there are C_1 values:

$$\tilde{C_1^*} = \ddot{C_1^*} - (\ddot{Y_1} - \dot{Y_1})$$

$$\tilde{C_1^{**}} = \ddot{C_1^{**}} - (\ddot{Y_1} - \dot{Y_1})$$

which, if consumed by the less wealthy individual, generate the same opportunity cost as the wealthier individual's interior solutions, since $Debt$ is the same:

$$opp\_cost(\tilde{C_1^*}, \dot{Y_1}) = opp\_cost(\ddot{C_1^*}, \ddot{Y_1})$$

$$opp\_cost(\tilde{C_1^{**}}, \dot{Y_1}) = opp\_cost(\ddot{C_1^{**}}, \ddot{Y_1})$$

and we have:

$$\int\limits_{\ddot{C_1^*}}^{\ddot{C_1^{**}}} opp\_cost(C_1, \ddot{Y_1})dC_1 = \int\limits_{\tilde{C_1^*}}^{\tilde{C_1^{**}}} opp\_cost(C_1, \dot{Y_1})dC_1$$

The value of the "default possible" relative to the "default impossible" solution for the less wealthy individual is:[2]

$$U(\dot{C}_1^{**}) - U(\dot{C}_1^{*}) = \int_{\dot{C}_1^{*}}^{\dot{C}_1^{**}} (u'(C_1) - opp\_cost(C_1, \dot{Y}_1))dC_1 =$$

$$= \int_{\ddot{C}_1^{*}}^{\ddot{C}_1^{**}} (u'(C_1)dC_1 - \int_{\tilde{C}_1^{*}}^{\tilde{C}_1^{**}} opp\_cost(C_1, \dot{Y}_1))dC_1 \tag{C.0.2}$$

$$+ \int_{\tilde{C}_1^{*}}^{\ddot{C}_1^{*}} opp\_cost(C_1, \dot{Y}_1)dC_1 - \int_{\tilde{C}_1^{**}}^{\ddot{C}_1^{**}} opp\_cost(C_1, \dot{Y}_1)dC_1 \tag{C.0.3}$$

$$- \int_{\dot{C}_1^{*}}^{\ddot{C}_1^{*}} (opp\_cost(C_1, \dot{Y}_1) - (u'(C_1))dC_1 + \tag{C.0.4}$$

$$+ \int_{\ddot{C}_1^{**}}^{\dot{C}_1^{**}} (u'(C_1) - opp\_cost(C_1, \dot{Y}_1))dC_1 \tag{C.0.5}$$

Since (C.0.2) is equal to $U(\ddot{C}_1^{**}) - U(\ddot{C}_1^{*})$, ((C.0.3) + (C.0.4) + (C.0.5)) is the difference between the less wealthy and the wealthier individuals in the relative value of the "default possible" solution. I now show that this difference is positive.

Consider (C.0.3):

$$\int_{\tilde{C}_1^{*}}^{\ddot{C}_1^{*}} opp\_cost(C_1, \dot{Y}_1)dC_1 - \int_{\tilde{C}_1^{**}}^{\ddot{C}_1^{**}} opp\_cost(C_1, \dot{Y}_1)dC_1 > (\ddot{C}_1^{*} - \tilde{C}_1^{*})(opp\_cost(\tilde{C}_1^{*}, \dot{Y}_1) - opp\_cost(\tilde{C}_1^{**}, \dot{Y}_1)) =$$
$$= (\ddot{Y}_1 - \dot{Y}_1)(u'(\ddot{C}_1^{*}) - u'(\ddot{C}_1^{**}))$$

where the inequality follows from the opportunity cost of $\dot{Y}_1$ increasing around $\ddot{C}_1^{*}$ and decreasing around $\ddot{C}_1^{**}$ and from $\ddot{C}_1^{*} - \tilde{C}_1^{*} = \ddot{C}_1^{**} - \tilde{C}_1^{**} = \ddot{Y}_1 - \dot{Y}_1$, and the equality follows from the definitions of $\tilde{C}_1^{*}$ and $\tilde{C}_1^{**}$ and from the optimality of $\ddot{C}_1^{*}$ and $\ddot{C}_1^{**}$.

Now consider (C.0.4):

$$- \int_{\dot{C}_1^{*}}^{\ddot{C}_1^{*}} (opp\_cost(C_1, \dot{Y}_1) - (u'(C_1))dC_1 > -(\ddot{Y}_1 - \dot{Y}_1)(opp\_cost(\ddot{C}_1^{*}, \dot{Y}_1) - u'(\ddot{C}_1^{*})) =$$
$$= -(\ddot{Y}_1 - \dot{Y}_1)(opp\_cost(\ddot{C}_1^{*}, \dot{Y}_1) - opp\_cost(\ddot{C}_1^{*}, \ddot{Y}_1))$$

[2] I assume that $\dot{Y}_1$ and $\ddot{Y}_1$ are close enough s.t. $\tilde{C}_1^{*} < \dot{C}_1^{*} < \ddot{C}_1^{*} < \tilde{C}_1^{**} < \ddot{C}_1^{**} < \dot{C}_1^{**}$

where the inequality follows from the value function decreasing past the optimum $\dot{C}_1^*$ and having a negative second derivative around $\dot{C}_1^*$ and from $\ddot{C}_1^* - \dot{C}_1^* < \ddot{C}_1^* - \tilde{C}_1^* = \ddot{Y}_1 - \dot{Y}_1$, and the equality follows from the optimality of $\ddot{C}_1^*$

It follows that:

$(C.0.3) + (C.0.4) > (\ddot{Y}_1 - \dot{Y}_1)[(u'(\ddot{C}_1^*) - u'(\ddot{C}_1^{**})) - (opp\_cost(\ddot{C}_1^*, \dot{Y}_1) - opp\_cost(\ddot{C}_1^*, \ddot{Y}_1))]$

For arbitrarily small $(\ddot{Y}_1 - \dot{Y}_1)$, $opp\_cost(\ddot{C}_1^*, \dot{Y}_1) - opp\_cost(\ddot{C}_1^*, \ddot{Y}_1)$ is arbitrarily close to zero and the expression is positive (since $u'(\ddot{C}_1^*) - u'(\ddot{C}_1^{**}) > 0$).

Adding (C.0.5), which is positive since $\dot{C}_1^{**}$ is a local maximum for $\dot{Y}_1$, we get that $U(\dot{C}_1^{**}) - U(\dot{C}_1^*) > U(\ddot{C}_1^{**}) - U(\ddot{C}_1^*)$, and that the value of the "default impossible" solution relative to any interior "default possible" solution increases in Y_1. The existence of $\ddddot{Y}_1$ follows.

Now, consider the corner solution (borrowing up to the borrowing constraint). In Proposition 4 I show that the value of the corner solution relative to any interior solution is decreasing in Y_1. It follows that there must be some $\widehat{Y_1} \geq \ddddot{Y}_1$ value above which the unique "default impossible" solution is preferred to any other interior solution as well as to the corner solution. $\square$

Proof of Proposition 4

Proof. Borrowing up to the borrowing constraint (choosing the corner solution) is optimal in case no interior solution exists and in case interior solutions exist but are inferior to borrowing up to the constraint.

Proposition 3 shows that there is some initial wealth threshold above which a "default impossible" solution exists and below which it does not. I now show that there is a similar threshold for "default possible" solutions. Consider an initial wealth level $\tilde{Y}_1$ for which there is a "default possible" solution $\tilde{C}_1 > \widehat{C_1}$ but not a "default impossible" solution. At $\tilde{C}_1$, the FOC is satisfied:

$$u'(\tilde{C}_1) = \beta R \int_{\widehat{Y_2}(\tilde{Y}_1)}^{\overline{Y_2}} u'(\tilde{Y}_2 - (\tilde{C}_1 - \tilde{Y}_1)R - \bar{D}_0) d\tilde{Y}_2 \tag{C.0.6}$$

It is easy to show that when the opportunity cost (the RHS of (C.0.6)) is decreasing in C_1, it is increasing in Y_1. It follows that for any initial wealth level $\widetilde{Y_1}$ s.t. $\widetilde{Y_1} > \tilde{Y}_1$ for which no "default impossible" solution exists, the RHS of (C.0.6) at $\tilde{C}_1$ is higher than $u'(\tilde{C}_1)$, i.e. the value function is decreasing at $\tilde{C}_1$. Since the value function is increasing at low values of C_1, an interior "default possible" solution $\widehat{C_1} < \widetilde{C_1} < \tilde{C}_1$ must exist for $\widetilde{Y_1}$. It follows that there is some initial wealth threshold above which an interior solution exists.

The corner solution is also chosen for some initial wealth values for which an interior solution exists. I now show that the value of the corner solution relative to the value of any interior solution decreases in Y_1, so that there is a Y_1 threshold below which the corner solution is always chosen and above which it is never chosen.

Denote first period consumption corresponding to the corner solution by $\overline{C_1} = Y_1 + \overline{B_1}$, where $\overline{B_1}$ is the borrowing constraint, and the value (expected utility) of consuming C_1 by $U(C_1)$. Since the corner solution dictates that an extra unit of initial wealth be entirely consumed in the first period, $\frac{dU(\overline{C_1})}{dY_1} = u'(\overline{C_1})$. As shown in Proposition 6, $\frac{dC_1^*}{dY_1} \neq 1$ for any interior solution C_1^*, i.e. $\frac{dU(C_1^*)}{dY_1} > u'(C_1^*)$ (consuming an extra unit of wealth in its entirety is suboptimal at an interior solution).

Since $C_1^* < \overline{C_1}$ and since marginal utility is decreasing, we have:

$$\frac{dU(\overline{C_1})}{dY_1} = u'(\overline{C_1}) < u'(C_1^*) < \frac{dU(C_1^*)}{dY_1}$$

That is, the value of any interior solution increases by more than does the value of the corner solution as we increase Y_1. It follows that there is an initial wealth threshold $\check{Y}_1$ above which an interior solution C_1^* is preferred to the corner solution $\overline{C_1}$ and below which the corner solution is preferred.[3] $\square$

Proof of Proposition 5

Proof. In Proposition 6 it is shown that $MPC < 0$ for individuals with initial wealth $Y_1 < \widehat{Y_1}$ and $0 < MPC < 1$ for those with $Y_1 > \widehat{Y_1}$. For $Y_1 < \check{Y}_1$ individuals, who hit the borrowing constraint, $MPC = 1$. Borrowing $(C_1 - Y_1)$ therefore weakly decreases in income for all individuals.

It follows that for any given interest rate R at which funds are both borrowed and saved, there is a threshold income level $\grave{Y}_1(R)$ above which the individual saves and below which he borrows.

[3]Note that for some parameters, we might have $\check{Y}_1 = \widehat{Y_1}$, i.e. $Y_1 < \widehat{Y_1}$ individuals borrow up to the constraint and have positive probability of default, and $Y_1 > \widehat{Y_1}$ individuals consume an amount such that they never default. I assume that this is not the case, i.e. that there is an initial wealth region where interior "default possible" solutions are chosen.

The FOC at this threshold (where $Debt = \bar{D_0}$) is:[4]

$$u'(\grave{Y_1}(R)) = \beta R[\int_{\widehat{Y_2}}^{\overline{Y_2}} u'(\tilde{Y_2} - \bar{D_0})d\tilde{Y_2} + \int_{\underline{Y_2}}^{\widehat{Y_2}} \Lambda'(\bar{D_0}))d\tilde{Y_2}]$$

The RHS is opp_cost. Its derivative wrt R is $\frac{opp\_cost}{R}$, since $\frac{d\widehat{Y_2}}{dR} = 0$ when $B_1 = 0$. The RHS therefore increases in R. Since marginal utility is decreasing, $\grave{Y_1}(R)$ must decrease in R for the LHS to increase in R as well. It follows that for $R_S < R_B$ we have $\grave{Y_1}(R_B) < \grave{Y_1}(R_S)$. Individuals with initial wealth between the 2 thresholds $\grave{Y_1}(R_B) < Y_1 < \grave{Y_1}(R_S)$ neither borrow nor save, wishing to borrow at R_S and save at R_B. □

Proof of Proposition 6

Proof. Applying the implicit function theorem on the FOC and noting that $\frac{\partial opp\_cost}{\partial Y_1} = -\frac{d}{dC_1} opp\_cost$, I obtain the following general expression for MPC:

$$MPC = \frac{-\frac{d}{dC_1} opp\_cost}{U''(C_1)}$$

Where $U''(C_1)$ is the second derivative of the value function, always negative at an optimum. It follows that the sign of MPC is determined by the sign of the derivative of the opportunity cost wrt C_1 at the optimum. Noting that $U''(C_1) = u''(C_1) - \frac{d}{dC_1} opp\_cost$ and assuming that the cost of default is fixed (which leads to $\frac{d}{dC_1} opp\_cost < 0$ for potential defaulters at the optimum), the proposition follows. □

[4] I assume $\bar{D_0} > \underline{Y_2}$ for all individuals, i.e. the fixed pre-existing obligation is large enough so that an individual must be saving in order to ensure that he is not forced to default in the case that $\underline{Y_2}$ realizes. This gives $\grave{Y_1}(R_S) < \widehat{Y_1}$, and means that both $\grave{Y_1}(R_S)$ and $\grave{Y_1}(R_B)$ are in the region where the effective first order condition is the one for "default possible" consumption levels ($C_1 < \widehat{C_1}$).

Proof of Proposition 7

Proof. The first point can be seen by recalling the proof of Proposition 3. A "default impossible" solution is chosen at initial wealth values for which:[5] [6]

$$u'(\widehat{C_1}) \leq \beta R_S E(u'(\tilde{Y_2} - \bar{D_0} - (\widehat{C_1} - Y_1)R_S)) \tag{C.0.7}$$

When the cost of default is fixed, $\widehat{C_1}(Y_1)$ is defined by:

$$u(\underline{Y_2}) - \Lambda = u(\underline{Y_2} - (\widehat{C_1} - Y_1)R_S - \bar{D_0})$$

Since the LHS is constant and $\widehat{C_1} - Y_1 < 0$ by assumption (see footnote 4), increasing R_S results in a higher $\widehat{C_1}$ for any given Y_1 level, s.t. $Debt = (\widehat{C_1} - Y_1)R_S - \bar{D_0}$ remains constant. The RHS of (C.0.7) subsequently increases and the LHS decreases for any given Y_1 level. It follows that (C.0.7) is relaxed and $\widehat{Y_1}$, the initial wealth threshold at which (C.0.7) holds with equality, decreases.

The second point can be seen by considering the individual with initial wealth $\check{Y_1}$ who is indifferent between hitting the borrowing constraint (choosing the corner solution) and choosing the interior "default possible" solution (see Proposition 4). Denote first period consumption at the borrowing constraint by $\overline{C_1} = \check{Y_1} + \overline{B_1}$. As the borrowing constraint is defined by $u(\overline{Y_2}) - \Lambda = u(\overline{Y_2} - (\overline{B_1}R + \bar{D_0}))$, a decrease in R leads to an increase of $\frac{\overline{B_1}}{R}$ in $\overline{B_1}$ so that $\overline{B_1}R$ stays unchanged ($\frac{\overline{B_1}}{R}$ is the negative of the derivative of $\overline{B_1}$ wrt R).[7] Since second period utility is not affected, the value of the corner solution increases by $\frac{\overline{B_1}}{R}u'(\overline{C_1})$ when R decreases.

Consider now the pre-change interior solution for $\check{Y_1}$, denoted by C_1^* (note that it has the same value as borrowing up to the constraint, by definition of $\check{Y_1}$). Let us examine the change in value if the individual maintains C_1^* in response to a decrease in R_B. In that case, the only change in value occurs in the second period:

$$-\frac{\partial \beta E(U_2^*)}{\partial R} = -\beta[-B_1^* \int_{\widehat{Y_2}}^{\overline{Y_2}} u'(\tilde{Y_2} - Debt)d\tilde{Y_2}] = -\frac{dE(U_2^*)}{dC_1}\frac{B_1^*}{R} = \frac{B_1^*}{R}u'(C_1^*)$$

[5] I assume here that where a "default impossible" solution exists, it is always chosen over the corner solution as well as over a "default possible" solution, if one exists. See the proof of Proposition 3 for more on this issue.

[6] R_S is the relevant rate here since I assume $\bar{D_0} > \underline{Y_2}$, so that at $\widehat{C_1}$ the individual must be saving (see footnote 4).

[7] Other specifications of the borrowing constraint might respond differently to a decrease in R, or might be determined jointly with R.

Where the last equality follows from the first order condition. It is likely that $\frac{B_1^*}{R}u'(C_1^*) > \frac{\overline{B_1}}{R}u'(\overline{C_1})$, i.e. that the increase in value implied by maintaining the pre-change interior solution C_1^* in response to a decrease in R_B (which is not the optimal response) is higher than the increase in the value of the corner solution. To see this, note that the difference of the changes in value $\frac{\overline{B_1}}{R}u'(\overline{C_1}) - \frac{B_1^*}{R}u'(C_1^*)$ can be written as:

$$\frac{1}{R}\int_{C_1^*}^{\overline{C_1}} \frac{d}{dC_1}(C_1 - \check{Y}_1)u'(C_1)dC_1 = \frac{1}{R}\int_{C_1^*}^{\overline{C_1}} ((C_1 - \check{Y}_1)u''(C_1) + u'(C_1))dC_1$$

$(C_1 - \check{Y}_1)u''(C_1) + u'(C_1)$ is likely negative over the range $(C_1^*, \overline{C_1})$. To see this, note that this expression is negative when:

$$\frac{C_1 - \check{Y}_1}{C_1} > EIS(C_1) \tag{C.0.8}$$

where $EIS(C_1)$ is the elasticity of intertemporal substitution at C_1 and the LHS is the inverse of the fraction of consumption that is borrowed (this is identical to the condition in Proposition 9). For the high borrowing levels implied by consuming in the range $(C_1^*, \overline{C_1})$, it is likely that this condition holds[8] and that $\frac{B_1^*}{R}u'(C_1^*) > \frac{\overline{B_1}}{R}u'(\overline{C_1})$.

Since maintaining C_1^* in response to a decrease in R_B is generally not the optimal response for an individual who always chooses the interior solution (see Proposition 8), the value of the corner solution increases by less than does the value of the interior solution at $\check{Y}_1$, and the threshold is lowered.

The third point can easily be seen from the proof of Proposition 5. □

Proof of Proposition 8

Proof. As no individuals transition between saving and borrowing (see Proposition 7), we can derive the general effect of a change in the interest rate on consumption and apply it to savers and borrowers separately. I apply the implicit function theorem on the first order condition to get:

$$\frac{dC_1}{dR} = \frac{1}{RU''(C_1)}[opp\_cost + (C_1 - Y_1)\frac{d}{dC_1}opp\_cost] \tag{C.0.9}$$

[8] For example, for $CRRA$ preferences with a coefficient of relative risk aversion equal to 3, the condition in (C.0.8) holds for any C_1 value which implies that borrowing accounts for more than $\frac{1}{3}$ of consumption. This seems reasonable for consumption levels in this range, which are relatively very high.

which, incorporating the expression for MPC derived in Proposition 6 and the definition of the elasticity of intertemporal substitution $EIS(C_1) = -\frac{u'(C_1)}{C_1 u''(C_1)}$ can also be written as equation (2.5.1):

$$\frac{dC_1}{dR}\frac{R}{C_1} = -(1 - MPC)EIS + \frac{Y_1 - C_1}{C_1}MPC$$

The first term (the substitution effect) is always negative at an interior solution since $MPC < 1$ (see Proposition 6). The second term (the wealth effect) is positive for those who initially save (for whom $Y_1 > C_1$) and have positive MPC (savers with zero probability of default) as well as for those who initially borrow and have negative MPC (borrowers with positive probability of default). It is negative for those who initially borrow and have positive MPC (borrowers with zero probability of default (see footnote 44 in Chapter 2) as well as for those who initially save and have negative MPC (savers with positive probability of default). Parts 1 and 2 of the proposition directly follow.

Part 3 of the proposition follows from considering the wealth effects just described as well as the dependence of the magnitude of the substitution effect (the first term of (2.5.1)) on MPC.

$\square$

Proof of Proposition 9

Proof. The probability of default is $(\widehat{Y_2} - \underline{Y_2})$ and the expected defaulted-upon amount is:

$$E(Def) = Pr.(Def) * Debt = (\widehat{Y_2} - \underline{Y_2})Debt$$

The quantities that we are interested in are:

$$\frac{d\widehat{Y_2}}{dR} = \frac{d\widehat{Y_2}}{dDebt}\frac{dDebt}{dR}$$

and:

$$\frac{dE(Def)}{dR} = \frac{d\widehat{Y_2}}{dR}Debt + (\widehat{Y_2} - \underline{Y_2})\frac{dDebt}{dR} = \frac{dDebt}{dR}(\frac{d\widehat{Y_2}}{dDebt}Debt + (\widehat{Y_2} - \underline{Y_2}))$$

In the proof of Proposition 2, I show that $\frac{d\widehat{Y_2}}{dDebt} = \frac{u'(\widehat{Y_2}-Debt)-\Lambda'(Debt)}{u'(\widehat{Y_2}-Debt)-u'(\widehat{Y_2})} > 0$. It follows that the signs of both $\frac{d\widehat{Y_2}}{dR}$ and $\frac{dE(Def)}{dR}$ depend on the sign of:

$$\frac{dDebt}{dR} = (C_1 - Y_1) + \frac{dC_1}{dR}R$$

which is positive if and only if $\frac{C_1-Y_1}{C_1} > -\frac{dC_1}{dR}\frac{R}{C_1}$.

To derive the equivalent condition, plug in $\frac{dC_1}{dR}$ from (C.0.9) and use the first order condition $u'(C_1) = opp\_cost$ to get:

$$\frac{dDebt}{dR} = \frac{1}{U''(C_1)}[u'(C_1) + (C_1 - Y_1)u''(C_1)]$$

Which can be rewritten as:

$$\frac{dDebt}{dR} = C_1(1 - MPC)(\frac{C_1 - Y_1}{C_1} - EIS(C_1)) \tag{C.0.10}$$

where $EIS(C_1) = -\frac{u'(C_1)}{C_1 u''(C_1)}$ is the elasticity of intertemporal substitution at C_1, equal to $\frac{1}{\rho}$ for the case of $CRRA$ preferences with coefficient of relative risk aversion ρ.

Since $MPC < 1$ at any interior solution (see Proposition 6), (C.0.10) is always negative for initial savers (for which $C_1 - Y_1 < 0$). It follows that increasing R_S always decreases the probability of default as well as the expected defaulted-upon amount of initial savers. For initial borrowers, (C.0.10) is positive (decreasing R_B decreases expected default) for $CRRA$ preferences if and only if:

$$\frac{C_1 - Y_1}{C_1} > \frac{1}{\rho} \tag{C.0.11}$$

The derivative of the LHS (the fraction of C_1 that is borrowed) with respect to Y_1 is $\frac{MPC*Y_1 - C_1}{C_1^2}$, which is always negative for borrowers (for whom $Y_1 < C_1$) given that $MPC < 1$. It follows that the condition in (C.0.11) is a condition on initial wealth: with $CRRA$ preferences, individuals less wealthy than some threshold will have $\frac{dDebt}{dR} > 0$, and individuals wealthier than that threshold will have $\frac{dDebt}{dR} < 0$. Since borrowing decreases monotonically in initial wealth ($MPC < 1$ for interior solutions), this initial wealth threshold corresponds to an initial borrowing threshold.

The claim that the probability of default and the expected defaulted-upon amount decrease for those who start saving as a result of favorable rate changes and increase for those who start borrowing is trivial when considering $\frac{dDebt}{dR}$. $\square$

Proof of Proposition 10

Proof. The LHS of (2.6.3) is constant, while the RHS, which is the marginal cost of borrowing (equal to the opportunity cost of C_1), increases in B_1.[9]

Assuming that at $B_1 = 0$ (which implies $\widehat{Y_2} = \underline{C_2} + \bar{D}_0$) the RHS is lower than Λ_1:

$$\Lambda_1 > \beta R[\int_{\underline{Y_2}}^{\underline{C_2}+\bar{D}_0} \Lambda_2 d\tilde{Y_2} + \int_{\underline{C_2}+\bar{D}_0}^{\overline{Y_2}} u'(\tilde{Y_2} - \bar{D}_0)d\tilde{Y_2}]$$

it is clear that for borrowing levels $B_1 < \breve{B}_1$, the marginal cost of borrowing is lower than the marginal cost of delinquency Λ_1 and the constraint $Del_1 \geq 0$ is hit. For borrowing levels $B_1 > \breve{B}_1$, the marginal cost of borrowing is higher than Λ_1 and the constraint $Del_1 \leq \bar{D}_0$ is hit.

It follows that it is optimal to borrow up to $\breve{B}_1$ while paying $\bar{D}_0$ in full. That is, consumption levels $C_1 < Y_1 - \bar{D}_0 + \breve{B}_1$ involve only borrowing and no delinquency. Once $\breve{B}_1$ is borrowed, the marginal costs of borrowing and delinquency are equal, and it is optimal to finance additional consumption by being partially delinquent on $\bar{D}_0$ since higher levels of borrowing imply a higher marginal cost. Once delinquent on the full amount (at $C_1 = Y_1 + \breve{B}_1$), additional consumption can only be financed by additional borrowing. □

Proof of Proposition 11

Proof. Let us examine the opportunity cost of C_1 (the RHS of (2.6.5)) for the 3 delinquency-borrowing tradeoff regions defined in Proposition 10, noting that it is equal to the RHS of (2.6.3).

Region 1: $C_1 < Y_1 - \bar{D}_0 + \breve{B}_1$: Opportunity cost is increasing in this region, since B_1 increases in C_1.[10] It is lower than Λ_1.

Region 2: $Y_1 - \bar{D}_0 + \breve{B}_1 \leq C_1 < Y_1 + \breve{B}_1$: Opportunity cost is flat in this region, since B_1 does not change in C_1 ($\breve{B}_1$ is borrowed throughout the region). It is equal to Λ_1 (see (2.6.4)).

Region 3: $Y_1 + \breve{B}_1 \leq C_1$: Opportunity cost is increasing in this region since B_1 increases in C_1. It is higher than Λ_1.

[9] Where $\widehat{Y_2} \leq \underline{Y_2}$, i.e. where the ex-ante probability of delinquency is zero, the derivative of the RHS of (2.6.3) wrt B_1 (equal to the opportunity cost) is equal to $-\beta R^2 E(u''(\tilde{Y_2} - Debt)) > 0$. Where $\widehat{Y_2} > \underline{Y_2}$, this derivative is equal to $\beta R^2(\Lambda_{\underline{2}} - u'(\overline{Y_2} - Debt))$. Assuming that at the highest possible income realization $\overline{Y_2}$ it is always optimal to pay in full ($\overline{Y_2} - Debt > \underline{C_2}$), we have that $\Lambda_2 > u'(\overline{Y_2} - Debt)$ and the derivative is positive in this region as well.

That is, with a constant marginal cost of delinquency the opportunity cost curve increases in C_1 everywhere and the implications of a decreasing opportunity cost established above, such as negative MPC, do not occur.

[10] See footnote 9 for the derivative of the opportunity cost of C_1 wrt B_1

Since the RHS of (2.6.5) is equal to Λ_1 for C_1 levels that are in region 2, it is optimal to choose C_1 in this region if and only if $u'(C_1) = \Lambda_1$. That is, if and only if $C_1 = Y_1 + \breve{B}_1 - \bar{D}_0 + Del_1 = \underline{C_1}$, where $(u')^{-1}(\Lambda_1) \equiv \underline{C_1}$. Examining the two extreme C_1 values that define the region (for which $Del_1 = 0$ and $Del_1 = \bar{D}_0$), it follows that individuals with first period income $\underline{C_1} - \breve{B}_1 < Y_1 < \underline{C_1} + \bar{D}_0 - \breve{B}_1$ satisfy $u'(C_1) = \Lambda_1$ while borrowing $\breve{B}_1$. First period consumption in this region is $Y_1 + \breve{B}_1 + (Del_1 - \bar{D}_0) = \underline{C_1}$, regardless of the value of Y_1 ($MPC = 0$), and delinquency is equal to $C_1 - (Y_1 - \bar{D}_0 + \breve{B}_1)$.

Consider now individuals with first period income $Y_1 > \underline{C_1} + \bar{D}_0 - \breve{B}_1$. For these levels of income we have $u'(Y_1 - \bar{D}_0 + \breve{B}_1) < \Lambda_1$. It follows that for these levels of Y_1, first period consumption must be lower than implied by borrowing $\breve{B}_1$ and paying $\bar{D}_0$ in full (i.e. consuming $C_1 = Y_1 - \bar{D}_0 + \breve{B}_1$) in order to satisfy (2.6.5). It follows that individuals with these levels of first period income choose C_1 in the first delinquency-borrowing tradeoff region from Proposition 10, borrowing $B_1 < \breve{B}_1$ and paying $\bar{D}_0$ in full. They consume $Y_1 - \bar{D}_0 + B_1 > \underline{C_1}$.

For individuals with first period income $Y_1 < \underline{C_1} - \breve{B}_1$, we have $u'(Y_1 + \breve{B}_1) > \Lambda_1$. First period consumption must be higher than that implied by borrowing $\breve{B}_1$ and being delinquent on the entire $\bar{D}_0$ (i.e. consuming $C_1 = Y_1 + \breve{B}_1$) in order to satisfy (2.6.5). It follows that individuals with these levels of income choose C_1 in the third delinquency-borrowing tradeoff region from Proposition 10, borrowing $B_1 > \breve{B}_1$ and being delinquent on the entire $\bar{D}_0$. They consume $Y_1 + B_1 < \underline{C_1}$. $\square$

Appendix D

Model of OTC Stock Pricing

Our stylized model of OTC stock prices features costly short selling and differences in investors' opinions. We analyze the price of a single equity-financed firm in three periods: 0, 1, and 2. At date 0, the firm has assets in place normalized to \$1. At date 2, the firm liquidates all assets and pays all cash flows. The share price of the stock (p) endogenously adjusts to clear the market. We normalize the supply of stock to one and the return on the risk-free asset to zero.

We assume short selling costs are related to the cost of locating shares to borrow. Short sellers borrow shares from share lenders, such as brokers or custodians, who incur deadweight quadratic costs of finding shares $(c/2)(\text{shares lent})^2$, where $c > 0$. Share lenders pass these costs on to short sellers who can borrow shares and pay total dollar fees of $(c/2)(\text{shares short})^2$. Based on this total fee, the average borrowing fee per share is $f = f(\text{shares short}) = (c/2)(\text{shares short})$. This lending fee ($f$) is akin to a negative rebate rate earned on collateral posted to borrow shares. We assume share owners do not receive payment when share lenders lend their shares.

There are two types of risk-neutral overconfident investors and N investors of each type. Each investor owns $1/(2N)$ of the firm's shares at date 0. At date 1, investors observe two public signals, S_A and S_B, about the firm's date 2 earnings (π_2). Earnings satisfy $\pi_2 = S_A + S_B + u_1 + u_2$, where S_A, S_B, u_1 and u_2 are independently uniformly distributed from $[-\sigma, +\sigma]$ and $\sigma \geq 0$ is a measure of fundamental volatility. Stockholders receive $1 + \pi_2$ at date 2.

The two types of investors differ in which signal they believe more, where the parameter $\eta \in [0, 1]$ represents agents' overconfidence in their preferred signal. Specifically, the investors mistakenly perceive the u_t components of earnings to be correlated with their preferred signals. Type $X \in \{A, B\}$ believes that these components of earnings satisfy $u_t = \eta S_X + (1 - \eta^2)^{1/2} \nu_t$, where $t = 1$ or 2, and the ν_t are uniformly distributed from $[-\sigma, +\sigma]$ and independent of each other, S_A, and S_B. Both types' beliefs are correct if and only if $\eta = 0$.

We consider two variants of the model: one in which the firm publicly discloses financial information ($e_1 = S_A + S_B + u_1$) about date 2 earnings at date 1, and one without such disclosure. We denote the date 1 earnings beliefs of investor type $X \in \{A, B\}$ by E_X. Based on only the two signals, the rational expectation of the firm's date 2 earnings is $S_A + S_B$. At date 1, investors' earnings expectations in the cases with and without financial disclosure are given by:

No disclosure:

$$\begin{aligned} E_A &= (1 + 2\eta)S_A + S_B \\ E_B &= (1 + 2\eta)S_B + S_A \end{aligned} \tag{D.0.1}$$

Disclosure:

$$\begin{aligned} E_A &= (1 + \eta)S_A + S_B + u_1 \\ E_B &= (1 + \eta)S_B + S_A + u_1. \end{aligned} \tag{D.0.2}$$

Define the difference in opinion between investors to be $DO = |E_A - E_B||$. From the above expressions, financial disclosure decreases difference in opinion as follows:

No disclosure:

$$DO = 2\eta|S_A - S_B| \tag{D.0.3}$$

Disclosure:

$$DO = \eta|S_A - S_B|. \tag{D.0.4}$$

For simplicity, we analyze the model's symmetric rational expectations equilibrium in which each investor takes the market price as given and investors within each type use the same strategies. Type $X \in \{A, B\}$ chooses q_X at date 1 to maximize expected profit, implying

$$q_X \in argmax\{q_X(1 + E_X – p_1) – I(q_X < 0)(c/2)q_X^2\}, \tag{D.0.5}$$

where $I(\ )$ is an indicator function. The more optimistic type, for which $E_X = max(E_A, E_B)$, chooses a long position, has a linear profit function, and buys stock until the price satisfies

$$1 + max(E_A, E_B) – p_1 = 0. \tag{D.0.6}$$

This condition implies the stock price reflects *only* the beliefs of the optimistic investors:

$$p_1 = 1 + max(E_A, E_B) = 1 + (E_A + E_B)/2 + DO/2. \tag{D.0.7}$$

Because prices reflect optimistic investors' beliefs, the pessimistic investor type chooses to short the stock and has a quadratic profit function. The pessimistic type's demand satisfies

$$min(q_A, q_B) = (1 + min(E_A, E_B) – p_1)/c = –DO/c < 0 \quad if \quad \eta > 0. \tag{D.0.8}$$

The second-order condition for pessimistic investors is satisfied because their expected profit is quadratic in q_X and $–c < 0$. Optimistic investors are also maximizing because their expected profit is zero for all $q_X > 0$. Market clearing $[N(q_A + q_B) = 1]$ implies optimists' demand is

$$max(q_A, q_B) = 1/N + DO/c \quad if \quad \eta > 0. \tag{D.0.9}$$

The resulting average stock lending/borrowing fee per share is

$$f = (c/2)|min(q_A, q_B)| = DO/2. \tag{D.0.10}$$

In expectation, the equilibrium price at date 1 (p_1) exceeds the efficient price (p_{1e}) that would prevail if there were no overconfidence. The efficient price is

No disclosure:

$$p_{1e} = 1 + S_A + S_B \tag{D.0.11}$$

Disclosure:

$$p_{1e} = 1 + S_A + S_B + u_1. \tag{D.0.12}$$

We define overpricing (Ovp) as the equilibrium price minus the efficient price (p_1–p_{1e}):

No disclosure:

$$Ovp_1 = 2\eta max(S_A, S_B) \tag{D.0.13}$$

Disclosure:

$$Ovp_1 = \eta max(S_A, S_B). \tag{D.0.14}$$

At date 0, before the signals are known, expected overpricing is

$$E[Ovp_1] = E[DO]/2 > 0 \quad if \quad \eta > 0. \tag{D.0.15}$$

At date 0, all investors anticipate the date 1 equilibrium, so the price is

$$p_0 = 1 + E[DO]/2 > 1 \quad if \quad \eta > 0. \tag{D.0.16}$$

The date 0 price is higher than its efficient value of 1 because expected overpricing is positive due to expected differences in opinion. As a consequence of overpricing, at date 0 the stock's expected return $E[r]$ from date 1 to date 2 is negative and given by

$$E[r] = E[p_2–p_1] = –E[DO]/2 < 0 \quad if \quad \eta > 0. \tag{D.0.17}$$

Expected return decreases with expected difference in opinion, which arises from overconfidence. The overconfidence bias causes the stock's expected return to be lower than the risk-free rate of zero even though investors are risk-neutral.

Equilibrium trading volume from date 0 to date 1 is:

$$Volume = |Nmax(q_A, q_B)–1/2| = 1/2 + (N/c)DO \quad if \quad \eta > 0. \tag{D.0.18}$$

where 1/2 is the initial share endowment of type A investors. Expected trading volume is thus:

$$E[Volume] = 1/2 + (N/c)E[DO] \quad if \quad \eta > 0. \tag{D.0.19}$$

Return volatility at date 1 is the standard deviation of the change in price, which is

$$\sqrt{Var[p_1 - p_0]} = (1/2)\sqrt{Var[E_A + E_B + DO]} = (1/2)\sqrt{Var[E_A + E_B] + (1/2)E[DO]^2} \tag{D.0.20}$$

where the second equality is based on the properties of the two uniformly distributed signals.

In summary, ex ante overpricing increases with expected difference in opinion, which is consistent with Miller (1977) and related theories. The equilibrium relies on the assumptions that the cost of short selling is positive ($c > 0$) and convex and that investors are overconfident ($\eta > 0$). Firm disclosure of financial information reduces differences in investors' opinions.

We now establish seven model predictions based on the above equilibrium.

Proposition 12. *If $\eta > 0$, expected return is negative and decreases with expected difference in opinion. If $\eta = 0$, expected return is zero, and an equilibrium with no trading exists.*

Proof. If $\eta > 0$, expected return is $–E[DO]/2$, so it decreases with$E[DO]$. If $\eta = 0$, then $DO = 0$ regardless of disclosure; and all traders believe firm value is $1 + E_A = 1 + E_B$, so this must be the equilibrium date 1 price. In this case, the price $p_1 = 1 + E(\pi_2)$ is efficient and equal to $E(p_2)$, implying that expected return is the risk-free rate of zero. At the price p_1, all traders are content to hold their initial endowments, implying an equilibrium with no trading exists. □

Proposition 13. *If $\eta > 0$, expected trading volume increases with expected DO and is thus negatively related to expected return.*

Proof. If $\eta > 0$, $E[Volume] = 1/2+(N/c)E[DO]$, which increases with DO. By substituting $E[r] = –E[DO]/2$, we obtain $E[Volume] = 1/2–(2N/c)E[r]$, which shows the negative relation. □

Proposition 14. *If $\eta > 0$, an increase in σ leads to an increase in expected DO, a decrease in expected return, and an increase in return volatility.*

Proof. E[DO] is proportional to $\eta E[|s_A–s_B|] = (2/3)\eta\sigma$, where the equality is based on the expectation of a random variable with a uniform difference distribution $[(2/3)\sigma]$. Thus, E[DO] increases proportionally with σ. Because expected return is $–E[DO]/2$, it decreases proportionally with σ. Return volatility is proportional to σ because both the $E[DO]^2$ term and the $Var[E_A + E_B]$ in the return variance expression in Equation (D.0.20) are proportional to σ^2, and volatility is the square root of variance. □

Proposition 15. *If $\eta > 0$, market equity (M) and the ratio of market-to-book equity (M/B) increase with expected DO and thus size and M/B are negatively related to expected return.*

Proof. Because the firm's book value is 1, its $M = M/B = p_0 = 1 + E[DO]/2$. Thus, M/B and M depend linearly on E[DO], which is negatively related to expected return. □

Proposition 16. *The average stock lending fee per share (f) increases with expected DO and is negatively related to expected return.*

Proof. From Equation (D.0.10), the average lending fee is $f = DO/2$, which increases proportionally with DO. Expected return decreases with E[DO] and thus with the lending fee. □

Proposition 17. *An increase in overconfidence (η) increases expected DO and decreases expected return. In addition, higher η amplifies each of the effects in Propositions 1 to 5.*

Proof. Regardless of disclosure, expected DO is proportional to $\eta E[|S_A–S_B|]$, which increases with η. Expected return is –E[DO]/2, which must decrease with η. Because the effects in Propositions 1 to 5 all rely on the expression for expected DO and this expression increases with η, an increase in η amplifies each of these effects. □

Proposition 18. *Expected difference in opinion is higher and expected return is lower with no firm disclosure; and a lack of disclosure amplifies the effects in Propositions 1 to 5.*

Proof. From Equations (D.0.3) and (D.0.4), non-disclosure increases DO by $\eta|S_A–S_B|$ and increases E[DO] by $\eta E[|S_A–S_B|]$. Because Propositions 1 to 5 rely on the expression for expected DO, which decreases with disclosure, a lack of disclosure amplifies these effects. □

The model delivers several intuitive results. Proposition 12 shows that difference in opinion (DO) decreases expected return if agents are overconfident ($\eta > 0$). If agents are not overconfident, the model predicts no trading and no overpricing because agents agree on the firm's value. Thus, Proposition 12 formally justifies our PNT (non-trading) proxy for no DO and its positive relation with expected return. Proposition 13 extends this idea to trading volume. An increase in expected DO increases expected shorting demand from the pessimistic investor type, which generates high trading volume. Because agents trade more when they disagree more and disagreement causes overpricing, stocks with high volume tend to be more overpriced.

Propositions 14 and 15 show that expected differences in opinion are also positively related to return volatility, firm size, and firms' ratios of market-to-book equity. Intuitively, an increase in the firm's fundamental volatility (σ) increases return volatility and expected DO because the public signals that generate disagreement are more volatile. In addition, an increase in expected DO increases overpricing and thus the firm's market capitalization, justifying size as a proxy for DO. Similarly, an increase in expected DO produces a higher stock price, holding book value constant, thereby raising the firm's M/B ratio, which justifies M/B as a proxy for DO. In this stylized model, size and M/B are the same because book value is normalized to 1. Allowing firms' book values (B) to differ would generate cross-sectional variation in M/B ratios and overpricing even among firms with identical size (M).

Proposition 16 shows that markets with higher lending fees, such as OTC markets, will exhibit larger overpricing. This proposition is consistent with studies such as D'avolio (2002) that interpret lending fees as arising from differences in investors' opinions.

Proposition 17 shows that an increase in investors' overconfidence (η) increases DO because disagreement results from placing excessive weight on different public signals. This overconfidence channel justifies DO proxies based on retail trading if retail traders are especially prone to overconfidence. In addition, Proposition 17 implies that stocks held primarily by retail investors are more subject to the overpricing effects in Propositions 12 to 16. This motivates our double-sorting methodology in which the initial sort is based on the presence of institutional (non-retail) investors.

Proposition 18 shows that a lack of firm disclosure increases differences in opinion because investors agree on how to interpret basic financial disclosures made by the firm. As a result, non-disclosure is associated with higher overpricing. Intuitively, lack of disclosure increases the uncertainty over which investors can disagree, which increases expected overpricing. Furthermore, non-disclosure amplifies the overpricing effects in Propositions 12 to 16, which motivates our double sorts using disclosure.

Appendix E

Estimating Betas and Accounting for Nonsynchronous Trading

To estimate a stock's betas in month t on return factors, we use a time series regression of the stock's monthly return on the monthly return factors from month t–24 to month t–1. In cases in which a stock is not traded for one month or longer, we cumulate monthly factors during the entire non-trading period to align the stock and factor returns. We compute stocks' betas on the MKT, SMB, and HML factors using the three-factor Fama and French (1993) regression. We compute betas with respect to the UMD momentum factor constructed by Kenneth French, which was originally used by Carhart (1997), and the illiquidity factor (ILQ) of Pástor and Stambaugh (2003) using regressions of returns on MKT, SMB and HML in addition to the respective factor. We require at least 10 observations in each regression.

Because many OTC stocks do not trade every day, we correct stocks' raw betas for non-synchronous trading by extending the method in Lo and MacKinlay (1990). Suppose that the unobservable, "true" return process for stock i is

$$R_{it} = \alpha_i + F_t\beta_i + \varepsilon_{it} \tag{E.0.1}$$

where F_t is a $1 \times m$ vector of factor returns. The econometrician only observes prices and returns in periods when trading occurs. We denote the probability that stock i does not trade by p_i and assume this probability is constant across periods. If a security does not trade for several periods, the observed return when it eventually does trade is the sum of all unobserved true returns per period. Formally, we define a variable $X_{it}(k)$ as follows:

$$X_{it}(k) = \begin{cases} 1 & \text{if stock } i \text{traded in period } t \text{but did not trade in all } k \text{periods prior to } t \\ 0 & \text{otherwise} \end{cases}$$

This definition implies that $X_{it}(k) = 1$ with probability $(1 - p_i)p_i^k$. Now we can write the observed return process (R_{it}^o) as

$$R_{it}^o = \sum_{k=0}^{\infty} X_{it}(k) R_{it-k}. \tag{E.0.2}$$

We assume that factor returns (F_t) are independent and identically distributed over time with $E(F_t) = \mu_F$ and

$$Var(F_t) = \Sigma_F = \begin{pmatrix} \sigma_1^2 & . & . & . & \sigma_{1m} \\ . & . & & & . \\ . & & . & & . \\ . & & & . & . \\ \sigma_{m1} & & & & \sigma_m^2 \end{pmatrix} \tag{E.0.3}$$

We estimate regressions of observed monthly returns on observed monthly factors. The observed beta vectors that we estimate are

$$\beta_i^o = [E(F_t^{o'} F_t^o) - E(F_t^{o'})E(F_t^o)]^{-1}[E(F_t^{o'} R_{it}^o) - E(F_t^{o'})E(R_{it}^o)]. \tag{E.0.4}$$

Simplifying and rearranging Equation (E.0.4) yields a relation between stock i's true beta and its observed beta and alpha:

$$\beta_i = \beta_i^o - \frac{2p_i}{1-p_i}\alpha_i^o[1 - \frac{2p_i}{1-p_i}\mu_F(\Sigma_F + \frac{2p_i}{1-p_i}\mu_F^{'}\mu_F)^{-1}\mu_F^{'}]^{-1}(\Sigma_F + \frac{2p_i}{1-p_i}\mu_F^{'}\mu_F)^{-1}\mu_F^{'} \tag{E.0.5}$$

When F_t is a scalar, such as an intercept in a factor regression, this formula simplifies to

$$\beta_i = \beta_i^o - \frac{2p_i}{1-p_i}\alpha_i^o \frac{\mu_F}{\sigma_F^2} \tag{E.0.6}$$

We obtain the parameters required for computing β_i as follows. First, we estimate the observed betas and alphas (β_i^o and α_i^o) for each firm for each month with regressions using the 24 previous months. Next, we estimate the factor means and covariances (μ_F and Σ_F) for each regression during the same 24 months. Lastly, we estimate the probability of a stock not trading p_i using the proportion of months in which the stock did not trade during the regression period. We then substitute these parameter estimates into Equation (E.0.6) to estimate stock i's true beta.