ESSAYS ON LABOR ECONOMICS AND PUBLIC POLICY

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Michael R. Strain

May 2012

ABSTRACT

Essays on Labor Economics and Public Policy

Michael R. Strain, Ph.D.

Cornell University 2012

The main component of this thesis, found in the first chapter, is an investigation of earnings instability, which can be thought of as the fluctuations around permanent earnings over time of a worker's labor market earnings. This chapter reflects my interests in labor economics and in economic analysis using longitudinal, linked worker-firm data.

The instability of labor earnings in the United States contributes to earnings inequality and may diminish household welfare. Despite the importance of earnings instability little is known about its correlates or causes. This study seeks to better understand whom earnings instability affects and why it affects them. Using both parametric and semi-parametric techniques, I provide an in-depth investigation into the relationship between earnings instability and worker skill. I find that earnings instability follows a U-shape over skill, with low-skill workers experiencing the least stable earnings, middle-skill workers experiencing the most stable, and high-skill workers falling in between the two. This finding is robust to a number of controls, sample selections, and other statistical concerns, and is not driven by workers entering and leaving employment, changing jobs, or holding multiple jobs. I then investigate whether firm characteristics affect the stability of worker earnings. I am the first to directly test the relationship between earnings instability and firm employment instability using linked employer-employee data. I find a positive and statistically significant relationship between the two that remains when the effect is estimated using only within-firm variation. This suggests that the effect is a feature of the way

workers are being paid by their employer. The size of the effect varies by a worker's position in the earnings distribution: low-earning worker are passed a greater share of firm employment instability than higher-earning workers. This finding helps to explain the left tail of the U-shape of earnings instability over skill. I find significant heterogeneity in the magnitude and significance of the effect across industries and explore how the competitiveness of an industry relates to the size of the industry-specific effect.

My interest in the economics of education is reflected in the second essay of this thesis, which studies a public policy innovation using administrative records. The effects of single-sex education are hotly contested, both in academic and policy circles. Despite this heated debate, there exists little credible empirical evidence of the effect of a U.S. public school's decision to offer single-sex classrooms on the educational outcomes of students. This study seeks to fill this hole. Using administrative records for third through eighth graders in North Carolina public schools, the chapter finds evidence that the offering of single-sex mathematics courses is associated with lower performance on end-of-grade math exams, and finds no evidence that the offering of single-sex reading scores increases performance on reading exams. Evidence of significant heterogeneity in the effect across schools is also presented.

Finally, my interest in public policy is further reflected in the third chapter of this thesis, coauthored with Donald Morgan and Ihab Seblani. Despite a dozen studies, the welfare effects of payday credit are still debatable. We contribute new evidence to the debate by studying how payday credit access affects bank overdrafts (such as returned checks), bankruptcy, and household complaints against lenders and debt collectors. We find some evidence that Chapter 13 bankruptcy rates decrease after payday credit bans, but where we find that, we also find that complaints against lenders and debt collectors increase. The welfare implications of these offsetting movements are unclear. Our most robust finding is that returned check numbers and overdraft fee income at banks increase after payday credit bans. Bouncing a check may cost more than a payday loan, so this finding suggests that payday credit access helps households avoid costlier alternatives. While our findings obviously do not settle the welfare debate over payday lending, we hope they resolve it to some extent by illuminating how households rearrange their financial affairs when payday loan supply changes.

In summary, this thesis nicely reflects my interests in labor economics, public policy, economic analysis using linked and administrative data, and education economics, and the econometric and research skills I have acquired during my five years as a graduate student in economics at Cornell.

BIOGRAPHICAL SKETCH

I was born in Kansas City, and was lucky enough to have a mother and father who valued education and who had as their goal that I live a good and full life. They encouraged my interest in books as a young boy, spent a lot of time reading to me, made sure that I had enough to read on my own when the time came, and sent me to good elementary and secondary schools. One of the best things they did for me was to send me to Rockhurst, the Jesuit preparatory school in Kansas City. I owe quite a bit to the Jesuits. Among many things, they taught me that knowledge, like life, has a just and proper end, and that we should pursue as complete an understanding of ourselves and society as possible. That they valued knowledge and education so much helped me to do the same.

I continued my association with the Jesuits as an undergraduate at Marquette University. I stumbled into economics through a couple of chance experiences. Looking back on it, I was hooked during the first five minutes of the first economics class I took. I had some great friends who studied economics with me, which made it all the better, and I had some absolutely phenomenal professors. I majored in history as well as economics, and minored in philosophy. I took way too few courses in mathematics, as I painfully discovered at Cornell.

I went off to New York City after college, pursued a terminal master's degree in economics at New York University and worked in the research group of the Federal Reserve Bank of New York. It was there that I decided to pursue a doctorate, working with great economists who were deeply interested in how the actual world works and in what that understanding can do to create better policy.

I decided to come to Cornell for my Ph.D., and looking back on it I made the right decision. I started working for the U.S. Census Bureau at the end of my first year of graduate school, and that work constituted a large share of my Ph.D. experience. It is reflected in this thesis. Five years later I am putting the finishing touches on this document, wrapping up at Census and Cornell, eager to discover what's next.

To my teachers.

ACKNOWLEDGMENTS

I should like to begin by thanking John Abowd, the man who made it all possible. Though we have never explicitly discussed it, I am very confident that it was Professor Abowd who let me into the Cornell Ph.D. program in economics. I know that it was Professor Abowd who paid my tuition and health insurance for four years at Cornell, and I certainly owe him thanks for that. And, especially in my first years, John was one of the only faculty members at Cornell who believed in me and who was willing to go to the mat for me, and I will always remain deeply grateful to him. John always treated me as if I knew more than I actually did, was more skilled and competent and capable than I actually was. This created a great challenge which pushed me to be better and to work harder. The effort to meet John's expectations made this document much better than it would have otherwise been.

Professor Abowd is a true innovator in economics — and that is even harder to be than it sounds. His contributions to social science, information science, and statistics will be recognized for many years to come. I am proud to say that I played a small role in advancing his research program during my years at Cornell and at the Census Bureau. But the fact remains that I got much more out of John's long work in creating and analyzing linked data than I contributed to it. I selfishly hope to continue to benefit from John's vision, labor, and leadership in the years to come.

I thank Kevin Hallock very, very much. Professor Hallock is an outstanding economist. He encouraged me in my third semester of graduate school to always make sure that I know something about what I am studying beyond what is in my dataset. This is a lesson that many economists never learn, and it is something that I aspire to in my research and aspired to in this thesis. The investment I made to follow his advice during my years at Cornell paid large dividends when I was on the job market. Professor Hallock's strong desire to know how things work both in the data

and in the actual world allowed him to be the inaugural recipient of the Joseph R. Rich '80 Chair, established to recognize a scholar who makes outstanding contributions to compensation studies. I was very happy to have been one of Kevin's students when he was awarded this great honor.

Looking back on it, Professor Hallock spent way too much time reading early drafts of my papers and giving me extensive and invaluable feedback. He also spent way too much time meeting with me to talk about all sorts of things that were on my mind, research-related and otherwise. Kevin may be the most decent, down to earth, just, empathetic, and generous person in galactic history. A stretch? Perhaps, but not by much, and it sure felt true countless times during the production of this thesis and my years at Cornell. If there were any justice in the world then Kevin would be running it. Without Kevin's support and advice and encouragement I probably would have concluded years ago that my best strategy would have been to sprint towards the nearest exit.

I thank Ronald Ehrenberg. When I was admitted to Cornell I told my colleagues at the New York Fed, and one of them told me that when I saw Professor Ehrenberg I should tell him that though they had never met he is the reason my colleague went into economics. That story nicely captures the effect that Professor Ehrenberg has had on me and on dozens and dozens and dozens of economists and thousands of economics undergraduates — many of whom only know him through his fantastic textbook.

Professor Ehrenberg is one of the great labor economists — during my time at Cornell he was given the lifetime achievement award by the Society of Labor Economists (of which Professor Abowd will soon be president) for his decades of world-class scholarship — but I think of Ron first as a teacher and a mentor. One of the most important lessons Ron taught me is that any list of the most important things

in life does not include the length of your CV. For a graduate student, whose world can easily shrink to the size of the economics department, these were welcome words to hear. Ron's openness and honesty and willingness to let his students know his own struggles, challenges, and doubts have made so many of us all the more confident that we would overcome our own. I was very lucky that Ron took both a personal and professional interest in me, and that I can call him a teacher and a mentor.

That's it for the thesis committee, but I was very lucky at Cornell to have received help and advice from more than three members of the faculty. I thank George Jakubson. Professor Jakubson is one of my favorite people at Cornell. He is one of the best applied econometricians around. He gave me great advice and a lot of help on the statistics that I used in this thesis. Maybe more than that, he gave me confidence that what I was doing wasn't stupid and wrong — I felt safe enough with George to tell him where the bodies were buried, and he told me that relative to the rest I was not populating a large graveyard.

I sincerely thank Larry Blume. Professor Blume is my ideal of an economics professor — well, at least one of my ideals! — right down to the corduroy pants and cable knit sweaters. With the possible exception of my friend and colleague Kiel Albrecht, he is the most intuitive economist I have ever met. Despite his handful of *Econometrica* publications, leading Ph.D. textbook, Econometric Society fellowship, and chair, Larry treated me and the other graduate students in his orbit exactly as he would treat a Nobel laureate — we were judged exclusively by the quality and substance of the words coming out of our mouths. I'm very glad I challenged myself and took his Ph.D. elective on networks, and am even more glad that the way we learned in his course was by sitting at a table and talking about papers — agreeing only when absolutely necessary. ("Mike, as an empirical economist your job isn't to tell me why you think this theory is stupid! Your job is to tell me how to estimate the parameter!" I'm lucky that I was usually able to meet that challenge.) Those discussions gave us a window into how Larry thinks — the most important lesson he could have taught his students. I greatly enjoyed and benefited from the many nights I spent with Larry drinking beer and talking into the early morning hours. Our conversations were great, but our disagreements and arguments were even better. And in many ways, our arguments about things outside economics were the best of all.

I thank Professor Kirabo Jackson for among the most useful and practical models of how to put together an economics research paper. Without Bo, the second chapter of this thesis would have been much harder for me to write. Professor Jackson was extremely generous and read everything in that chapter from start to finish, even including my NCERDC project proposal. Bo is a great guy, and I'm happy to call him a friend.

I thank Professor Francesca Molinari. Francesca is simply awesome. As with Professor Blume's course, I am very lucky that I manned up and took her Ph.D. elective. Without her course and without her help, the first chapter of this thesis would not include any semiparametric econometrics. She is a large part of the future of Cornell economics, and the future looks bright. She read my paper and she met with me and she cared whether what I was doing was correct and she cared about my success. She also cared about my happiness, and encouraged me — very publicly! — to put my work aside and find a woman! And she threatened to call my mother — again, very publicly! — when I jokingly told her my strategy for scoring drugs in inner cities. (It's a good strategy.)

I thank Jordan Matsudaira for his surpassingly useful and practical course. I thank Matt Freedman for talking to me about what I was doing and giving me advice and encouragement along the way. Professor Freedman was very generous and read the first chapter of this thesis in its entirety. I thank Dan Benjamin for his encouragement during the mandatory seminar. And I thank Emily Owens for her concern about my mental health, particularly as I was about to enter the job market.

Someone told me before I came to Cornell that in a Ph.D. program you learn more from your classmates than from the faculty. It's hard to judge, but I can say that I benefited very much from my classmates and friends here at Cornell. Without them, it's hard to imagine that this document would exist.

I thank Ian Schmutte, who played the Virgil to my Dante here at Cornell, helping me to navigate successfully the often-strange and confusing and sometimes-hostile world of graduate school. In addition, Professor Schmutte provided invaluable comments on the first chapter of this thesis, particularly on the writing of the introduction. Ian left Cornell at the end of my third year of graduate school, as did Professor Jessica Bean, who taught me by her marvelous example that you can have fun and be a graduate student, but never at the same time.

I thank my friend Eamon Molloy, who is one of the few people to have actually read every word in this thesis more than once. Eamon didn't like me at first, but my wit, charm, and good looks soon proved irresistible, and our friendship has been a professional and personal highlight of my time at Cornell. Dr. Molloy helped make this thesis much better than it would have been, and I am glad that he and I will both be working and living in Washington next year. I thank Chris Handy, without whom I may very well still be in the first year of this program. I thank the aforementioned Kiel Abrecht, conversations with whom taught me more useful microeconomic theory than anything I learned in the first year. I hope that the world can benefit from Kiel's considerable talent in the years to come. I thank Kevin Roth, who is just a good guy and a bright and talented economist. Countless drinks and conversations with Kevin helped to make me realize that I wasn't alone in my misery. And, especially in our first year, it was nice to be able to spend time with someone whose life experience consisted of more than midterms and problem sets. I thank Doug Webber and Catherine Maclean, and am glad to have gotten to know them and am more glad that they got to know each other, and marry. I thank Peter Brummund for being an all-around good guy and for introducing me to my favorite Brummund, his son Jake. And I thank Carolyne very much, who for the last year held my hand and assured me that it would all be over soon.

I thank my coauthors and collaborators. First, Don Morgan. Don and I coauthored the third chapter of my thesis, along with Ihab Seblani. I spent a lot of time on that chapter in my second and third years of graduate school, and I learned so much about doing research from working with Don. My collaboration with Dr. Morgan helped make the writing of the first and second chapters much easier. I hope to collaborate with Don in the future. I thank my collaborator and friend Alex Rees-Jones. Alex has talent and ambition, and I predict great things. Hopefully I can benefit from his success. I am happy to have Ian Schmutte as a collaborator as well, and my friend Ben Ost. When we started together at Cornell I wanted to finish graduate school in four years and move to Chicago, and Ben wanted to stay in graduate school for six years and never leave Ithaca. Well, Ben finished his Ph.D. in four years and is now a professor in Chicago, and I feel like I've been in Ithaca my entire life.

I thank my colleagues at the U.S. Census Bureau. My friend Chen Zhao read every word in the first chapter of this thesis, joining a small club. Chen was great to have in the office every day, and was actually a pleasure to work with on Census projects. My friend Liliana Sousa provided great company, particularly during the memorable summer of the education impute, and also gave me hope that one day my Ph.D. experience would actually end. I'm glad to be joining Liliana and Mario in Washington next year. I worked with Lars Vilhuber during my first year of graduate school, and being thrown into the deep end of the pool so early actually proved to be a helpful experience. Erika McEntarfer was my boss for a while, and her help and support with the production of the first chapter of this thesis are very appreciated. And last but certainly not least, I thank Kevin McKinney. Dr. McKinney taught me that if thirty records out of eight billion in your dataset don't behave exactly as you would expect, then it is important, goddammit, to dig in deep and figure out why. In seriousness, I learned as much about using SAS and working with data in general and about the LEHD data specifically — used in the first chapter of this thesis —from Kevin as from any other person. Kevin's care and attention to detail are lacking in too many economists, and his example helped me to understand the importance of knowing what's going on and doing things right. Kevin was usually very patient with me and my mistakes, and I enjoyed our conversations about everything under the sun — over video, with him in Los Angeles and me in Ithaca — very much. One day I hope to be able to tell him why those thirty records are a little weird.

Economics understands the importance of initial conditions, and so in acknowledging people who helped with the production of this thesis I must thank my former colleagues at the New York Fed, my teachers from over the years, the great essayists and polemicists and novelists I have been lucky to have read, the Jesuits, my friends, my mother and father, and, of course, the Lord. It was a fluke of scheduling that my thesis defense was held on Good Friday. While the literal comparison is *prima facie* absurd, I will say that I am glad I met a better fate that day than the man to whom all thanks are owed.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

UNDERSTANDING EARNINGS INSTABILITY:
SKILLS AND EMPLOYERS[1]

Michael R. Strain[2]

ABSTRACT

The instability of labor earnings in the United States contributes to earnings inequality and may diminish household welfare. Despite the importance of earnings instability little is known about its correlates or causes. This study seeks to better understand whom earnings instability affects and why it affects them. Using both parametric and semi-parametric techniques, I provide an in-depth investigation into the relationship between earnings instability and worker skill. I find that earnings instability follows a U-shape over skill, with low-skill workers experiencing the least stable earnings, middle-skill workers experiencing the most stable, and high-skill workers falling in between the two. This finding is robust to a number of controls, sample selections, and other statistical concerns, and is not driven by workers entering and leaving employment, changing jobs, or holding multiple jobs. I then investigate whether firm characteristics affect the stability of worker earnings. I am the first to

---

[1] Any opinions and conclusions expressed herein are those of the author and do not necessarily represent the views of the U.S. Census Bureau. All results have been reviewed to ensure that no confidential information is disclosed. This research uses data from the Census Bureau's Longitudinal Employer Household Dynamics Program, which was partially supported by the following National Science Foundation Grants
[2] Department of Economics, Cornell University, Ithaca, NY 14853. Author’s email: mrs274@cornell.edu

directly test the relationship between earnings instability and firm employment instability using linked employer-employee data. I find a positive and statistically significant relationship between the two that remains when the effect is estimated using only within-firm variation. This suggests that the effect is a feature of the way workers are being paid by their employer. The size of the effect varies by a worker's position in the earnings distribution: low-earning worker are passed a greater share of firm employment instability than higher-earning workers. This finding helps to explain the left tail of the U-shape of earnings instability over skill. I find significant heterogeneity in the magnitude and significance of the effect across industries and explore how the competitiveness of an industry relates to the size of the industry-specific effect.

I. Introduction

The earnings of American workers have grown significantly more volatile since the 1970s. This *earnings instability* — the fluctuations over time in a worker's earnings — plays an important role in earnings inequality: the rise in earnings instability accounts for nearly one-third of the increase in earnings inequality from 1970 to 2004 (Moffitt and Gottschalk, 2008). Earnings instability lowers household welfare because risk averse households prefer stable to unstable earnings, even if average earnings are the same. There is evidence that households have trouble smoothing consumption in the face of earnings instability (Dynarski and Gruber, 1997 and Gorbachev, 2011). Households of lesser means may only be able to partially insure their consumption against transitory income shocks (Blundell *et al.*, 2008). Income volatility is associated with volatile food consumption, particularly for traditionally-vulnerable households (Gorbachev, 2011).

Despite the importance of earnings instability, little empirical evidence has been documented on its causes and correlates — most of the research to date has focused on documenting the trends in aggregate earnings instability.[3] With this study I add to our understanding of earnings instability by providing evidence of whom earnings instability affects and why it affects them. Specifically, I examine whether workers of differing skill levels have differing amounts of time-series variance in their earnings. The results from this analysis suggest that the relationship between earnings instability and skill may be a function of the way that workers are paid by their employing firms, as opposed to being driven by workers entering and leaving employment, changing jobs, or holding multiple jobs. This finding invites the second component of my paper, wherein I investigate the relationship between earnings instability and an important firm characteristic: the instability of firm employment. To carry out this investigation I use linked employer-employee data for the U.S. labor market — the first time linked data have been used to estimate this relationship.

Earnings instability is a complex and interesting issue because it may mean different things to different workers. The risk preferences of higher-skilled workers and the compensation schemes into which they select may suggest that earnings instability and skill are positively related. At the same time, if earnings instability is driven by a breakdown in implicit contracts fueled by increasingly competitive global markets, then one might conjecture that lesser-skilled workers would have less stable earnings as labor markets increasingly operate like spot markets.

[3] There have been some studies in this literature which do not focus on the trends. For example, changing jobs does affect the stability of a worker's earnings, but the trends found in the data remain even when only workers who do not change jobs are studied (Gottschalk and Moffitt, 1994). Job instability has been linked to earnings instability — displaced workers experienced substantially increased earnings instability in the years following the loss of their job (Huff Stevens, 2001). Comin *et al.* (2009) find evidence that the instability of the average wage paid by a firm is associated with firm employment and sales instability, and interpret this as evidence that firms are passing volatility in their sales and employment to their workers in the form of more volatile wages.

My study of earnings instability and skill begins by drawing on simple models in the compensation literature. Workers of different skill level sorting into different compensation schemes can naturally generate a relationship between skill and earnings instability — differences in skill across workers predict not only differences in *mean* earnings, but also differences in the *variance* of earnings over time.

A simple extension to a model from Lazear (1995) explores this with two different piece-rate compensation schemes and finds that higher skill workers will sort into schemes characterized by less stable earnings. While this motivates the relationship between skill and earnings instability, the true labor market is significantly more complex. A rigorous empirical investigation is needed to establish the nature of the relationship between earnings instability and skill.

In this paper, I study the relationship between a worker's skill (i.e., her unobserved person-specific heterogeneity associated with earnings, or "ability") and her earnings instability using data from the National Longitudinal Survey of Youth 1979, a well known and commonly used panel. To proxy for the worker's unobserved heterogeneity I use the worker's score on the Armed Forces Qualifying Test (AFQT), a variable often used for this purpose in the economics literature.

Instead of a monotonic relationship between earnings instability and AFQT scores, I find a U-shape, with middle-skill workers experiencing more stable earnings than either high- or low-skill workers. The right tail of the U is less pronounced than the left tail: low-skill workers are found to have the least stable earnings, workers of middling ability are found to have the most stable earnings, and high-skill workers fall between the two. This relationship is estimated using a fully parametric model and using semi-parametric techniques, and is robust to different sample restrictions, control variables, and estimation strategies.

This finding is most closely related to the examination of earnings instability trends for different education subgroups. While Gottschalk and Moffitt (1994) focus on documenting the trends in *aggregate* earnings instability, they do examine the trends for three education subgroups: less than twelve years of school, twelve or more, and sixteen or more. They find that the least educated workers have the least stable earnings and that the most educated workers have the most stable earnings. Gottschalk and Moffitt (2009) conduct a similar exercise, except they restrict their attention to two mutually exclusive subgroups: workers with twelve or fewer years of school and workers with more than twelve years. They compute average earnings instability for each subgroup, as before, and find that in the 1970s both groups had about the same level of earnings instability, in the 1980s the less educated group had less stable earnings than the more educated group, but in the 1990s it was the more educated workers with less stable earnings. Cameron and Tracy (1998) also present trends by subgroup, and find that transitory earnings variance declines with a worker's level of education, and that most of the decline occurs if a worker is a high school graduate. Earnings instability rose for all education categories in the 1980s in their study, with the largest relative increase experienced by high school dropouts.

In addition to education, permanent earnings are another cousin of skill. Gottschalk and Moffitt (2009) place individuals into three groups based on the percentile rank of their permanent earnings: bottom quartile, top quartile, and middle quartiles. They find that the bottom quartile had the least stable earnings in all the time periods, and that the largest increase in instability occurred for this group between the 1970s and 1980s.

I build on these subgroup-trend analyses in a number of ways, and provide a more thorough and reliable analysis of the link between skill and earnings instability. First, I use a worker's AFQT score as a measure of his skill. AFQT is a much finer measure than has been previously used, and will allow me to examine the relationship between earnings instability and skill at a much more detailed level than would looking at two education categories or three categories of permanent earnings. Second, AFQT is plausibly exogenous — workers are not selecting their AFQT score in order to attain a desired level or variance of earnings, as they may be with education. Third, instead of studying subgroup trends I study the underlying relationship between skill and earnings instability. The literature has focused on the trends because the original motivation for studying earnings instability was to understand better the growth over time in earnings inequality. My primary interest is in understanding whom earnings instability affects and why it affects them. Fourth, instead of simply computing average earnings instability by subgroup and time period, I study the relationship in more depth, using both parametric and semi-parametric regression methods and a variety of economically interesting samples.

In addition to the U-shape of earnings instability over skill, this paper presents the following results on the relationship between earnings instability and skill: (1) Low-skill workers have less stable earnings than either middle- or high-skill workers — the right tail of the U is less pronounced than the left. (2) The U-shaped relationship is stronger for workers who do not change jobs, and so is not driven by earnings fluctuations associated with switching employers or job loss, and implying that the pattern may be related to a feature of compensation. (3) While single-job holders have more stable earnings than the sample of workers who are allowed to hold multiple jobs, the U-shape is strongest for single-job holders. This suggests that the U may be a feature of compensation in "better" jobs. (4) The result that low-skill workers experience less stable earnings than high-skill workers occurs in both the 1990s and 2000s, but the result that high-skill workers have less stable earnings than middle-skill workers only holds for the 2000s. (5) The instability of total family income does not follow a U-shape over skill. Instead, it seems to follow a W.

What might explain these findings? The U-shape is robust to workers entering and leaving employment, changing jobs, and holding multiple jobs. I provide a speculative discussion in the paper which relates the U-shape of earnings instability over skill to the hollowing out of the skill distribution. Autor and Dorn (2011) find that from 1980 to 2005 both employment growth and wage growth follow a U-shape over skill. If employment and *average* wages are growing in low-skill and high-skill occupations at faster rates than middle-skill occupations, then it may follow that the *variance* of earnings for workers of low- and high-skill will be relatively larger. Alternatively, or perhaps in addition, earnings variance may be driven by fundamentally different forces for low- and high-skill workers, with high-skill workers hoping for a large positive shock and low-skill workers operating in something similar to a secondary labor market characterized by a breakdown of implicit contracts governing earnings.

Results from the analysis of earnings instability and skill suggest that the U-shape may be driven in part by compensation practices — by the way that workers are paid by their employers. This suggestion invites the study of the firm-level characteristics which are associated with their workers' earnings instability. To investigate this, I study the relationship between earnings instability and an important characteristic of a firm: the instability of employment at the worker's employing firm. Firm employment is a natural measure of the scope of economic activity undertaken by the firm, and firm employment instability is a natural measure of fluctuations in the scope of economic activity.

Models of perfect competition predict that a worker's earnings are unrelated to firm performance: the price of labor is set in the market, and both firms and workers take the price of labor as given. Other models — including contract models, bargaining models, and monopsony models — predict that firms have some control over the earnings they pay their workers, and suggest a link between the market outcomes of a firm and the earnings dynamics of the firm's employees. The core of this issue is the motivation behind implicit contracts: Do firms insure the earnings of their risk-averse workers from shocks by smoothing the earnings of their workers from period to period?

Bertrand (2004) finds that a firm may decide to reduce the protection from external labor market conditions which it provides to its workers' earnings when competition from imports increases. This suggests a possible link between firm instability and earnings instability: firms which operate in increasingly hostile markets are less willing to insure the earnings of their workers from shocks. To my knowledge, Comin *et al.* (2009) provide the only direct test of whether firms pass volatility onto workers' earnings. Using COMPUSTAT data on publicly traded firms, they find a robust relationship between the instability of sales and employment of a firm and the instability of the average earnings paid by that firm, implying that firms may be passing instability onto workers in the form of more volatile earnings.

While Comin *et al.* (2009) is compelling, there remains more work to do on the question. Comin *et al.* (2009) use a sample of firms restricted to those which are publicly traded, but publicly traded firms have very different patterns of volatility than do privately held firms (Davis *et al.*, 2006). In addition, Comin *et al.* (2009) use average earnings paid by the firm, calculated by dividing a firm's total annual wage bill by total employment — they study the instability of *average* earnings at the firm. While this is perhaps the best measure that can be found in COMPUSTAT, worker-level earnings data are needed to fully explore this question.

To study the relationship between firm employment instability and worker earnings instability I use worker-level earnings linked to firm-level variables. The linked employer-employee data I use come from the LEHD program of the U.S. Census Bureau. These linked employer-employee data provide me with the ability to directly investigate the question at hand because I have information about individual workers and the firms in which they are employed.

I present three main findings on the effect of firm employment instability on earnings instability. (1) I find a positive relationship between the instability of a firm's employment and the instability of the earnings of workers in that firm. This result is robust to a number of demographic and firm controls and remains when estimated using within-firm variation, suggesting that the effect is a feature of the way workers are being paid by firms. (2) I find that low-earning workers are being passed more than twice as much instability as high-earning workers — the size of the effect decreases over earnings quintile. This helps to explain one of the major results of my study of earnings instability and skill: that low-skill workers experience more earnings instability than middle- or high-skill workers. And (3), I find significant heterogeneity in the magnitude and statistical significance of the effect across industries.

The third finding is particularly interesting. In the canonical implicit contract model, the fully-insured wage enjoyed by the worker is driven by the assumption that firms are risk neutral and workers risk averse. Is an increasingly competitive economy introducing risk aversion into firms' utility over profits? Does increasing competition drive down corporate profits, and thus weaken the ability of a firm to honor implicit contracts? If so, then variation across industries in competitive pressure may explain the variation across industries in the magnitude of the effect. At the same time, variation across industries in the power of firms to affect the earnings of their workers may also shed some light on the inter-industry heterogeneity of the effect. I present preliminary evidence of a relationship between the magnitude of the pass-through effect and the competitiveness of the industry.

This paper proceeds as follows. The next section will discuss earnings instability with more rigor, and specify the measure used in this paper. Section III will present the investigation of earnings instability and skill. Section IV will present the analysis of the effect of firm employment instability on earnings instability, and the fifth section will conclude.

II. Earnings Instability: Motivation & Measurement

At least since Milton Friedman (1957), economists have found it natural to think of a worker's earnings as consisting of a time-invariant permanent component and a time-varying transitory component. The transitory component can be thought of as a "random" shock to earnings, causing earnings to be unstable over time. Attempting to better understand this *earnings instability* is the focus of this paper.

To help build intuition, think of these transitory shocks in a number of labor market settings. Among high-skill workers, a lawyer who wins a once-in-a-decade case or a corporate vice president who receives an unusually generous annual bonus has received a positive transitory shock to her permanent earnings, and a hedge fund manager who makes a bad investment decision and earns no bonus in a given year can be thought of as experiencing a negative transitory shock. Among low-skill workers, a negative transitory shock to earnings could come in the form of a factory layoff, a business slowdown that results in weekly hours being temporarily cut from forty to thirty, or an unusually slow summer in a landscaping occupation. In each case, the worker's average or permanent earnings are being hit by a shock which dies out quickly — which would not be expected to affect earnings into the future.

Earnings instability is an important component of many interesting economic questions. It is a critical component of cross-sectional earnings inequality. Assuming that transitory shocks to earnings are a random process which is uncorrelated with permanent earnings, the cross-sectional variance of earnings at any point in time is simply the sum of the cross-sectional variance of permanent earnings and the cross-sectional variance of transitory earnings.

A conceptual example helps to clarify this point. Imagine that you observe the distribution over earnings at two points in time, t and t'. At t all workers earn exactly the same amount of money. There is no spread in the earnings distribution, and thus there is no earnings inequality. At t', one worker receives a random shock to his earnings, causing him to earn more than the other workers in the distribution. All other workers retain the same earnings as in t. Now, at t', there is non-zero variance in the distribution, and we conclude that earnings inequality has increased from t to t', though nothing in the structure of the labor market has actually changed.

Peter Gottschalk and Robert Moffitt have studied the trends in earnings instability in a series of papers (1994, 2002, 2008, and 2009). Using the Michigan Panel Study of Income Dynamics, they find that the rise in transitory earnings variance accounts for 31 percent of the total rise in cross-sectional earnings variance from 1970 to 2004 (Moffitt and Gottschalk, 2008). Standard explanations of increasing earnings inequality — increasing demand for high-skill workers, increasing supply of low-skill workers, and structural changes such as a decrease in union membership — apply to permanent earnings, but we would not expect them to affect transitory shocks. Supply and demand factors evolve slowly over time, and while a gradual increase in the demand for college-educated workers would be expected to raise the average wage of a college-educated worker over time, it seems less likely that this change in demand would affect deviations from the worker's average wage. To better understand earnings inequality, then, we must better understand earnings instability.

Earnings instability also plays a significant role in consumer welfare. Abstract away from savings behavior and assume that a household consumes all of its earnings in the form of goods and services. For households with a concave utility function — a standard assumption — we know that the utility of the expected value of earnings is greater than the expected utility of the gamble:

$$u(\delta y + (1 - \delta) y') \geq \delta u(y) + (1 - \delta) u(y'), \text{ where } \delta \in (0,1).$$

If $y = y'$ — if the household has perfectly stable earnings — then the two expressions are equivalent. But if not, household welfare is higher in the scenario to the left of the inequality, when the household income is not uncertain.[4]

Household welfare may also be affected by earnings instability as it relates to consumption smoothing. Dynarski and Gruber (1997) find that households may have trouble smoothing in the face of earnings instability, particularly with respect to durable goods. Blundell *et al.* (2008) find that low-wealth households are only able to partially insure their consumption against transitory income shocks. Similarly, Gorbachev (2011) finds that income volatility is associated with volatile consumption of food, particularly for traditionally-vulnerable households.

Instability of earnings also has a clear relationship with earnings mobility; depending on the size of the transitory shock, a worker's relative position in the

[4] In addition to preferring the expected value of the gamble to the gamble, it is the case that the household prefers smaller variance in expected earnings to larger variance. Let y be a random level of earnings for a given year, with $\mu_y = E[y]$ and $\sigma_y = var(y) > 0$. Let U be utility over earnings, with $U''(y) < 0$. Consider $y' = y + w$, where w is a mean-zero random variable which is independent of y, such that $cov(y, w) = 0$. Then $E[y'] = E[y] = \mu_y$, but $var(y') = \sigma_y + \sigma_w > \sigma_y$. It is a well known result that $E[U(y)] > E[U(y')]$.

earnings distribution could change. Finally, earnings dynamics are interesting in and of themselves, and earnings instability plays a clear role in the path of earnings over time.

The decomposition of annual earnings into a permanent and transitory component can be summarized by the following equation:

$$y_{it} = u_i + v_{it},$$

where y_{it} is annual earnings for person i in year t, u_i is a time-invariate component specific to individual i, and v_{it} is the transitory shock experienced by worker i in year t. The dynamics of v_{it} cause annual earnings to change from year to year, introducing instability.

The canonical model can be further developed. For example, it is natural to think that permanent earnings may change over time — investments in human capital surely affect permanent earnings, or changes in occupation and industry due to a layoff, or a severe illness. We also might expect a transitory shock to linger for more than one period.

These features can be incorporated into a richer model:

$$y_{i\tau t} = u_{i\tau} + v_{i\tau t} \tag{a}$$

$$u_{i\tau} = u_{i,\tau-1} + \delta_i + \omega_{i\tau} \tag{b}$$

$$v_{i\tau t} = \rho v_{i\tau,t-1} + \varepsilon_{i\tau t} \tag{c}$$

Here, equation (a) is the same as the canonical model, except all three terms are indexed by τ. The symbol τ identifies a group of years. So permanent earnings are assumed to be constant during that time window, and the sum of permanent earnings and a transitory shock equals observed annual earnings.

Permanent earnings are assumed to follow a random walk with a growth process. The growth factor, δ_i, allows for some individuals to have permanently higher or lower earnings than others, perhaps due to a genetic ability endowment or to environmental factors. Permanent earnings are hit by a shock during time interval τ, $\omega_{i\tau}$, which does not die out over time. The transitory component of earnings is assumed to be a first-order autoregressive process. During each year t in each time interval τ, the transitory component is hit by a shock, $\varepsilon_{i\tau t}$, which fades away according to $0 < \rho < 1$.

Moffitt and Gottschalk (2008) estimate a parametric error components model based on a three-equation system very similar to the one presented above (along with distributional assumptions governing the means, variances, and covariances of the components), and conclude that the method of estimating earnings instability used, for example, in Gottschalk and Moffitt (1994), Comin *et al.* (2009), and this paper approximates the trends from the richer model closely.[5] That method is to use the following formula:

$$instability(y_{it}) = \frac{1}{a+b-1} \sum_{j=-a}^{j=b} \left(y_{i,t+j} - \bar{y}_{i,[\mathrm{t}-a,\mathrm{t}+b]}\right)^2.$$

[5] Similar earnings decompositions are very common in the labor economics literature. See Abowd and Card (1989) as an example. DeBacker, Heim, Panousi, and Vidangos (2010) estimate earnings instability trends using two modeling choices: they allow the transitory component of earnings to follow an ARMA(1,1) process and an AR(1) process. In addition, they estimate earnings instability using the method used in this paper, and using a similar method from Kopczuk, Saez, and Song (2010). They find that the *trends* in earnings instability are sufficiently similar across the four methods, but that the *shares* of cross-sectional earnings variance explained by the permanent and transitory components differ. Haider (2001) estimates a parametric model of life-cycle earnings to study whether the increase in earnings inequality during the 1970s and 1980s was caused by lifetime earnings inequality increasing or the receipt of lifetime earnings becoming less stable. He finds that earnings instability increased significantly over the period, with most of the increase occurring in the 1970s.

Here, y_{it} is log annual earnings for worker i in year t. The parameters a and b represent years before and after year t, respectively. $\bar{y}_{i,[\mathrm{t}-a,\mathrm{t}+b]}$ is the average earnings for worker i during the time interval defined by a and b.

Essentially, this is a rolling variance window. A compelling feature of this methodology is that is allows for the calculation of worker-level earnings instability. This is critical for this study because I seek to examine the relationship between worker-level earnings instability and both worker skill and the instability of employment of the worker's employer.

To measure the instability of worker i's earnings in year t using, say, a nine-year window, compute the variance of his log annual earnings over the nine-year period starting four years before t and ending four years after t ($a = b = 4$). The permanent component of earnings during that nine-year window is simply the average earnings during the window, and the transitory component for each year is the deviation from the average.

Computing this measure for every worker i in every year t generates a time series of earnings instability for each worker — earnings instability is calculated for, say, the year 2000 by computing the variance of earnings from 1996 through 2004, and then earnings instability is calculated for 2001 by computing the variance of earnings from 1997 through 2005, etc.

The length of the time window is very important in this calculation. Following the previous literature, I use windows of two lengths: nine years, and five years. Five years allows the permanent component to shift frequently, and captures very high-frequency variation in earnings (Comin *et al.*, 2009). Nine years is the original length suggested by Gottschalk and Moffitt (1994). The results in this paper are substantively the same regardless of which window is used.

III. Earnings Instability & Skill

Economists are used to thinking about averages, and it is well known that differences in skill across workers can produce differences in average earnings. Whether a relationship exists between skill and the *variance* over time of workers' earnings is less clear. In this section, I present two simple exercises to demonstrate that the relationship between earnings instability and skill can be found by extending standard models.

The idea that low-skill workers will sort into fixed-pay jobs while high-skill workers will sort into piece-rate jobs is intuitive, and is prevalent in the compensation literature. Lazear (1986) derives a model which implies this result, and Lazear (2000) finds evidence supporting the prediction using data from the Safelite Glass Corporation, an automobile glass installation company.

In the mid-1990s Safelite changed its compensation scheme from hourly wages to piece-rate pay, with a guaranteed minimum rate. If the worker installed enough windows such that his piece-rate pay was higher than the guaranteed minimum, he was given the piece-rate pay. If not, he was given the hourly minimum. Lazear finds that higher-quality workers sorted into Safelite after the piece-rate scheme was implemented, and that average worker ability rose (as well as average output and output variance) after the change in compensation structure.

To present a simple argument based on Lazear (1995), imagine that a worker is deciding between two firms. Firm 1 offers a high base salary and a low performance pay component. Firm 2 offers a low base salary and a high performance pay component.[6] This situation can be summarized by the following equations:

[6] This could just as easily be two compensation schemes within the same firm, or two industries, provided that the shock affects workers within firm or across industries in the same manner.

$$pay_{1it} = a_1 + b_1 q_{it}$$
$$pay_{2it} = a_2 + b_2 q_{it}$$

where $a_1 > a_2 > 0$, reflecting the higher base salary offered by Firm 1; $0 < b_1 < b_2$, reflecting the higher performance-pay component offered by Firm 2; and $q_{it} > 0$ is the output of the worker during period t. Let the worker's output be a linear function of the worker's skill and a random component which affects both workers: $q_{it} = f(s_i) = \alpha s_i + \xi_t$, where s is skill, α is a parameter, and ξ_t is a shock.[7] It may be most intuitive to think of ξ_t as a technology shock.

Then we have the following:

$$pay_{1it} = a_1 + b_1(\alpha s_i + \xi_t)$$
$$pay_{2it} = a_2 + b_2(\alpha s_i + \xi_t)$$

For some level of skill, call it $\bar{s}$, $pay_{1it} = pay_{2it}$. Worker with skill level $s < \bar{s}$ will receive higher pay with Firm 1, while workers with skill level $s > \bar{s}$ will receive higher pay with Firm 2. Thus, workers of higher skill will sort into the firm with a larger performance pay component, and less-skilled workers will sort into the firm with a larger base salary component.

This compensation scheme captures the intuition that workers of relatively higher skill may experience less stable wages than lower skill workers. Note that output is more variable than base pay due to the presence of the time-varying shock. This is by construction, but it may be reasonable to think that base pay will experience smaller year-to-year changes than output. The pay scheme implies that

---

[7] For the result to hold, the assumption of linearity is unnecessary. If $q_{it} = h(s_i) + \xi_t$, the result will still hold provided that $h()$ is increasing in skill and unbounded.

$var(pay_{fit}) = b_f^2 var(\xi_t), f = 1,2$. Since $b_2^2 > b_1^2$, the firm which offers the larger performance pay component will see workers with less stable earnings.

In addition to this stylized example, the result that workers of different skill will experience different levels of earnings instability can be found in a standard general equilibrium model of a perfectly competitive market.

Suppose that two workers, a higher-skilled worker and a lesser-skilled worker, maximize utility from leisure and consumption and are endowed with nothing but time. Suppose further that workers' labor supply is inelastic, and that both workers supply one unit of labor to a firm.

The firm chooses the quantity of higher-skilled labor and the quantity of lesser-skilled labor to maximize profit. The production function is Cobb-Douglas. The profit function is:

$$\vartheta_t N_{Ht}^{\alpha} N_{Lt}^{\beta} - w_{Ht} N_{Ht} - w_{Lt} N_{Lt},$$

where N_{Ht} is the quantity of higher-skilled labor the firm chooses in period t, N_{Lt} is the quantity of lesser-skilled labor chosen, α and β are the marginal products of labor, $\alpha > \beta$, and w_{Ht} and w_{Lt} are the wage rates. The parameter ϑ_t is a shock. This can be thought of as a time-varying output price or as a time-varying productivity shock which affects both types of workers.

Labor demand is a function of the wage, the marginal product, and the shock, and is as follows: $N_{Ht} = \alpha \frac{\vartheta_t}{w_{Ht}}$ and $N_{Lt} = \beta \frac{\vartheta_t}{w_{Lt}}$. Setting labor supply equal to labor demand ($N_{Ht} = N_{Lt} = 1$) gives equilibrium wages of: $\overline{w}_{Ht} = \alpha\vartheta_t$ and $\overline{w}_{Lt} = \beta\vartheta_t$. Since $\alpha^2 > \beta^2$, we have the result that $var(w_{Ht}) > var(w_{Lt})$.

To be clear, I do not intend to test these models in this paper. The purpose of presenting these exercises is to demonstrate how readily a relationship between earnings instability and skill can be found using standard models, both stylized and general equilibrium. In these models, differences in skill produce not only difference in mean earnings, but also differences in the time-series variance of earnings. Having developed some intuition regarding the relationship, I now turn to the data and descriptive statistics.[8]

3.1. Data, Samples, & Descriptive Statistics

The data used in this paper come from the National Longitudinal Survey of Youth 1979. The NLSY79 follows a cohort of American youths aged fourteen to twenty-one at the time of first interview in 1979 and records detailed information about their labor market outcomes. The cohort is interviewed every year from 1979 through 1994, and every other year thereafter. The NLSY79 oversampled Hispanic, black, and economically-disadvantaged whites, so sample weights are used throughout this paper.

The annual earnings variable used is the log of total income from wages and salaries. The respondent is asked to recall his total earnings for the previous year from working, which includes wages, salaries, commissions, and tips from all jobs, before taxes or other deductions. Measurement error is clearly a concern here, but provided that the share of measurement error in the individual's annual report does not vary significantly over time, the earnings instability measure should be largely unaffected.[9]

---

[8] In ongoing research I am directly testing the extent to which the increasing prevalence of performance pay is related to the macro earnings instability trends.

[9] Consider the case where a respondent misreports his earnings by x percent per year, each year of the rolling variance window. Reported earnings are actual earnings y plus error, equal to $y(1+x)$. The earnings instability formula can be written such that the term which is being squared (and then summed) is $\ln\left[y_t(1+x)/\left(\prod_T (y(1+x))^{1/T}\right)\right]$. As the measurement error is in both the numerator and denominator of the fraction, the measurement error cancels out, and the instability of reported earnings is identical to the instability of actual earnings.

The “skill” variable in the NLSY79 used in this paper is a worker’s AFQT score. In the summer and fall of 1980, the Armed Services Vocational Aptitude Battery — the standardized aptitude test used by the military to assist in assigning service members to jobs — was administered to the NLSY79 cohort, and was completed by approximately ninety-four percent of the sample. The AFQT score is calculated using the arithmetic reasoning, word knowledge, paragraph comprehension, and numerical operations components of the ASVAB, and is frequently used in labor economics to proxy for unobserved ability or skill (see, for example, Heckman and LaFontaine, 2006). A benefit of using this particular measure of unobserved skill is that it is plausibly exogenous to future labor market experiences — i.e., workers are not choosing their AFQT score in order to affect their earnings and earnings dynamics. In addition, there is no reason to be concerned that a worker’s earnings are causing the worker’s AFQT score.

A unit of observation is a worker-year, and the analysis sample consists of males older than twenty-five years of age who are not enrolled in school and who have positive earnings for the year, and runs from 1979 through 2008. The data are annual, and males are studied to avoid confounding labor supply issues. Workers older than twenty-five are studied to avoid the confounding volatility which may be associated with beginning a career. Students are excluded because their labor supply patterns are likely quite different than non-students. Workers with positive earnings are studied because they are the observations in the sample which are being paid by a firm, and the primary purpose of this paper is to study the instability of labor income — to see if the instability of earnings paid to workers by firms is related to the skill level of workers. (Total family income, described later, is also studied.)

Earnings are deflated using the CPI-U. Following the literature, I delete worker-years in the bottom one percent and the top one percent of the log earnings distribution. This is done to eliminate reporting errors, to deal with the problem of top-coding, and to ensure that outliers are not driving the regression results. When interpreting the results, note that they only apply to the middle ninety-eight percent of the log earnings distribution.

I follow Gottschalk and Moffitt (2002) by computing the instability measure on log earnings, and not on the residuals from a log earnings regression. This generalizes the results and ensures that the choice of first-stage specification is not driving the results. Standard control variables (years of schooling, experience and experience squared, race, region of residence, marital status, urban residence, hours, occupation, and industry) are included in the regressions.

An issue with using the NLSY79 is that after 1994 the respondents cease being surveyed annually and are instead surveyed every other year. For odd-numbered years following 1994 there are no data on earnings. Following Comin *et al.* (2009), I deal with this issue by modifying the earnings instability measure by calculating it only on every other year of data. (Comin *et al.* refer to this as a "skipping years" methodology.) To ensure consistency, I use the skipping years methodology on the entire sample. All results in this paper use this methodology.

This paper studies earnings instability measured by rolling variance windows of two different lengths; consequently, there are two different analysis samples. In addition, much of the analysis is conducted on a subsample of observations which contain worker-years where the worker did not change his primary employer at any time during the window — e.g., for the nine-year window centered on the year 2000, the earnings instability measure is only computed if the worker did not change jobs from 1996 through 2004. Job stayers are interesting to study because we can be sure that the instability of their earnings is a function of the way they are being paid by their employer, and not a function of changing employers. Consequently, there are four analysis samples for total labor earnings: the nine-year window and the five-year window, both for job stayers only and for both job stayers and job changers.

All four samples contain workers with strong labor market attachment, and the two samples of job stayers contain workers with quite strong attachment to a particular employer. It is intuitive to think that some of these workers are less risk taking than those who experience periods of unemployment and job changing, and that many of them are more likely to have less variable earnings than the job changers and the sometimes unemployed. Studying these workers is interesting because doing so helps to control for entry and exit from employment and entry and exit from employers when estimating the relationship between earnings instability and skill. Studying workers who work continuously throughout the window, and studying those who work continuously for the same primary employer, helps to ensure that what we are learning about is the way these workers are paid, without the confounding effects of job changing and entry/exit from working.

Figure 1.1 plots the distribution over annual earnings for each of the four subsamples. The four distributions look very similar. The mean for the five-year window consisting of both job stayers and job changes is slightly less than its nine-year counterpart; the same holds for the subsamples of only job stayers. Stayers have slightly higher earnings than stayers and changers.

Figure 1.2 plots the distribution over AFQT percentile score for each of the four subsamples. A potential concern might be that in selecting the four analysis samples a disproportionate number of observations from a particular part of the AFQT distribution may not be included — e.g., that by examining job stayers in the nine-year window we may effectively eliminating low-skill workers from the sample.

Looking at Figure 1.2, it is clear that there is mass above each point of support. The mean values of the distributions are all very close to each other. Overall, the

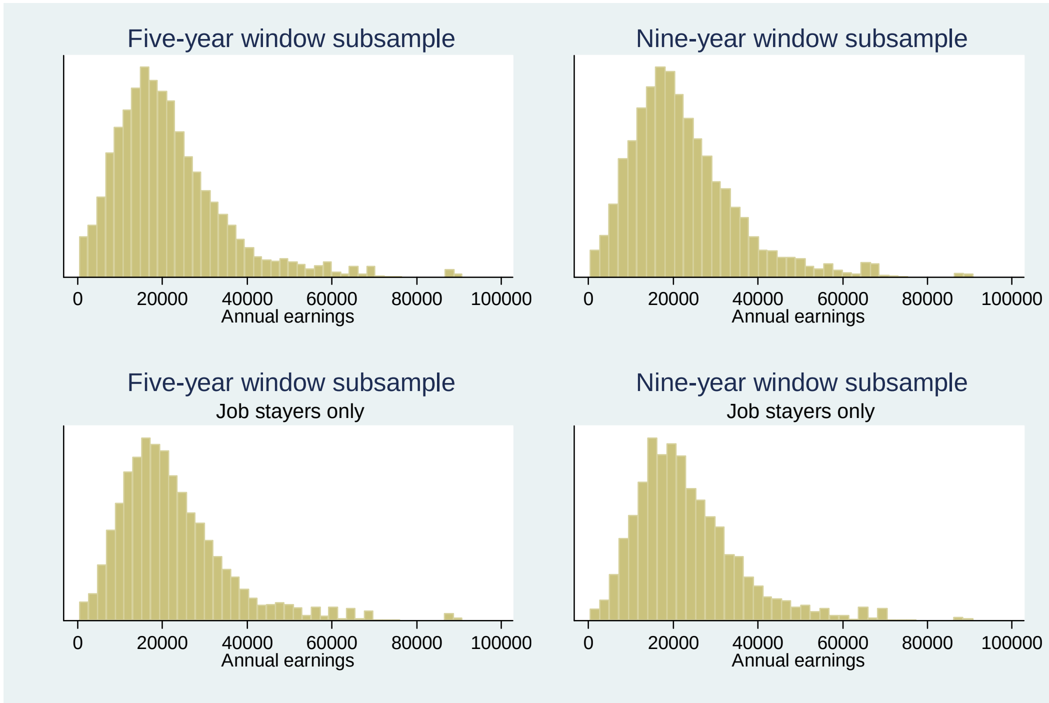

FIGURE 1.1: Distribution over annual earnings by subsample.

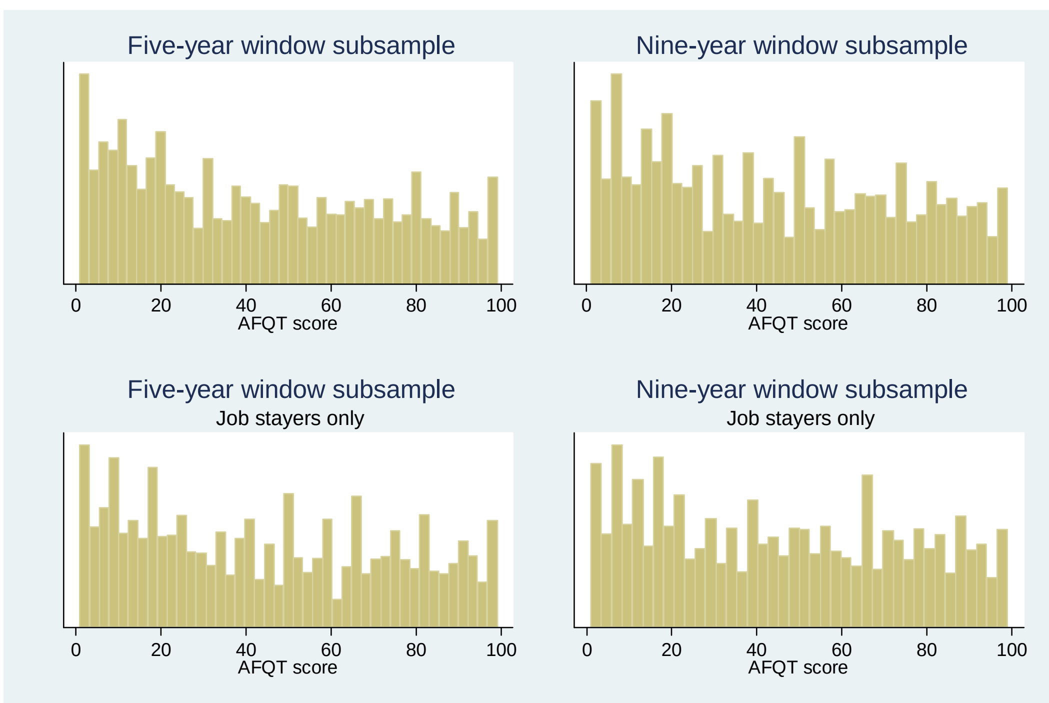

FIGURE 1.2: Distribution over AFQT score by subsample.

TABLE 1.1: Summary statistics

Panel A: Primary sample: Total labor earnings								
	Subsamples							
	Five-year window		Nine-year window		Five-year window Job stayers		Nine-year window Job stayers	
	Mean	**Std. Dev.**	**Mean**	**Std. Dev.**	**Mean**	**Std. Dev.**	**Mean**	**Std. Dev.**
Earnings instability	0.146	0.390	0.136	0.287	0.117	0.339	0.108	0.249
Labor earnings	21674.840	13221.700	22556.190	12822.700	22718.370	13206.120	23624.350	12787.900
Total income	34605.280	52108.020	35327.360	49085.830	36166.080	53823.590	36939.940	50423.280
AFQT	0.433	0.295	0.443	0.291	0.447	0.296	0.456	0.292
Years of school	13.047	2.383	13.086	2.308	13.147	2.395	13.191	2.314
Percent Hispanic	17.930		17.670		17.620		17.250	
Percent black	25.690		24.550		24.460		23.770	
Observations	25287		13515		22262		11504	

TABLE 1.1 (Continued)

Panel B: Single-job holder sample

	Subsamples							
	Five-year window		**Nine-year window**		**Five-year window Job stayers**		**Nine-year window Job stayers**	
	Mean	**Std. Dev.**	**Mean**	**Std. Dev.**	**Mean**	**Std. Dev.**	**Mean**	**Std. Dev.**
Earnings instability	0.088	0.297	0.072	0.200	0.069	0.249	0.059	0.160
Labor earnings	24741.650	14110.300	26033.940	13415.830	25341.670	14047.460	26439.700	13405.760
Total income	38950.030	56148.420	41433.880	61289.930	39652.380	56391.410	41818.710	60740.040
AFQT	0.459	0.300	0.472	0.300	0.467	0.299	0.478	0.299
Years of school	13.148	2.381	13.140	2.296	13.196	2.388	13.182	2.308
Percent Hispanic	17.540		16.670		17.250		16.370	
Percent black	22.420		21.090		21.520		20.580	
Observations	11028		4320		10294		4057	

pictures indicate that the entire support of the AFQT distribution is well represented in each of the four analysis subsamples.

Panel A of Table 1.1 reports some summary statistics of key variables for the four analysis samples. Computing earnings instability using a five-year window results in a slightly higher value for earnings instability than using the nine-year window. This is true for both the sample of only job stayers and the sample of both job stayers and job changers. Using either window length, stayers have more stable earnings than the sample of both stayers and changers. This is intuitive — earnings instability should be less when employer changes are controlled for. Demographically, all four samples look very similar. Stayers seem to have about 0.1 years of schooling more than the sample of stayers and changers. In all four samples, 17 to 18 percent of the sample is Hispanic, and 24 to 26 percent is black.

Single-Job Holders

In addition to studying the relationship between earnings instability and skill using the total labor earnings variable described above, another interesting sample consists of those workers who only work at one employer in a given year — single-job holders.

Workers select into holding multiple jobs at a time for a variety of reasons. Hours constraints at a primary job have been found to be an important reason why some workers may take a second job (Paxson and Sicherman, 1996). Longer hours and higher income in a primary job have been found to deter workers from holding a second job (Krishnan, 1990). This evidence is consistent with the hypothesis that workers select into multiple job holding to supplement income, which may increase earnings instability for those workers. At the same time, if workers have concave utility over earnings, it may also be the case that some workers select into holding a second job to smooth their income over time — to decrease their earnings instability.

Panel B of Table 1.1 reports descriptive statistics for the sample of worker-years where the worker is a single-job holder. The same four subsamples and measures are generated from this sample as in the sample previously discussed: five-year earnings instability, nine-year earnings instability, five-year earnings instability for job stayers only, and nine-year earnings instability for job stayers only. Interestingly, average earnings instability for the sample of worker-years with only one job is nearly half that of the sample for all worker-years, with both single and multiple job holding. For single-job holders, the five-year instability measure averages 0.088; for single- and multiple-job holders, the measure averages 0.147. This disparity also holds for job stayers. The average earnings instability for single-job holders who do not change jobs during the window equals 0.069. For single- and multiple-job holders, the average of the measure is 0.117.

The sample of single-job holders is different than the sample of single- and multiple-job holders. For the five-year and five-year-stayers subsamples, average earnings are 14 and 12 percent higher, respectively. This is consistent with the literature, which suggests that higher-income workers are less likely to select into a second job. In addition, single-job holders have higher AFQT scores and are less likely to be a minority.

All the sample cuts studied in this paper consist of workers with very strong labor market attachment. For the subsamples of job stayers only, the workers additionally have very strong employer attachment. The sample of worker-years who hold only one job are perhaps the most stable: not only do they work every year in the window, but also they work for only one employer, and for the job-stayer subsamples they work only for the same employer. If the relationship between skill and earnings instability is the same for this group as for the group previously discussed, that is evidence that the relationship between skill and earnings instability is a strong and persistent feature of the way workers are paid. In addition, studying the earnings instability of these workers controls for instability generated by multiple income streams.

Total Family Income

The total family income variable in the NLSY79 captures receipts from labor earnings, described above, and in addition from business and farm income, social insurance programs, including unemployment compensation, AFDC payments, food stamps, education and disability benefits, and other welfare and SSI payments.

From the descriptive statistics in Table 1.1, we see that the earnings instability measure calculated on total family income and the earnings instability measure calculated on total labor income have similar average magnitudes. However, this says nothing about the relationship between the instability of total family income and skill. If labor earnings are systematically different for workers of different skill level, then investigating whether the relationship between labor earnings and skill is different than the relationship between total income and skill is of interest.

3.2. Empirical Strategies & Regression Results

Using sample means, interesting evidence on earnings instability was presented in the previous section. This section will study the relationship between earnings instability and skill by employing regression analysis. A flexible technique using categorical variables is presented first, followed by a semi-parametric technique, local linear regression, and finally a parametric model. The analysis will first be conducted on the total labor earnings variable, and then subsequently the labor earnings of single-job holders and total family income will be studied.

Categorical Variable Model

To gain an understanding of the shape of the relationship, I create AFQT quintile indicator variables. AFQT20 equals 1 if the worker has an AFQT percentile score less than or equal to twenty, and 0 otherwise. AFQT40 equals 1 if the worker's score is both greater than twenty and less than or equal to forty, and 0 otherwise. AFQT60, AFQT80, and AFQT100 are analogously defined. The empirical specification is presented below as Eq. 1:

$$instability(y_{it}) = \alpha + \sum_{k \in \psi} \beta_k AFQT_k + X_{it}\delta + \varepsilon_{it}, \qquad (1)$$
$$\psi \stackrel{\text{def}}{=} \{20, 40, 80, 100\},$$

where the dependent variable is the instability of the log of total labor earnings for worker i in the window centered on year t, α is a constant, ε_{it} is the residual, δ is a parameter vector, and X_{it} is a matrix of control variables. The symbol ψ signifies the four-element set $\{20, 40, 80, 100\}$, so β_k is the coefficient on each of the four AFQT categorical variables. Standard errors are robust, and the regressions are weighted. AFQT60 is omitted and used as the comparison group.

Before turning to the regression results, it is helpful to examine the average level of earnings instability for each AFQT group. Column 1 of Table 1.2 reports the means by AFQT quintile for total labor earnings. For each of the four subsamples — five-year, five-year stayers, nine-year, and nine-year stayers — the quintile with the least stable earnings is the lowest quintile. Interestingly, the quintile with the *most* stable earnings is the middle quintile, AFQT60.

TABLE 1.2: Average instability of earnings and income by AFQT quintile

	Five-year window measure/subsample		
	Labor earnings	Labor earnings Single-job holders	Total income
0 < AFQT ≤ 20	0.186	0.120	0.200
20 < AFQT ≤ 40	0.150	0.095	0.151
40 < AFQT ≤ 60	0.118	0.056	0.139
60 < AFQT ≤ 80	0.118	0.071	0.114
80 < AFQT ≤ 100	0.126	0.080	0.130
	Five-year window measure/subsample: Job stayers only		
	Labor earnings	**Labor earnings Single-job holders**	**Total income**
0 < AFQT ≤ 20	0.146	0.090	0.168
20 < AFQT ≤ 40	0.117	0.076	0.132
40 < AFQT ≤ 60	0.092	0.046	0.122
60 < AFQT ≤ 80	0.097	0.059	0.101
80 < AFQT ≤ 100	0.113	0.067	0.122
	Nine-year window measure/subsample		
	Labor earnings	**Labor earnings Single-job holders**	**Total income**
0 < AFQT ≤ 20	0.161	0.086	0.192
20 < AFQT ≤ 40	0.138	0.082	0.157
40 < AFQT ≤ 60	0.118	0.046	0.144
60 < AFQT ≤ 80	0.121	0.060	0.116
80 < AFQT ≤ 100	0.124	0.077	0.133
	Nine-year window measure/subsample: Job stayers only		
	Labor earnings	**Labor earnings Single-job holders**	**Total income**
0 < AFQT ≤ 20	0.123	0.064	0.155
20 < AFQT ≤ 40	0.110	0.074	0.144
40 < AFQT ≤ 60	0.090	0.038	0.119
60 < AFQT ≤ 80	0.097	0.052	0.102
80 < AFQT ≤ 100	0.112	0.062	0.124

These quintile-specific means imply that earnings instability follows a U-shape over skill. Importantly, the left tail is much larger than the right. For the five-year subsample, for example, the average earnings instability for AFQT20 is approximately 57 percent larger than the average earnings instability for AFQT60, while AFQT100 is only approximately 7 percent larger than the middle quintile. This pattern holds for the subsample of job stayers. The nine-year stayers subsample finds average earnings instability of AFQT20 to be approximately 36 percent larger than the middle quintile, with AFQT100 approximately 24 percent larger.

The magnitude of the difference is economically significant. Imagine two workers, each of whom earns 22,000 dollars in year t, which is approximately the mean inflation-adjusted value of earnings in the regression samples. Between years $t-2$ and $t+2$, the first worker starts at 26,000 dollars, suffers a 2,000 dollar pay cut each year, and so is earning 18,000 dollars in the last year of the five-year window. The second worker receives a pay cut of 4,000 dollars per year, starts at 30,000 dollars in year $t-2$, and finishes at 14,000 dollars in year $t+2$. The earnings instability measure for the second worker is larger than the earnings instability measure for the first worker by a magnitude of 0.07. This difference of 0.07 is the same as the difference for the five-year subsample between the middle AFQT quintile and the lowest AFQT quintile.

To control for other economic and demographic factors, I now turn to regression estimates of this relationship. First, consider the subsamples of job stayers and changers. Table 1.3 reports the results from these regressions. The first five columns of the table report results using the five-year window, and the last five columns report results using the nine-year window. For each subsample, the first regression is uncontrolled. The second regression includes a control for the log of

TABLE 1.3: Earnings instability on AFQT indicator variables

The table reports estimates of Eq. 1. The dependent variable for each regression is the earnings instability measure using either a nine-year or five-year rolling window. (See text for details.) In addition to the control variables shown in the table, regressions which are labeled as including demographic controls include dummies for year, race, region of residence, marital status, and urban residence. Regressions are weighted. Standard errors are robust. The AFQT indicators equal 1 if the worker's AFQT score falls within the specified interval, otherwise 0.

	Five-year window measure/subsample					**Nine-year window measure/subsample**				
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**	**(10)**
0 < AFQT ≤ 20 Indicator	0.051***	0.040***	0.029***	0.024***	0.020**	0.030***	0.027***	0.012	0.014	0.009
	[0.008]	[0.008]	[0.009]	[0.008]	[0.009]	[0.008]	[0.008]	[0.009]	[0.008]	[0.009]
20 < AFQT ≤ 40 Indicator	0.033***	0.027***	0.021**	0.018**	0.013	0.026***	0.023**	0.016*	0.016*	0.013
	[0.009]	[0.009]	[0.009]	[0.009]	[0.010]	[0.009]	[0.009]	[0.009]	[0.009]	[0.010]
60 < AFQT ≤ 80 Indicator	0.002	0.005	0.009	0.009	0.013	0.003	0.006	0.01	0.011	0.015
	[0.008]	[0.008]	[0.008]	[0.008]	[0.008]	[0.008]	[0.008]	[0.008]	[0.008]	[0.009]
80 < AFQT ≤ 100 Indicator	0.007	0.007	0.022**	0.018**	0.021**	0.001	0.003	0.013	0.012	0.018*
	[0.009]	[0.008]	[0.009]	[0.009]	[0.009]	[0.009]	[0.008]	[0.009]	[0.009]	[0.010]
Log hours worked		-0.140***		-0.123***	-0.114***		-0.088***		-0.077***	-0.074***
		[0.014]		[0.014]	[0.013]		[0.011]		[0.011]	[0.012]
Years of school			-0.007***	-0.005**	-0.005*			-0.009***	-0.008***	-0.007**
			[0.002]	[0.002]	[0.002]			[0.002]	[0.002]	[0.003]
Experience			-0.004	-0.005	-0.008**			-0.016***	-0.017***	-0.019***
			[0.003]	[0.003]	[0.004]			[0.004]	[0.004]	[0.005]
Experience squared			0.000	0.000	0.000*			0.000**	0.000***	0.000***
			[0.000]	[0.000]	[0.000]			[0.000]	[0.000]	[0.000]
Constant	0.113***	1.192***	0.298***	1.222***	1.125***	0.115***	0.793***	0.486***	1.056***	0.926***
	[0.005]	[0.106]	[0.065]	[0.127]	[0.138]	[0.006]	[0.089]	[0.060]	[0.109]	[0.140]
Demographic controls	No	No	Yes	Yes	Yes	No	No	Yes	Yes	Yes
Occ and industry controls	No	No	No	No	Yes	No	No	No	No	Yes
Adjusted R-squared	0.003	0.021	0.02	0.033	0.125	0.002	0.016	0.024	0.034	0.175
Number of observations	25287	24850	25029	24599	24446	13515	13292	13432	13210	13122

Standard errors in brackets. *** signifies statistical significance at the one-percent level, ** signifies the five-percent level, and * the one-percent level.

hours worked. Earnings instability could be driven by hours variation, so this is an important control to consider. The third regression removes the hours control, and adds controls for years of schooling, experience and its square, and indicator variables for race, region of residence, marital status, urban residence, and year. The fourth regression returns the hours control to the estimating equation. And the fifth regression adds controls for industry and occupation.

The coefficient for the log of hours worked is negative and statistically significant in each equation in which it is included — longer hours are associated with more stable earnings. Years of schooling also consistently has a negative and statistically significant coefficient. The coefficients suggest that the instability-experience profile has a negative slope.

On balance, using the five-year subsample, there appears to be a U-shaped relationship between earnings instability and AFQT. The excluded AFQT category in the regressions is AFQT60 — AFQT scores between 40 and 60. The coefficients on AFQT20 are statistically significant and positive in all five regressions, and the coefficients on AFQT40 are statistically significant and positive in all but one, implying that low-skill workers experience greater earnings instability than middle-skill workers. In the three fully controlled regressions, AFQT100 has a statistically significant and positive coefficient. Taken together, this subsample indicates that both high- and low-skill workers have less stable earnings than middle-skill workers. When earnings instability is calculated using the nine-year window, the shape of the relationship is less clear. The control variables eliminate the statistical significance of the AFQT20 variable (although the statistical significance of AFQT40 survives in two of the three), and AFQT100 is statistically significant in only one of the five regressions.

Some of the earnings instability captured in these two subsamples may come from workers changing jobs. Table 1.4 presents regression results again using the five- and nine-year windows, but now on a sample of workers who not only worked every year in the window, but also who did not change their primary employer during the window.

Compellingly, every AFQT coefficient is statistically significant and positive in sign, providing fairly strong evidence that low- and high-skill workers have less stable earnings than middle-skill workers. Using the five-year window, the low-skill workers have less stable earnings than the high-skill workers. However, using the longer window, the reverse is true: it is the high-skill workers with the least stable earnings of all the quintiles.

Taken together, the results from these four subsamples indicate three findings: First, that the instability in total labor earnings follows a U-shape over skill, with low-skill workers experience less stable labor earnings than high-skill workers, and middle-skill workers experiencing the most stable earnings. Second, the relationship is more robust for the subsample of workers who do not change their primary job during the rolling window. And third, low-skill workers experience less stable earnings than high-skill workers.[10] In addition, note that relative instability is more equal between low- and high-skill workers when the longer window is used to calculate earnings instability.

[10] There is some evidence that relative to middle-skill workers and after regression adjustment, low-skill workers have less stable short-term earnings than high-skill workers, but high-skill workers have less stable longer-term earnings. However, as the next two sections will show, this result is not robust.

TABLE 1.4: Earnings instability on AFQT indicator variables: Job stayers

The table reports estimates of Eq. 1. The dependent variable for each regression is the earnings instability measure using either a nine-year or five-year rolling window. (See text for details.) In addition to the control variables shown in the table, regressions which are labeled as including demographic controls include dummies for year, race, region of residence, marital status, and urban residence. Regressions are weighted. Standard errors are robust. The AFQT indicators equal 1 if the worker's AFQT score falls within the specified interval, otherwise 0. Only workers who do not change their primary employer during the rolling window are included in the sample. (See text for details.)

	Five-year window measure/subsample					Nine-year window measure/subsample				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
0 < AFQT ≤ 20 Indicator	0.046***	0.042***	0.037***	0.035***	0.032***	0.028***	0.026***	0.022***	0.021***	0.017**
	[0.007]	[0.007]	[0.008]	[0.008]	[0.009]	[0.007]	[0.007]	[0.007]	[0.007]	[0.008]
20 < AFQT ≤ 40 Indicator	0.033***	0.030***	0.027***	0.025***	0.022**	0.032***	0.030***	0.028***	0.027***	0.026**
	[0.008]	[0.008]	[0.009]	[0.009]	[0.010]	[0.009]	[0.009]	[0.009]	[0.009]	[0.010]
60 < AFQT ≤ 80 Indicator	0.012*	0.012*	0.016**	0.015**	0.021***	0.012*	0.012*	0.016**	0.015**	0.021***
	[0.007]	[0.007]	[0.007]	[0.007]	[0.008]	[0.007]	[0.007]	[0.007]	[0.007]	[0.008]
80 < AFQT ≤ 100 Indicator	0.022***	0.020***	0.028***	0.025***	0.026***	0.020***	0.020***	0.026***	0.025***	0.029***
	[0.008]	[0.008]	[0.008]	[0.008]	[0.009]	[0.007]	[0.007]	[0.008]	[0.008]	[0.009]
Log hours worked		-0.064***		-0.055***	-0.049***		-0.030***		-0.025**	-0.021*
		[0.013]		[0.013]	[0.013]		[0.011]		[0.011]	[0.012]
Years of school			-0.004**	-0.004**	-0.003			-0.007***	-0.007***	-0.005**
			[0.002]	[0.002]	[0.002]			[0.002]	[0.002]	[0.003]
Experience			-0.007**	-0.007**	-0.011***			-0.015***	-0.015***	-0.018***
			[0.003]	[0.003]	[0.003]			[0.004]	[0.004]	[0.005]
Experience squared			0.000	0.000	0.000**			0.000**	0.000**	0.000***
			[0.000]	[0.000]	[0.000]			[0.000]	[0.000]	[0.000]
Constant	0.085***	0.584***	0.303***	0.717***	0.601***	0.084***	0.316***	0.420***	0.569***	0.375***
	[0.004]	[0.097]	[0.090]	[0.137]	[0.140]	[0.004]	[0.087]	[0.105]	[0.107]	[0.139]
Demographic controls	No	No	Yes	Yes	Yes	No	No	Yes	Yes	Yes
Occ and industry controls	No	No	No	No	Yes	No	No	No	No	Yes
Adjusted R-squared	0.002	0.006	0.014	0.017	0.119	0.002	0.004	0.019	0.02	0.185
Number of observations	22262	21957	22049	21750	21628	11504	11360	11441	11297	11227

Standard errors in brackets. *** signifies statistical significance at the one-percent level, ** signifies the five-percent level, and * the one-percent level.

Local Linear Regression

The OLS regressions with skill indicator variables presented above offer a very flexible technique towards discovering the relationship between earnings instability and skill. As an even more flexible technique, I turn to a semi-parametric econometric method, local linear regression, which I will now discuss following Hansen (2009) and Li and Racine (2007).

The goal is to find a function which relates the expected value of earnings instability (calculated using the log of total labor earnings) to each value of AFQT. Call this function $g()$, where:

$$g(AFQT) = E(earnings\ instability | AFQT = AFQT_0).$$

$AFQT_0$ is a specific value of AFQT.

To this end a series of local regressions are estimated, of the form:

$$instability(y_i) = \alpha + \beta(AFQT_i - AFQT_0) + \varepsilon_i.$$

Note that the AFQT variable is defined as deviations from the specific level of AFQT, called $AFQT_0$. The estimators are the solution to the following minimization problem:

$$\min_{\alpha,\beta} \sum_{i=1}^{n} [instability(y_i) - \alpha - \beta(AFQT_i - AFQT_0)]^2\ 1(|AFQT - AFQT_0| \leq h),$$

where h is the bandwidth, and $1()$ is an indicator function. A kernel is substituted for the indicator function, and the estimators are algebraically equivalent to the weighted least squares estimators, where in this case the weight is the kernel function.

Once we have estimates of the parameters α and β, for each local regression we can set $AFQT_i = AFQT_0$, which gives our estimate of $g(AFQT_0)$: $\hat{g}(AFQT_0) = \hat{\alpha}(AFQT_0)$. In other words, for each local regression, the estimate of the intercept in the local linear regression gives us our estimate of the conditional expectation.

In addition to looking at the local linear plots, it may be helpful to look at some potentially more-familiar kernel regressions. Figure 1.3 plots results using standard kernel regression for each of the four subsamples. Figure 1.4 plots semi-parametric results using the local linear technique described above. The pictures were generated using an epanechnikov kernel and the optimal rule-of-thumb bandwidth.

The plots largely confirm the conclusion from the categorical variable models. For the portion of the AFQT distribution to the right of the middle-fifties, earnings instability is rising in AFQT. For the portion to the left, earnings instability is falling in AFQT. Over the entire AFQT distribution earnings instability is U-shaped. Job stayers exhibit a stronger U-pattern than the sample which allows workers to change jobs during the rolling window. Low-skill workers have greater earnings instability than high-skill workers — indeed, in some of the pictures, the right tail is rather weak. And low- and high-skill workers are most equal using the longer window.

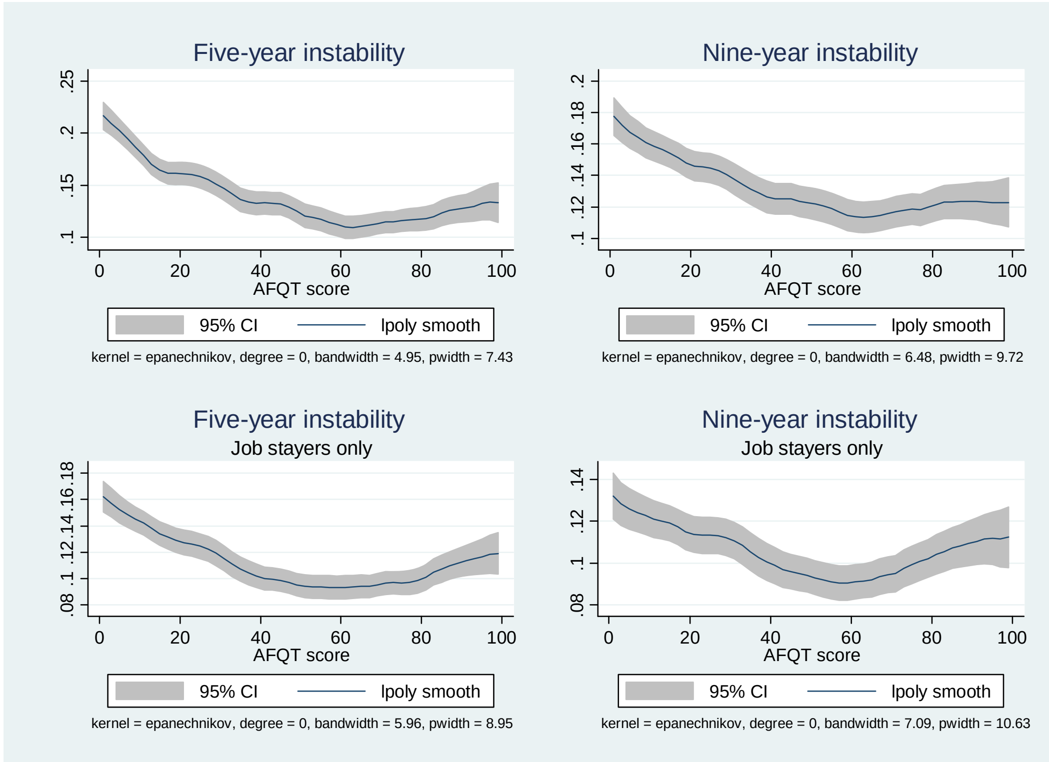

FIGURE 1.3: Kernel regression results by subsample.

FIGURE 1.4: Local linear regression results by subsample.

Quadratic Model

The categorical variable model and semi-parametric techniques discussed above are quite flexible. It is interesting to see if a parametric specification can be fit to the data successfully. Since the previous results suggest a U-shaped relationship between earnings instability and AFQT, I estimate a model with a quadratic in AFQT. Specifically, the following equation, labeled Eq. 2, is estimated:

$$instability(y_{it}) = \alpha + \beta_1 AFQT + \beta_2 AFQT^2 + X_{it}\delta + \varepsilon_{it}, \qquad (2)$$

where everything is defined as in Eq. 1, and with a quadratic in AFQT replacing the four categorical variables. The pattern of control variables from the categorical variable regressions in Tables 1.3 and 1.4 is repeated for Tables 1.5 and 1.6, which present the results from Eq. 2.

The quadratic specification confirms the evidence found by the previous two methods. In every regression the coefficient β_1 is negative and the coefficient β_2 is positive, implying a U-shape of earnings instability over AFQT. Both coefficients are statistically significant in each regression.

This result is more readily seen by examining Figure 1.5 and Figure 1.6. Both figures plot the predicted level of earnings instability by using estimates of Eq. 2: $\widehat{instability}(y_{it}) = \hat{\alpha} + \hat{\beta}_1 AFQT + \hat{\beta}_2 AFQT^2$. Figure 1.5 plots predicted earnings instability using the five-year and nine-year windows for the subsamples of job stayers and job changers, while Figure 1.6 plots predicted earnings instability for job stayers only. The dashed line (colored blue) in both pictures are predicted earnings instability using the five-year window, and the solid line (colored red) is predicted earnings instability using the nine-year window.

TABLE 1.5: Earnings instability on AFQT score and AFQT score squared

The table reports estimates of Eq. 2. The dependent variable for each regression is the earnings instability measure using either a nine-year or five-year rolling window. (See text for details.) In addition to the control variables shown in the table, regressions which are labeled as including demographic controls include dummies for year, race, region of residence, marital status, and urban residence. Regressions are weighted. Standard errors are robust. AFQT ranges from 0.01 to 0.99.

	Five-year window measure/subsample					**Nine-year window measure/subsample**				
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**	**(10)**
AFQT	-0.248***	-0.206***	-0.192***	-0.162***	-0.146***	-0.135***	-0.130***	-0.085**	-0.087**	-0.074*
	[0.036]	[0.036]	[0.038]	[0.038]	[0.043]	[0.035]	[0.035]	[0.036]	[0.036]	[0.040]
AFQT^2	0.184***	0.157***	0.168***	0.142***	0.141***	0.088**	0.091***	0.070**	0.072**	0.074*
	[0.037]	[0.037]	[0.038]	[0.038]	[0.041]	[0.035]	[0.035]	[0.035]	[0.034]	[0.039]
Log hours worked		-0.139***		-0.122***	-0.114***		-0.088***		-0.076***	-0.074***
		[0.014]		[0.014]	[0.013]		[0.011]		[0.011]	[0.012]
Years of school			-0.006***	-0.005**	-0.004*			-0.008***	-0.007***	-0.006**
			[0.002]	[0.002]	[0.003]			[0.002]	[0.002]	[0.003]
Experience			-0.003	-0.005	-0.008**			-0.016***	-0.017***	-0.018***
			[0.003]	[0.003]	[0.004]			[0.004]	[0.004]	[0.005]
Experience squared			0	0	0.000*			0.000**	0.000***	0.000***
			[0.000]	[0.000]	[0.000]			[0.000]	[0.000]	[0.000]
Constant	0.195***	1.254***	0.335***	1.252***	1.156***	0.165***	0.838***	0.497***	1.069***	0.941***
	[0.008]	[0.105]	[0.063]	[0.125]	[0.135]	[0.007]	[0.087]	[0.059]	[0.107]	[0.137]
Demographic controls	No	No	Yes	Yes	Yes	No	No	Yes	Yes	Yes
Occ and industry controls	No	No	No	No	Yes	No	No	No	No	Yes
Adjusted R-squared	0.004	0.022	0.02	0.034	0.125	0.003	0.017	0.025	0.034	0.175
Number of observations	25287	24850	25029	24599	24446	13515	13292	13432	13210	13122

Standard errors in brackets. *** signifies statistical significance at the one-percent level, ** signifies the five-percent level, and * the one-percent level.

TABLE 1.6: Earnings instability on AFQT score and AFQT score squared: Job stayers

The table reports estimates of Eq. 2. The dependent variable for each regression is the earnings instability measure using either a nine-year or five-year rolling window. (See text for details.) In addition to the control variables shown in the table, regressions which are labeled as including demographic controls include dummies for year, race, region of residence, marital status, and urban residence. Regressions are weighted. Standard errors are robust. AFQT ranges from 0.01 to 0.99. Only workers who do not change their primary employer during the rolling window are included in the sample. (See text for details.)

	Five-year window measure/subsample					**Nine-year window measure/subsample**				
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**	**(10)**
AFQT	-0.223***	-0.207***	-0.200***	-0.188***	-0.159***	-0.137***	-0.134***	-0.120***	-0.120***	-0.097**
	[0.033]	[0.033]	[0.035]	[0.035]	[0.039]	[0.032]	[0.032]	[0.033]	[0.033]	[0.039]
AFQT^2	0.183***	0.171***	0.176***	0.164***	0.144***	0.114***	0.113***	0.109***	0.108***	0.098***
	[0.033]	[0.033]	[0.035]	[0.035]	[0.037]	[0.032]	[0.032]	[0.032]	[0.032]	[0.037]
Log hours worked		-0.064***		-0.055***	-0.049***		-0.030***		-0.025**	-0.022*
		[0.013]		[0.013]	[0.013]		[0.011]		[0.011]	[0.012]
Years of school			-0.004*	-0.003*	-0.002			-0.006***	-0.006***	-0.005*
			[0.002]	[0.002]	[0.002]			[0.002]	[0.002]	[0.003]
Experience			-0.006**	-0.007**	-0.011***			-0.015***	-0.015***	-0.017***
			[0.003]	[0.003]	[0.003]			[0.004]	[0.004]	[0.005]
Experience squared			0	0	0.000**			0.000**	0.000**	0.000***
			[0.000]	[0.000]	[0.000]			[0.000]	[0.000]	[0.000]
Constant	0.158***	0.649***	0.355***	0.765***	0.648***	0.133***	0.364***	0.451***	0.599***	0.395***
	[0.007]	[0.097]	[0.089]	[0.136]	[0.138]	[0.007]	[0.086]	[0.105]	[0.106]	[0.137]
Demographic controls	No	No	Yes	Yes	Yes	No	No	Yes	Yes	Yes
Occ and industry controls	No	No	No	No	Yes	No	No	No	No	Yes
Adjusted R-squared	0.003	0.007	0.014	0.017	0.119	0.002	0.003	0.018	0.019	0.184
Number of observations	22262	21957	22049	21750	21628	11504	11360	11441	11297	11227

Standard errors in brackets. *** signifies statistical significance at the one-percent level, ** signifies the five-percent level, and * the one-percent level.

The U-shape shown in the plots is evident. For all four lines, the U bottoms out over the high-fifties and low-sixties of AFQT score. The left tail is taller than the right tail, implying that extremely low-skill workers experience greater earnings instability than do extremely high-skill workers. Figure 1.2 shows that these tails are not merely artifacts of the functional form; the distribution over AFQT approximates a uniform, and there is a good deal of mass over the very-low and very-high AFQT scores.

Results For The 1990s & 2000s

Average, labor-market-wide earnings instability increased during the 1980s, remained at its new, higher level during the 1990s, and may have begun rising again in the early 2000s (Moffitt and Gottschalk, 2008).[11] While the focus of this paper is not on the macro trends of earnings instability, it is interesting if the U-shaped relationship between earnings instability and skill documented above looks different in different decades, and if these differences correspond to periods in which the trends were changing.

Figures 1.7 and 1.8 plot the kernel regression and local linear regression estimates for the 1990s, estimated as described previously. While a few of the pictures do seem to have an increasing right tail, on balance the evidence is that earnings instability was not larger in magnitude for high-skill workers than for middle-skill workers. Figure 1.8, the local linear regression plots, shows that earnings instability declines in AFQT score until approximately AFQT = 40, and then remains

[11] Some authors have found that earnings instability did not increase over the 1980s — e.g., Dynarski and Gruber (1997). These authors employ a differences model to calculate earnings instability, using as their measure of earnings instability the variance of year-to-year changes in earnings. See Moffit and Gottschalk (2008) for a discussion of the differences between these methods.

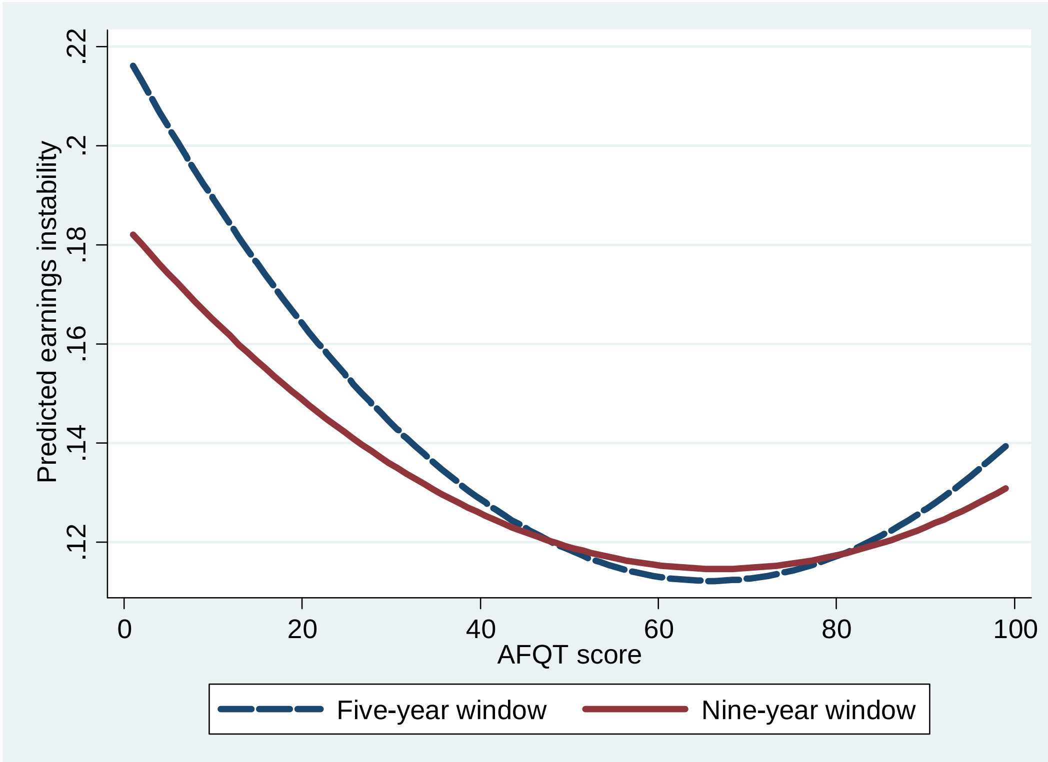

FIGURE 1.5: Predicted earnings instability

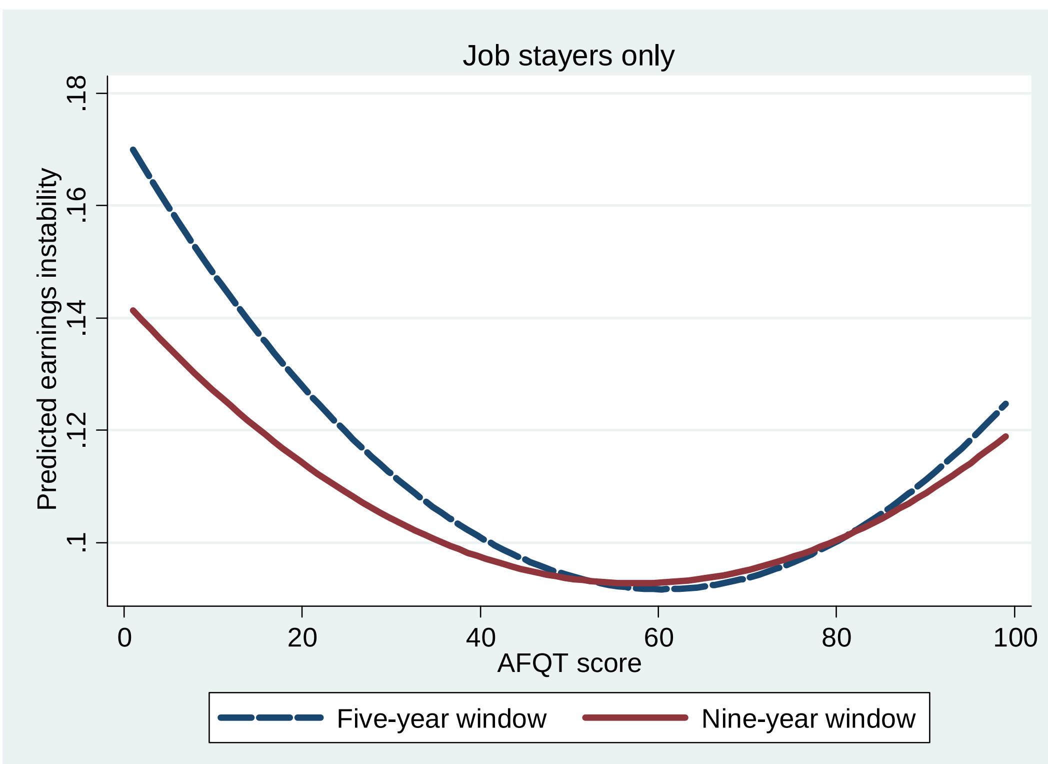

FIGURE 1.6: Predicted earnings instability. Job Stayers only.

flat for the remainder of the AFQT distribution. Figure 1.7 shows the same result in the top two panels, and shows a something of an upward bend for job stayers. Figures 9 and 10 produce analogous plots for the 2000s. For nearly all of the plots for the 2000s, earnings instability is rising in AFQT for AFQT scores greater than 60.[12]

To conclude, the result that low-skill workers experience less stable earnings than high-skill workers occurs in both the 1990s and 2000s. However, the result that high-skill workers have less stable earnings than middle-skill workers only holds for the 2000s. This implies the possibility that instability of earnings for high-skill workers is driving the increase in the macro trend for the 2000s. A more detailed investigation of this result and its potential causes is the subject of future work.

Single-Job Holders

The results discussed above are generated using total labor earnings for workers from all jobs. Workers select into multiple-job holding to supplement income, which may have an effect on earnings instability. In addition, it is easy to imagine that workers with volatile income may take a second job in order to smooth their total labor earnings. Studying workers with only one job is interesting because it may be the case that these workers are in a more stable labor market situation than multiple-job holders. If the U-shaped relationship found using total labor earnings holds for single-job holders, then we may be more convinced that this is a stable

[12] In some of the (unreported) categorical variable regressions for the 2000s, the coefficient on AFQT20 is insignificant, while the coefficient on AFQT100 is positive and significant. In several of the categorical variable regressions, the magnitude of the AFQT100 coefficient is nearly as large or larger than for AFQT20. A table of these results is available upon request.

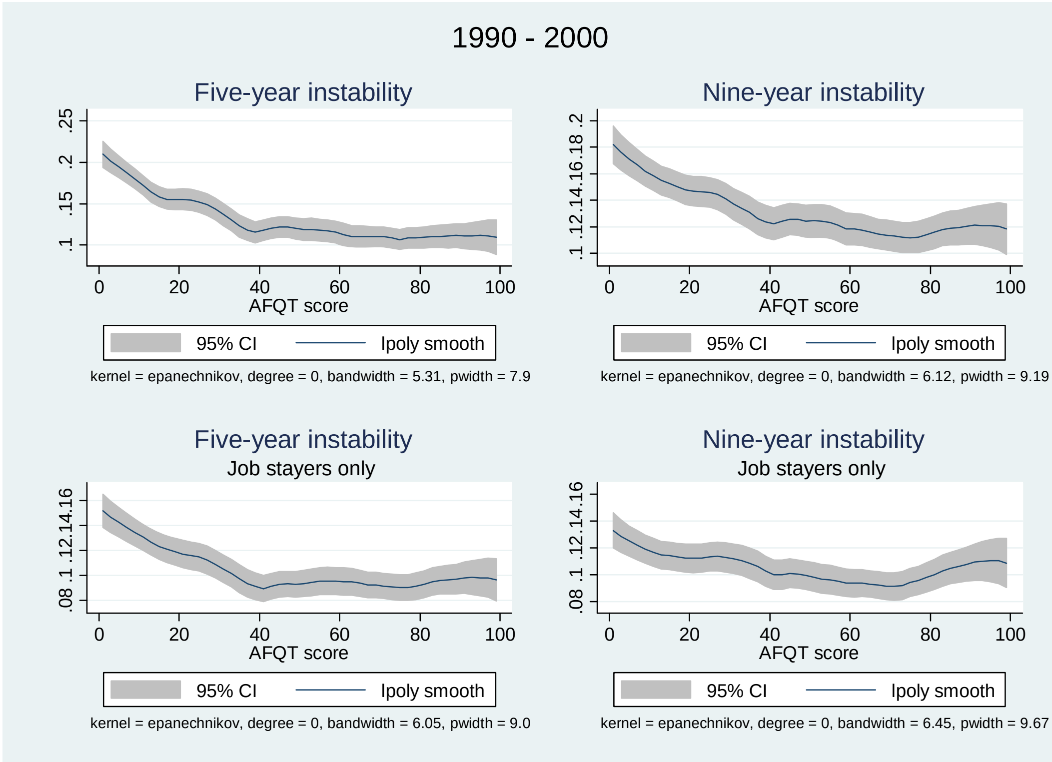

FIGURE 1.7: Kernel regression results by subsample.

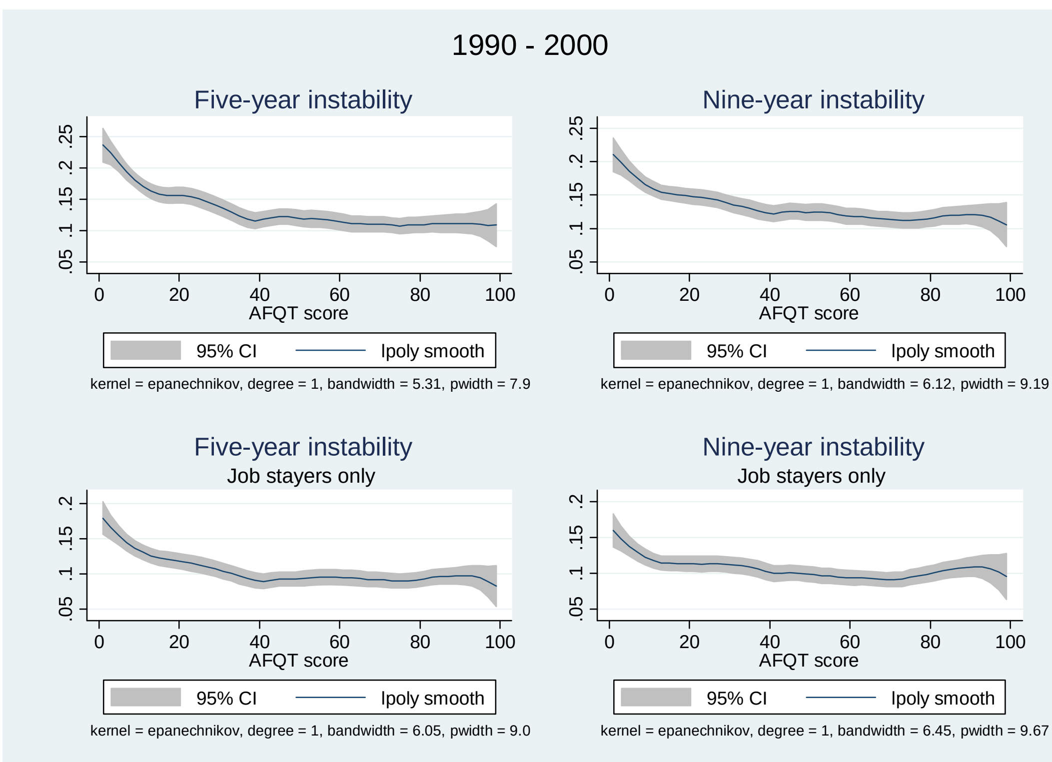

FIGURE 1.8: Local linear regression results by subsample.

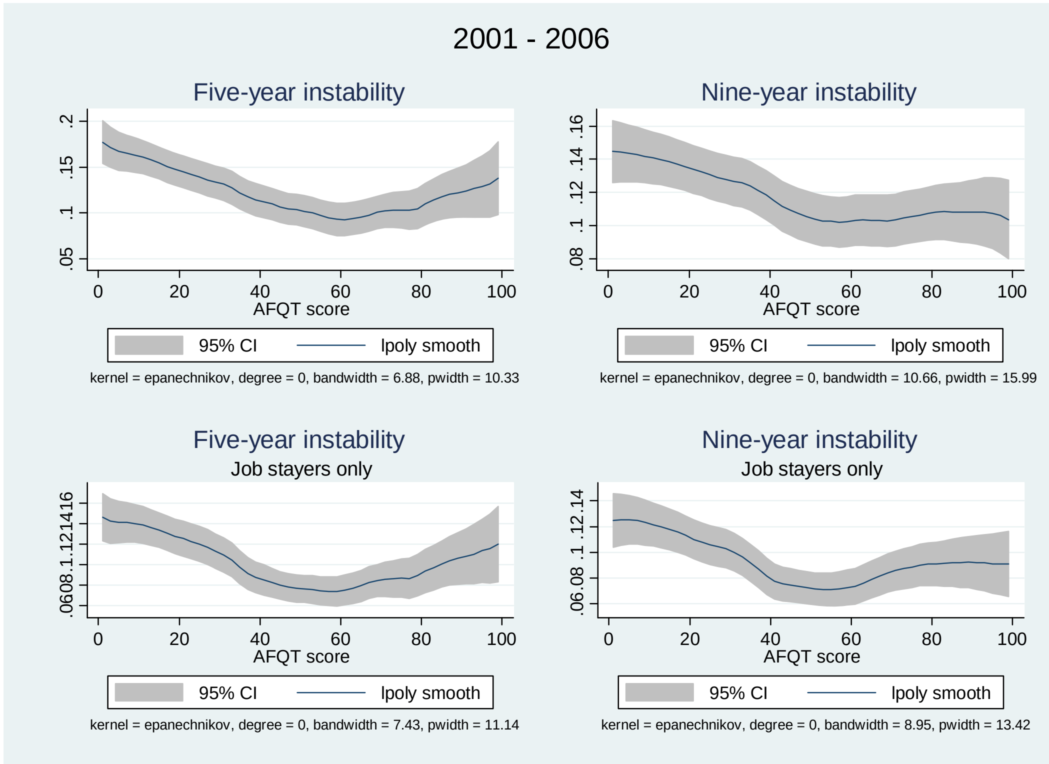

FIGURE 1.9: Kernel regression results by subsample.

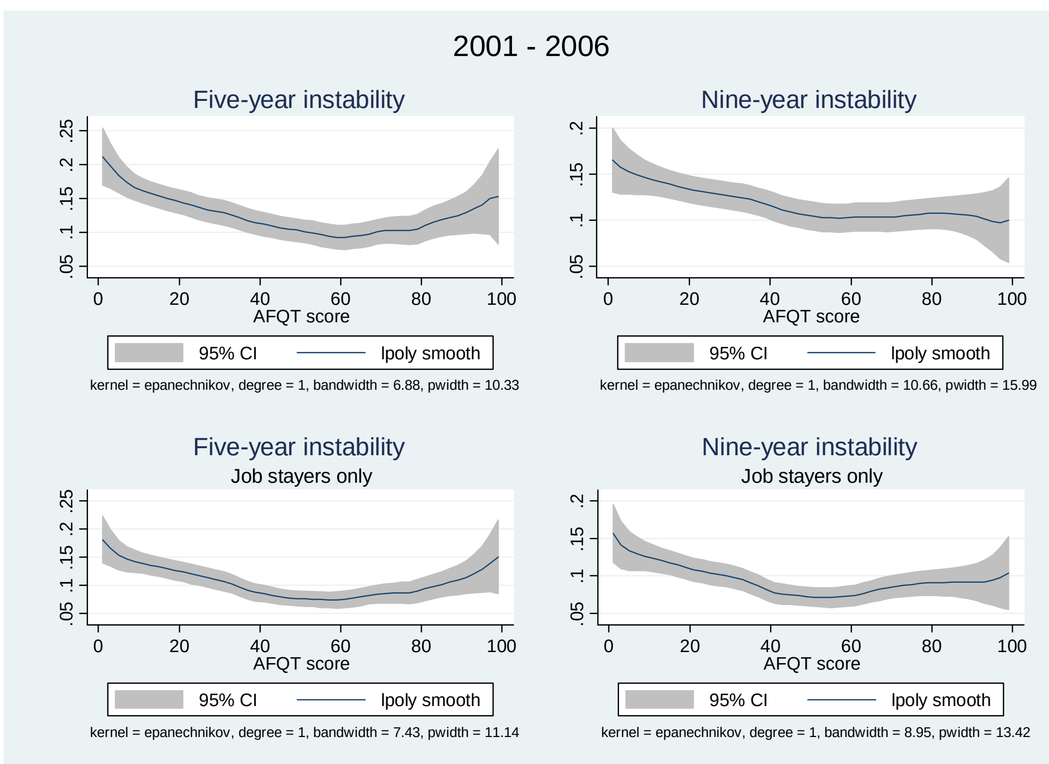

FIGURE 1.10: Local linear regression results by subsample.

feature of compensation, and not an artifact of income from multiple streams or from low-paying, low-hours jobs.

Table 1.1, previously discussed, shows that average earnings instability for the sample of single-job holders is nearly half that of the sample of single- and multiple-job holders. The second column of Table 1.2 shows that, although single-job holders have relatively more stable earnings, earnings instability still follows a U-shape pattern over skill. For all four subsamples of single-job holders, workers in the middle AFQT quintile have the most stable earnings.

Table 1.7 confirms the simple means from Table 1.2 by estimating Eq. 1, the categorical variable model, for the sample of single-job holders. Everything is the same in these regressions as before, except that the sample used to estimate the coefficients is now worker-years where the worker held only one job. Using either the nine- or five-year earnings instability measure, for the subsample of single-job holders who are allowed to change jobs during the window, the U-shape relationship is very strong. Every coefficient in the top panel of Table 1.7 is positive, and nearly every coefficient is statistically significant. The bottom panel reports results for job stayers. Using the five-year measure, the U-shape remains. However, the controls wipe out the right tail of the U using the nine-year measure.

Figures 1.11 and 1.12 present the plots from kernel and local linear regression for single-job holders. These plots confirm the results from the categorical variable regression. Taken together, the sample of single-job holders allows us to conclude that the U-shape of earnings instability over skill is not a function of multiple-job holding. In addition, these results suggest that high- and low-skill workers have more

TABLE 1.7: Earnings instability on AFQT indicator variables: Single-job holders

The table reports estimates of Eq. 1. The dependent variable for each regression is the earnings instability measure using either a nine-year or five-year rolling window. (See text for details.) In addition to the control variables shown in the table, regressions which are labeled as including demographic controls include dummies for year, race, region of residence, marital status, and urban residence, and controls for years of school, experience, and experience squared. Regressions are weighted. Standard errirs are robust. The AFQT indicators equal 1 if the worker's AFQT score falls within the specified interval, otherwise 0. Only worker-years for which the worker was a single-job holder are included in the estimation samples. (See text for details.)

	Job stayers and job changers									
	Five-year window measure/subsample					**Nine-year window measure/subsample**				
0 < AFQT ≤ 20 Indicator	0.056***	0.048***	0.047***	0.045***	0.042***	0.026***	0.024***	0.021***	0.023***	0.032***
	[0.009]	[0.009]	[0.010]	[0.010]	[0.012]	[0.006]	[0.006]	[0.007]	[0.007]	[0.011]
20 < AFQT ≤ 40 Indicator	0.034***	0.034***	0.030***	0.031***	0.026**	0.034***	0.033***	0.034***	0.034***	0.039***
	[0.008]	[0.008]	[0.009]	[0.009]	[0.012]	[0.009]	[0.009]	[0.010]	[0.010]	[0.012]
60 < AFQT ≤ 80 Indicator	0.017**	0.018***	0.015**	0.016**	0.017*	0.015**	0.014**	0.010	0.008	0.014
	[0.007]	[0.007]	[0.007]	[0.007]	[0.009]	[0.007]	[0.006]	[0.007]	[0.006]	[0.010]
80 < AFQT ≤ 100 Indicator	0.030***	0.031***	0.028***	0.026***	0.024**	0.035***	0.034***	0.022**	0.019**	0.037***
	[0.008]	[0.008]	[0.008]	[0.008]	[0.010]	[0.009]	[0.009]	[0.009]	[0.009]	[0.013]
Adjusted R-squared	0.004	0.016	0.019	0.028	0.181	0.005	0.006	0.031	0.032	0.367
Number of observations	11028	10931	10921	10825	10781	4320	4287	4296	4263	4252

TABLE 1.7 (Continued)

	Job stayers only									
	Five-year window measure/subsample					**Nine-year window measure/subsample**				
0 < AFQT ≤ 20 Indicator	0.043***	0.041***	0.044***	0.044***	0.040***	0.017***	0.017***	0.019***	0.018***	0.030***
	[0.008]	[0.008]	[0.010]	[0.010]	[0.010]	[0.005]	[0.005]	[0.006]	[0.006]	[0.010]
20 < AFQT ≤ 40 Indicator	0.027***	0.027***	0.026***	0.027***	0.020**	0.035***	0.035***	0.036***	0.036***	0.038***
	[0.007]	[0.007]	[0.008]	[0.008]	[0.010]	[0.010]	[0.010]	[0.010]	[0.010]	[0.012]
60 < AFQT ≤ 80 Indicator	0.015***	0.015***	0.012**	0.012**	0.015**	0.011**	0.011**	0.005	0.005	0.011
	[0.006]	[0.006]	[0.006]	[0.006]	[0.008]	[0.005]	[0.005]	[0.006]	[0.005]	[0.009]
80 < AFQT ≤ 100 Indicator	0.025***	0.026***	0.018**	0.018**	0.01	0.024***	0.024***	0.009	0.009	0.018
	[0.007]	[0.007]	[0.007]	[0.007]	[0.008]	[0.007]	[0.007]	[0.007]	[0.007]	[0.011]
Adjusted R-squared	0.003	0.005	0.016	0.017	0.193	0.006	0.006	0.036	0.036	0.392
Number of observations	10294	10237	10193	10137	10105	4057	4038	4036	4017	4008
Log hours control	No	Yes	No	Yes	Yes	No	Yes	No	Yes	Yes
Demographic controls	No	No	Yes	Yes	Yes	No	No	Yes	Yes	Yes
Occ and industry controls	No	No	No	No	Yes	No	No	No	No	Yes

Standard errors in brackets. *** signifies statistical significance at the one-percent level, ** signifies the five-percent level, and * the one-percent level.

volatile earnings than middle-skill workers, even among workers with extremely strong labor market attachment and who are in employment situations which do not compel them to select into multiple-job holding.

Total Family Income

The total family income variable in the NLSY79 captures receipts from labor earnings from all jobs, from business and farm income, and from social insurance programs, including unemployment compensation, AFDC payments, food stamps, education and disability benefits, and other welfare and SSI payments. One of the purposes of government transfer payments is to smooth labor market shocks — if you are injured and become disabled, your labor earnings will take a hit, so the government attempts to offset this by giving you a disability payment.

While the focus of this paper is on compensation — how workers are paid in the labor market — it is interesting to study whether the U-shape of earnings instability over skill holds for total family income, and not just for labor earnings. If, say, low-skill and middle-skill workers are more likely to experience adverse, instability-inducing labor market shocks than high-skill workers, then this might explain the left-tail of the U. If broadening labor earnings to total family income eliminates the U-shape, then this is evidence that transfer payments (and/or other forms of income) are working with respect to the smoothing of labor income.

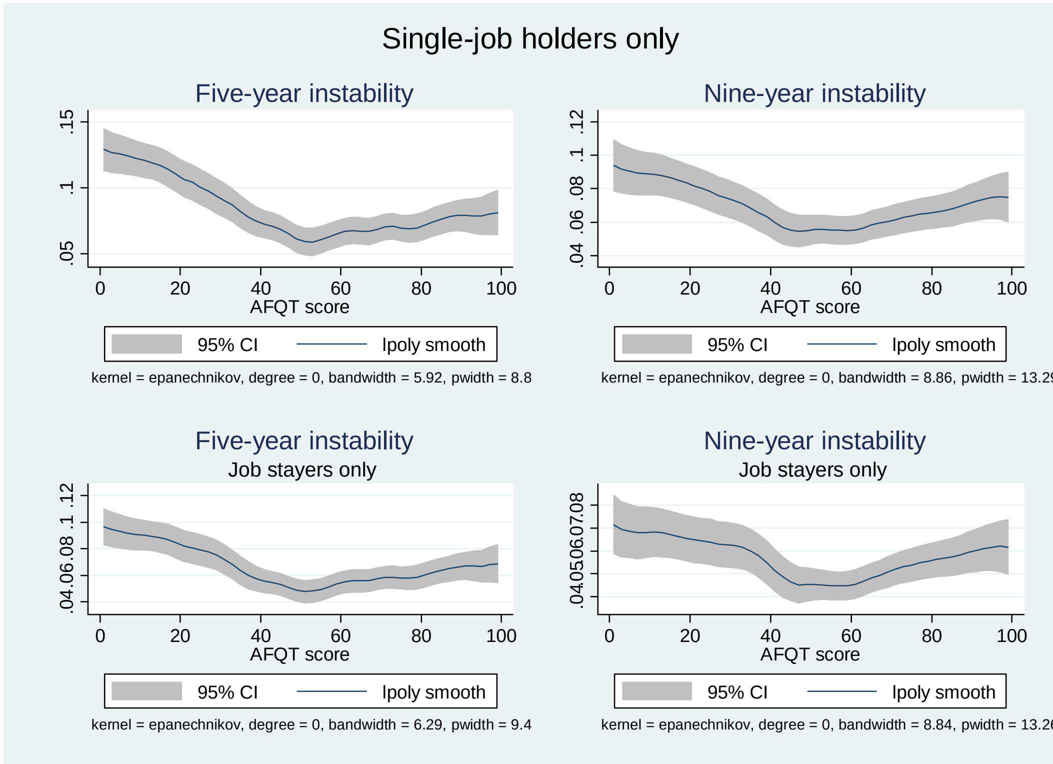

FIGURE 1.11: Kernel regression results by subsample.

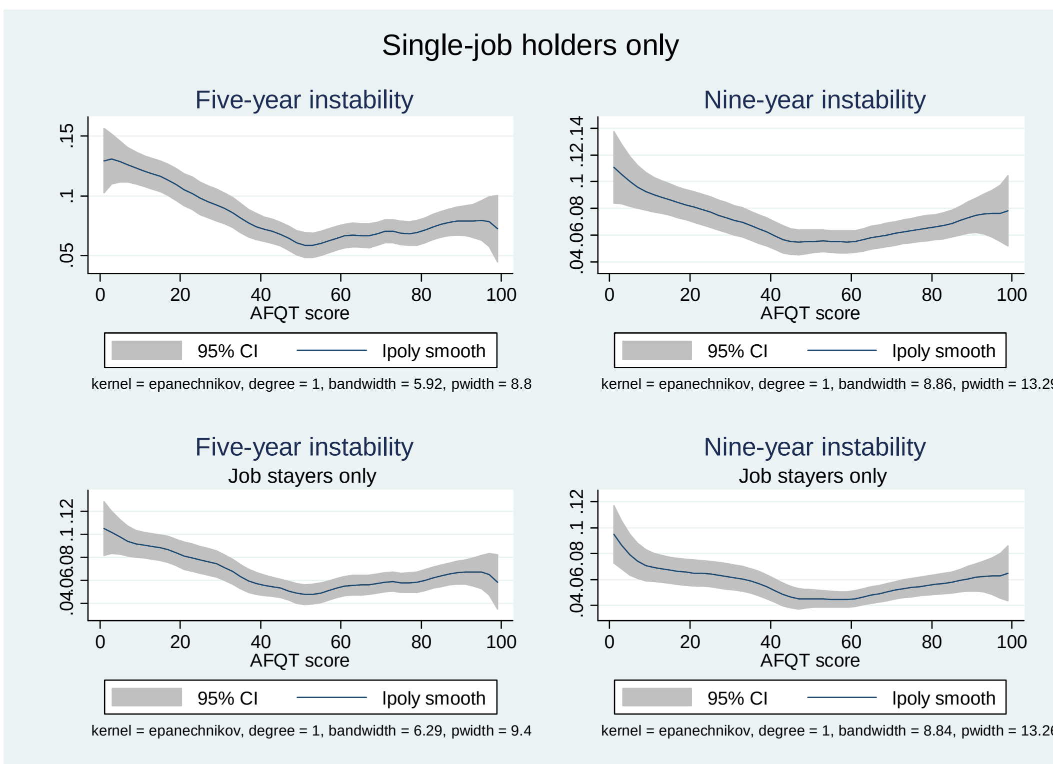

FIGURE 1.12: Local linear regression results by subsample.

Turning again to Table 1.2, we see three interesting items. First, the instability of total family income is almost always larger in magnitude than the instability of labor earnings. Second, for each of the four subsamples, the instability of total family income is most different from the instability of labor earnings for the middle quintile of the AFQT distribution. Third, the U-shape disappears from the simple sample means. Instead, the instability of total family income declines over the first four quintiles of the AFQT distribution, with AFQT100 always larger in magnitude than AFQT80, and sometimes larger than AFQT60.

Figures 1.13 and 1.14 show kernel and local linear regression plots of the relationship between the instability of total family income and AFQT score. They are estimated as described above, the same as with labor earnings. Family income instability declines steadily until around AFQT = 40, then rises over the next ten points, then falls again until AFQT = 60, and then rises. This could be described as a W-shape of total income instability over AFQT, generated by the disproportionate increase in income instability over labor earnings instability for the middle quintile of the AFQT distribution. Further investigation of this result is warranted.

3.3. Discussion

Using three different empirical strategies, this paper finds that the instability of total labor earnings follows a U-shape over AFQT scores, which are used here to proxy for skill. Much of earnings instability is driven by workers changing jobs, but this paper finds that the U-shaped relationship holds even when earnings instability is only calculated for a worker during a time window when he is not changing employers. Furthermore, the U-shaped relationship remains for single-job holders, who may be in an even more stable labor market situation.

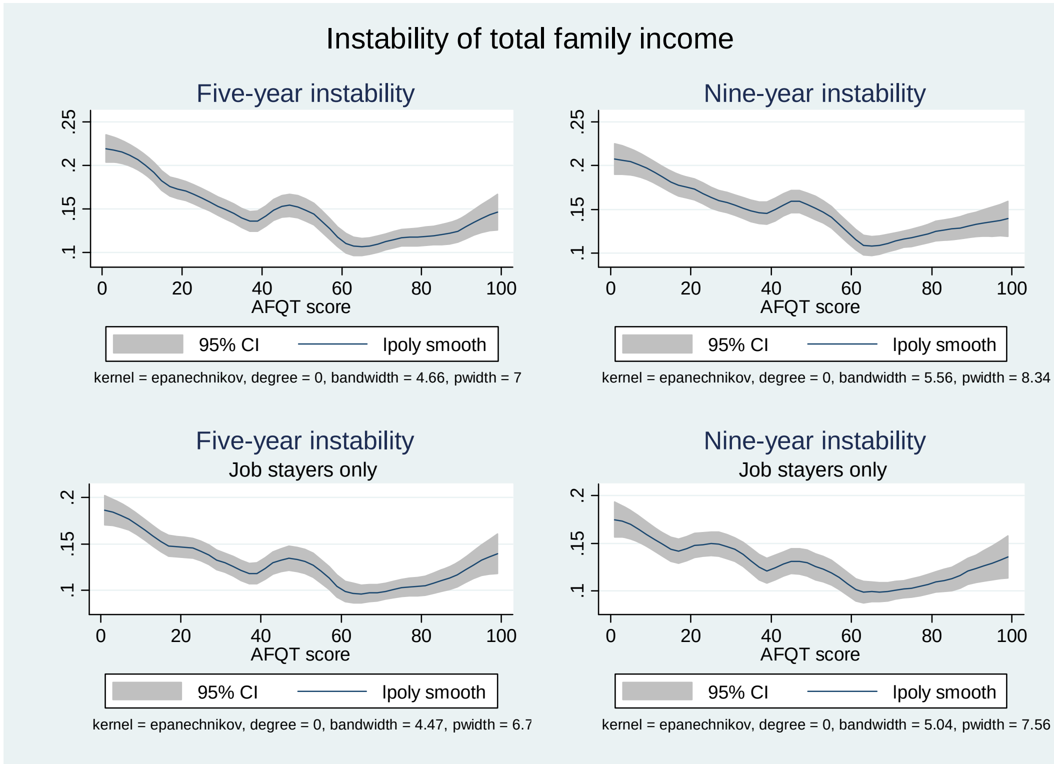

FIGURE 1.13: Kernel regression results by subsample.

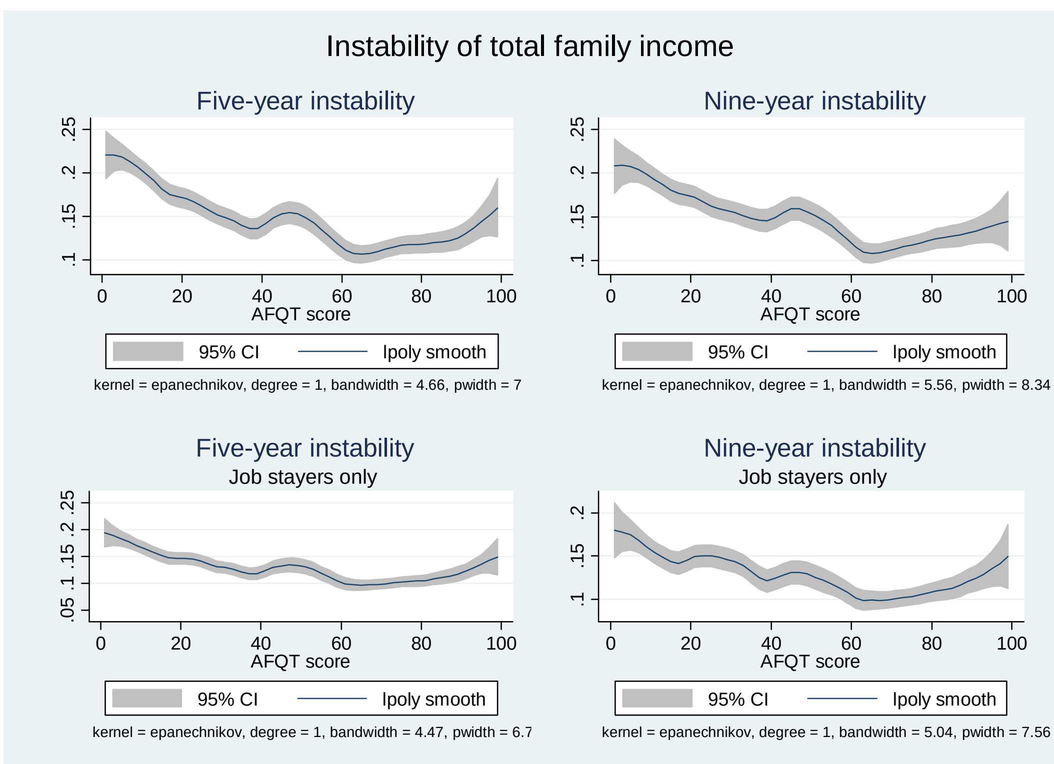

FIGURE 1.14: Local linear regression results by subsample.

With entry and exit from employment, job changing, and multiple-job holding accounted for, one might think that hours of work explains much of the time-series variation in individual earnings: perhaps high-skill workers have relatively more volatile pay because they choose to work longer and more variable hours than middle-skill workers, and perhaps low-skill workers have relatively more volatile pay because they are subject to hours restrictions. However, we find that the U-shape holds both when hours are controlled for and when hours are excluded from the empirical model.

Another likely candidate is occupation and industry — perhaps some occupations and industries have more volatile earnings than others. There is some evidence for this: in some of the categorical variable regressions using the longer window, the specification which controls for occupation and industry has imprecise coefficients. However, on balance, the weight of the evidence is that controlling for occupation and industry does not eliminate the U-shape.

One possible explanation for the finding is that earnings instability means something different for high-skill and low-skill workers. Perhaps low-skill workers experience earnings instability because they are in "bad jobs", both within and across industry and occupation. The existence of a continuum of job quality has been postulated by many economists (for example, Abowd *et al.*, 1999 and Schmutte, 2010), and there is evidence that labor market segmentation results in some workers holding jobs with scant returns to education and experience (Dickens and Lang, 1985). These "bad jobs" have been thought to be characterized by low returns to education and experience, periods of unemployment, and little opportunity for advancement. The results from this paper suggests that we may want to add unstable earnings to this list of characteristics.

High-skill workers, on the other hand, may welcome unstable earnings due to a mean-variance tradeoff — they may be hoping for a big payday — and middle-skill workers may simply fall between the two.

An alternative explanation builds on the research examining the hollowing out of the skill distribution. Autor *et al.* (2003) finds evidence that computers substitute for middle-skill workers who are employed to perform routine tasks, like bookkeepers or bank tellers. These jobs were never the lowest paid, and they were never offered to the lowest-skilled workers. They required precision and repetition, so middle-skill, middle-income workers were employed in these jobs. But they are easily replaced by computers. (For example, bookkeepers are replaced by Quicken and bank tellers by ATM machines.) At the same time, computers largely had no effect on the lowest-skill jobs, such as janitors, and they had a complementary effect on high-skill workers. Autor *et al.* finds evidence that as the price of computers fell over recent decades the relative demand for middle-skill workers (whose jobs could be replaced by computers) fell.

Goos and Manning (2007) add to this evidence by documenting that the pattern of employment changes in Britain from 1975 through 1999 shows a relative increase in employment in low- and high-paying jobs, associated with nonroutine tasks, and a relative decrease in employment in middle-paying jobs — employment changes exhibit a U-shaped pattern over skill level. This is precisely what one would expect to see if computers were replacing middle-skill workers working in jobs which require routine tasks. Autor *et al.* (2006) finds similar results for the United States. Autor and Dorn (2011) find that from 1980 to 2005 both employment growth and wage growth follow a U-shape over skill. If employment and *average* wages are growing in low-skill and high-skill occupations at faster rate than middle-skill occupations, then it may follow that the *variance* of earnings for workers of low- and high-skill will be relatively larger.

Perhaps the relationship between earnings instability and skill level documented in this paper is related to the hollowing out of the skill distribution. Perhaps all the action is outside of the middle — employment growth is concentrated among high- and low-skill jobs, middle-skill workers are sorting into those jobs, and this activity is generating relatively higher earnings instability for low- and high-skill workers. More research is needed to support or refute this hypothesis.

IV. Earnings Instability & Firm Instability

The results we have seen so far suggest that firms pay workers of different skill earnings which differentially vary over time. This suggestion invites a deeper study of why firms pay unstable earnings. Specifically, I now turn to a study of the relationship between the instability of firm employment and the instability of the earnings paid to the firm's workers.

In the perfectly competitive model of the labor market firms take the price of labor as given. The price of labor is set at the market level, and responds to *market level* demand and supply conditions. The firm faces a perfectly elastic labor supply curve, pays the going rate for labor, and derives its labor demand from profit maximization.

In this model of the labor market, the firm does not have the option to pass volatility onto its workers, or to protect workers from volatile economic conditions. The firm enters the labor market as a price taker facing a perfectly elastic labor supply curve. If the firm pays above the market price it will earn negative profits and go out of business. If the firm pays below the market price then it will lose all its workers and shut down. A firm experiencing volatility would have to adjust along margins other than the price of labor, and a firm wishing to shield its workers from external conditions would be unable to do so.

Other models of the labor market offer different predictions. Perhaps the simplest model is one in which a firm's profits or output are distributed among its workers in each period. If profits or outputs experience time series variation, then so too will worker's earnings. As another simple exercise, consider a general equilibrium economy with one worker and one firm. The worker supplies one unit of labor to the firm, irrespective of his earnings. The firm produces output according to $f(N) = N_t^{\gamma_t}$, where N_t is labor input and $0 < \gamma_t < 1$. Interpret the parameter γ_t as the firm's time-varying technology which converts labor into output. The firm maximizes profit, taking the price of labor as given, and demands labor according to the following function: $N_t^d = \left(\gamma_t / w_t\right)^{1/1-\gamma_t}$. In this model, the equilibrium wage will be a function of the time-varying technology parameter. So the introduction of new technology to the firm (say, computers) will affect the earnings of workers in that firm.

Richer models also suggest that the firm will pass volatility onto its workers. Consider a firm-employee negotiation. Both firm and worker know that the firm will experience shocks. The firm and worker want to write a contract which specifies both the wage the worker will receive and the hours the worker will work in each state of the world. Call the shock α. Call the contract $C(\alpha) = \{w(\alpha), h(\alpha)\}$, where $h(\alpha)$ are the hours worked conditional on α and $w(\alpha)$ is the wage paid conditional on α. Suppose that both the firm and the worker have concave utility, the firm over profits and the worker over the wage and leisure. Suppose that the worker is endowed with one unit of time.

The prospective employee must be induced to work at the firm — at the time the contract is signed, the expected utility of the contract must exceed the workers opportunity cost. Let the workers next-best offer be summarized by $\overline{U}$. Then the contract ensures the survival of both the firm and the worker by satisfying the following:

$$\max_{C(\alpha)} EV[\pi(\alpha)] \quad s.t. \quad EU[w(\alpha), 1 - h(\alpha)] \geq \bar{U}.$$

The contract maximizes the expected utility of profit for the firm $EV[\pi(\alpha)]$ subject to the participation constraint of the worker.

Let the worker's utility function be $\log w(\alpha) + \log(1 - h(\alpha))$. A first order condition for the maximization problem is $V'[\pi(\alpha)] = \lambda / w(\alpha)$, where λ is the Lagrange multiplier. For state of the world z and z', we have the following result:

$$\frac{V'[\pi(z)]}{V'[\pi(z')]} = \frac{w(z')}{w(z)}.$$

It is evident that the firm and the worker share the risk associated with the shock. Specifically, if $z' \neq z$ then $w(z') \neq w(z)$. The shock α which hits the firm will be absorbed partly by the firm, but the firm will pass some of this risk onto the worker, inducing time series variation in the worker's wage.[13]

In bargaining models, increasing firm performance is correlated with higher wages for workers — the time series variance of firm performance is positively correlated with the time series variance of worker earnings. Finally, in models of

[13] See Sherwin Rosen (1985) for a survey of the implicit contracts literature. If the firm in this example were risk neutral — i.e., if it did not have concave utility over profits — then the model would result in a constant wage for the worker. However, the firm would still adjust hours in response to the shock, which would cause time-series variation in labor earnings, the variable of interest in this paper. The example also assumes that workers face prohibitive costs in switching firms, as the only parameter which matters in writing the contract is the opportunity cost at the time the contract is written. For an implicit contract model with costly mobility see Beaudry and DiNardo (1991). For empirical investigations of contract models, see Card (1986), Abowd and Card (1987), and Ham and Reilly (2002).

monopsony, the monopsonistic firm faces the upward sloping, market level labor supply curve, inducing a positive correlation between firm size and worker earnings.

Empirically, there is evidence that firms have some measure of control over the earnings of their workers — that firms are not pure price takers in the labor market. The existence of sizable firm effects in earnings regressions using linked employer-employee data is evidence in favor of the hypothesis that firms have some discretion in the compensation of their workers (Abowd *et al.*, 1999; Goux and Maurin, 1999). Recent research by Brummund (2011) and Webber (2011) finds a distribution over market power at the firm level, implying that variance exists over the ability of firms to influence the earnings of their workers. If firms do have this ability, then some may choose to vary worker earnings with firm performance.

Bertrand (2004) studies the extent to which product market competition influences a firm's decision to shield its workers from external labor market conditions. Specifically, she studies whether increased competition from imports induced by globalization increases the sensitivity of workers' earnings to the current unemployment rate and decreases the sensitivity of workers' earnings to the unemployment rate at the time of hire. She finds evidence of both effects, and concludes that the labor market is more like a spot market than it used to be.

While Bertrand (2004) does not specifically study the relationship between the instability of firm performance and the instability of worker earnings, her results imply that the correlation may exist: if increasing foreign competition makes firms less likely to shield workers from external conditions, then the earnings of those workers will be characterized by an increase in time series variation. Provided that increasing foreign competition makes the product market riskier for firms, the two may be correlated.

To my knowledge, Comin *et al.* (2009) provide the only direct test of whether firms pass volatility onto workers' earnings. Using COMPUSTAT data on publicly traded firms, they study whether the instability of the average wage paid by a firm is correlated with the instability of firm sales and employment. They find a robust relationship between the two, implying that firms are passing instability onto workers in the form of more volatile earnings.

While Comin *et al.* (2009) has compelling results, the paper is not without limitations. The sample of firms is restricted to publicly traded firms. It has been shown that firm instability follows very different patterns for privately-held firms than for publicly traded firms (Davis *et al.*, 2006). In addition, since Comin *et al.* (2009) do not have linked data, they study the instability of *average* earnings at a firm-year, and not the instability of *worker-level* earnings. Furthermore, Comin *et al.* (2009) use a noisy measure of average earnings — they divide the firms total annual wage bill by total employment, and are not able to control for the entry and exit of workers. Finally, only twenty percent of COMPUSTAT firms report their wage bill, and those which do report have less firm volatility than those which do not.

Linked employer-employee data are needed for a more complete investigation of the question at hand. In this paper, I investigate the relationship between firm employment instability and earnings instability using data from the Longitudinal Employer-Household Dynamics (LEHD) program of the U.S. Census Bureau.[14] In the next section, I describe the data.

---

[14] The LEHD data have been used to study earnings instability topics prior to this paper. Gottschalk, McEntarfer, and Moffitt (2008) estimate the trend in earnings instability from 1991 to 2003 in the LEHD data and compare it to the estimated trend in the PSID. They find that the trend is very similar in the two data sets. (They also find that the cross-sectional variance of earnings is quite different.) Celik *et al*. (2009) use LEHD data to study the importance of employment fluctuations and job changes in explaining the trends in earnings instability.

4.1. LEHD Data & Sample

The LEHD program is a federal-state partnership between the U.S. Census Bureau and all fifty states.[15] The states supply LEHD with Unemployment Insurance (UI) administrative files, providing LEHD with a report of worker-level UI-covered quarterly earnings. UI records cover approximately ninety-eight percent of wage and salary payments in private, non-farm jobs. Each earnings record is associated with a state UI account number which identifies the employing entity of the worker. In addition, the states supply an extract of their ES-202 report, providing LEHD with information on the firms in which the workers are employed, including employment and industry.

The UI and ES-202 files are the core of the LEHD data. The UI records are a worker-employer link — a job. So the LEHD data are a job frame. The unique person identifier for each record allows for the workers' demographic characteristics to be linked from other administrative and survey records. Demographic information in the LEHD data now includes sex, date and place of birth, citizenship status, race, ethnicity, and education. The unique identifiers for workers and firms allow for the study of a wide variety of topics, including job-to-job transitions, worker earnings histories, and coworker characteristics. Abowd *et al.* (2009) provides a comprehensive overview and description of the LEHD data.[16]

---

[15] LEHD currently has data for every state except Massachusetts, the most recent state to join.

[16] Interested researchers can access these data through the network of Census Research Data Centers. There are currently twelve Census RDCs in the United States. The RDC network is part of the Center for Economic Studies of the U.S. Census Bureau. The RDCs are Census Bureau facilities housed in partner institutions. For more information on conducting research using the LEHD data (or other data) in a Census RDC, please see this website: http://www.census.gov/ces/rdcresearch/index.html.

Sample

The sample of LEHD data on which the earnings and firm employment instability measures were computed consists of all male long-form Census 2000 records, an (approximately) random one-in-six sample of the U.S. population, from the states of California, Colorado, Idaho, Illinois, Indiana, Kansas, North Carolina, Oregon, Washington, and Wisconsin, for the years 1992 through 2009.[17] As with the previous analysis of earnings instability and worker skill, workers younger than twenty-five and older than sixty-five are dropped in order to avoid the time series variation in earnings associated with beginning and ending a career. To ensure that outliers are not driving the regression results, worker-years with earnings in the first and ninety-ninth percentile of the earnings distribution are dropped. For each worker, only the dominant job in each quarter is studied. Quarterly earnings from dominant jobs are summed to create annual earnings, which are then logged. The earnings instability measure is calculated on log annual earnings as previously described.

The nine-year earnings instability measure for year t requires four years of data on each side of t. With earnings data starting in 1992, the earliest year for which a worker's earnings instability could be calculated is thus 1996. Likewise, the last year for which the worker's earnings instability could be calculated is 2005. Unlike with the previous analysis of earnings instability and skill, I keep only those records for which both the five-year and the nine-year earnings instability measure are calculated.[18] Also unlike the previous analysis, only job stayers are studied. The research question of interest is whether firms are passing instability to workers'

[17] The LEHD data are relatively new, and many states do not start until later in the 1990s. These states were chosen because their data begin in either 1991 or 1992, and because many of them are large and representative. Gottschalk, McEntarfer, and Moffitt (2008) use the LEHD data to study earnings instability, and use a subset of the states in this paper. Celik *et al.* (2009) use a similar set of states.
[18] This was done to avoid internal Census complications associated with releasing analysis conducted on two different but very similar samples of confidential Census data. In addition, unlike the NLSY79, the sample sizes in the LEHD data are so large that there is little cost to doing this.

earnings, so looking at workers who change jobs is not helpful. This leaves a baseline sample of approximately five million worker-years composed of approximately one million unique workers and 250,000 unique firms running from 1996 through 2005, and from the states listed previously.[19]

As discussed previously, this sample of workers is characterized by very strong labor market and employer attachment. To have earnings instability calculated for a given year, the worker must have worked for the same employer for four years before and four years after the year in question. These workers may be among the most stable employed by their respective firms. Studying them will go a long way towards eliminating the concern that a relationship between earnings and firm stability is being driven by workers changing jobs, or workers entering and leaving employment, or workers who are not strongly committed to the labor market. We can be reasonably confident, then, that the relationship between earnings and firm employment instability calculated on this sample of workers reflects a feature of the way workers are paid, and is not driven by other factors.

Tables 1.8a and 1.8b present summary statistics from the regression sample. Average earnings instability is roughly comparable in magnitude to the average from the NLSY79. Unlike the instability measured calculated using the NSLY79, though, earnings instability is larger in magnitude using the longer window.

Firm employment instability is measured using log annual employment, and is calculated in exactly the same way as earnings instability. Employment is a natural measure of the scope of economic activity taking place at the firm, and it is intuitive to think that a firm with relatively greater time series variance in employment — in economic activity — is in some sense relatively less stable.

[19] Observation numbers are rounded for confidentiality protection.

TABLE 1.8a: LEHD sample summary statistics

	Mean	Std. Dev.
Log annual earnings	10.132	0.601
Firm size	4540.987	12561.390
Earnings instability		
Nine-year	0.100	0.275
Five-year	0.042	0.192
Firm instability		
Nine-year	0.044	0.134
Five-year	0.023	0.092

Table 1.8b presents the distributions for the sample over North American Industry Classification System (NAICS) sector, race, state, and education categories. Over one-quarter of the observations come from the manufacturing sector. Over ten percent come from public administration, with an additional ten percent from educational services. Construction, wholesale trade, retail trade, and transportation and warehousing are the remaining industries which constitute over five percent of the sample. The sample is 83.3 percent white. Interestingly, the number of observations with an African American worker is roughly the same as the number with an Asian worker. Over sixty percent of the samples have at least some college education.

TABLE 1.8b: LEHD sample summary statistics

Distribution over:

NAICS sector	Percent	Race	Percent
Agriculture (11)	1.85	White	83.3
Mining (21)	0.54	Black	4.47
Utilities (22)	2.41	U.S. Indian or Alaskan Native	0.91
Construction (23)	7.14	Asian	4.55
Manufacturing (31-33)	26.03	Pacific Islander	0.18
Wholesale Trade (42)	6.27	Two or more	6.6
Retail Trade (44-45)	7.84		
Transportation and Warehousing (48-49)	5.13		
Information (51)	2.42		
Finance and Insurance (52)	2.26		
Real Estate (53)	0.94		
Professional Services (54)	3.7		
Management (55)	1.04		
Administrative (56)	1.84		
Educational Services (61)	10.3		
Health Care and Social Assistance (62)	3.99		
Arts, Entertainment, and Recreation (71)	1.03		
Accommodation and Food Services (72)	1.8		
Other Services (81)	2.24		
Public Administration (92)	11.23		

Education categories	Percent	State	Percent
No school	1.02	California	31.6
Nursery to 4th grade	0.57	Colorado	4.7
5th or 6th grade	1.64	Idaho	1.83
7th or 8th grade	1.17	Illinois	19.02
9th grade	1.18	Indiana	3.65
10th grade	1.41	Kansas	4.78
11th grade	1.56	North Carolina	7.73
12th grade, no diploma	2.72	Oregon	4.61
High school graduate	26.22	Washington	7.28
< 1 year of college	7.77	Wisconsin	14.81
1+ years of college	16.55		
Associate degree	8.61		
Bachelor's degree	18.64		
Master's degree	7.27		
Professional degree	1.89		
Doctorate degree	1.79		

4.2. Empirical Strategies & Regression Results

To investigate the relationship between the instability of firm employment and the instability of the earnings of the firm's workers, the following equation, Eq. 3, is estimated:

$$instability(y_{ift}) = \alpha + \varphi\ instability(e_{ft}) + X_{ift}\delta + \varepsilon_{ift}, \qquad (3)$$

where y_{ift} is the earnings of worker i employed by firm f in year t; e_{ft} is employment of firm f which employs worker i in year t; X_{ift} is a matrix of firm and worker characteristics, and includes controls for industry, race, education category, age and age squared, year, and state. Because the size of the firm may have an effect on the volatility of the earnings it pays its workers, the log of employment is included in X_{ift} as well. ε_{ift} is the error term, and δ and α are parameters. The parameter of interest is φ, which measures the effect of firm employment instability on earnings instability.

Columns 1 through 8 of Table 1.9 report the results. The odd-numbered columns use the five-year instability measure for both earnings and firm employment instability, and the even-numbered columns use the nine-year measure. The first two columns are estimated with no control variables. They show a positive and statistically significant relationship between earnings instability and firm employment instability. Columns 3 and 4 add a control for firm size. The coefficient is negative and statistically significant, suggesting that larger firms pay more stable earnings to their workers. The sign and statistical significance of the coefficient on firm employment instability, φ, is robust to this control. Columns 5 through 8 add controls

TABLE 1.9: Earnings instability and firm instability

The table reports estimates of Eq. 3 and Eq. 4. The dependent variable for each regression is the earnings instability measure using either a nine-year or five-year rolling window. The odd-numbered columns use the five-year window, and the even use the nine-year window. (See text for details.) Firm and worker controls include race, education category, industry, year, and state dummies, and age and age squared. Standard errors are clustered on firm.

	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**	**(10)**
Firm instability (Five-year)	0.152***		0.143***		0.142***		0.138***		0.074***	
	[0.007]		[0.008]		[0.007]		[0.007]		[0.008]	
Firm instability (Nine-year)		0.189***		0.183***		0.180***		0.178***		0.148***
		[0.007]		[0.007]		[0.007]		[0.007]		[0.010]
Log number of employees			-0.003***	-0.002***			-0.002***	-0.001	-0.008***	0.005
			[0.001]	[0.001]			[0.001]	[0.001]	[0.004]	[0.004]
Constant	0.038***	0.092***	0.053***	0.106***	0.259***	0.740***	0.264***	0.744***	0.303***	0.765***
	[0.001]	[0.001]	[0.003]	[0.004]	[0.013]	[0.016]	[0.015]	[0.018]	[0.034]	[0.034]
Firm fixed effects	No	No	No	No	No	No	No	No	Yes	Yes
Firm and worker controls	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Observations	~5000000	~5000000	~5000000	~5000000	~5000000	~5000000	~5000000	~5000000	~5000000	~5000000
R-squared	0.005	0.009	0.006	0.009	0.012	0.029	0.012	0.03	0.222	0.197

Standard errors in brackets. *** signifies statistical significance at the one-percent level, ** signifies the five-percent level, and * the one-percent level.

for workers and firms, described above. Columns 5 and 6 do not include the firm-size control, though it is returned in Columns 7 and 8. As before, the sign and significance of φ is robust to these controls. Across the four specifications, the five-year firm employment instability coefficient φ ranges in size from 0.138 to 0.152. The nine-year coefficient ranges from 0.178 to 0.189. Its magnitude is largely unaffected by the inclusion of the controls.

Consider two states of the world for a given firm. The firm experiences greater time series variance in employment in the first state, and less in the second state. The thought experiment of interest here is whether, *ceteris paribus*, the firm's workers will experience less stable earnings in the first state of the world than in the second. This line of thinking suggests including a firm fixed effect in Eq. 3, allowing us to estimate φ using within-firm variation. Call this Eq. 4:

$$instability(y_{ift}) = \alpha + \varphi\ instability(e_{ft}) + X_{ift}\delta + \theta_f + \varepsilon_{ift}, \qquad (4)$$

where everything is as in Eq. 3 except θ_f, a firm effect. This firm effect controls for persistent, time-invariant features of firms, including compensation practices, other HR practices, preferences over worker characteristics, occupational composition, and other factors which may be of importance to both earnings and employment instability.

Columns 9 and 10 of Table 1.9 report estimates of Eq. 4. The coefficient retains its statistical significance and positive sign in both regressions. The five-year regression sees the coefficient's magnitude nearly cut in half, suggesting that the estimates of Eq. 3 were driven in part by between-firm comovements in earnings and employment instability. The coefficient on the nine-year coefficient is reduced in magnitude, but not nearly as much as the five-year coefficient.

A potential concern with the interpretation that firms are passing instability onto workers earnings is reverse causality. As Comin *et al.* (2009) argue, reverse causality — that unstable earnings of workers are causing their employer to experience unstable employment — is unlikely here. Earnings are determined by the supply of and demand for labor. Changes in labor supply are usually gradual, driven by population growth and other factors. At the frequencies studied in this paper, it is unlikely that unstable labor supply causes unstable earnings, which in turn cause unstable employment levels of employing firms. Volatile labor demand may influence earnings volatility, but labor demand is not determined by earnings — rather, the opposite is true. (I attempt to control for these aggregate factors by including year effects and industry effects in the regressions.) Finally, it is unlikely that the worker's *personal* demand for the products of the firm influences the size of the firm, so earnings instability almost surely does not induce employment instability along this margin.

The weight of the evidence presented in Table 1.9 suggests that firms are passing instability in the scope of undertaken economic activity onto their workers in the form of less stable earnings. While I want to stress that these findings do not rise to the level of causal, it is the case that the finding is robust to a host of control variables, including demographic controls of workers, industry effects which control for aggregate, industry-level conditions, year effects, and firm size. In addition, the effect exists when estimated using within-firm variation, providing strong support for the hypothesis that what is being estimated is a feature of the way workers are paid by their employers.

Results By Earnings Quintile

Related to the results presented previously of the relationship between earnings instability and AFQT, it is of interest to see if the effect of firm employment instability on earnings instability is different for workers of different skill. To this end, I group each worker-year into an earnings quintile and estimate Eq. 4 separately on each quintile.

Table 1.10 reports the results. Each regression in the table is an estimate of Eq. 4 — the model with full worker and firm controls, and firm effects. The odd-numbered columns estimate firm and earnings instability using the five-year measure, and the even-numbered columns use the nine-year measure. Columns 1 and 2 report results estimated on the first earnings quintile, Columns 3 and 4 on the second earnings quintile, and so on.

The results are interesting. In each regression, the coefficient on firm employment instability is positive and statistically significant. The coefficient magnitude is considerably larger for the lowest earnings workers than for the highest. For the nine-year measure, the coefficient φ strictly decreases over earnings quintile. The magnitude of the coefficient estimated on the lowest quintile of earnings (0.230) is over double the magnitude of φ estimated for the highest earnings quintile. For the five-year measure, except for a slight increase between the second and third quintile, the coefficient magnitude decreases over earnings quintile as well. The magnitude of the first earnings quintile's coefficient is over 3.5 times as large as the magnitude of the highest earnings quintile.

TABLE 1.10: Earnings instability and firm instability by earnings quintile

The table reports estimates of Eq. 4. The dependent variable for each regression is the earnings instability measure using either a nine-year or five-year rolling window. The odd-numbered columns use the five-year window, and the even use the nine-year window. (See text for details.) Firm and worker controls include race, education category, industry, year, and state dummies, and age and age squared. Standard errors are clustered on firm. Regressions estimated separately by earnings quintile.

	Quintile 1		Quintile 2		Quintile 3		Quintile 4		Quintile 5	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Firm instability (Five-year)	0.157***		0.057***		0.059***		0.050***		0.044***	
	[0.016]		[0.008]		[0.015]		[0.018]		[0.008]	
Firm instability (Nine-year)		0.230***		0.169***		0.159***		0.134***		0.114***
		[0.018]		[0.018]		[0.018]		[0.022]		[0.012]
Log number of employees	-0.007**	0.001	0	0.008***	0.007	0.020**	0.006	0.020*	-0.004	0.011***
	[0.003]	[0.003]	[0.003]	[0.003]	[0.009]	[0.008]	[0.011]	[0.012]	[0.003]	[0.004]
Constant	0.306***	0.651***	0.161***	0.532***	0.141**	0.601***	0.135	0.699***	0.280***	0.908***
	[0.029]	[0.030]	[0.023]	[0.025]	[0.067]	[0.063]	[0.087]	[0.100]	[0.026]	[0.046]
Firm fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm and worker controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	~1000000	~1000000	~1000000	~1000000	~1000000	~1000000	~1000000	~1000000	~1000000	~1000000
R-squared	0.426	0.417	0.292	0.279	0.238	0.231	0.184	0.191	0.252	0.203

Standard errors in brackets. *** signifies statistical significance at the one-percent level, ** signifies the five-percent level, and * the one-percent level.

These results strongly suggest that low-earnings workers have more instability passed to them from their employers than do high-earnings workers, and may help to explain the finding presented previously that low-skill workers have much higher earnings instability than workers of other skill groups.

Industry Analysis

Due to across-industry variation in competition, monitoring technology, and a host of factors, we may expect that the magnitude of φ — the amount of firm employment instability passed to employees in the form of earnings instability — may vary across industries. I compute estimates of Eq. 4 — the fully controlled model with firm effects — for each North American Industry Classification System (NAICS) sector using both the five- and nine-year measures and report the estimates of φ in Table 1.11.

There is significant heterogeneity in the statistical significance and magnitude of the coefficient φ.[20] Using the five-year measure, the construction industry has the largest coefficient, equal to 0.187, while transportation and warehousing has the smallest (precisely estimated) coefficient, which is less than one-third the magnitude of construction. Utilities, manufacturing, administrative, and construction have large estimates of φ, while mining, management, accommodation and food service, and education have small or imprecise estimates of φ.

[20] There is also significant heterogeneity in the number of observations in each sector-specific regression. The sector with the largest number of observations is the manufacturing sector, with approximately 1.3 million records. The smallest sector is mining, with approximately 27,000 records. For the number of records in a regression, refer to Table 8b for the distribution over industries.

TABLE 1.11: Earnings instability and firm instability by industry

This table reports the coefficient on firm instability estimated from Eq. 4, by industry. The dependent variable in each regression is the earnings instability measure using either a nine-year or five-year rolling window. (See text for details.) Each regression controls for race, education category, industry, year, age and age squared, and includes a firm fixed effect. Standard errors are clustered on firm. Regressions estimated separately by industry.

	Firm instability	
NAICS sector	Five-year	Nine-year
Agriculture (11)	0.059***	0.111***
Mining (21)	0.021	0.050
Utilities (22)	0.152**	0.158
Construction (23)	0.187***	0.326***
Manufacturing (31-33)	0.059***	0.167***
Wholesale Trade (42)	0.075***	0.154***
Retail Trade (44-45)	0.074***	0.103***
Transportation and Warehousing (48-49)	0.056***	0.131***
Information (51)	0.085*	0.097***
Finance and Insurance (52)	0.061**	0.146***
Real Estate (53)	0.080**	0.130***
Professional Services (54)	0.138***	0.185***
Management (55)	0.015	0.041***
Administrative (56)	0.092***	0.225***
Educational Services (61)	0.025	0.043
Health Care and Social Assistance (62)	0.071***	0.133***
Arts, Entertainment, and Recreation (71)	0.104	0.132***
Accommodation and Food Services (72)	0.057***	0.096***
Other Services (81)	0.082**	0.181***
Public Administration (92)	0.03	0.054*

*** signifies statistical significance at the one-percent level, ** signifies the five-percent level, and * the one-percent level.

As discussed previously, Bertrand (2004) finds that firms which experience greater import competition employ workers whose wages are more sensitive to current labor market conditions. She interprets this as evidence that firms in increasingly competitive markets are providing less shielding to workers' earnings from external conditions. If so, then we may expect that the size of the coefficient φ would be positively correlated with the amount of competition in an industry — greater competition leads to less shielding, which leads to a stronger relationship between firm and earnings instability. On the other hand, classes of models which assume that firms have market power allow for firms to control the earnings of their workers. Under this framework, we may expect that the coefficient φ is negatively correlated with competition — less competitive industries may have larger values of φ.

To conclude the investigation of the effect of firm employment instability on the instability of worker earnings, I offer a preliminary test of the relationship between the magnitude of the *subsector-specific* pass-through coefficient φ and the competitiveness of the industry. I estimate Eq. 4 for each NAICS subsector (often referred to as three-digit NAICS) and collect the subsector-specific estimates of the φ coefficients. I measure subsector competitiveness as the ratio of receipts, revenue, sales, or value added of the top N firms in a subsector to the total receipts, revenue, sales, or value added of that subsector. N equals four, eight, twenty, or fifty. The concentration ratio data come from the 2002 Economic Census. *Sector-level* concentration ratios are reported in Table 1.12; *subsector-level* concentration ratios are used in the analysis.

A major caveat is in order. First, there is no concentration ratio data available for the following five sectors: management, mining, public administration, agriculture, and construction. In addition, there are a small number of other NAICS subsectors for which there are no data. This leaves data on concentration ratios available for 77 of

TABLE 1.12: Concentration ratios by NAICS sector, 2002 Economic Census

This table contains concentration ratios for each NAICS sector from the 2002 Economic Census. There are four concentration ratios for each industry. Each represents the percent of sector-level revenue, sales, or receipts accounted for by the N largest firms, where N = 4, 8, 20, or 50.

NAICS sector	Percent of total industry receipts accounted for by N largest firms, where N = …			
	Four	Eight	Twenty	Fifty
Management (55)	N/A	N/A	N/A	N/A
Mining (21)	N/A	N/A	N/A	N/A
Educational Services (61)	6.5	10.6	15.6	21.4
Public Administration (92)	N/A	N/A	N/A	N/A
Transportation and Warehousing (48-49)	14.8	18.3	25.2	33
Accommodation and Food Services (72)	5.1	8.9	16.5	23.1
Agriculture (11)	N/A	N/A	N/A	N/A
Manufacturing (31-33)	N/A	N/A	N/A	N/A
Finance and Insurance (52)	9.9	16.1	28.2	44.9
Health Care and Social Assistance (62)	3.9	5.4	9	14.7
Retail Trade (44-45)	11	15.3	23.9	31.7
Wholesale Trade (42)	7.5	11.6	18.7	27.2
Real Estate (53)	6.5	10.4	17.1	24.4
Other Services (81)	N/A	N/A	7.1	11.2
Information (51)	23.2	34.4	48.5	62
Administrative (56)	6	9	14.9	21.9
Arts, Entertainment, and Recreation (71)	5.4	7.7	12.4	19.6
Professional Services (54)	3.9	6.4	11.1	16.2
Utilities (22)	13.4	24.6	44.9	69
Construction (23)	N/A	N/A	N/A	N/A

These data can be found at http://www.census.gov/econ/concentration.html the 100 NAICS subsectors. The concentration ratios for these 77 subsectors were matched to their respective subsector-specific φ coefficient. Two of the 77 φ coefficients could not be used in the analysis due to Census Bureau disclosure avoidance rules — the two subsectors were too small, and so statistical output relating to those subsectors could not be released. These leaves 75 φ-concentration ratio matches.

Using the five-year measure, 32 of the 75 estimates of φ are statistically significant, demonstrating that the sector-level heterogeneity reported in Table 1.11 is also present at the subsector level. Fifty-three of the 75 estimates of φ are statistically significant using the nine-year measure.

The question at hand is whether the level of competition in a subsector, as measured by the concentration ratio, is predictive of the size of the coefficient φ for that subsector. To that end, I plot φ against the four concentration ratios. I only use those coefficients with associated t-statistics greater than two, as those are the coefficients estimated with precision, so the five-year plots have 32 observations, and the nine-year plots have 53.

Figures 1.15 and 1.16 present the scatter plots of the subsector-specific estimate of φ against the corresponding subsector's concentration ratio. Figure 1.15 plots the coefficients estimated using the five-year measure. The line of best fit is upward sloping, suggesting that the amount of firm employment instability passed to employees in the form of earnings instability is increasing in the share of the market accounted for by the top firms. I regress φ against the concentration ratio to learn more about the relationship. The coefficient on the concentration ratio is statistically significant when the concentration ratio is measured using $N = 8$ and $N = 20$. Figure 1.16 plots the coefficients estimated using the nine-year measure against the

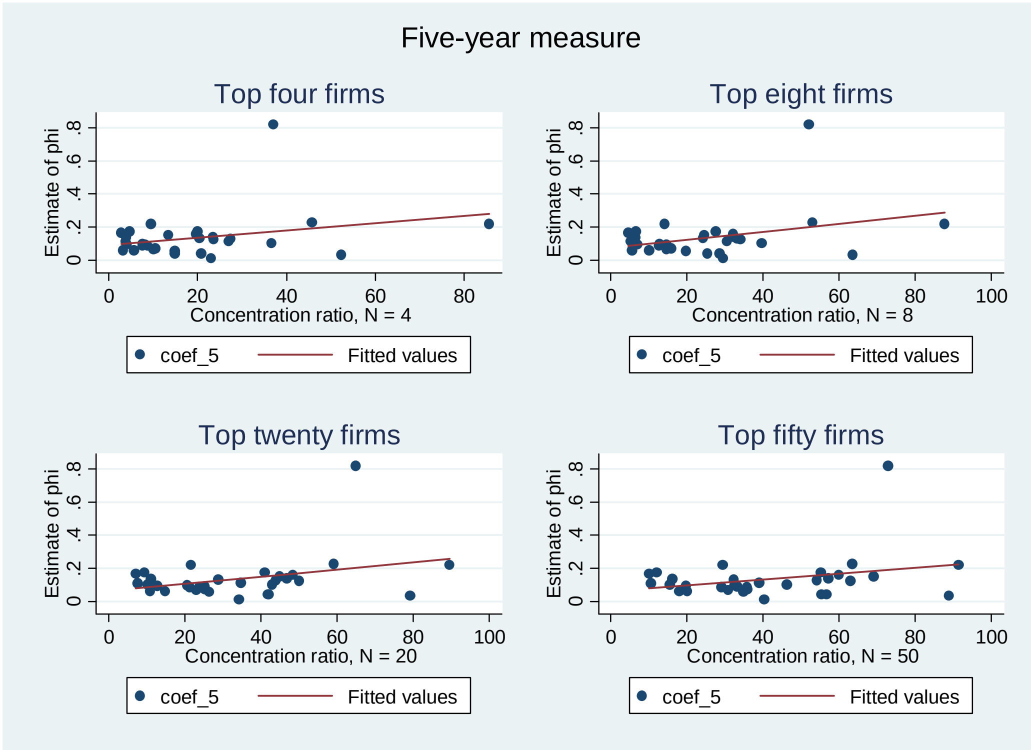

FIGURE 1.15: Estimated phi against subsector concentration ratio – 5-year measure.

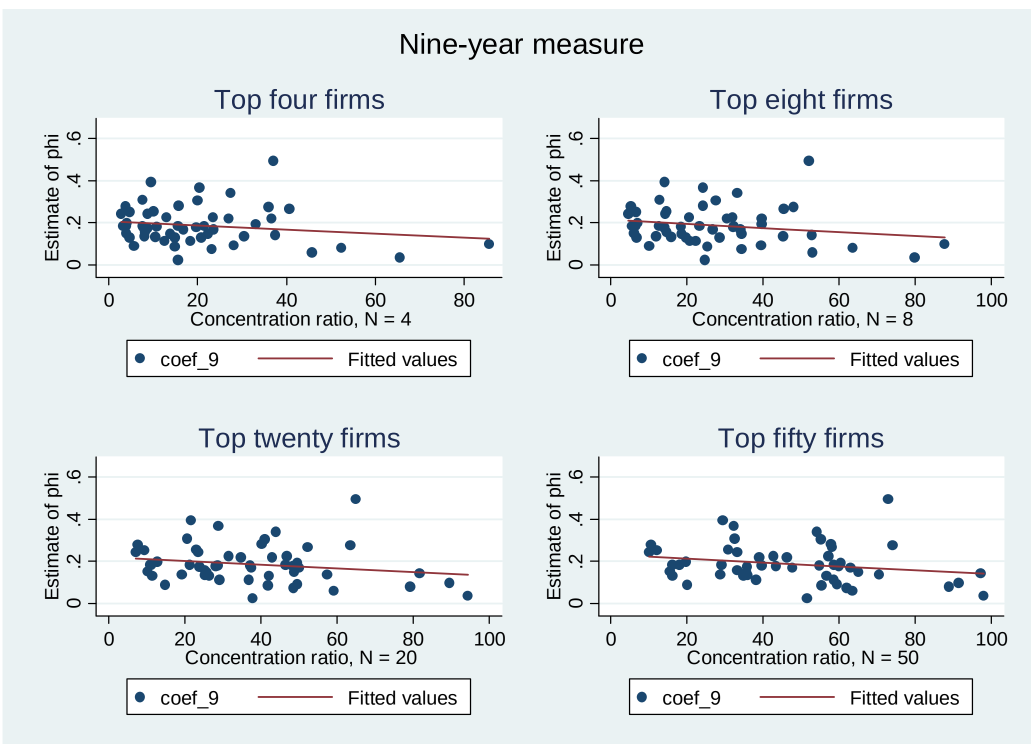

FIGURE 1.16: Estimated phi against subsector concentration ratio – 9-year measure.

concentration ratios. Unlike the plots with the five-year measures, the line of best fit here is negative, not positive. However, in a regression of φ using the nine-year measure against the concentration ratios, the coefficient on the concentration ratio is statistically insignificant in all four regressions.

To conclude, there is significant heterogeneity in the size and significance of φ across both NAICS subsectors and NAICS sectors. I present preliminary evidence that the level of competition in an industry as measured by concentration ratios is predictive of the size of φ. In ongoing research I am attempting to come to a better understanding of the size of φ — of the amount of firm employment instability passed to employees in the form of earnings instability.

V. Conclusion

Earnings instability affects earnings inequality, may lower household welfare, and impacts other important economic phenomena, yet little evidence has been documented on its causes and correlates. This paper adds to our understanding of earnings instability by studying whom earnings instability affects and why. Specifically, its relationship with both worker skill and firm employment instability is investigated.

This paper argues that differences in skill across workers contribute not only to differences in *mean* earnings, but also differences in the *time-series variance* of earnings. A detailed investigation of the relationship between worker skill (proxied by a worker's AFQT score) and earnings instability is presented. Using three different empirical methods a U-shaped relationship is documented, with low-skill workers experiencing the least stable earnings, middle-skill workers experiencing the most stable earnings, and high-skill workers falling in between the two.

This U-shaped relationship is more robust for workers who do not change jobs than for workers who may. Although there is evidence that the U-shape is more pronounced for the sample of single-job holders than for the sample of workers who are allowed to hold multiple jobs, single-job holders have more stable earnings than the less restrictive sample. The left-tail of the U is present during both the decade of the 1990s and the decade of the 2000s. However, the right-tail only exists in the 2000s. This implies the possibility that instability of earnings for high-skill workers is driving the increase in the macro trend for the 2000s. Finally, the instability of total family income does not follow a U-shape over skill. Instead, evidence is presented that it follows a W.

The evidence presented of the relationship between earnings instability and skill suggests that firms may pay workers of different skill earnings which fluctuate differently over time. This invites a study of why firms may pay unstable earnings.

Models of the labor market offer different predictions as to whether firms can affect the volatility of their workers' earnings. This paper is the first to directly test the relationship between firm employment instability and earnings instability using linked worker-firm data. The data are created by the Longitudinal Employer-Household Dynamics (LEHD) program of the U.S. Census Bureau. LEHD earnings data come from unemployment insurance earnings records, and firm data come from ES-202 reports. Firm employment instability is defined as the instability of employment, as employment is a natural measure of the scope of economic activity undertaken by the firm.

Three main findings are presented in this paper. First, the effect of firm employment instability on the instability of its workers' earnings is positive, statistically significant, robust to a number of firm and worker controls, and remains when estimated using within-firm variation. This suggests that the effect is a feature of the way workers are being paid by firms. Second, the effect is much stronger for low-earning workers than it is for high-earning workers. This helps to explain the left tail of the U-shape of earnings instability over skill. Finally, I find significant heterogeneity in the magnitude and statistical significance of the effect across NAICS sectors. I present preliminary evidence of the relationship between the competitiveness of the industry and the size of the effect, which suggests that the presence of large firms in an industry is positively associated with the amount of firm employment instability passed to employees in the form of earnings instability.

ACKNOWLEDGMENTS

I thank John Abowd, Philip Armour, Ronald Ehrenberg, Matthew Freedman, Kevin Hallock, George Jakubson, Erika McEntarfer, Kevin McKinney, Francesca Molinari, Eamon Molloy, Alex Rees-Jones, Kevin Roth, Ian Schmutte, Lars Vilhuber, Douglas Webber, Chen Zhao, and seminar participants at the American Enterprise Institute, the Board of Governors of the Federal Reserve System, Cornell University, and Marquette University for their extremely helpful comments and suggestions.

REFERENCES

Abowd, John M., Bryce E. Stephens, Lars Vilhuber, Fredrik Andersson, Kevin L. McKinney, Marc Roemer, and Simon Woodcock (2009). "The LEHD infrastructure files and the creation of the Quarterly Workforce Indicators," in T. Dunne, J. B. Jensen and M. J. Roberts (eds), *Producer Dynamcs: New Evidence from Micro Data*, vol. 68 of Studies in Income and Wealth, The University of Chicago Press, chapter 5, pp. 149-234.

Abowd, John M. and David Card (1987). "Intertemporal Labor Supply and Long Term Employment Contracts," *American Economic Review*, vol. 77, no. 1, March 1987, pp. 50—68.

Abowd, John M. and David Card (1989). "On the Covariance Structure of Earnings and Hours Changes," *Econometrica*, vol. 57, no. 2, March 1989, pp. 411—445.

Abowd, John M., Francis Kramarz, and David N. Margolis (1999). "High Wage Workers and High Wage Firms," *Econometrica*, vol. 67, no. 2, March 1999, pp. 251—333.

Autor, David H. and David Dorn (2011). "The Growth of Low-Skill Service Jobs and the Polarization of the U.S. Labor Market," *M.I.T. Working Paper*, June 2011.

Autor, David H., Lawrence F. Katz, and Melissa S. Kearney (2006). "The Polarization of the U.S. Labor Market," *American Economic Review*, vol. 96, no. 2, May 2006, pp. 189—194.

Autor, David H., Frank Levy, and Richard J. Murname (2003). "The Skill Content of Recent Technological Change: An Empirical Exploration," *Quarterly Journal of Economics*, vol. 118, no. 4, November 2003, pp. 1279—1333.

Beaudry, Paul and John DiNardo (1991). "The Effect of Implicit Contracts on the Movement of Wages Over the Business Cycle: Evidence from Micro Data," *Journal of Political Economy*, vol. 99, no. 4, August 1991, pp. 665—688.

Bertrand, Marianne (2004). "From the Invisible Handshake to the Invisible Hand? How Import Competition Changes the Employment Relationship," *Journal of Labor Economics*, vol. 22, no. 4, October 2004, pp. 723—765.

Blundell, Richard, Luigi Pistaferri, and Ian Preston (2008). "Consumption Inequality and Partial Insurance," *American Economic Review*, vol. 98, no. 5, pp. 1887—1921.

Brummund, Peter (2011). "Variation in Monpsonistic Behavior Across Establishments: Evidence from the Indonesian Labor Market," mimeo.

Cameron, Stephen and Joseph Tracy (1998). “Earnings Variability in the United States: An Examination Using Matched-CPS Data,” mimeo.

Card, David (1986). “Efficient Contracts with Costly Adjustment: Short-Run Employment Determination for Airline Mechanics,” *American Economic Review*, vol. 76, no. 5, December 1986, pp. 1045—1071.

Celik, Sule, Chinhui Juhn, Kristin McCue, and Jesse Thompson (2009). “Understanding Earnings Instability: How Important Are Employment Fluctuations and Job Changes?”, *U.S. Census Bureau Center for Economic Studies Working Paper*, no. CES-WP-09-20.

Comin, Diego, Erica L. Groshen, and Bess Rabin (2009). “Turbulent firms, turbulent wages?” *Journal of Monetary Economics*, vol. 56, pp. 109—133.

Davis, Steven J., John Haltiwanger, Ron Jarmin, and Javier Miranda (2006). “Volatility and Dispersion in Business Growth Rates: Publicly Traded versus Privately Held Firms,” *NBER Macroeconomics Annual*.

DeBacker, Jason Matthew, Bradley T. Heim, Vasia Panousi, and Ivan Vidangos, (2010). “Rising Inequality: Transitory or Permanent? New Evidence from a Panel of U.S. Tax Returns 1987-2006,” *Indiana University-Bloomington: School of Public & Environmental Affairs Research Paper Series*, no. 2011-01-01, December 2010.

Dickens, William T. and Kevin Lang (1985). “A Test of Dual Labor Market Theory,” *American Economic Review*, vol. 75, no. 4, September 1985, pp. 792—805.

Dynarski, Susan and Jonathan Gruber (1997). “Can Families Smooth Variable Earnings?” *Brookings Papers on Economic Activity,* no. 1, pp. 229—303.

Friedman, Milton (1957). *A Theory of the Consumption Function*. Princeton University Press.

Goos, Maarten, and Alan Manning (2007). “Lousy and Lovely Jobs: The Rising Polarization of Work in Britain,” *Review of Economics and Statistics*, vol. 89, no. 1, February 2007, pp. 118—133.

Gorbachev, Olga (2011). “Did Household Consumption Become More Volatile?” *American Economic Review*, vol. 101, no. 5, August 2011, pp. 2248—2270.

Gottschalk, Peter, Erika McEntarfer, and Robert Moffitt (2008). "Trends in the Transitory Variance of Male Earnings in the U.S., 1991-2003: Preliminary Evidence from LEHD data," *Boston College Working Papers in Economics*, no. 696.

Gottschalk, Peter and Robert Moffitt (1994). "The Growth of Earnings Instability in the U.S. Labor Market," *Brookings Papers on Economic Activity*, no. 2, pp. 217—272.

Gottschalk, Peter and Robert A. Moffitt (2002). "Trends in the Transitory Variance of Earnings in the United States," *The Economic Journal*, vol. 112, no. 478, Conference Papers, March 2002, pp. C68—C73.

Gottschalk, Peter and Robert Moffitt (2009). "The Rising Instability of U.S. Earnings," *Journal of Economic Perspectives*, vol. 23, no. 4, Fall 2009, pp. 3-24.

Goux, Dominique and Eric Maurin (1999). "Persistence of Interindustry Wage Differentials: A Reexamination Using Matched Worker-Firm Panel Data," *Journal of Labor Economics*, vol. 17, no. 3, July 1999, pp. 492—533.

Haider, Steven J. (2011). "Earnings Instability and Earnings Inequality of Males in the United States: 1967-1991," *Journal of Labor Economics*, vol. 19, no. 4, pp. 799—836.

Ham, John C. and Kevin T. Reilly (2002). "Testing Intertemporal Substitution, Implicit Contracts, and Hours Restriction Models of the Labor Market Using Micro Data," *American Economic Review*, vol. 92, no. 4, September 2002, pp. 905—927.

Hansen, Bruce E. (2009). "Lecture Notes on Nonparametrics," University of Wisconsin, Madison, accessed Jan. 2011. http://www.ssc.wisc.edu/~bhansen/718/NonParametrics2.pdf.

Heckman, James J. and Paul A. LaFontaine (2006). "Bias-Corrected Estimates of GED Returns," *Journal of Labor Economics*, vol. 24, no. 3, pp. 661—700.

Huff Stevens, Ann (2001). "Changes in Earnings Instability and Job Loss," *Industrial and Labor Relations Review*, vol. 55, no. 1, October 2001, pp. 60—78.

Kopczuk, Walter, Emmanuel Saez, and Jae Song (2010). "Earnings Inequality and Mobility in the United States: Evidence from Social Security Data since 1937, *Quarterly Journal of Economics*, vol. 125, no. 1, February 2010, pp. 91—127.

Krishnan, Pramila (1990). "The Economics of Moonlighting: A Double Self-Selection Model," *Review of Economics and Statistics*, vol. 72, no. 2, May 1990, pp.361—367.

Lazear, Edward P. (1986). "Salaries and Piece Rates," *Journal of Business*, vol. 59, no. 3, July 1986, pp.405—431.

Lazear, Edward P. (1995). *Personnel Economics*. The Wicksell Lectures series. The MIT Press: Cambridge, Massachusetts.

Lazear, Edward P. (2000). "Performance Pay and Productivity," *American Economic Review*, vol. 90, no. 5, December 2000, pp. 1346—1361.

Li, Oi and Jeffrey Scott Racine (2007). *Nonparametric Econometrics: Theory and Practice*. Princeton University Press: Princeton, New Jersey.

Moffitt, Robert and Peter Gottschalk (2008). "Trends in the Transitory Variance of Male Earnings in the U.S., 1970-2004," *Boston College Working Papers in Economics*, no. 697.

Paxson, Christina H. and Nachum Sicherman (1996). "The Dynamics of Dual Job Holding and Job Mobility," *Journal of Labor Economics*, vol. 14, no. 3, July 1996, pp.357—393.

Rosen, Sherwin (1985). "Implicit Contracts: A Survey," *Journal of Economic Literature*, vol. 23, no. 3, September 1985, pp. 1144—1175.

Schmutte, Ian M. (2010). "Job Referral Networks and the Determination of Earnings in Local Labor Markets," mimeo.

Webber, Douglas (2011). "The Impact of Firm Market Power on the Earnings Distribution," mimeo.

CHAPTER 2

SINGLE-SEX CLASSES & STUDENT OUTCOMES: EVIDENCE FROM NORTH CAROLINA

Michael R. Strain[21]

ABSTRACT

The effects of single-sex education are hotly contested, both in academic and policy circles. Despite this heated debate, there exists little credible empirical evidence of the effect of a U.S. public school's decision to offer single-sex classrooms on the educational outcomes of students. This study seeks to fill this hole. Using administrative records for third through eighth graders in North Carolina public schools, the paper finds evidence that the offering of single-sex mathematics courses is associated with lower performance on end-of-grade math exams, and finds no evidence that the offering of single-sex reading scores increases performance on reading exams. Evidence of significant heterogeneity in the effect across schools is also presented.

I. Introduction

The No Child Left Behind Act of 2001 (NCLB) instituted many reforms in public education. Among them was a provision which allowed public school districts to use funds to offer single-sex schools and single-sex classes. In October 2006, the U.S. Department of Education followed up on NCLB by amending Title IX, thereby

[21] Department of Economics, Cornell University, Ithaca, NY 14853.

granting school districts even greater flexibility to offer single-sex schools and classes (U.S. Department of Education, 2008).

These reforms seem to be having an effect — school districts are responding by offering single-sex programs. The National Association for Single Sex Public Education (NASSPE), a nonprofit whose purpose is to advance single-sex programs, reports that as of January 2009 there were at least 518 public schools in America which offered single-sex programs. At least 95 of the 518 schools were single-sex schools, as opposed to schools with single-sex classes (NASSPE, 2009). During the 1990s, only a handful of public schools with single-sex programs existed (U.S. Department of Education, 2008).

Despite the proliferation of single-sex programs and a heated debate about the efficacy and discriminatory effects of single-sex education, there is little credible empirical evidence on whether enrollment in single-sex programs enhances the educational outcomes of students in public schools.

There are many reasons to believe these programs would enhance educational outcomes. Single-sex programs might enhance educational outcomes by decreasing distractions in the classroom; by allowing teachers to cater teaching methods and style to personality differences between genders; by increasing teachers' ability to maintain order in and control of the classroom; by facilitating better peer interactions among students; by giving students greater freedom to pursue activities and goals which are stereotypically assigned to members of the opposite sex; by removing the need for teachers to take into account the different maturity levels of elementary-school-aged boys and girls; by allowing students to have teachers of their own gender who could serve as a more effective role model to the students; less sex-bias in student/teacher interactions; and by facilitating a greater sense of community in the classroom.

At the same time, there are many reasons to believe that single-sex programs would be detrimental to educational outcomes. For example, instead of giving students greater freedom to pursue activities and goals stereotypically assigned to the opposite gender, segregating schools/classes by sex might easily have the opposite effect of enforcing those stereotypes. Instead of enhancing teachers' ability to maintain order and control of their classes, single-sex classes might decrease that ability by concentrating unruly students in the same classroom.

Perhaps most importantly, if single-sex programs increase educational outcomes for one gender but decrease them for the opposite gender, then it may be discriminatory to allow single-sex programs to continue.[22]

It is also important to note that single-sex programs are not generic. Some schools are reportedly implementing innovative teaching techniques to compliment their single-sex classes. An all-boys mathematics class, for example, might find the students standing in a circle throwing a football to each other while the teacher quizzes the students on their multiplication tables. At the same time, other schools may simply be separating students by gender while continuing traditional instruction. It may be that the effects of these two single-sex programs are different.

For a starting point, we may turn to the literature on Catholic schools (some of which are single-sex) and the literature on peer effects in education. Studies have found a positive effect of Catholic school attendance on educational outcomes (see, for example, Neal, 1997 and Evans and Schwab, 1995). However, we may not want to extrapolate from these results to single-sex programs since it is unknown (a) whether the schools in these samples were single-sex, and (b) whether the Catholic effect or the single-sex effect is driving the results.

[22] This paper will use the word discriminatory to describe a policy which benefits students of one gender at the expense of the other. The word is not being used in its legal sense, nor is it used to ascribe motivation.

There is evidence that peer effects exist in classes. An increase in the proportion of girls in a classroom has been shown to significantly increase educational outcomes — classroom disruptions and violence are decreased, inter-student and student-teacher relationships improve, teacher fatigue lessens, and student satisfaction increases (Lavy and Schlosser, 2007). Hoxby (2000) finds that classes with a higher proportion of girls perform better in writing and math, and attributes this to classroom conduct. Peer effects also seem to be present at the college-level: Students at women's colleges are more likely to study traditionally male subjects (Solnick, 1995) and women were more likely to study traditionally male subjects before their all-female college switched to coeducation (Billger, 2002). Peer effects in single-sex schools have been shown to drive student performance in mathematics: Girls in Thailand see their math scores increase when they enroll in single-sex schools, while the single-sex environment decreases boys' math scores (Jimenez and Lockheed, 1989). However, there may be a large difference between a classroom with a high proportion of female students and a classroom with *only* female students. As such, the peer effects literature can only take us so far in understanding the effects of single-sex education.

Billger (2009) studies private, single-sex schools and finds that girls are less likely to go to college and that graduates have the least sex-segregated college major choices. In contrast, Jackson (2010) finds that attending a single-sex school has no effect on exam performance, course selection, and secondary school leaving among students in Trinidad and Tobago.

This paper contributes to the small number of existing studies on single-sex education by estimating its effect on students enrolled in North Carolina in grades three through eight. The effect is identified using a difference-in-differences estimation strategy which exploits the previously-discussed policy changes in value-added regression models.

Data on which schools offer single-sex classes in which years and for which grades are taken from the Student Activity Reports (SAR) database from the North Carolina Education Research Data Center (NCERDC), which records the distribution over gender for all activities (included mathematics and reading/English classes) in North Carolina public schools. The dependent variables in the regressions are the number of days a student is suspended from school and the standardized end-of-grade reading and mathematics scores of North Carolina public school students, also provided by the NCERDC.

Five main finding are presented in this paper. (1) There is evidence that offering single-sex mathematics classes hurts the performance of students on their end-of-grade math exams. (2) While the evidence that single-sex reading/English classes hurt student outcomes is less strong than the evidence for math, we can say that there is no evidence that offering single-sex reading/English and mathematics classes in a school-year-grade is associated with *higher* average end-of-grade reading scores. (3) There is significant heterogeneity in the effect of single-sex classes across schools. (4) There is weak evidence of a discriminatory effect due to these single-sex programs. (5) There is no evidence that these classes have an effect on the number of days a student is suspended from school.

The rest of this paper is organized as follows. Section 2 discusses the data, and presents a detailed discussion of the method of coding the treatment variable. Section 3 discusses the empirical strategy and the regression models estimated in this study. Section 4 presents sample statistics, the baseline results of the estimation, investigates the sensitivity of the results to different control variables and to different methods of coding the treatment variable, and presents a falsification test. Section 5 offers a concluding discussion with suggestions for future research.

II. Data and Treatment-Variable Construction

The data for this paper come from the North Carolina Education Research Data Center (NCERDC), a data source which has been increasingly used in education research over the past several years.[23] From the NCERDC website: "This ongoing project was established in 2000 through a partnership with the N.C. Department of Public Instruction to store and manage data on the state's public schools, students and teachers." This longitudinal dataset allows researchers to follow a student's end-of-grade test scores over time, knowing her grade level and the school in which she was enrolled.

This paper studies a sample of students in the third through eighth grades in North Carolina, drawn from a population of roughly 2.2 million between 1997 and 2009. (There were approximately 7.7 million student-years over this time period.) The outcomes studied are end-of-grade test scores in mathematics and reading, and the number of days which a student was suspended from school. I create a standardized test score for each student-year by subtracting the year-specific, grade-specific mean from each observation and dividing that difference by the standard deviation.[24] This ensures comparability across years in the test score variable.

The NCERDC maintains a School Activity Report (SAR) database which consists of student demographic counts for each activity occurring in the school on a

---

[23] For example, Clotfelter, Ladd, and Vigdor (2006), Clotfelter, Glennie, Ladd, and Vigdor (2006), and Jackson (2009). The NCERDC maintains a list of projects which use their data which can be accessed on their website: http://www.childandfamilypolicy.duke.edu/research/publications.php

[24] The mean of this variable equals zero and the standard deviation equals one. This is a common strategy when using test-score data. See, for example, Jackson and Bruegmann (2009), Ost (2010), and Rothstein (2010).

given school day. A record in this dataset is at the year-school-activity-section-grade level; for example, a record could by a third-period mathematics course for seventh graders in school *s* in year *y*. The variables for each record include student demographic counts, so for the mathematics course above the data will tell you how many male students are in the classroom and how many female students are in the classroom. (Note that the data do not tell you *which* students are in the classroom, only that there were, say, seven boys and six girls.)

From these data I am able to code the single-sex classes treatment variable at the year-school-grade level — I am able to identify which schools offer single-sex classes in a given year, and for which grades the classes are offered. A year-school-grade cell is at risk of being coded as offering single-sex classes if it contains classrooms with more than fifteen students and if the number of single-sex math and reading/English classes in that school-year pair exceeds a threshold level. A detailed description of the data exercise is now presented.

The purpose of this data exercise is to determine which North Carolina schools offered same-sex math and reading/English classes in which years and for which grade levels, using the SAR data. The final output from this data exercise will be a list of school-year-grades for which same-sex classes were offered.

The SAR, also known as the Student Count file, is actually a series of datasets, one for each year, where a record is at the district-school-activity-section-grade level. The information recorded for each record is demographic: the number of, say, white females participating in that activity, or Asian males. Activities in this dataset include academic courses, so the data can be used to identify which district-school-course-section is composed entirely of male/female students.

I begin by collecting all the year-specific files and creating a master dataset spanning all years. I then select the sample: I keep only activities for students in third through eighth grade. In addition, I use only records which identify the activity as being in the first semester, dropping the other records. For activities assigned a grade level of third through eighth, 62.09 percent are coded with semester equal to one, 32.28 percent with semester equal to two, and roughly three percent each with semester equal to three and four. Assuming that these math and reading/English classes are year-long courses (whereas non-classroom activities and classes like band are often one semester in length) which start in the first semester, the distribution over semesters seems reasonable.

I checked the sensitivity of the results to this sample selection. Redoing the analysis with the classes coded based of records with semester equal to two results in an identical list of schools for math. For reading, coding the records using semester equal to two results in the same list as when using semester equal to one except for the addition of one school. For the analysis I will use the coding with semester equal to one, as a class which starts in the second semester is likely to be a non-traditional reading/English course.

A classroom activity in these data can span more than one observation in the dataset, as the data are at the year-district-school-activity-section-grade level. This would only happen if a classroom activity had, say, both fourth and fifth grade students enrolled. Care must be taken to arrive at accurate counts of these records.

I generate a variable which records class size by taking the sum of the total student count variable for each year-district-school-activity-section cell. I then generate variables which record the total number of males and the total number of females for each cell. This variable is calculated by summing over the race-sex count variables. The sum of the male and female totals should equal the value of the class size variable for each observation in the dataset. The sum of the male and female variables equals the class size variable for all but nine records out of approximately 3.5 million.

At this point, I identify math and reading/English classes using subject codes provided in the data.[25] Same-sex classes are identified as those classes with a ratio of the total number of male students to the total size of the class equal to either one or zero. Records for which this ratio does not equal one or zero are dropped.

Many of the same-sex classrooms identified up to this point are not standard classrooms, and are thus not the type of classrooms which this paper is intended to study. Fully eighty-five percent of the same-sex math classrooms, for example, have less than six students enrolled. Over one-third of the same-sex math classrooms contain only one student. To deal with this issue, only classrooms with more than fifteen students are kept.[26]

Table 2.1 shows the distribution over years of same-sex classes, conditional on the sample selection discussed above. Note that there are same-sex math and reading/English classes for each year from 1996 through 2009. For both reading/English and math there is a large spike in 2003 — this is consistent with the 2001 No Child Left Behind law. There is also a large spike in 2008, consistent with the 2006 changes to Title IX. This table is taken as evidence that the policy change had an effect on the offering of same-sex classes in North Carolina.

---

[25] Math classes are identified as those activities for which the first two characters of the subject code (variable name: subject) equal 20. Reading/English classes are identified as those activities with subject code equal to 1001 or 1010.

[26] Imposing class-size cutoffs is necessary when using the NCERDC data due to the problem of small classes — e.g., Rothstein (2010), Rothstein (2009), and Ost (2010). These classes could be special education classes, tutoring classes, measurement error, or actual classes which had to be small due to extenuating circumstances.

TABLE 2.1: Number of single-sex classes with more than 15 students, by year

Math				Reading/English			
Year	**Freq.**	**Percent**	**Cum.**	**Year**	**Freq.**	**Percent**	**Cum.**
1996	1	0.25	0.25	1996	1	0.18	0.18
1997	4	1.01	1.27	1997	5	0.89	1.07
1998	1	0.25	1.52	1998	5	0.89	1.96
1999	2	0.51	2.03	1999	2	0.36	2.31
2000	5	1.27	3.29	2000	9	1.6	3.91
2001	2	0.51	3.8	2001	1	0.18	4.09
2002	6	1.52	5.32	2002	5	0.89	4.98
2003	38	9.62	14.94	2003	88	15.66	20.64
2004	36	9.11	24.05	2004	31	5.52	26.16
2005	26	6.58	30.63	2005	18	3.2	29.36
2006	25	6.33	36.96	2006	27	4.8	34.16
2007	38	9.62	46.58	2007	42	7.47	41.64
2008	106	26.84	73.42	2008	165	29.36	71
2009	105	26.58	100	2009	163	29	100

However, there are same-sex classes in these data prior to the 2001 policy change. It is possible that some of these classes occurred due to special circumstances at their respective school. It is also possible that they are a function of measurement error. Regardless of the reason, this must be dealt with. To do so, I construct a variable which records the number of same-sex math and reading/English classes for each school in each year. I require that a school-year offer more than φ same-sex classes in order to be included in the final list of school-year-grades which are coded as offering same-sex classes. (φ is a positive integer.) I calibrate φ such that there is no $\varphi' < \varphi$ where same-sex classes did not exist prior to 2001. This method ensures that a consistent standard is used across all years. For math φ equals four and for reading φ equals eight.

Table 2.2 shows the distribution over years for same-sex classes for math and reading after the sample selection criterion described in the preceding paragraph is applied. As intended, there are no same-sex classes prior to 2001. Table 2.3 shows the two-way distribution over year-grade pairs. The treatment variable will be defined at the year-school-grade level, and this table shows that there is variance both within grades across years and within years across grades.

TABLE 2.2: Number of single-sex classes by year -- all sample restrictions

Math				Reading/English			
Year	**Freq.**	**Percent**	**Cum.**	**Year**	**Freq.**	**Percent**	**Cum.**
2003	17	21.25	21.25	2003	45	35.71	35.71
2005	17	21.25	42.5	2006	16	12.7	48.41
2006	8	10	52.5	2008	55	43.65	92.06
2008	26	32.5	85	2009	10	7.94	100
2009	12	15	100				

This year-district-school-course-grade level file is collapsed to the school-year-grade level. The remaining school-year-grades are those which are coded as offering same-sex classes. There are twenty-eight such school-year-grades for math and twenty-two for reading/English. For the analysis of end-of-grade math tests, student-years which are matched to these twenty-eight school-year-grades are the math treatment group, and student-years which are not matched to these twenty-eight

TABLE 2.3: Number of single-sex classes by year-grade -- all sample restrictions

Math

Year	**3**	**4**	**5**	**6**	**7**	**8**	**Grade Total**
2003	1	2	7	2	3	2	17
2005	0	0	0	10	4	3	17
2006	0	0	0	8	0	0	8
2008	4	2	2	11	2	5	26
2009	2	2	1	4	1	2	12
Total	7	6	10	35	10	12	80

Reading/English

Year	**3**	**4**	**5**	**6**	**7**	**8**	**Grade Total**
2003	8	13	22	2	0	0	45
2006	0	0	0	16	0	0	16
2008	7	4	4	24	4	12	55
2009	4	4	2	0	0	0	10
Total	19	21	28	42	4	12	126

school-year-grades are the math control group. Likewise, for the end-of-grade reading analysis, student-years which are matched to the twenty-two school-year-grades are the reading treatment group, and all other students are the reading control group.

The first single-sex math and reading/English classes found in the data and subject to the sample restrictions occur in the year 2003, consistent with the 2001 No Child Left Behind policy change. Single-sex math classes are found in the years 2003, 2005, 2006, 2008, and 2009; single-sex reading classes are found in the years 2003, 2006, 2008, and 2009. Single-sex math classes are found in nine non-charter schools, and single-sex reading/English classes are found in seven. There are 2,104 total non-charter schools in the data.

That there are some years with no school-grades offering single-sex classes is consistent with reports from the NASSPE that schools are experimenting with this policy intervention — year-to-year variation in school leadership and parental cohorts could result in a school adopting single-sex classes for a year or two and then discontinuing them. The number of schools which offer single-sex classes should inspire confidence in the sample selection, as we would not expect a large number of schools offering these classes given the relatively small number of schools nationwide reported to do so by organizations such as the NASSPE.

III. Empirical Strategy

The first component of the empirical strategy is to select the analysis sample. I conduct the analysis only on those student-years which are associated with a school district which *ever* offered single-sex classes. I do this under the assumption that the characteristics (likely unobservable) which drive a school to offer single-sex classes are more similar between two schools in the same district than between two schools in different districts. To help keep "all else constant", I limit the control group to those schools which are located in school districts which ever offered single-sex classes. I conduct the analysis separately for math and reading/English, as the SAR data allow course subjects to be identified — i.e., the end-of-grade reading regressions and the end-of-grade math regressions are run on *different* samples. Conditional on a student-year being present in a school district which ever offered single-sex math classes, 0.60 percent of student-years are matched to year-school-grades which offered single-sex math classes. The corresponding statistic for single-sex reading/English classes is

0.42 percent. There are a total of 223 schools in the end-of-grade math analysis sample, and 185 schools in the end-of-grade reading analysis sample. As more schools offer single-sex math classes than reading/English, there are 621,620 student-years in the baseline end-of-grade math analysis sample, and 523,150 in the end-of-grade reading sample.

To study the effects of single-sex classes on student outcomes a differences-in-differences model is estimated using a value-added framework. Specifically, the following equation is estimated:

$$y_{igst} = \alpha + \beta y_{igst-1} + \gamma T_{gst} + X'_{igst}\theta + \omega_{st} + \varepsilon_{igts}. \quad (1)$$

where y_{igts} is a student-level outcome defined for student *i* at school *s* in year *t* in grade *g*, either an end-of-grade exam score or the number of days the student is suspended, and y_{igst-1} is its lag; ω_{st} is a school-by-year effect; X_{igst} is a matrix of control variables with associated coefficient vector θ; ε_{igts} is a standard residual; and α, β are parameters. The left-hand side of the equation can be rewritten as $(y_{igts} - \beta y_{igst-1})$, allowing the coefficients in the model to be interpreted as measuring the impact of the variables on *gains* in test scores over the previous year. Studying gains using a value-added specification is a widely-used technique in the literature and allows the researcher to control for initial conditions, including the prior history of school inputs and home inputs, which affect a student's year *t* test score (Rivkin *et al.*, 2005).

T_{gst} is a dummy variable equal to one if school *s* in year *t* offers single-sex classes for grade *g*, and equal to zero otherwise. This is the variable which will be used to identify the effect of *being in a school which offers* single-sex classes. The parameter γ is the estimate of the effect of the treatment on the student, and is equal to the average difference in y between student outcomes in school-year-grades which offer single-sex classes and those which do not, holding constant the other covariates.

Take care to note that the treatment is defined as being enrolled *in a school-year-grade which offers* single-sex classes, and not being enrolled *in* a single-sex classroom. Though the data link students to schools, grades, and years, the data are not sufficiently detailed as to indicate which students belong to which classes.

Apart from mechanical data issues described in the previous section, defining the treatment variable in this way is the policy-relevant definition. The control group is surely contaminated by the policy change — students who select into single-sex classes by definition will not be in mixed-sex classes, so their selection into single-sex classes changes the distribution over unobservables in the traditional, co-ed classes. Thus, the policy change affects all the students in the school, and not just those students who choose to select into single-sex classes.

In addition, this definition allows us to avoid the problem of students selecting into single-sex classes. The proper, policy-relevant way to view the treatment group is that it contains *all* students in a school-year-grade, regardless of which students actually remain in a co-ed class and which select into single-sex classes. Because of this, we need to know neither which students selected into single-sex classes nor why they chose to do so.

A school decides to offer single-sex classes, parents/students decide what is best for them, and sorting occurs. The parameter γ identifies the effect of the policy change, then, on *all* the students in the school, which is proper due to the fact that the treatment affects all the students in the school.

A potential concern with this analysis is that the estimated effect of single-sex classes could be confounded by other factors. Perhaps there is something about these schools — a new and dynamic principal, say, or a particularly involved cohort of parents — which both affects test scores and affects the school's decision to offer single-sex classes. If so, then this correlation between the error term and the treatment variable will bias the coefficient of interest.

Since the treatment effect is coded at the school-year-grade level, I am able to include ω_{st}, a school-by-year effect. This effect will go a long way towards mitigating the concern that the treatment coefficient is picking up the effect not of single-sex classes but instead of a new principal coming in to the school and instituting a set of changes. This specification, conditional on the covariates, looks within a school-year cell, using variation across grades to estimate the effect of the treatment. Using within-school-by-year variation to estimate the effect of a school offering single-sex classes helps to ensure that the coefficient γ is truly capturing the effect of the treatment — it is unlikely that the presence of a dynamic principal in a given year only affects sixth-grade students and not all the students in the school. By "comparing" two students in the same school-year but in different grades, we can ease the concern that the treatment coefficient is picking up a host of unobservable school-level changes.

In addition to using within-school-by-year variation, the school-year-grade treatment variable allows for the estimation of the treatment effect using within-year-by-grade variation. Specifically, the following equation can be estimated:

$$y_{igst} = \alpha' + \beta' y_{igst-1} + \gamma' T_{gst} + X'_{igst}\theta' + \tau_{tg} + \varepsilon'_{its}. \qquad (2)$$

The terms in equation (2) are the same as equation (1) with the exception of τ_{tg}, the year-by-grade effect. Under this specification, the treatment effect is estimated, all else equal, by looking across schools within a year-grade pair. Any common component effecting test scores for a year-grade pair is held constant, and the effect is estimated using variation across schools, some of which offer single-sex classes, and some of which do not.

Measurement error is a concern here, and could either be found in the microdata or be generated by the treatment-variable coding algorithm, or both. The procedure used to select the school-year-grades which are coded as offering single-sex classes is quite conservative, so on balance the more likely nature of the measurement error is having school-year-grades which should be in the treatment group incorrectly assigned to the control group.

What effect would this misassignment-generated measurement error have on the treatment coefficient? Under the assumption of a relatively homogenous treatment effect across schools, having schools-year-grades which actually offer single-sex classes but which are coded as not offering single-sex classes should bias the coefficient on the treatment variable towards zero — contaminating the control group with treated observations will attenuate the average difference between the two group.

However, imagine that there are some schools which enthusiastically adopt single-sex classes, and some schools which are more lukewarm in their adoption. If the lukewarm schools are systematically misassigned to the control group while the enthusiastic schools are systematically and correctly assigned to the treatment group, then under the assumption that the enthusiastic schools will have a stronger treatment effect than the lukewarm schools, this misassignment could actually lead to a treatment coefficient which *overstates* the magnitude of the true effect. It is important to keep this in mind when reviewing the results, to which we now turn.[27]

[27] This concern is investigated in Section 4. The baseline results are calculated using the treatment-variable coding scheme described in this section. An additional set of results is reported in Section 4 which removes the requirement of a school-year-grade offering multiple single-sex classes in a given school-year to be eligible to be assigned to the treatment group. Effectively, the additional results require that $\varphi = 1$.

IV. Results

Table 2.4a presents descriptive statistics for the end-of-grade math sample. The table is split into three panels. The top panel presents summary statistics for the entire analysis sample. After sample selection, the mean test score is 0.034, with a standard deviation of 1.04. The distribution over student race shows that approximately half the students are white, a little over one-third are black, with the remainder spread between Asian, Hispanic, American Indian, multi-racial, and other.

Of interest is whether the performance of students compelled a school to offer single-sex classes — whether the pre-treatment control and treatment groups look similar. No schools offered single-sex classes prior to 2003, so to examine the pre-treatment control and treatment group distributions I will use pre-2003 data. Using the common pre-treatment period of pre-2003, the mean end-of-grade math score for the schools which will be treated in the future is 0.001, and for school which will be control in the future is -0.050. There is no evidence that the treated schools were performing poorly in the pre-treatment period. In fact, these statistics suggest that the treated schools were outperforming the untreated schools.[28]

Table 2.4b presents the same statistics for the end-of-grade reading sample. For the total analysis sample, the mean end-of-grade reading score is 0.007, with a standard deviation of 1.030. The distribution over race looks similar to the end-of-grade math sample. As with the end-of-grade math sample, the reading sample shows that the treated schools are outperforming the control schools, pre-treatment.

Table 2.5 presents estimates of equations (1) and (2). There are twenty-one regressions represented in Table 2.5, and each coefficient and standard error estimate is the estimate of the treatment effect for that regression. Columns (A) through (G) determine the control variables used in the regression. For example, the regressions

---

[28] A *t*-test rejects the null of equality with a *t*-statistic of 6.53.

TABLE 2.4a: Summary statistics for math specifications.

This table reports summary statistics for specifications with end-of-grade mathematics score as the outcome variable. The first panel contains summary statistics for the total sample used in the regressions. The next two panels report summary statistics for years prior to 2003.

Total Sample

	Obs	**Mean**	**Std. Dev.**	**Min**	**Max**
End-of-grade math score	621620	0.034	1.040	-3.610	3.640
Number of days suspended	621620	0.270	2.280	0.000	405.000
Distribution over race					
Asian	3.940				
Black	36.300				
Hispanic	7.020				
American Indian	0.380				
Multi-racial	1.580				
Other	0.000				
White	50.790				

Treated schools prior to 2003

	Obs	**Mean**	**Std. Dev.**	**Min**	**Max**
End-of-grade math score	18940	0.001	1.030	-3.020	3.470
Number of days suspended	18940	0.110	1.150	0.000	50.000
Distribution over race					
Asian	2.940				
Black	37.580				
Hispanic	3.720				
American Indian	0.170				
Multi-racial	0.600				
Other	0.000				
White	54.990				

Untreated schools prior to 2003

	Obs	**Mean**	**Std. Dev.**	**Min**	**Max**
End-of-grade math score	225145	-0.050	1.050	-3.610	3.640
Number of days suspended	225145	0.140	1.450	0.000	124.000
Distribution over race					
Asian	3.730				
Black	35.590				
Hispanic	3.170				
American Indian	0.350				
Multi-racial	0.510				
Other	0.000				
White	56.660				

TABLE 2.4b: Summary statistics for reading specifications

This table reports summary statistics for specifications with end-of-grade reading score as the outcome variable. The first panel contains summary statistics for the total sample used in the regressions. The next two panels report summary statistics for years prior to 2003.

Total Sample

	Obs	Mean	Std. Dev.	Min	Max
End-of-grade reading score	523150	0.007	1.030	-4.040	3.100
Number of days suspended	523150	0.260	2.090	0.000	186.000
Distribution over race					
Asian	4.210				
Black	39.160				
Hispanic	6.880				
American Indian	0.410				
Multi-racial	1.650				
Other	0.000				
White	47.690				

Treated schools prior to 2003

	Obs	Mean	Std. Dev.	Min	Max
End-of-grade reading score	14421	0.091	1.020	-3.270	2.800
Number of days suspended	14421	0.130	1.260	0.000	50.000
Distribution over race					
Asian	3.180				
Black	39.130				
Hispanic	1.370				
American Indian	0.210				
Multi-racial	0.590				
Other	0.000				
White	55.510				

Untreated schools prior to 2003

	Obs	Mean	Std. Dev.	Min	Max
End-of-grade reading score	190837	-0.030	1.050	-4.040	3.100
Number of days suspended	190837	0.160	1.540	0.000	124.000
Distribution over race					
Asian	3.970				
Black	38.160				
Hispanic	3.130				
American Indian	0.370				
Multi-racial	0.500				
Other	0.010				
White	53.870				

TABLE 2.5: Offering single-sex math classes reduces math test scores

This table reports estimates of equations (1) and (2). The dependent variable is a student's end-of-grade test score in mathematics. Treatment equals one if school s in year t offered single-sex classes for grade g, and equals zero otherwise. All regressions control for the race of the student and for the student's lagged test score. Standard errors are clustered on school.

	(A)	**(B)**	**(C')**	**(D)**	**(E)**	**(F)**	**(G)**
	Male and female students						
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**
Single-sex math classes	-0.072	-0.109	-0.146	-0.136	-0.099	-0.108	-0.134
	[0.044]*	[0.020]***	[0.059]**	[0.053]**	[0.035]***	[0.021]***	[0.027]***
Observations	621620	621620	621620	621620	621620	621620	621620
Adjusted R-squared	0.75	0.76	0.76	0.76	0.75	0.76	0.77
	Only male students						
	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Single-sex math classes	-0.077	-0.111	-0.135	-0.116	-0.101	-0.113	-0.117
	[0.044]*	[0.020]***	[0.070]*	[0.066]*	[0.034]***	[0.021]***	[0.037]***
Observations	313925	313925	313925	313925	313925	313925	313925
Adjusted R-squared	0.75	0.76	0.76	0.76	0.75	0.76	0.77

TABLE 2.5 (Continued)

	(A)	(B)	(C')	(D)	(E)	(F)	(G)
	Only female students						
	(15)	(16)	(17)	(18)	(19)	(20)	(21)
Single-sex math classes	-0.068	-0.108	-0.158	-0.155	-0.096	-0.102	-0.153
	[0.045]	[0.023]***	[0.050]***	[0.043]***	[0.038]**	[0.024]***	[0.022]***
Observations	307695	307695	307695	307695	307695	307695	307695
Adjusted R-squared	0.75	0.76	0.76	0.76	0.75	0.76	0.77
School effects	No	Yes	No	No	No	Yes	No
Grade effects	No	Yes	No	Yes	No	No	No
Year effects	No	Yes	No	No	No	No	No
School-by-year effects	No	No	Yes	Yes	No	No	Yes
Year-by-grade effects	No	No	No	No	Yes	Yes	Yes

represented in column (F) — regressions (6), (13), and (20) — include school effects and year-by-grade effects, but not grade effects and school-by-year effects. Likewise, the three regressions represented by column (B) — regressions (2), (9), and (16) — include school effects, year effects, and grade effects, but neither school-by-year nor year-by-grade effects. The table is split into three panels. The first panel presents results estimated on the pooled sample of male and female students; the second panel presents results estimated on male students only; and the third, on female students only. The student-year level outcome variable is either number of days suspended in year *t* or an end-of-grade test score in year *t*.

Turning first to the pooled sample (regressions (1) through (7)), the coefficient on the treatment variable is negative and statistically significant in each specification. The size of the coefficient ranges from -0.072 to -0.146. The preferred estimating equations are represented by columns (D) through (G). Rounded to two significant digits, the coefficient on the treatment variable is equal to at least -0.1 standard deviations. For perspective, consider that the black-white test score gap measured in these regressions ranges from approximately 0.14 to 0.16.

The bottom two panels of Table 2.5 report estimates using a single-sex sample. The coefficients on the treatment variable are statistically significant in all the controlled specifications, and are negative in all the specifications. Given a set of controls, the coefficients on the male-only regressions are about the same magnitude as the female-only regressions. These results imply that there is not a discriminatory effect on end-of-grade math scores.[29]

The results for specifications with end-of-grade reading are shown in Table 2.6. The coefficient is negative in twenty of the twenty-one regressions, and in all the

[29] As noted above, what is implied by the word discriminatory is a policy which hurts students of one gender at the expense of the other. The word is not intended to have a legal or motivational connotation.

TABLE 2.6: Offering single-sex reading/English classes reduces reading test scores

This table reports estimates of equations (1) and (2). The dependent variable is a student's end-of-grade test score in reading. Treatment equals one if school s in year t offered single-sex classes for grade g, and equals zero otherwise. All regressions control for the race of the student and for the student's lagged test score. Standard errors are clustered on school.

	(A)	(B)	(C')	(D)	(E)	(F)	(G)
	Male and female students						
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**
Single-sex reading/English classes	-0.057	-0.054	-0.051	-0.021	-0.085	-0.057	-0.007
	[0.020]***	[0.022]**	[0.014]***	[0.023]	[0.019]***	[0.024]**	[0.012]
Observations	523150	523150	523150	523150	523150	523150	523150
Adjusted R-squared	0.69	0.70	0.70	0.70	0.70	0.70	0.70
	Only male students						
	(8)	**(9)**	**(10)**	**(11)**	**(12)**	**(13)**	**(14)**
Single-sex reading/English classes	-0.043	-0.049	-0.072	-0.022	-0.066	-0.046	0.005
	[0.025]*	[0.025]**	[0.010]***	[0.013]*	[0.025]***	[0.027]*	[0.026]
Observations	263061	263061	263061	263061	263061	263061	263061
Adjusted R-squared	0.69	0.69	0.70	0.70	0.69	0.70	0.70

TABLE 2.6 (Continued)

	(A)	(B)	(C')	(D)	(E)	(F)	(G)
	Only female students						
	(15)	**(16)**	**(17)**	**(18)**	**(19)**	**(20)**	**(21)**
Single-sex	-0.073	-0.058	-0.016	-0.007	-0.103	-0.066	-0.007
reading/English classes	[0.021]***	[0.026]**	[0.023]	[0.038]	[0.018]***	[0.027]**	[0.021]
Observations	260089	260089	260089	260089	260089	260089	260089
Adjusted R-squared	0.69	0.70	0.70	0.70	0.70	0.70	0.71
	(A)	**(B)**	**(C')**	**(D)**	**(E)**	**(F)**	**(G)**
School effects	No	Yes	No	No	No	Yes	No
Grade effects	No	Yes	No	Yes	No	No	No
Year effects	No	Yes	No	No	No	No	No
School-by-year effects	No	No	Yes	Yes	No	No	Yes
Year-by-grade effects	No	No	No	No	Yes	Yes	Yes

regressions with a statistically-significant treatment coefficient. Not all the coefficients are estimated with precision, and the magnitude of many of the coefficients is smaller than in the end-of-grade mathematics regressions. The estimates suggest that the effect on end-of-grade reading scores of being enrolled in a school-year-grade which offers single-sex reading/English classes is a decrease of approximately 0.05 to 0.09 standard deviations, though the statistically-insignificant coefficient in the school-by-year specification (controlling for grade) is evidence that the treatment has no effect on test scores. Consistent with much of the education literature, the policy effect on reading scores is smaller than on math scores.

Unlike the results for end-of-grade math, there does appear to be evidence of a discriminatory effect of single-sex reading/English classes, although the results are contradictory as to which sex is being penalized. The two specifications with school-by-year effects (columns (C) and (D)) have statistically insignificant estimates of the treatment coefficient for female students, while the estimates for male students imply a decrease of 0.07 and 0.02 standard deviations. The specifications with year-by-grade effects (columns (E) and (F)), however, both imply a larger decrease for female students than for male students (-0.1 versus -0.07, and -0.07 versus -0.05). (The treatment effect is estimated imprecisely in the model with both year-by-grade and school-by-year effects, shown in column (G)).

Table 2.7 reports results for the conduct model. The treatment variable is defined according to the following rule: If a school-year-grade offered either single-sex math classes or single-sex reading/English classes, then single-sex suspensions treatment equals one; otherwise, single-sex suspensions treatment equals zero. The coefficient on the treatment variable is positive and statistically insignificant in every regression. This is perhaps unsurprising since being suspended is a fairly serious

TABLE 2.7: Offering single-sex classes has no effect on suspensions

This table reports estimates of equations (1) and (2). The dependent variable is the number of days the student was suspended in year t. Treatment equals one if school s in year t offered single-sex classes for grade g in either reading/English or mathematics, and equals zero otherwise. All regressions control for the race of the student. Standard errors are clustered on school.

	(A)	**(B)**	**(C')**	**(D)**	**(E)**	**(F)**	**(G)**
	Male and female students						
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**
Single-sex suspensions treatment	1.982	1.585	1.855	1.904	1.321	1.335	1.938
	[1.532]	[1.417]	[2.068]	[2.051]	[1.344]	[1.351]	[1.977]
Observations	629253	629253	629253	629253	629253	629253	629253
Adjusted R-squared	0.02	0.07	0.17	0.17	0.07	0.08	0.17
	Only male students						
	(8)	**(9)**	**(10)**	**(11)**	**(12)**	**(13)**	**(14)**
Single-sex suspensions treatment	2.796	2.202	2.968	3.014	1.858	1.870	3.038
	[2.133]	[1.963]	[3.095]	[3.078]	[1.865]	[1.873]	[2.975]
Observations	320519	320519	320519	320519	320519	320519	320519
Adjusted R-squared	0.02	0.08	0.19	0.19	0.09	0.10	0.19

TABLE 2.7 (Continued)

	(A)	(B)	(C')	(D)	(E)	(F)	(G)
	Only female students						
	(15)	**(16)**	**(17)**	**(18)**	**(19)**	**(20)**	**(21)**
Single-sex suspensions treatment	1.103	0.900	0.879	0.935	0.731	0.744	0.990
	[0.879]	[0.815]	[1.069]	[1.052]	[0.769]	[0.773]	[1.013]
Observations	308734	308734	308734	308734	308734	308734	308734
Adjusted R-squared	0.01	0.05	0.17	0.17	0.05	0.07	0.17
School effects	No	Yes	No	No	No	Yes	No
Grade effects	No	Yes	No	Yes	No	No	No
Year effects	No	Yes	No	No	No	No	No
School-by-year effects	No	No	Yes	Yes	No	No	Yes
Year-by-grade effects	No	No	No	No	Yes	Yes	Yes

infraction for students in grades three through eight. A policy intervention would have to have a large effect on student conduct for it to affect suspensions.

Heterogeneity Across Schools

To the extent that the effect of single-sex education is a function of sex-specific teaching methods, we might expect there to be variance in the treatment effect across schools which offer single-sex classes. Tables 2.8 and 2.9 report school-specific estimates of the treatment effect for end-of-grade math and reading, respectively. Equation (2) is estimated. Equation (1) is omitted because colinearity in the effects renders the equation inestimable.

Column (F) of Table 2.8 shows that the treatment effect of a school-year-grade offering single-sex math classes varies significantly across schools. For two of the schools, the treatment effect is estimated without precision. Among the seven schools measured with precision, the size of the effect ranges from -0.033 standard deviations to -0.281 standard deviations. Column (F) of Table 2.9 looks at end-of-grade reading scores, and finds that two of the seven school-specific treatment coefficients are positive. One of the two positive coefficients is statistically significant at the five-percent level. The statistically significant coefficients range from 0.018 on the high end to -0.166 on the low end.

If we interpret a statistically-insignificant treatment effect coefficient as evidence that the policy is neither increasing nor decreasing test scores, then the results from the end-of-grade math regressions suggest that in some schools single-sex classes are seriously lowering test scores (a coefficient of -0.281 is *very* large, probably implausibly so), in some schools single-sex classes are having a marginally negative effect, and in some the effect is nonexistent. The end-of-grade reading results take this heterogeneity a qualitative step further: in one school the policy seems

TABLE 2.8: Heterogeneity across schools in math treatment effect

This table reports estimates of equations (1) and (2). The dependent variable is a student's end-of-grade test score in math. Treatment equals one if school s in year t offered single-sex classes for grade g, and equals zero otherwise. All regressions control for the race of the student and for the student's lagged test score. Standard errors are clustered on school.

	(A)	(B)	(E)	(F)
School-specific treatment:				
School 278	-0.013	-0.092	-0.069	-0.093
	[0.008]	[0.009]***	[0.013]***	[0.012]***
School 287	0.052	0.008	0.011	0.009
	[0.007]***	[0.009]	[0.011]	[0.009]
School 989	-0.11	-0.125	-0.118	-0.123
	[0.006]***	[0.008]***	[0.015]***	[0.010]***
School 1191	-0.267	-0.258	-0.318	-0.281
	[0.007]***	[0.007]***	[0.010]***	[0.011]***
School 1194	0.075	-0.126	0.003	-0.12
	[0.008]***	[0.008]***	[0.017]	[0.010]***
School 1274	-0.15	-0.192	-0.25	-0.223
	[0.006]***	[0.009]***	[0.020]***	[0.012]***
School 1288	-0.12	-0.002	-0.004	-0.005
	[0.006]***	[0.011]	[0.018]	[0.015]
School 1307	-0.112	-0.036	-0.119	-0.033
	[0.007]***	[0.008]***	[0.015]***	[0.011]***
School 1472	-0.181	-0.143	-0.177	-0.119
	[0.006]***	[0.010]***	[0.021]***	[0.014]***
Observations	621620	621620	621620	621620
Adjusted R-squared	0.75	0.76	0.75	0.76
School effects	No	Yes	No	Yes
Grade effects	No	Yes	No	No
Year effects	No	Yes	No	No
School-by-year effects	No	No	No	No
Year-by-grade effects	No	No	Yes	Yes

TABLE 2.9: Heterogeneity across schools in reading/English treatment effect

This table reports estimates of equations (1) and (2). The dependent variable is a student's end-of-grade test score in reading. Treatment equals one if school s in year t offered single-sex classes for grade g, and equals zero otherwise. All regressions control for the race of the student and for the student's lagged test score. Standard errors are clustered on school.

	(A)	(B)	(E)	(F)
School-specific treatment:				
School 271	-0.077	0.013	-0.118	0.005
	[0.007]***	[0.007]*	[0.009]***	[0.008]
School 278	0.045	0.009	0.003	-0.001
	[0.008]***	[0.007]	[0.008]	[0.007]
School 287	0.01	0.026	-0.035	0.018
	[0.007]	[0.007]***	[0.009]***	[0.007]**
School 1191	-0.095	-0.081	-0.125	-0.083
	[0.005]***	[0.007]***	[0.008]***	[0.007]***
School 1274	-0.089	-0.093	-0.121	-0.095
	[0.005]***	[0.007]***	[0.012]***	[0.011]***
School 1288	-0.081	-0.144	-0.119	-0.166
	[0.005]***	[0.007]***	[0.013]***	[0.010]***
School 1472	-0.093	-0.052	-0.065	-0.023
	[0.005]***	[0.007]***	[0.015]***	[0.010]**
Observations	523150	523150	523150	523150
Adjusted R-squared	0.69	0.70	0.70	0.70
School effects	No	Yes	No	Yes
Grade effects	No	Yes	No	No
Year effects	No	Yes	No	No
School-by-year effects	No	No	No	No
Year-by-grade effects	No	No	Yes	Yes

to be *increasing* scores, while in others it is decreasing scores. While outside the scope of this paper, an investigation into why single-sex classes seem to work for some schools and not for others would clearly be interesting.

Controlling For Parental Education

The baseline results presented above do not control for parental education due to the fact that this variable was only collected through 2006. It is reasonable to think that this is an important omitted variable — parental education is likely to affect both end-of-grade test scores and the decision of a school to adopt single-sex classes.

Tables 2.10 and 2.11 investigate the sensitivity of the results to the omission of parental education. The analysis sample is identical in each regression in Table 2.10 and in each regression in Table 2.11 (therefore, each regression in both tables only has observations through 2006); the only difference between the top panel and the bottom panel of each table is that the seven regressions in the top panel do not control for parental education while the seven regressions in the bottom panel do.

Table 2.10 presents end-of-grade math results. First, we will compare the baseline results in Table 2.5 with the results in Table 2.10 which do not control for parental education. As with the baseline results presented in Table 2.5, every coefficient in Table 2.10 is negative and statistically significant. The magnitude of the coefficients is very similar between Tables 2.5 and 2.10. For example, the specification with only main effects (year, school, and grade effects) features a coefficient of -0.109 in the baseline model and -0.090 in Table 2.10; the school-by-year with grade models feature coefficients of -0.136 (Table 2.5) versus -0.149 (Table 2.10); and the year-by-grade with school models feature coefficients of -0.108 (Table 2.5) versus -0.085 (Table 2.10). The object changing between these samples is the omission of post-2006 observations.

TABLE 2.10: Math results robust to parental education control

This table reports estimates of equations (1) and (2). The dependent variable is a student's end-of-grade test score in mathematics. Treatment equals one if school s in year t offered single-sex classes for grade g, and equals zero otherwise. All regressions control for the race of the student and for the student's lagged test score. Standard errors are clustered on school.

	(A)	(B)	(C')	(D)	(E)	(F)	(G)
	No parental education control						
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**
Single-sex math classes	-0.094	-0.090	-0.123	-0.149	-0.113	-0.085	-0.132
	[0.029]***	[0.025]***	[0.008]***	[0.020]***	[0.027]***	[0.025]***	[0.021]***
Observations	466704	466704	466704	466704	466704	466704	466704
Adjusted R-squared	0.76	0.77	0.77	0.77	0.76	0.77	0.77
	Parental education control						
	(8)	**(9)**	**(10)**	**(11)**	**(12)**	**(13)**	**(14)**
Single-sex math classes	-0.083	-0.086	-0.121	-0.148	-0.093	-0.082	-0.132
	[0.029]***	[0.025]***	[0.005]***	[0.016]***	[0.027]***	[0.024]***	[0.018]***
Observations	466704	466704	466704	466704	466704	466704	466704
Adjusted R-squared	0.76	0.77	0.77	0.77	0.76	0.77	0.77
School effects	No	Yes	No	No	No	Yes	No
Grade effects	No	Yes	No	Yes	No	No	No
Year effects	No	Yes	No	No	No	No	No
School-by-year effects	No	No	Yes	Yes	No	No	Yes
Year-by-grade effects	No	No	No	No	Yes	Yes	Yes

TABLE 2.11: Reading results robust to parental education control

This table reports estimates of equations (1) and (2). The dependent variable is a student's end-of-grade test score in reading. Treatment equals one if school s in year t offered single-sex classes for grade g, and equals zero otherwise. All regressions control for the race of the student and for the student's lagged test score. Standard errors are clustered on school.

	(A)	(B)	(C')	(D)	(E)	(F)	(G)
	No parental education control						
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**
Single-sex	-0.015	0.006	-0.049	-0.033	-0.043	0.007	0.012
reading/English classes	[0.030]	[0.013]	[0.013]***	[0.009]***	[0.027]	[0.009]	[0.016]
Observations	392336	392336	392336	392336	392336	392336	392336
Adjusted R-squared	0.69	0.70	0.70	0.70	0.70	0.70	0.71
	Parental education control						
	(8)	**(9)**	**(10)**	**(11)**	**(12)**	**(13)**	**(14)**
Single-sex	-0.001	0.004	-0.058	-0.043	-0.018	0.006	0.000
reading/English classes	[0.023]	[0.017]	[0.013]***	[0.009]***	[0.016]	[0.013]	[0.017]
Observations	392336	392336	392336	392336	392336	392336	392336
Adjusted R-squared	0.70	0.70	0.71	0.71	0.71	0.71	0.71
School effects	No	Yes	No	No	No	Yes	No
Grade effects	No	Yes	No	Yes	No	No	No
Year effects	No	Yes	No	No	No	No	No
School-by-year effects	No	No	Yes	Yes	No	No	Yes
Year-by-grade effects	No	No	No	No	Yes	Yes	Yes

Comparing the top and bottom panels of Table 2.10, we see that the coefficients are remarkably similar. Model (D), with school-by-year and grade effects, shows that the coefficient on the treatment moves from -0.149 without controlling for parental education to -0.148 with the addition of the parental education control. The coefficient in model (G), with year-by-grade and school effects, moves from -0.085 without the control to -0.082 with the control. Model (F), with both school-by-year and year-by-grade effects, shows that the addition of the parental education control has no effect on the coefficient estimate: both coefficients equal -0.132.

Turning to the reading results, we will again first compare the baseline reading results from Table 2.6, estimated on all available data, to the top panel of Table 2.11, estimated on pre-2007 data. The results for reading are less robust to the change in sample than the results for math. In the model with main effects, for example, the Table 2.6 baseline results has a coefficient of -0.054, while the coefficient in Table 2.11 flips signs and is equal to 0.006 (the Table 2.6 coefficient is statistically significant, while the Table 2.11 coefficient is not). Five of the seven regressions in the baseline results have statistically significant coefficients, while only the two regressions with school-by-year effects have statistically significant coefficients in the restricted sample. Having said that, a comparison of those two models across the different samples shows that the coefficients are similar in magnitude (-0.051 in the baseline sample versus -0.049 in the restricted sample, and -0.033 versus -0.021).

As with the math results in Table 2.10, comparing the top and bottom panels of Table 2.11 shows that the addition of parental education as a control variable does not have a large effect on the results when using a consistent sample. In both panels, only models (C’) and (D) have statistically significant coefficients. The addition of the parental education control changes the magnitude of the coefficient by about 0.01 standard deviations.

While it is reassuring that the math results are robust to the change in sample, it is not particularly troubling that the reading results are not. Returning to Table 2.2, we see that over half the single-sex reading/English classes and around half of the single-sex math classes occurred post-2006 — in the restricted samples of Tables 2.10 and 2.11, none of these classes are included.

With respect to the same-sample comparison, the results shown in Tables 2.10 and 2.11 suggest that the omission of parental education is not substantially biasing the estimate of the treatment effect. A possible explanation for this is that the inclusion of the lagged test score is effectively controlling for the student's prior history of home inputs (Rivkin *et al.*, 2005).

Different Sample Selection Criterion

Section 2 details the algorithm used to assign school-year-grades to treatment status. $T_{gst} = 1$ if a school-year-grade is determined to offer single-sex classes; $T_{gst} = 0$ otherwise. Student-years enrolled in a school-year-grade with $T_{gst} = 1$ are the treatment group, and student-years enrolled in a school-year-grade with $T_{gst} = 0$ are the control group. In order for T_{gst} to equal one, it is required that the school-year associated with that school-year-grade offer more than φ single-sex classes. So even if a classroom composed of, say, exclusively male students is found in the data for a particular school-year-grade, in order for that school-year-grade to have $T_{gst} = 1$, it must be the case that at least φ single-sex classes are found in the data for that school-year.

Table 2.1 shows the distribution over years of single-sex classrooms prior to imposing that requirement. Notice that single-sex classes are found prior to the 2001 policy change. Recall that the additional requirement is imposed as a check against measurement error in the SAR data, and to ensure that there are no schools offering single-sex classes prior to the enactment of the 2001 NCLB law.

While I argue that the additional requirement is the correct coding strategy given the data, it is clearly possible that for at least some schools I am finding evidence of pre-2001 single-sex classes not because of measurement error, but instead because single-sex classes were actually offered in those years at those schools. Analyzing the data with φ equal to one is a useful check against the baseline results.

Using φ equal to one, the number of schools which offer single-sex classes is implausibly large. In the preceding analysis, school-year-grades were coded as offering single-sex math classes in nine schools, and school-year-grades were coded as offering single-sex reading classes in seven. Using φ equal to one — removing the requirement that a school-year offer multiple single-sex classes as a method of dealing with measurement error — results in school-year-grades being coded as offering single-sex math classes in 122 schools and in school-year-grades being coded as offering single-sex reading/English classes in 125 schools.

Recall from the introduction that the NASSPE reports that as of January 2009 there were at least 518 schools which were either single-sex or which were co-ed but offering some single-sex classes operating in the United States. It is unlikely that approximately twenty-five percent of those schools were in North Carolina.

With that caveat in mind, let us turn to the results. Tables 2.12 and 2.13 present estimates of the baseline models where the treatment variable, T_{gst}, has been redefined as described above, with φ equal to one. Notice first that the sample size for each regression increases dramatically. The sample size for the baseline math regressions is 621,620 student-years, and for reading is 523,150. However, when φ equals one, the new math and reading sample sizes are 3,220,463 and 3,003,708 student-years, respectively. This is because the control group is still defined as those student-years enrolled in school-year-grades which did not offer single-sex classes, but which are in school districts which ever had a school which offered single-sex classes. Because the number of schools which are coded as offering single-sex classes has increased from a one-digit number to a three-digit number, the number of school districts included in the control group has increased as well, and hence the number of student-years.

Turning first to Table 2.12, we see that under the new, less conservative definition of the treatment variable, the treatment coefficient is dramatically attenuated, and in all twenty-one regressions the coefficient is statistically insignificant. The coefficient estimate is as small as 0.000 in one regression, and the largest the coefficient gets in absolute value is less than three-hundredths of a standard deviation. The results for reading, reported in Table 2.13, are very similar. None of the coefficients is statistically significant, and the magnitude of the coefficients is very close to zero.

If it is the case, as the number of treated schools suggests, that the coding of T_{gst} used in the baseline results is closer to the truth than the coding using φ equal to one, then this type of attenuation is what you might expect to see. Essentially, what is happening is the misassignment of control observations to the treatment group.
The control group should experience no effect due to the policy, since it is by definition untreated. If the policy has an effect, then the treatment group will experience the effect of the policy. The result from incorrectly assigning control observations to the treatment group, then, is to bias the treatment coefficient to zero. Comparing the baseline results to the results presented in Tables 2.12 and 2.13 shows attenuation which is consistent with the hypothesis that (a) there is an effect of the treatment, as evidenced by the baseline results, and (b) using φ equal to one to define

TABLE 2.12: No effect for math under alternative definition of treatment status

This table reports estimates of equations (1) and (2). The dependent variable is a student's end-of-grade test score in mathematics. Treatment equals one if school s in year t offered single-sex classes for grade g, and equals zero otherwise. All regressions control for the race of the student and for the student's lagged test score. Standard errors are clustered on school.

	(A)	**(B)**	**(C')**	**(D)**	**(E)**	**(F)**	**(G)**
	Male and female students						
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**
Single-sex math classes	0.002	0.000	0.016	0.013	-0.014	-0.015	-0.018
	[0.019]	[0.015]	[0.028]	[0.027]	[0.016]	[0.012]	[0.018]
Observations	3220463	3220463	3220463	3220463	3220463	3220463	3220463
Adjusted R-squared	0.73	0.74	0.74	0.75	0.73	0.74	0.75
	Only male students						
	(8)	**(9)**	**(10)**	**(11)**	**(12)**	**(13)**	**(14)**
Single-sex math classes	0.000	0.004	0.021	0.018	-0.011	-0.011	-0.014
	[0.019]	[0.015]	[0.028]	[0.027]	[0.016]	[0.012]	[0.019]
Observations	1628301	1628301	1628301	1628301	1628301	1628301	1628301
Adjusted R-squared	0.73	0.74	0.74	0.75	0.73	0.74	0.75

TABLE 2.12 (Continued)

	(A)	(B)	(C')	(D)	(E)	(F)	(G)
	Only female students						
	(15)	**(16)**	**(17)**	**(18)**	**(19)**	**(20)**	**(21)**
Single-sex math classes	0.004	-0.003	0.010	0.007	-0.017	-0.019	-0.024
	[0.019]	[0.016]	[0.028]	[0.028]	[0.015]	[0.013]	[0.019]
Observations	1592162	1592162	1592162	1592162	1592162	1592162	1592162
Adjusted R-squared	0.73	0.74	0.74	0.75	0.73	0.74	0.75
School effects	No	Yes	No	No	No	Yes	No
Grade effects	No	Yes	No	Yes	No	No	No
Year effects	No	Yes	No	No	No	No	No
School-by-year effects	No	No	Yes	Yes	No	No	Yes
Year-by-grade effects	No	No	No	No	Yes	Yes	Yes

TABLE 2.13: No effect for reading under alternative definition of treatment status

This table reports estimates of equations (1) and (2). The dependent variable is a student's end-of-grade test score in reading. Treatment equals one if school s in year t offered single-sex classes for grade g, and equals zero otherwise. All regressions control for the race of the student and for the student's lagged test score. Standard errors are clustered on school.

	(A)	**(B)**	**(C')**	**(D)**	**(E)**	**(F)**	**(G)**
	Male and female students						
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**
Single-sex	-0.005	-0.004	0.000	-0.001	-0.012	-0.006	-0.005
reading/English classes	[0.012]	[0.009]	[0.012]	[0.012]	[0.011]	[0.008]	[0.011]
Observations	3003708	3003708	3003708	3003708	3003708	3003708	3003708
Adjusted R-squared	0.68	0.69	0.69	0.69	0.68	0.69	0.69
	Only male students						
	(8)	**(9)**	**(10)**	**(11)**	**(12)**	**(13)**	**(14)**
Single-sex	-0.007	-0.004	0.002	-0.001	-0.015	-0.007	-0.007
reading/English classes	[0.014]	[0.010]	[0.016]	[0.015]	[0.013]	[0.009]	[0.013]
Observations	1515070	1515070	1515070	1515070	1515070	1515070	1515070
Adjusted R-squared	0.68	0.68	0.69	0.69	0.68	0.68	0.69

TABLE 2.13 (Continued)

	(A)	(B)	(C')	(D)	(E)	(F)	(G)
	Only female students						
	(15)	**(16)**	**(17)**	**(18)**	**(19)**	**(20)**	**(21)**
Single-sex reading/English classes	-0.003	-0.004	-0.002	-0.002	-0.010	-0.006	-0.005
	[0.012]	[0.009]	[0.011]	[0.011]	[0.011]	[0.009]	[0.011]
Observations	1488638	1488638	1488638	1488638	1488638	1488638	1488638
Adjusted R-squared	0.68	0.69	0.69	0.69	0.68	0.69	0.69
School effects	No	Yes	No	No	No	Yes	No
Grade effects	No	Yes	No	Yes	No	No	No
Year effects	No	Yes	No	No	No	No	No
School-by-year effects	No	No	Yes	Yes	No	No	Yes
Year-by-grade effects	No	No	No	No	Yes	Yes	Yes

the treatment group results in misassignment of observations from control to treatment, thereby attenuating the coefficient estimate.

Period (*t* + 1) Treatment & Period *t* Value Added

Perhaps the principle concern regarding the validity of this analysis is a correlation between the error term and the treatment variable. While the school-by-year and year-by-grade effects are certainly powerful controls, it still may be the case that the effect measured by the coefficient on T_{gst} is confounding unobservables with a student being enrolled in a school-year-grade which offers single-sex classes.

Suppose that a particular cohort of students is “bad”, and that their “badness” is unobservable to the researcher. A principal or group of parents may offer single-sex classes for this cohort as a way to help improve their test scores, the single-sex classes may have no effect on test scores, the cohort may continue to perform poorly, and the researcher will find a negative treatment effect due to the correlation between unobservable “badness” and the offering of single-sex classes.

The school-by-year effects are designed to “compare” students in the same school in the same year but in different grades, and the year-by-grade effects are designed to “compare” students in the same year in the same grade in different schools. These are powerful controls, and they remove much of the concern about unobservables biasing the coefficient, but of course they do not completely remove this concern. A check against this bias is to estimate the following equations:

$$y_{igst} - \beta'' y_{igst-1} = \alpha'' + \gamma'' T_{gst+1} + X'_{igst}\theta'' + \omega_{st} + \varepsilon''_{its} \quad (3)$$

$$y_{igst} - \beta''' y_{igst-1} = \alpha''' + \gamma''' T_{gst+1} + X'_{igst}\theta''' + \tau_{tg} + \varepsilon'''_{its}. \quad (4)$$

Everything is defined in these equations as in equations (1) and (2). The only difference is that the treatment variable is advanced in time one year. So the regressions estimate the effect of *future* treatment on *current* test score value added.[30]

If $cov(\varepsilon_{its}, T_{gst}) \neq 0$, then it stands to reason that $cov(\varepsilon_{its}, T_{gst+1})$ may not be equal to zero as well — if treatment status in the current-period is correlated with some unobservable characteristic of a student or cohort of students, and if that unobservable characteristic lasts more than one period, then we might expect treatment status in the next-period to be correlated with the present-period unobservable.

Of course, being enrolled in a school-year-grade which offers single-sex classes *next year* cannot affect *this year's* end-of-grade test scores. So a statistically significant coefficient on T_{gst+1} would raise the concern that the baseline regressions are not adequately controlled, and that the treatment coefficient in the baseline regressions is biased.

The results of these models are presented in Table 2.14. The top panel presents results using end-of-grade math scores, and the bottom panel presents results using end-of-grade reading scores. Columns (A) through (G) represent the same configurations of control variables used in the previous tables.

The results for end-of-grade math scores are encouraging: of the five models with two-way effects, only the year-by-grade and school effects model has a statistically significant treatment coefficient. Interestingly, the model without two-

[30] This test is based on Avery, Kenkel, Lillard, and Mathios (2007). They study how smoking behavior is affected by the advertisement of smoking cessation products. To explore potential reverse causality driven by the possibility that cessation product advertisers target smokers who are relatively more likely to quit smoking, Avery *et al* regress current-period smoking behavior against future advertisements.

TABLE 2.14: Future-period treatment & current-period value added

This table reports estimates of equations (3) and (4). The dependent variable is a student's end-of-grade test score in either reading or mathematics in period t. Treatment equals one if school s in year (t+1) offered single-sex classes for grade g, and equals zero otherwise. All regressions control for the race of the student and for the student's period (t-1) test score. Standard errors are clustered on school.

	(A)	**(B)**	**(C')**	**(D)**	**(E)**	**(F)**	**(G)**
	Dependent variable: End-of-grade mathematics score						
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**
Single-sex math classes	-0.037	-0.051	-0.023	-0.033	-0.047	-0.048	-0.036
	[0.032]	[0.022]**	[0.024]	[0.026]	[0.042]	[0.023]**	[0.026]
Observations	407035	407035	407035	407035	407035	407035	407035
Adjusted R-squared	0.75	0.76	0.76	0.76	0.75	0.76	0.77
	Dependent variable: End-of-grade reading score						
	(8)	**(9)**	**(10)**	**(11)**	**(12)**	**(13)**	**(14)**
Single-sex	-0.117	-0.073	-0.060	-0.057	-0.110	-0.072	-0.052
reading/English classes	[0.015]***	[0.020]***	[0.032]*	[0.032]*	[0.018]***	[0.020]***	[0.033]
Observations	340744	340744	340744	340744	340744	340744	340744
Adjusted R-squared	0.69	0.70	0.70	0.70	0.69	0.70	0.70
School effects	No	Yes	No	No	No	Yes	No
Grade effects	No	Yes	No	Yes	No	No	No
Year effects	No	Yes	No	No	No	No	No
School-by-year effects	No	No	Yes	Yes	No	No	Yes
Year-by-grade effects	No	No	No	No	Yes	Yes	Yes

way effects (but with school, year, and grade effects) has a statistically significant treatment coefficient. This is evidence that only controlling for main effects is not adequate. The results for end-of-grade reading scores should raise concerns about the baseline results. Every coefficient which is significant in the baseline results (reported in Table 2.6) is significant in the falsification test. Since it cannot be that treatment next period affects scores this period, there must important unobserved factors biasing the treatment coefficient.

V. Concluding Discussion

This paper studies the effect of a school offering single-sex classes on the test scores and conduct of students. Estimates are produced using both within-school-by-year and within-year-by-grade variation. Four results from this study are highlighted in this discussion: (1) the heterogenous effect of the treatment on test scores, (2) the average effect of the treatment, (3) the discriminatory effect across genders, and (4) the sensitivity of the results.

There is wide heterogeneity in the effect across schools, with some school-year-grades having higher average end-of-grade gains to reading scores (a positive coefficient) associated with their offering single-sex classes and most school having lower average scores.

Advocates of single-sex classes are quick to emphasize that simply segregating students by gender is no panacea — teachers must be trained to teach in a single-sex environment, and some advocates even argue that different teaching methods should be used for boys and for girls. It is possible, then, that the heterogeneity documented in this paper is evidence in support of the advocates' assertion: some schools are doing a good job implementing single-sex classes, and some are not. More research is needed on the specific practices offered by the schools to determine whether this speculative observation is grounded in fact.

Using either empirical model, school-year-grades which offer single-sex reading and math classes have lower end-of-grade scores as measured by a common treatment effect variable. In several models the magnitude of the effect is quite large, similar in size to the black-white test score gap. While the school-specific treatment variables suggest that some schools may be successfully enhancing student outcomes by offering single-sex classes, the evidence presented in this study suggests that the average effect is not positive.

A concern among opponents of single-sex education is that such programs will be discriminatory — relative to co-ed classes, single-sex classes will help students of one gender to learn better while hurting the ability of the other gender to learn. I find weak evidence that this discriminatory effect is present.

A fruitful direction for future research is to quantify the actual implementation of single-sex classes. Through a survey or through interviews, a researcher could first ascertain with more certainty which schools are offering these programs and which are not. In addition, the researcher could record exactly what is happening in these classes: are boys and girls simply separated and taught using traditional teaching methods, or are innovative, gender-specific methods used? If gender-specific methods are used, what are they? Quantifying these survey results and linking them to test score data would allow an empirical investigation not only of the effects of a school offering single-sex classes, as I have done here, but also of the types of gender-specific instruction offered at these schools. This would go a long way to settling the heated debate over single-sex education.

There are at least two caveats which need to be addressed in interpreting the results of this study. The first is the sensitivity of the results to the coding of the treatment variable. Under the more conservative coding scheme (which requires a school-year to offer several single-sex classes before students enrolled in its associated school-year-grade can be coded as receiving the treatment) this paper finds a negative and sizeable effect of single-sex classes on student outcomes. However, when the conservative requirement is removed, the treatment effect coefficient is statistically insignificant and its magnitude is very close to zero. While I argue that the baseline coding of the treatment variable is the more correct of the two, the evidence from the robustness check does not support the hypothesis that the treatment effect is negative.

The second caveat is that *future* treatment affects *current* end-of-grade reading scores, though this is not true for math. This implies that even the powerful school-by-year and year-by-grade effects are not adequately controlling for unobservables in the end-of-grade reading regressions.

On balance, then, perhaps the most responsible conclusions from this paper are these: there is evidence that schools which offer single-sex math classes are hurting the end-of-grade math scores of their students, there is evidence that schools which offer single-sex classes in reading/English are not helping their students to increase achievement on end-of-grade reading tests, and some schools are implementing single-sex classes more successfully than others.

ACKNOWLEDGMENTS

I thank John Abowd, Kiel Albrecht, James Cowan, Ronald Ehrenberg, Kevin Hallock, George Jakubson, Eamon Molloy, Ben Ost, Ian Schmutte, Douglas Webber, and Kenneth Whelan for very helpful comments and discussion. Special thanks are owed to Kirabo Jackson. I am grateful to the staff of the North Carolina Education Research Data Center for providing the data, and thank Pam Baxter and Carol Murphee of the Cornell Institute for Social and Economic Research for providing and maintaining the computing environment in which this research was carried out.

REFERENCES

Avery, Rosemary, Donald Kenkel, Dean R. Lillard, and Alan Mathios (2007). "Private Profits and Public Health: Does Advertising of Smoking Cessation Products Encourage Smokers to Quit?" *Journal of Political Economy*, vol. 115, no. 3, pp. 447-481.

Billger, Sherrilyn M. (2002). "Admitting Men Into a Women's College: A Natural Experiment," *Applied Economics Letters*, vol. 9, no. 7, pp. 479-483.

Billger, Sherrilyn M. (2009). "On reconstructing school segregation: The efficacy and equity of single-sex schooling," *Economics of Education Review*, vol. 28, no. 3, June 2009, pages 393-402.

Clotfelter, Charles, Elizabeth Glennie, Helen Ladd, and Jacob Vigdor (2006). "Would Higher Salaries Keep Teachers in High-Poverty Schools? Evidence from a Policy Intervention in North Carolina," *Journal of Public Economics*, vol. 92, no. 5-6, June 2008, pp. 1352-1370.

Clotfelter, Charlies T., Helen F. Ladd, and Jacob L. Vigdor (2006). "Teacher-Student Matching and the Assessment of Teacher Effectiveness," *Journal of Human Resources*, vol. 41, no. 4, Fall 2006, pp. 778-820.

Evans, William N. and Robert M. Schwab (1995). "Finishing High School and Starting College: Do Catholic Schools Make a Difference?" *Quarterly Journal of Economics*, vol. 110, no. 4, November 1995, pp. 941-974.

Hoxby, Caroline M. (2000). "The effects of class size and composition on student achievement: New evidence from population variation," *Quarterly Journal of Economics*, vol. 115, no. 4, November 2000, pp. 1239-1285.

Jackson, C. Kirabo (2009). "Student Demographics, Teacher Sorting, and Teacher Quality: Evidence from the End of School Desegregation," *Journal of Labor Economics*, vol. 27, no. 2, April 2009, pp. 213-256.

Jackson, C. Kirabo (2010). "Single-Sex Schools, Student Achievement, and Course Selection: Evidence from Rule-Based Student Assignments in Trinidad and Tobago", Working paper, retrieved April 2010: http://works.bepress.com/c_kirabo_jackson/19

Jackson, C. Kirabo and Elias Bruegmann (2009). "Teaching Students and Teaching Each Other: The Importance of Peer Learning for Teachers," *American Economic Journal: Applied Economics*, vol. 1, no. 4, 2009, pp. 85-108.

Jimenez, Emmanuel and Marlaine E. Lockheed (1989). “Enhancing Girls’ Learning Through Single-Sex Education: Evidence and a Policy Conundrum,” *Educational Evaluation and Policy Analysis*, vol. 11, no. 2, Summer 1989, pp. 117-142.

Lavy, Victor and Analia Schlosser (2011). “Mechanisms and Impacts of Gender Peer Effects at School,” *American Economic Journal: Applied Economics*, vol. 3, no. 2, April 2011, pp. 1-33.

National Association for Single Sex Public Education, http://www.singlesexschools.org/home.php, accessed April 2011.

Neal, Derek (1997). “The Effects of Catholic Secondary Schooling on Educational Achievement,” *Journal of Labor Economics*, vol. 15, no. 1, January 1997, pp. 98-123.

Ost, Ben (2010). “How Do Teachers Improve? The Relative Importance of General and Specific Human Capital,” working paper.

Rivkin, Steven G., Eric A. Hanushek, and John F. Kain (2005). “Teachers, Schools, and Academic Achievement,” *Econometrica*, vol. 73, no. 2, March 2005, pp. 417-458.

Rothstein, Jesse (2009). “Student Sorting and Bias in Value Added Estimation: Selection on Observables and Unobservables,” *Education Finance and Policy*, vol. 4, no. 4, Fall 2009, pp. 537-571.

Rothstein, Jesse (2010). “Teacher Quality in Educational Production: Tracking, Decay, and Student Achievement,” *Quarterly Journal of Economics*, vol. 125, no., 1, February 2010, pp. 175-214.

Solnick, Sara J. (1995). “Changes in women’s majors from entrance to graduation at women’s and coeducational colleges,” *Industrial and Labor Relations Review,* vol. 48, no. 3, April 1995, pp. 505-514.

U.S. Department of Education, Office of Planning, Evaluation and Policy Development, Policy and Program Studies Service, *Early Implementation of Public Single-Sex Schools: Perceptions and Characteristics*, Washington, D.C., 2008.

CHAPTER 3

PAYDAY CREDIT, OVERDRAFTS, AND BANKRUPTCY, BOTH FORMAL AND INFORMAL[31]

Donald P. Morgan,[32] Michael R. Strain,[33] and Ihab Seblani

ABSTRACT

Despite a dozen studies, the welfare effects of payday credit are still debatable. We contribute new evidence to the debate by studying how changes in payday credit supply affect bank overdrafts (such as returned checks), bankruptcy, and "informal bankruptcy," where debtors are in default without the benefit of bankruptcy protection from debt collectors. We use household complaints against debt collectors as a proxy for informal bankruptcy. We find some evidence that households switch from informal bankruptcy to bankruptcy when payday credit supply expands. That finding provides an alternative interpretation of a recent finding that payday credit access "tips" households into bankruptcy. Our most robust finding is that returned check rates and overdraft fee income at banks tend to decline when payday credit supply expands. A hundred dollar payday loan is probably cheaper than a hundred dollar overdraft, so this finding may indicate payday credit access helps some consumers avoid costlier alternatives.

---

[31] JEL classification: G21, G28, I38
Key words: payday credit, consumer welfare, bounced checks, overdrafts, debt collectors, dunning, bankruptcy, informal bankruptcy.
[32] Morgan and Seblani are Research Officer and Assistant Economist at the Federal Reserve Bank of New York. Address correspondence to don.morgan@ny.frb.org.
[33] Strain is a Ph.D. student in Economics at Cornell University.

I. Introduction

Payday lenders supply credit by cashing and holding (without depositing) customers' personal checks for a few weeks. Their product can be seen as a less automated, possibly cheaper, version of the overdraft credit depository institutions supply when they cover depositors' overdrafts.

Whether payday lending helps or hurts its users depends on who asks the question and who gets asked. A survey of payday credit consumers commissioned by a payday lending trade group found relatively high rates of customer satisfaction, although dissatisfied customers wished the service were cheaper (Elliehausen and Lawrence 2008). Consumer advocates tend to see payday lending as harmful because it is expensive, because some users borrow repeatedly, because payday lenders may congregate in Hispanic and African American neighborhoods (Damar 2009), and because payday lenders have gone to great lengths – even partnering with banks – to circumvent state usury limits. In their own defense, payday lenders claim their loans help households with short-term debt problems to avoid further problems.

The disparate views of payday lending are *writ large* in state payday loan laws. A few states are more or less *laissez faire* while others, mostly in the Northeast, have never allowed payday lending. Eight states have closed active payday credit markets, including Arizona just this year.

Economic theory is ambiguous as to whether the invention of another form of overdraft credit should raise or lower household welfare. Textbook theory predicts households should be weakly better off having an alternative available if for no other reason than increased competition. Moreover, if the new product enables better consumption smoothing, households benefit as a result. Of course, textbook models assume symmetric information between lenders and borrowers and perfectly rational, time-consistent borrowers, and violations of those assumptions are easily imaginable.

To the extent these problems plague overdraft credit users, the invention of another form of overdraft credit could make users worse off.

Academic empirical research on the welfare effects of payday credit access, though highly energetic, began just a few years ago. Except for Zinman (2010), that research to date is still in working papers.[34] These early findings from the dozen studies thus far are mixed. Our paper extends the literature by looking at new outcomes that seem particularly germane to the debate.

Broadly speaking, we study two outcomes: bankruptcy and overdrafts. While the former is more dramatic, the latter seems at least as pertinent. Payday and overdraft credit are essentially the same product so it is easy to imagine how someone with a job but a temporarily empty checking account might take out a cash payday loan to avoid bouncing checks or overdrawing their account in other ways. Avoiding overdrafts, it turns out, is a common reason why payday credit users say they turn to payday credit (Cerillo 2004, Stegman and Faris 2003). We measure overdrafts in two ways: by returned checks and by overdraft fee income at banks. Those variables are direct counterparts; every fee paid on a returned check or ATM overdraft is revenue for the counterparty depository institution. Both variables are measured with error (because of aggregation) but as we show, the error does not bias our estimates of the coefficients of interest, though it does inflate the standard error of the estimates.

We also study multiple bankruptcy outcomes: Chapter 7, Chapter 13, and "informal bankruptcy" (Dawsey and Ausubel 2004), where debtors are in default without the protection from debt collectors that bankruptcy provides. Given default, borrowers have to decide whether to seek bankruptcy protection or to endure debt

[34] See Morgan (2007), Morgan and Strain (2008), Skiba and Tobacman (2008a, 2008b), Campbell et al. (2008), Stoianovici and Maloney (2008), Morse (2009), Carrell and Zinman (2009), Melzer (2009), Wilson et al. (2008), Melzer and Morgan (2010), and Hynes (2010).

collectors. One hypothesis we consider is whether changes in payday credit supply cause substitution between bankruptcy and unofficial bankruptcy.

Our proxy for informal bankruptcy is complaints against lenders and debt collectors filed by households with the Federal Trade Commission.[35] Those data are new to this study; we obtained them under the Freedom of Information Act. According to the FTC: "Abusive (debt) collection practices … are known to cause substantial consumer injury" (Commission 2006, p.1), so complaints might be associated with consumer welfare, the ultimate outcome of interest. We provide auxiliary evidence that complaints are a reasonable proxy for informal bankruptcy, but even if they were not, it still seems worthwhile to see how they vary with payday loan supply.

Our data reflect market events, not field-experiment or laboratory outcomes, so we face the usual problem of distinguishing supply from demand. The goal is to estimate how exogenous changes in payday credit supply affect outcomes; the problem is that those same outcomes may drive payday credit demand. To avoid bias, we use changes in states' payday loan laws to identify changes in payday credit supply. We study how the outcomes change after law changes in nineteen states, including bans in eight states (including D.C.) and the passage of enabling legislation in eleven states. Given those events, we estimate fixed effect difference-in-difference regression models where delta (the diff-in-diff) measures the change in outcomes after a ban or the passage of enabling legislation relative to the change in states where laws were constant. The models control for state economic conditions and demographic characteristics that likely correlate with outcomes and payday loan demand.

We find some evidence of substitution along two margins. Consistent with one study, we find some evidence that Ch. 13 bankruptcy rates decrease when payday

[35] People can complain by calling 1877-FTC-HELP on line.

loans are banned. That first finding suggests payday lending is harmful. However, in those models where we find lower rates of Ch. 13 after payday loan bans, we observe higher rates of informal bankruptcy/complaints against lenders and debt collectors. That suggests some households may use payday credit to avoid debt collectors by filing for Ch. 13. The welfare implications of that substitution are not obvious. Perhaps more importantly, we find no evidence that complaints against lenders and debt collectors increase when payday loan supply increases. If anything, the relationship between complaints and payday credit supply is negative.

Our most robust finding is that returned check rates and overdraft fee income at banks decrease when payday credit supply expands. That finding suggests that borrowers substitute payday lenders' variety of overdraft credit for the bank variety when the former is available. A payday loan can be cheaper than a small overdraft so that substitution could save households money. In fact, our estimates suggest that households served by a given Federal Reserve Check Processing Center save about $38 million per year on average in overdraft fees after states pass enabling legislation. Falsification tests suggest the link between payday credit supply and those overdraft outcomes is not merely coincidence.

Our findings add to the nascent literature on the welfare effects of payday credit access. In particular, our evidence of substitution from informal to formal bankruptcy suggests an alternative interpretation of Skiba and Tobacmans' (2008a) finding that payday credit "tips" marginal payday credit users into bankruptcy; if the informal substitution hypothesis is correct, marginal borrowers that are already in default may use payday credit to "step" into bankruptcy to avoid debt collectors. Our findings also reinforce Zinman (2010). He finds that households in Oregon expected to bounce more checks after payday loans were banned there, but in the event they did not report bouncing more checks relative to their counterparts in Washington. Our findings using actual returned checks and overdraft fee income data suggest that households do indeed overdraw their accounts less frequently when payday credit supply contracts.

The next section presents background on the overdraft credit market. Section III discusses our regression strategy and taxonomy of state payday loan laws. Section IV presents the regression results. Section V comprises robustness tests, falsification tests, and potential bias. VI concludes with policy observations and suggestions for future research.

II. The Overdraft Credit Market and Its Players

Payday loans are said to be small, short, and insecure. The typical loan is commodity-like: $300 for two weeks of credit secured by proof of employment and a personal, post-dated check for $345 drawn on the borrowers' checking account. Two weeks later the lender deposits the check and the credit is extinguished. At those terms, the annual percentage rate on a payday loans is 390 percent.

Given that business model, we know for sure that payday credit consumers are employed and "banked." That suggests they are not the poorest of the poor. On the contrary, the comparative survey by Lawrence and Elliehausen (2008) found that 51 percent of payday loan users earned $25,000 to $50,000 annually.[36] Payday customers were more likely than the average household to have attended college (36 percent versus 21 percent) but less likely to have graduated (19 percent versus 35 percent).

Payday loan demand seems to have a demographic profile, although this profile is blurry. Damar (2009) finds payday lenders are more likely to enter predominately Hispanic, well-banked zip codes in Oregon, but he finds no difference for predominately African American zip codes. Looking across counties, Prager

[36] Twenty-five percent earned more than $50,000 annually.

(2009) finds just the opposite racial pattern. (We control for racial shares in our regressions.)

It is important to recognize that some, perhaps most, payday credit users had debt problems that predate their first payday loan. After all, who else but over-indebted, credit constrained consumers would borrow at such terms? In fact, past problems with debt collectors and bounced checks, two outcomes we study, were primary predictors of whether lower income households in North Carolina demanded payday loans (Stegman and Faris 2003). Elliehausen and Lawrence (2001) found that 61 percent of payday customers were "maxed out" on their credit cards. We stress the possibility that payday credit users may have had pre-existing debt problems because it is crucial to the hypothesis that households may switch from informal bankruptcy (where they are already in default) to formal bankruptcy when payday credit supply expands.

Payday credit is closely akin to the overdraft credit ("protection") supplied by depository institutions. Both financial intermediaries supply credit by postponing depositing a check or debiting an account for a time, providing float in the interim. Certain usage patterns are common as well; as with payday credit, some depositors overdraft repeatedly and revenues from those "core" depositors accounts for a disproportionate share of overall overdraft revenues (FDIC 2008, Campbell et al. 2008).

The welfare benefits of payday credit are debatable only if one departs from textbook models of consumer credit with symmetric information and fully rational, time-consistent borrowers. If lenders are better informed than borrowers, for example, unsuspecting borrowers can be made worse off by a voluntary credit transaction (Morgan 2007). Alternatively, if payday credit users are naïve hyperbolic discounters that systematically overestimate their commitment to repay "short-term" loans, access to payday credit might make them worse off (Paige and Tobacman 2008b). In fact, repeat borrowing by some payday credit users may be indicative of information asymmetries and/or behavioral biases. Moreover, counseling prospective payday credit borrowers about the possibility of repeat usage does reduce demand for some borrowers (Bertrand and Morse 2009). Thus, one is theoretically and empirically justified in doubting the welfare benefits of payday credit.

In fact, the literature to date is quite mixed. For every study suggesting payday credit help users (Morgan and Strain 2008; Morse 2009; Zinman 2010), there is another study that finds harm (Melzer 2009; Carrel and Ziman 2009; Skiba and Tobacman 2008a). Thus, the question of whether payday credit helps or hurts users, and how, remains unsettled.

III. Regression Models and Payday Loan Laws

To identify the link between changes in payday credit supply and the outcomes we study, we estimate difference-in-difference regressions of the form:

$$Outcome_{st} = \delta Ban_{st} + \delta' Enabled_{st} + a + a_s + a_t + Controls_{st}\,\gamma + \varepsilon_{st}. \qquad \text{(A)}$$

The dependent variable is one of several outcomes we study, suitably scaled, in state *s* in month or quarter *t*. We discuss and source the outcomes below and in the appendix.

The key variables in model (A) are the dummy variables Ban_{st} and $Enabled_{st}$. Ban_{st} equals one (zero) after (before) state *s* banned payday lending. $Enabled_{st}$ equals one (zero) after (before) state *s* passed enabling legislation.[37] The coefficients on those variables, δ and δ', measure the difference-in-difference in $Outcome_{st}$ associated with a change in Ban_{st} or $Enabled_{st}$. Under any hypothesis about the relationship

[37] For laissez faire states that allow payday lending without explicit enabling legislation, $Enabled_{st}$ equals one for all *t*.

between payday credit supply and the outcomes, we would expect δ and δ' to have opposite signs. We do not force $|\delta| = |\delta'|$, however, in case one or other corresponding variables is a better proxy for supply. Note that because (A) includes the usual state fixed effects, δ and δ' are identified off within-state variation. Many panel data studies dimensioned like ours stop with (A), but we also estimate versions with a state-specific trend for robustness. Some results depend on whether the state specific trend is included.

We maintain the standard assumption that *Ban* and *Enabled* are exogenous with respect to the outcomes. In truth, both variables may reflect the relative interests and power of the four main stakeholders in the overdraft credit market: consumers, consumer advocates, payday lenders, and payday lenders' competitors. Given the confluence of so many, possibly offsetting, forces, it seems natural to follow the literature and take the law changes as exogenous. We discuss that assumption in more detail later in the paper.

$Controls_{st}$ is a vector of economic variables – log income, income growth, unemployment rate, home price index -- and demographic characteristics -- the share of population that are Black, Hispanic, Asian, and have college degrees. The economic variables are monthly or quarterly. The demographic characteristics are annual. All the data range between 1998 and 2008.

Table 3.1 reports our taxonomy of payday loan laws. Our coding largely follows but extends the coding in Morgan and Strain (2008), Melzer (2009), Melzer and Morgan (2010), Zinman (2010), Hynes (2010), and our own research of state laws. We do not claim our coding captures every binding law change, or that we have not included any non-binding changes. In particular, payday lenders were operating in many states even before enabling legislation was passed. Our supposition is that supply may have increased after enabling legislation as the new laws may have

TABLE 3.1: Payday lending laws in the 50 states and DC: January 1998 - December 2008*

States without Law Changes:			*States with Law Changes:*			
Always Banned	*Always Legal*		*Banned*	*Date*	*Enabled*	*Date*
Connecticut	California	Montana	Arkansas[b]	December 2007	Alabama[b]	June 2003
Maine	Colorado	Nebraska	D.C.	November 2007	Alaska	June 2004
Massachusetts	Delaware	Nevada	Georgia	May 2004	Arizona	April 2000
New Jersey	Florida	New Mexico	Maryland	June 2000	Arkansas	April 1999
New York	Idaho	Ohio	North Carolina	December 2005	Hawaii	July 1999
Vermont	Ilinois	South Carolina	Oregon	July 2007	Michigan	November 2005
	Indiana[a]	South Dakota	Pennsylvania	November 2007	New Hampshire	January 2000
	Iowa	Tennessee	West Virginia	June 2006	North Dakota	April 2001
	Kansas	Texas			Oklahoma[b]	September 2003
	Kentucky	Utah			Rhode Island	July 2001
	Louisiana	Washington			Virginia	April 2002
	Minnesota	Wisconsin				
	Mississippi	Wyoming				
	Missouri					

Sources: Morgan and Strain (2008), Melzer (2009), Morgan and Melzer (2010), Zinman (2010), Hynes (2010), Authors' research.

[a] Indiana passed a law "enabling" payday loans. We follow Hynes (2010) and code Indiana as always allowing payday lending. The results are robust to excluding Indiana

[b] The legality of payday lending was ambiguous in these states at times (See Fox and Mierzwinski (2001) and Carrel and Zinman (2009)). The results are robust to excluding those states.

provided safe harbors to payday lenders that were hesitant to enter without protection. We are confident that the bans are binding, however, based on the annual store counts by Stephens Inc., an investment bank that tracks the payday lending industry. Furthermore, to the extent the legal changes are *not* binding, it biases us against finding any relationship between the law changes and outcomes.

IV. Findings

We study bankruptcy, complaints, and then overdrafts. Background on each outcome comes at the beginning of each section. Summary statistics and sources are reported in the appendix. Means of the outcomes are also reported in the regression tables.

4.1.a Bankruptcy

Three papers have already examined the link between bankruptcy and payday credit access. Their findings are mixed. Stoivanici and Mahoney (2008) and Hynes (2010) find no relationship or a mixed relationship in their studies using state and county data. Using borrower-level data and a regression discontinuity design, Skiba and Tobacman (2008a) find that marginal applicants approved for a payday loan are more likely to file Ch. 13 than are marginal, rejected applicants.

In considering the hypothesis that variation in payday credit supply may cause substitution between bankruptcy and informal bankruptcy, it is important to remember that -- conditional on a borrower already being in default -- filing bankruptcy has costs *and* benefits. The costs are real and potentially substantial: court and attorney fees, diminished credit score, and stigma, a catch-all term for the non-pecuniary costs of filing bankruptcy. A main benefit of bankruptcy is protection, the automatic stay that halts all collection efforts by lenders and debt collectors, at least on unsecured debts.

Bankruptcy is clearly a bad outcome but it may not be the worst outcome; for a borrower already in default, bankruptcy may be preferable to informal bankruptcy.

We study personal bankruptcy filings per 10,000 persons at the state level between 1998:Q2 and 2008:Q4 by Chapter. Table 3.2 reports the bankruptcy regression models. The relationship between bankruptcy filings and the economic control variables seem sensible. For example, bankruptcy under either Chapter is increasing in the unemployment rate. Income growth is positively related to Ch. 7 rates and negatively related to Ch. 13 rates. Those opposing signs make sense; Ch. 13 filers have to share income with creditors, so growing income means a growing absolute obligation to creditors. Ch. 7 filers get to keep all their future income so all else the same, they may be more likely to use Ch. 7 when income is increasing.

A few of the racial variables are significant in some models. Ch. 7 rates are higher in states with proportionately more Asian households.[38] In the standard, fixed effects model, Ch. 7 rates are lower and Ch. 13 rates are higher in states proportionately more black households. We return to the latter result momentarily.

Turning to δ and δ', we observe no significant relationship between Ch. 7 and either *Ban* or *Enabled,* though we do see some link between *Ban* and Ch. 13. In the model without controls, δ is negative and significant at the five percent level. With controls, δ remains significant at the ten percent level. In the model with state specific trends, δ remains negative but is insignificant.

[38] Higher Ch. 7 demand in states with disproportionately more Asians could reflect filing by failed entrepreneurs that used credit cards to finance their business.

TABLE 3.2: Does Payday Credit Supply Affect Bankruptcy Demand?

Reported are OLS regression coefficients [robust t-statistics] estimated over 2244 state-quarter observations between 1998:Q1 and 2008:Q4. *Banned* equals one (zero) after (before) state *s* banned payday lending. *Enabled* equals one (zero) after (before) state *s* passed enabling legislation. All models include state and date fixed effects. Standard errors are clustered by state.

	Dependent variable (mean) = bankruptcy filings per 10000 persons under …					
	Ch. 7 (8.25)			Ch. 13 (2.95)		
	(1)	(2)	(3)	(4)	(5)	(6)
Banned	-0.65	-0.73	-1.16	-1.02	-0.93	-0.48
	[0.69]	[1.04]	[1.06]	[2.02]**	[1.71]*	[1.31]
Enabled	-0.44	-0.31	-0.49	0.06	0.01	-0.07
	[0.72]	[0.88]	[1.18]	[0.31]	[0.05]	[0.22]
Unemployment rate		0.37	0.40		0.26	0.19
		[1.84]*	[1.83]*		[2.36]**	[1.94]*
Log (income)		-8.69	-3.52		1.90	0.83
		[2.14]**	[0.93]		[1.09]	[0.61]
Income growth		6.30	3.81		-2.05	-1.53
		[2.77]***	[1.72]*		[2.05]**	[1.84]*
Home Prices		-0.01	-0.02		0.00	0.00
		[6.72]***	[4.90]***		[1.28]	[2.53]**
Black share		-22.60	-108.76		18.69	2.89
		[1.97]*	[1.51]		[2.45]**	[0.09]
Hispanic share		-10.49	-23.25		-20.35	21.53
		[1.19]	[0.66]		[3.94]***	[1.41]
Asian share		11.15	18.21		-0.19	0.60
		[4.03]***	[3.32]***		[0.15]	[0.22]
College share		3.97	0.15		3.89	5.80
		[0.75]	[0.03]		[1.44]	[2.29]**
Constant	4.52	91.99	35.09	0.56	-21.19	-9.01
	[14.82]***	[2.17]**	[1.03]	[5.13]***	[1.16]	[0.73]
State specific trend?	No	No	Yes	No	No	Yes
Adjusted R-squared	0.84	0.87	0.88	0.93	0.94	0.96

* significant at 10%; ** significant at 5%, *** significant at 1%

Though not particularly strong, the evidence here that Ch. 13 rates decrease with payday loan bans is consistent with Skiba and Tobacman (2008a). Before interpreting that result, we investigate how informal bankruptcy, as proxied by complaints against lenders and debt collectors, varies with payday credit supply. Recall that Stegman and Faris (2003) found that past problems with debt collectors was a primary reason (along with bounced checks) that lower income households demand payday credit. So it seems apt to study the relationship between complaints against debt collectors and payday credit supply.

4.1.b Complaints (Informal Bankruptcy)

The complaints data we study are collected by the Federal Trade Commission (FTC), the agency charged with enforcing the Fair Debt Collection Practices Act of 1978. The FTC keeps tabs for complaints against lenders and debt collectors.[39] We observe the data from January 1998 when the FTC created its hotline (1-877-FTC HELP) and there is a strong, positive trend in the series.[40]

The rate of complaints is low; the mean number of complaints against lenders and debt collectors collectively was only 1.41 per 100,000 persons per year. However, the FTC believes that only a "small percentage" (Commission 2006, p.4) of households being harassed by debt collectors actually complain to the FTC. We view the low rate of complaints as a scaling issue; presumably every defaulted debtor suffers some dunning unless and until they declare bankruptcy, so *latent* complaints might be the same order of magnitude as bankruptcy rates. Despite the low rate of

[39] "Lenders" comprises banks, credit unions, and other lenders (finance companies, mortgage lenders, installment lenders, health care lenders, and other lenders.) The FTC does not have a separate field for payday lenders.

[40] The litany of complaints received by the FTC in 2005 (percent of total): exaggerating amount or legal status of debts (43), calling continuously, before eight am, or after nine pm (25), obscene language (12), repeatedly calling family, friends, and neighbors (11), false threats of dire consequences (10), impermissible calls to employer (6), revealing debt to third parties (5), threatened violence (0.4).

complaints, we maintain that *changes* in the rate could still reliably measure changes in debt problems.

We also see complaints against lenders and debt collectors as a potential proxy for "informal bankruptcy," where borrowers are in default with the benefit of bankruptcy protection (Dawsey and Ausubel 2004). Defaulted-but-not-bankrupt debtors are subject to debt collectors' full treatment — dunning, wage garnishment, law suits, etc. — so it seems natural to assume that some debtors in that state will seek protection elsewhere, by complaining to the FTC for example.

Note that model (A) controls for some issues that might compromise the use of complaints as a proxy for informal bankruptcy. For example, Dawsey et al. (2008) show that bankruptcy and informal bankruptcy rates vary predictably with state-level protections against debt collectors and wage garnishment, but the state effects in (A) control for such differences to the extent they are constant within states.[41] While debt collectors' efforts may vary with business conditions, the economic control variables and date effects help account for cyclical effects.

Identification theft (ID theft) might also compromise complaints; some complainants may appeal to the FTC over harassment about debts incurred by an ID thief. If ID theft covaries with payday credit supply, our delta estimates would be biased. We control for state-specific, time-varying differences in ID theft by predicting ID theft by state using annual ID theft rates per capita by state between 2002 and 2008.[42] The prediction model includes all the variables on the right side of (A) including the state-specific trend.[43]

---

[41] Dawsey et al. (2008) find those laws are mostly constant within states.

[42] ID theft rates by state are available from various issues of the Federal Trade Commissions' Consumer Sentinel Network Data Book and Consumer Sentinel Annual Fraud and Identity Theft Reports http://www.ftc.gov/sentinel/reports.shtml

[43] Apart from the fixed effects and state trend, only education was (positively) correlated with ID theft, presumably because education is positively correlated with internet usage.

Before investigating how complaints vary with payday credit access, we digress to show that complaints seem to satisfy necessary conditions to proxy for informal bankruptcy in that they are positively related to defaulted debt and negatively related to Ch. 13. Table 3.3 reports auxiliary regressions to that effect, where defaulted debt is charge-offs of credit card loans by unit banks, i.e., banks operating in a single state. The pattern of coefficients is as predicted; complaints are positively related to credit card charge-offs and negatively related to Ch. 13 rates. In particular, complaints are significantly negatively related to Ch. 13 bankruptcy rates in the standard fixed effect models and they are positively related to credit card chargeoffs in the model that includes a state specific trend. In the latter model, we can reject at below the ten percent level that the coefficients on charge-offs and Ch. 13 are jointly equal to zero. Having provided some evidence that complaints are a proxy for informal bankruptcy, Table 3.4 reports regressions of informal bankruptcy/complaints. To emphasize substitution between bankruptcy and informal bankruptcy, we report corresponding regressions of Ch. 13 bankruptcy rates. Only credit card chargeoffs are significant in explaining informal bankruptcy/complaints in the model with state specific trends, but a number of variables are significant in the standard fixed effects model. In that model, informal bankruptcy/complaints are significantly related to income, income growth, and the home price index. We also observe significantly lower informal bankruptcy/complaints in states with a higher share of black households and lower informal bankruptcy/complaints in states with higher share of Hispanic households. Consistent with the substitution hypothesis, we observe just the opposite relationship between those demographic variables and formal bankruptcy/complaints under Ch. 13.

TABLE 3.3: Auxiliary Regressions: Do Complaints Increase with Credit Card Chargeoffs and Decrease in Ch. 13 Filing Rates?

Reported are OLS regression coefficients [robust t-statistics] estimated using state-monthly observations between January 1998 and December 2008. *Banned* equals one (zero) after (before) state banned payday lending. *Enabled* equals one (zero) after (before) state passed enabling legislation. Predicted ID theft is forecasted using state-year data over 2002-2008. All models include state and date fixed effects. Standard errors are clustered by state. Bottom line reports p-value of F-test whether Ch. coefficients on 13 filings and credit card chargeoffs are jointly zero.

	Dependent Variable (mean): Complaints against Lenders & Collectors per 100,000 (1.41)					
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Credit card net chargeoff rate	0.69	0.57	0.69	0.70	0.60	0.72
	[0.87]	[1.06]	[2.29]**	[0.90]	[1.13]	[2.41]**
Ch13 filings per 10000 persons	-0.27	-0.12	-0.03	-0.26	-0.11	-0.02
	[3.45]***	[1.81]*	[0.76]	[3.25]***	[1.60]	[0.45]
Ch 7 filings per 10000 persons	0.01	0.02	0.02			
	[0.22]	[1.13]	[1.43]			
Predicted ID theft		0.01	-1.73		0.01	-1.70
		[2.42]**	[1.35]		[2.46]**	[1.32]
Unemployment rate		0.00	-0.14		0.00	-0.14
		[0.13]	[1.46]		[0.02]	[1.40]
Log (income)		-1.43	1.24		-1.50	1.20
		[2.16]**	[1.02]		[2.24]**	[0.98]
Income growth		0.68	0.53		0.73	0.55
		[1.78]*	[1.43]		[1.90]*	[1.46]
Home Prices		0.00	-0.05		0.00	-0.05
		[2.62]**	[1.35]		[2.56]**	[1.32]
Black share		-21.71	773.36		-21.96	760.13
		[4.20]***	[1.33]		[4.24]***	[1.30]

TABLE 3.3 (Continued)

	Dependent Variable (mean): Complaints against Lenders & Collectors per 100,000 (1.41)					
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Hispanic share		9.67	349.91		9.62	344.08
		[2.50]**	[1.37]		[2.49]**	[1.33]
Asian share		5.46	-1335.10		5.65	-1313.28
		[2.31]**	[1.35]		[2.39]**	[1.32]
College share		-0.04	99.35		-0.03	97.73
		[0.03]	[1.36]		[0.02]	[1.32]
Constant	0.45	15.11	-276.37	0.46	15.81	-271.78
	[5.71]***	[2.18]**	[1.37]	[6.16]***	[2.27]**	[1.33]
State specific trend?	No	No	Yes	No	No	Yes
Observations	5877	5868	5868	5877	5868	5868
Adjusted R-squared	0.85	0.87	0.89	0.85	0.87	0.89
F-test p-value		0.149	0.065		0.187	0.058

* significant at 10%; ** significant at 5%, *** significant at 1%

TABLE 3.4: Payday Credit Supply, Bankruptcy, and Informal Bankruptcy/Complaints

Reported are OLS regression coefficients [robust t-statistics] estimated between January 1998 and December 2008. *Banned* equals one (zero) after (before) state *s* banned payday lending. *Enabled* equals one (zero) after (before) state *s* passed enabling legislation. Predicted ID theft is forecast using state-year data over 2002-2008. All models include state and date fixed effects. Standard errors are clustered by state.

	Complaints vs. Lenders & Debt Collectors per 100,000 (1.41)			**Ch.13 Filings per 10,000 (2.95)**			**Log (Complaints vs. Lenders & Debt Collectors / (Complaints vs. Lenders & Debt Collectors + Ch.13 Filings) (0.73)**		
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**
Banned	0.27	0.26	0.03	-1.02	-0.92	-0.47	0.09	0.08	0.073
	[1.82]*	[2.36]**	[0.58]	[2.02]**	[1.76]*	[1.30]	[0.97]	[0.94]	[2.13]**
Enabled	0.04	0.04	0.04	0.06	0.00	0.01	-0.06	-0.08	0.042
	[0.44]	[0.68]	[0.94]	[0.31]	[0.01]	[0.03]	[0.54]	[0.91]	[0.60]
Credit card net charge-off rate		0.69	0.72		-1.47	-1.23		0.43	0.609
		[1.42]	[2.47]**		[1.02]	[1.07]		[1.12]	[2.08]**
Predicted ID theft		0.01	-1.72		-0.01	-0.04		0.01	-2.025
		[2.83]***	[1.33]		[1.47]	[1.48]		[4.56]***	[2.26]**
Unemployment rate		0.00	-0.14		0.25	0.19		-0.03	-0.175
		[0.14]	[1.41]		[2.24]**	[1.86]*		[1.02]	[2.53]**
Log (income)		-1.48	1.22		1.89	0.88		-1.22	1.465
		[2.43]**	[1.00]		[1.03]	[0.65]		[2.05]**	[1.66]
Income growth		0.76	0.55		-2.01	-1.51		0.70	0.689
		[2.14]**	[1.47]		[1.90]*	[1.77]*		[2.26]**	[2.27]**
Home Prices		0.00	-0.05		0.00	-0.01		0.00	-0.050
		[2.84]***	[1.34]		[1.55]	[2.81]***		[1.56]	[2.16]**

TABLE 3.4 (Continued)

	Dependent Variable (mean): Complaints vs. Lenders & Debt Collectors per 100,000 (1.41)			Ch.13 Filings per 10,000 (2.95)			Log (Complaints vs. Lenders & Debt Collectors / (Complaints vs. Lenders & Debt Collectors + Ch.13 Filings) (0.73)		
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**
Black share		-22.55	769.88		21.27	22.54		-5.76	904.406
		[4.79]***	[1.31]		[2.67]**	[0.67]		[1.47]	[2.22]**
Hispanic share		10.13	348.05		-16.31	7.57		1.39	403.948
		[2.63]**	[1.34]		[2.77]***	[0.51]		[0.50]	[2.24]**
Asian share		6.03	-1329.10		-20.99	-91.32		6.75	-1561.657
		[2.71]***	[1.33]		[1.53]	[1.47]		[4.28]***	[2.26]**
College share		-0.56	98.80		4.20	7.98		-0.87	115.156
		[0.46]	[1.34]		[1.58]	[2.68]**		[0.98]	[2.25]**
Constant	0.40	15.73	-274.85	0.56	-20.33	-13.96	-2.61	9.94	-321.287
	[5.57]***	[2.47]**	[1.34]	[5.13]***	[1.05]	[1.20]	[28.02]***	[1.59]	[2.27]**
State specific trend?	No	No	Yes	No	No	Yes	No	No	Yes
Observations	5934	5868	5868	2244	2225	2225	5934	5868	5868
Adjusted R-squared	0.85	0.87	0.89	0.93	0.94	0.96	0.92	0.93	0.95

* significant at 10%; ** significant at 5%, *** significant at 1%

Turning to δ and δ', we observe that informal bankruptcy/complaints are positively related to *Ban* in the standard fixed effects models. Consistent with the substitution hypothesis, the models where δ is significantly negative in the informal bankruptcy/complaints regressions correspond precisely to the models where δ is significantly positive in the Ch. 13 bankruptcy regressions. The estimates in model (2) of Table 3.4 imply that complaints against lenders and debt collectors rise by 19 percent relative to average (1.41). That figure seems plausible, though we have no real metric to gauge plausibility. Perhaps the most notable aspect of Table 3.4 is the absence of evidence that complaints against lenders and debt collectors increase with payday credit supply. If anything, the relationship is negative.

The outcome in the final set of regressions in Table 3.4 is the log ratio of informal bankruptcy to Ch. 13 plus informal bankruptcy.[44] That ratio directly measures substitution between informal and formal bankruptcy. Neither δ nor δ' is significant in the standard, fixed effects model, but δ is positive and significant at the five percent level in the model with state specific trends. The estimate in model (9) implies informal bankruptcy as a proportion of total bankruptcy increases by ten percent after payday loan bans.

Overall, the bankruptcy findings suggest that payday loan bans have weakly significant and opposing effects on formal and informal bankruptcy rates/complaints, with the former tending to rise and the latter tending to fall after bans. The welfare implications of that substitution are ambiguous except that it happens under a constraint on payday loan supply. At a minimum, it does suggest a possible alternative interpretation of Skiba and Tobacmans' (2008a) finding that payday credit access causes or "tips" marginal loan applicants into bankruptcy. Marginal payday loan applicants (who barely qualify for loans) may be on the verge of bankruptcy in

[44] We omit Ch. 7 as we find no evidence of substitution between Ch. 7 and informal bankruptcy

the first place and thus face intense pressure from debt collectors. We can only speculate on the transmission, but perhaps the extra credit from payday lenders affords them the opportunity to buy bankruptcy protection.[45]

4.2 Overdrafts

Avoiding overdrafts is a common theme among payday credit users. In his study of the Oregon payday ban, Zinman (2010) found that payday credit users there expected to bounce more checks after the ban. In a survey of 2000 payday credit users, Cerillo (2004) found that 66 percent reported demanding payday loans to avoid bouncing checks. Morgan and Strain (2008) and Melzer and Morgan (2010) find that returned check rates decline after payday bans. We extend their findings by looking at more law changes, including enabling legislation, and by looking at the counterpart to returned checks, namely overdraft fee income at depository institutions.

4.2.a Returned Checks

The returned checks data are from Federal Reserve Check Processing Centers (Fed CPC). Although checks are declining as a medium of payment, the Fed was and is a major player in check processing. For example, the 43 CPC operating in 2003 cleared about 38 percent of the estimated 36.6 billion checks drawn on U.S. banks and credit unions that year.[46]

Checks are observed at the CPC level so the returned checks regression model is

$$R_{cst} = \delta Ban_{st} + \delta' Enabled_{st} + a + a_c + a_t + X_{st}\,\gamma + Z_{dt}\varkappa + \varepsilon_{cst,} \qquad \text{(B)}$$

[45]The \$274 fee for filing Ch. 13 filing is about the size of the typical payday loan. http://www.wiwb.uscourts.gov/ch13_filing_req.htm

[46] http://www.federalreserve.gov/boarddocs/press/other/2004/20040802/attachment2.pdf.

where R_{cst} denotes the number or dollar amount of returned checks at CPC c in s at t. A CPC can process checks drawn on depository institutions from other states in the Fed district it serves so we include economic control variables measured at the Fed District level (Z_{dt}) as well.[47] Standard errors are clustered at the CPC level.

The fact that a CPC may process checks drawn on depository institutions in other states creates error in the dependent variable. Under the assumption that changes in payday loan laws in state s do not affect returned checks in other states, our estimates of δ and δ' are unbiased. Their t-statistics are biased downward, however, so we are less likely to reject $\delta = 0$ and $\delta' = 0$ than if "pure" state level check data were available.[48]

In response to decreased aggregate demand for checks, the Fed began merging CPCs in response in 2004. In cases where the mergers involved a CPC in a state where payday loan laws changed, we adjust the checks data and the right-hand side variables in model (B) following a procedure intended to minimize potential attenuation bias created by mergers. The procedure depends on whether the legal change occurred before or after the CPC merger, and whether the CPC in the state where the law changed was the "target" or "acquirer" in the merger. We treat both the dependent variable and right-hand side variable differently in each case. In cases where the legal change preceded the merger and the acquiring CPC was located in the state where the law changed, we follow the bank merger literature and create pro forma series by adding the returned checks at the merging CPC at time zero. In all other cases, we did not create pro forma series as that would add unnecessary noise to

[47] Fed District level economic data are from the Federal Reserve Bank of St. Louis "Fred" dataset.
[48] To see where aggregation does and does not create bias, decompose the number or value of returned checks at CPC c into checks drawn on state s and checks drawn on another state $\sim s$: $R_c = R_s + R_{\sim s}$. We wish to estimate $R_s = \alpha + \delta B_s + \delta' E_s + \varepsilon_s$, where B denotes *Ban* and E denotes *Enabled.* Plugging the first equation into the second yields the estimating equation: $R_c = \alpha + \delta B_s + \delta' E_s + \varepsilon_s + R_{\sim s}$. Our estimates of δ and δ' are consistent so long as $\text{cov}(R_{\sim s}, B_s) = 0 = \text{cov}(R_{\sim s}, E_s)$. The t-statistics will be biased downward, however, because $var(\varepsilon_s) + var(R_{\sim s}) > var(\varepsilon_s)$.

the dependent variable. Instead we use a dummy variable to account for any shift in the mean in returned checks after the merger. In all cases, we created weighted values of the right hand side variables where the weights were the respective share of checks processed at the merging CPC.[49]

Returned checks are measured per number or per dollar amount. We predict the former will be more correlated with payday credit supply because depositors who use payday credit are likely apt to have lower income than depositors who do not and thus write (and possibly bounce) smaller checks.

Table 3.5 reports the returned check regressions. Several of the economic controls are significant in some models but we do no not discuss those in detail. We observe one demographic effect: the number and dollar amount of returned checks are higher at CPCs located in states with more blacks in the standard fixed effects model. That result is consistent with Campbell et al. (2008) who find that involuntary deposit closures are higher in counties with proportionately more black households.

Turning to δ and δ', we see no evidence that increased payday supply is associated with more returned checks. On the contrary, δ is positive in all models and δ' is negative. Furthermore, one or the other coefficients is significant at the five percent or one percent level in the regression of the number of returned checks (the outcome arguably more closely associated with small dollar check writers), even in

[49] We have confirmed the returned check results using logs of returned checks and the rate of returns per check process and per $100 of checks processed. We have also confirmed the results using pro forma values for all the CPC located in states that experienced a law change where we use control variables for the state with the "acquiring" CPC or if we create population shares of the states where the merging CPCs were located.

TABLE 3.5: Fewer Bounced Checks when Payday Loan Supply Increases

Reported are OLS regression coefficients [robust t-statistics] estimated between 1998:Q1 and 2008:Q3. *Banned* equals one (zero) after (before) state *s* banned payday lending. *Enabled* equals one (zero) after (before) state passed enabling legislation. All models include CPC and date fixed effects. Standard errors are clustered by CPC.

	Dependent Variable (mean):					
	Thousands of returned checks (1335)			$Millions of returned checks (1238)		
	(1)	(2)	(3)	(4)	(5)	(6)
Banned	301.81	272.43	360.32	238.70	105.75	115.64
	[1.05]	[1.63]	[3.49]***	[0.61]	[0.66]	[0.46]
Enabled	-191.42	-188.01	-150.11	-166.55	-102.17	-48.29
	[2.46]**	[2.84]***	[1.06]	[2.20]**	[1.70]*	[0.56]
Unemployment rate		41.05	-23.78		-6.69	-29.28
		[0.84]	[0.78]		[0.12]	[0.71]
Log (income)		3037.07	763.87		2490.26	1060.24
		[2.07]**	[1.08]		[1.47]	[1.04]
Income growth		-1750.01	-520.91		-1239.87	-505.47
		[2.07]**	[1.17]		[1.35]	[0.84]
Home Prices		-2.27	1.61		-1.15	2.23
		[2.31]**	[0.83]		[1.16]	[1.02]
Black share		20304.63	8285.89		38291.11	7264.80
		[2.12]**	[1.16]		[2.71]***	[0.84]
Hispanic share		5612.53	8036.95		13519.08	12982.47
		[1.01]	[1.25]		[2.46]**	[1.70]*
Asian share		-4924.94	-2772.50		15742.24	8311.83
		[0.24]	[0.18]		[0.68]	[0.42]
College share		209.32	-49.41		246.61	130.01
		[0.18]	[0.05]		[0.15]	[0.11]

TABLE 3.5 (Continued)

	Dependent Variable (mean): Thousands of returned checks (1335)			Dependent Variable (mean): $Millions of returned checks (1238)		
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
District log income		-6859.97	-4024.19		-3227.85	-1087.32
		[2.06]**	[1.15]		[0.97]	[0.31]
District Income growth		1003.20	1820.96		-1605.56	-107.93
		[0.40]	[0.53]		[0.49]	[0.03]
District unemployment		-57.70	40.63		-112.24	30.88
		[0.84]	[0.83]		[1.74]*	[0.76]
Constant	2288.06	35065.85	37734.72	1739.70	-2386.25	7230.84
	[21.65]***	[1.07]	[1.07]	[11.60]***	[0.07]	[0.21]
State specific trend?	No	No	Yes	No	No	Yes
Observations	1550	1550	1550	1550	1550	1550
Adjusted R-squared	0.83	0.92	0.95	0.83	0.94	0.97

* significant at 10%; ** significant at 5%; *** significant at 1%

the model with state specific trends. Note that, as predicted, δ and δ' are larger in absolute value and more significant in the regressions of the number of returned checks than in the regressions of the dollar amount of returned checks. That differential effect provides some evidence that the link between returned checks and payday credit supply is causal.

The implied magnitudes are substantial. The estimate of δ' in model (2) implies the number of returned checks increases by 14 percent relative to average after the passage of enabling legislation. The dollar magnitudes are also large; the number of returned checks falls by 188,000 per quarter after states pass enabling legislation. At $50 per returned check ($25 to the bank and $25 to the merchant), that implies a savings to households of $37.6 million per year per CPC in returned check fees.[50]

The returned checks regressions indicate households bounce fewer checks when payday credit is available. Below we look at overdrafts from the other side.

4.2.b Fee Income at Banks

Every overdraft, whether covered or not, generates revenue for the counterparty depository institution. Overdraft fees have become an important source of revenue for depository institutions. The median bank studied in FDIC (2008) earned 43 percent of its noninterest income and 21 percent of its net operating income from overdraft fees.[51]

We measure overdraft fee income by the "Services Charges on Deposit Accounts" item on the Call Reports banks file with their federal regulators. Unlike with checks, fee income is observed at the state level because we limit the sample to unit banks which operate in a single state. There is error in the variable nonetheless,

---

[50] 188000*4*$25
[51] FDIC (2008) Table VIII-2, p. 58. Data on the costs of providing overdraft credit is not available, so the revenue figures overstate the importance of overdraft profits relative to net income.

because the services charges item measures fee income from sources other than overdraft fee income. Under the assumption that income from those other components is uncorrelated with the payday loan supply, our estimates of delta will be unbiased but their statistical significance will be biased downward.[52]

The mean of service charges on deposit accounts over 1998 to 2009:Q1 at the 1917 unit banks in our sample was \$145 million. The mean fee per resident was \$31. The dependent variables of our regressions are the log of those variables: log of fee income or log of fee income per capita.[53]

Table 3.6 reports the fee income regressions. Fee income is almost entirely orthogonal to the economic control variables, but several demographic variables are significant in some models. In the standard fixed effects model with controls, fee income is lower in states with a high share of college educated residents, presumably because better educated households are better at avoiding overdrafts or contesting the fees. In those models, fee income also tends to be lower in states with a higher share of Asian households. In the fixed effects models with state-specific trends, fee income is higher in states with a higher share of Hispanic residents.

As with some earlier results, the estimates of δ and δ' depend on the model. In the model with state specific trends, none of the previous relationships just noted hold up, not even the sensible relationship between *College Share* and fee income. In those models δ and δ' are both insignificant. By contrast, in the fully controlled standard fixed effects model without a trend, δ and δ' have the opposite sign, as predicted, and both are significant at the ten percent level. The estimate of δ in model (5) implies the

[52] The logic for this claim follows from footnote 14. Simply imagine *R* denotes Revenue and ~s denotes fee income from other (non overdraft) sources.
[53] "Unlogged" fee income and income per capita were unpredictable by model (B).

TABLE 3.6: Fee Income at Bank Falls with Payday Loan Supply

Reported are OLS regression coefficients [robust t-statistics] estimated between 1998:Q1 and 2008:Q2. *Banned* equals one (zero) after (before) state *s* banned payday lending. *Enabled* equals one (zero) after (before) state passed enabling legislation. All models include state and date fixed effects. Standard errors are clustered by state.

	Dependent Variable (mean):					
	Log of Fee Income (10.7)			Log of Fee Income per Capita (2.56)		
	(1)	(2)	(3)	(4)	(5)	(6)
Banned	0.24	0.30	0.18	0.22	0.30	0.17
	[1.30]	[1.77]*	[1.44]	[1.26]	[1.75]*	[1.39]
Enabled	-0.30	-0.34	0.18	-0.29	-0.33	0.18
	[1.58]	[1.75]*	[1.05]	[1.51]	[1.70]*	[1.04]
Unemployment rate		0.09	0.06		0.10	0.07
		[1.16]	[1.00]		[1.25]	[1.01]
Log (income)		0.81	0.11		0.92	0.05
		[0.31]	[0.07]		[0.35]	[0.03]
Income growth		-0.28	0.15		-0.41	0.15
		[0.12]	[0.10]		[0.18]	[0.10]
Home Prices		0.00	0.00		0.00	0.00
		[1.80]*	[1.07]		[1.81]*	[1.13]
Black share		5.36	18.10		4.98	16.82
		[0.70]	[0.70]		[0.66]	[0.65]
Hispanic share		-3.13	20.73		-5.78	20.30
		[0.33]	[1.76]*		[0.62]	[1.72]*
Asian share		-2.76	-1.57		-2.67	-1.52
		[2.12]**	[1.05]		[2.04]**	[1.01]
College share		-3.45	-2.10		-3.50	-2.11
		[1.90]*	[1.30]		[1.95]*	[1.31]
Constant	8.25	0.51	13.73	1.81	-6.96	8.22
	[92.72]***	[0.02]	[0.96]	[19.73]***	[0.26]	[0.57]
State specific trend?	No	No	Yes	No	No	Yes
Observations	2113	2113	2113	2113	2113	2113
Adjusted R-squared	0.88	0.89	0.94	0.78	0.79	0.89

* significant at 10%; ** significant at 5%; *** significant at 1%

log of fee income increases about 12 percent relative to average. The corresponding estimate of δ' implies the log of fee income falls by 13 percent relative to average after the passage of enabling legislation, so the effects of bans and enabling legislation are roughly symmetric.

As we see them, the fee income results provide weak evidence confirming the more robust findings on actual overdrafts. They provide no evidence that payday credit access leads to significantly higher overdraft fee income for banks.

V. Falsification Tests, Robustness, and Potential Biases

This section reports falsification tests, shows that the results are robust to the exclusion of states with ambiguous payday loan laws, and discusses potential biases.

5.1 Falsification Tests

By construction, the difference-in-difference regressions we estimate are intended to identify causal connections between outcomes and payday loan supply. As further evidence, Table 3.7 reports falsification tests that indicate whether *Ban* and/or *Enabled* are correlated with outcomes that one would *not* expect to vary with payday loan supply.[54] To save space, we report only δ and δ' even though the regressions include the full set of controls. The falsification results indicate that *Ban* and *Enabled* are uncorrelated with home prices, the unemployment rate, and

[54] We thank a referee for suggesting falsification tests.

TABLE 3.7: Falsification Tests

Reported are coefficients on *Banned* and *Enabled* in regression of outcome indicated in column heading. First regression for each outcome includes state and date fixed effect. Second includes state specific trend and fixed effects. All regressions include controls reported in main regressions excluding the one used as the outcome variable. Row heading indicates sample and/or re-coding.

	Dependent Variable											
	Home prices		Unemployment		Log (income)		Black share		Hispanic share		Asian share	
	fe	fe + trend	fe	fe + trend	fe	fe + trend	fe	fe + trend	fe	fe + trend	fe	fe + trend
Banned	-12.90	-4.58	-0.38	-0.05	-0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	[0.63]	[0.47]	[1.64]	[0.23]	[1.30]	[0.14]	[0.44]	[0.90]	[0.22]	[1.23]	[0.93]	[0.43]
Enabled	-11.67	-11.29	0.12	-0.25	0.01	-0.01	0.00	0.00	0.00	0.00	-0.01	-0.01
	[0.58]	[1.43]	[0.82]	[1.25]	[0.97]	[0.65]	[0.51]	[0.75]	[0.23]	[0.60]	[1.11]	[1.23]

* significant at 10%; ** significant at 5%, *** significant at 1%

demographic shares. These falsification tests provide further evidence that the relationships between payday loan supply and outcomes identified above are not merely coincidental.

5.2 Robustness to excluding ambiguous states

There is, or was at times, some ambiguity in the status of payday lending in Alabama (AL), Arkansas (AR), and Oklahoma (OK).[55] Table 3.8 shows that the main results are mostly unchanged when those states are excluded. We still observe the weak negative link between Ch. 13 and *Banned* as in the main results, but with the three ambiguous states excluded, this link is not quite significant at the ten percent level. In the model of Ch. 13 with the state specific trend, we obtain a new result suggesting that Ch. 13 rates *fall* significantly when payday lending is enabled. While that new result may go against the informal-formal bankruptcy substitution hypothesis, it does not suggest that payday credit access is tipping households into bankruptcy. More generally, the somewhat unstable Ch. 13 results are consistent with the mixed findings in the literature.

The results on the number of returned checks with AL, AR, and OK excluded change slightly, but not in a uniform direction. The absolute size and significance of δ' falls in the model with fixed effects model, but estimate of δ in the model with state-specific trends becomes larger and more significant.

Excluding the ambiguous states strengthens the results on fee income per capita. In the main results, δ and δ' were only significant at the ten percent level. With AL, AR, and OK excluded, δ remains positive at the ten percent level while δ' is significant at the five percent level.

[55] See Fox and Mierzwinski (2001) for a discussion of ambiguity about payday loan status in Oklahoma and Carrell and Zinman (2009) for a discussion of ambiguity in Alabama and Arkansas. We thank a referee for bringing these ambiguous states to our attention.

TABLE 3.8: Robustness Checks

Reported are coefficients on *Banned* and *Enabled* in regression of outcome indicated in column heading [robust t-statistics]. First regression for each outcome includes state and date fixed effect. Second includes state specific trend and fixed effects. All regressions include controls reported in main regressions. Row heading indicates sample and/or re-coding.

		Ch. 13 per 1000 residents		**Complaints v lenders and debt collectors**		**Number of returned checks**		**log (fee income) at unit banks**		**log (fee income per capita) at unit banks**	
		fe	fe + trend	fe	fe + trend	fe	fe + trend	fe	fe + trend	fe	fe + trend
Main Results	Banned	-0.93	-0.48	0.26	0.03	272.4	360.3	0.30	0.18	0.30	0.17
		[1.71]*	[1.31]	[2.36]**	[0.58]	[1.63]	[3.49]***	[1.77]*	[1.44]	[1.75]*	[1.39]
	Enabled	0.01	-0.07	0.04	0.04	-188.0	-150.1	-0.34	0.18	-0.33	0.18
		[0.05]	[0.22]	[0.68]	[0.94]	[2.84]***	[1.06]	[1.75]*	[1.05]	[1.70]*	[1.04]
Excluding AL, AR, OK	Banned	-0.93	-0.39	0.28	0.00	271.4	356.9	0.32	0.19	0.31	0.18
		[1.62]	[0.97]	[2.49]**	[0.04]	[1.59]	[3.50]***	[1.79]*	[1.43]	[1.76]*	[1.38]
	Enabled	0.10	-0.47	0.01	0.03	-121.9	-34.8	-0.50	0.19	-0.50	0.19
		[0.72]	[2.25]**	[0.13]	[0.78]	[1.68]*	[0.60]	[2.21]**	[0.79]	[2.15]**	[0.79]
MD Recoded June 2002	Banned	-0.98	-0.59	0.25	0.02	275.7	305.0	0.24	0.12	0.23	0.11
		[1.92]*	[1.71]*	[2.34]**	[0.29]	[1.59]	[2.23]**	[1.18]	[0.84]	[1.16]	[0.80]
	Enabled	0.00	-0.07	0.05	0.00	-188.2	-146.6	-0.34	0.17	-0.34	0.17
		[0.02]	[0.25]	[0.79]	[0.12]	[2.85]***	[1.04]	[1.77]*	[1.02]	[1.72]*	[1.01]
Excluding Indiana	Banned	-0.91	-0.48	0.27	0.03	274.3	363.2	0.29	0.18	0.28	0.17
		[1.71]*	[1.33]	[2.42]**	[0.59]	[1.62]	[3.54]***	[1.69]*	[1.44]	[1.67]	[1.38]
	Enabled	0.03	-0.07	0.05	0.04	-192.0	-148.1	-0.35	0.18	-0.35	0.18
		[0.17]	[0.22]	[0.72]	[0.93]	[2.86]***	[1.05]	[1.79]*	[1.05]	[1.74]*	[1.04]

* significant at 10%; ** significant at 5%, *** significant at 1%

5.3 Potential Biases

As already noted, we follow the literature in taking the changes in payday loan laws as exogenous. If that assumption is violated, δ and δ' will be biased. For example, if payday lenders see an exogenous increase in household debt problems in a state, say increased overdrafts, as an opportunity to enter a state, they seem likely to lobby for enabling legislation. That would impart an upward bias on δ'. Note however, that their competitors in the overdraft credit market, depository institutions, would tend to lobby against enabling legislation (or for bans), so the two forces would tend to offset. *A priori,* the net bias seems ambiguous.[56]

Our estimates of δ and δ' will also be biased if we have excluded any state-specific, time-varying variables that are correlated with both the outcomes and the legal changes. We have included the most likely covariates, namely the economic and demographic variables, and we have confirmed our results when we control for the share of population that are female or in the military.[57] However, there may be other covariates that we have overlooked.

VI. Conclusion

Despite a dozen studies, the question of how payday credit affects its users remains unanswered. Economists do not even agree on the sign of the effect, much less the transmission. Our findings obviously do not settle the debate, but they do

[56] The interest of consumers and consumer advocates will also tend to offset to the extent they see payday credit differently. Consumers who, rightly or wrongly, see payday credit as palliative for their debt problems will tend to resist bans and lobby for enabling legislation. Consumer advocates, who tend to view payday credit as exacerbating users problems, will tend to support banks and lobby against enabling legislation.

[57] Neither the share of population that is female of the share in the military were correlated with the outcomes.

contribute to it by illuminating how households rearrange their financial affairs when payday loan supply changes.

We find some evidence that after payday credit bans expand households switch from formal bankruptcy, where they are protected to from debt collectors, to informal bankruptcy, where they are exposed. The welfare implications of that substitution are ambiguous except that it results from a tightened credit constraint. We should qualify by noting that none of our evidence about bankruptcy, either the decrease in Ch. 13 associated with payday credit bans or the substitution toward informal bankruptcy is particularly strong so further study of the formal-informal bankruptcy substitution hypothesis is clearly called for. Perhaps the more important point to note about the complaints we use to proxy for informal bankruptcy is that we find no evidence that complaints against lenders and debt collectors increase with payday credit supply. If anything, the relationship is negative.

We find stronger evidence that overdrafts decline when payday credit supply expands. That suggests households substitute the payday variety of overdraft credit for the bank version when the former is available. Small overdrafts are more expensive than payday loans, so that finding suggests payday credit bans force households to use costlier alternatives. In hindsight our overdraft findings should not come as too much of a surprise as avoiding overdrafts is a common theme among payday credit users (e.g. Zinman 2010, Cerillo 2004).

Some caveats are called for. First, none of our outcomes, except perhaps complaints, measure actual welfare and all of them, except perhaps bankruptcy, are measured with error. We stress that the measurement error in the returned checks and fee income data attenuates the statistical significance of our findings for those variables without biasing the estimates but bias in some other direction for some other reason is always a concern with data generated in the market instead of in a laboratory or in the field. Second, although the state fixed effect in our models control for constant differences, the possibility remains that we have omitted some time-varying variable that is correlated with payday loan supply and the outcomes we study. If so, our estimates of the effects of payday credit supply on overdrafts and informal bankruptcy will be biased.

Given those caveats, we avoid drawing welfare or policy conclusions except to observe that lawmakers seem to be ahead of economists on the welfare effects of payday credit access; some lawmakers are closing payday loan markets as though they know for certain payday credit lowers welfare. To help economists catch up, we end with some suggestions for additional research. First, a straightforward test of whether payday credit helps users smooth their consumption would be to test if purchases of certain items get postponed when payday credit becomes unavailable. As a first approximation, that could be tested by comparing same store sales of items payday credit users are prone to demand before and after payday credit bans. Second, not much attention has been paid to the payday credit enforcement mechanism, that is, how lenders get borrowers to repay. If the primary enforcement mechanism is the refusal of future credit, then repeat borrowing that looks so suspicious to some observers may be inevitable; we would never observe equilibria with payday credit where everyone borrows just once. Lastly, given our evidence that fee income at depository institutions falls when payday credit supply increases, the political economy of payday lending regulations seems ripe for study. Are the usury limits that bind payday lenders the type that reflect public interest and raise household welfare (Glaeser and Scheinkman 1998) or the kind that reflect competing interests (Benmelich and Moskowitz 2009; and Krozner and Strahan)?

ACKNOWLEDGMENTS

The authors thank Bob DeYoung (the editor) and two anonymous referees for helpful comments. Diego Aragon, Matt Botsch, Peter Hull, Ben Mandel, Lev Menand, and Anna Peterson provided excellent assistance. The views expressed in this paper are those of the authors and do not necessarily reflect the position of Cornell University, the Federal Reserve Bank of New York, or the Federal Reserve System.

APPENDIX: CODING PAYDAY LOAN BANS

Seven states and D.C. banned payday loans over our sample period. Georgia (GA) declared payday lending a felony in May 2004 (O.C.G.A. § 16-17-1). North Carolina (NC) closed its market in December 2005 after a series of law suits by NC Attorney General persuaded payday lenders to cease operation under the bank agency model.[58] West Virginia (WV) tried to ban payday lending *via* deferred presentment under statute W. Va. Code § 32A-3-1(passed in 1998) and a usury limit (W. Va. Code § 47-6-5b), but at least one firm, First American Cash Advance, continued operating under the bank agency model until June, 2006.[59] Maryland (MD) banned payday lending through restrictions on fees charged by check cashers (MD Financial Institutions Code § 12-120) and small loan interest rates (MD Commercial Law Code § 12-306) effective in 2000 and finally passed anti-loan brokering legislation (MD Commercial Law Code § 14-1902), effective June, 2002 to eliminate the agency payday lending model. Oregon closed its payday credit market by vigorously enforcing a 36 percent usury cap in July 2007 (Zinman 2010). D.C. prohibited payday lending in November, 2007, by limiting fees on check cashing and prohibiting post-dated check cashing (D.C. Code § 26-317 and 26-319). Despite a cap on small loan interest rates in Pennsylvania (P.A. 7 P.S. § 6201-6219), payday lenders were able to operate there via the bank agency model and a law that sanctioned loan brokering (P.A. 73 P.S. § 2181-2192). Store numbers began falling after the FDIC restricted bank-payday lender affiliations in 2006.[60] However, Advance America, the

[58] "Payday lending on the way out in NC." North Carolina Department of Justice Press release (March 1, 2006) http://www.ncdoj.gov/News-and-Alerts/News-Releases-and-Advisories/Press-Releases/Payday-lending-on-the-way-out-in-NC.aspx. Retrieved 2009-08-27

[59] "Attorney General McGraw Reaches Agreement with WV's Last Payday Lender, First American Cash Advance." West Virginia Office of the Attorney General press release (May 9, 2007).http://www.wvago.gov/press.cfm?ID=337&fx=more.Retrieved 2009-09-01.

[60] Sabatini, Patricia. 2006. "Days May Be Numbered for State's Payday Lenders." *Pittsburgh Post-Gazette* March 26.

largest national payday lender, did not stop lending and close its Pennsylvania stores until December, 2007.[61] Arkansas (AK) is a difficult state to code because a number of court rulings have affected the supply of payday lending there. Oklahoma and Alabama are also problematic states to code (see Fox and Mierzwinski 2001 and Carrell and Zinman 2008). In the robustness section, we confirm our results with Alabama, Arkansas, and Oklahoma excluded.

[61] "Advance America Announces Decision to Close 66 Remaining Centers in Pennsylvania." December 18, 2007 Press release http://investors.advanceamerica.net/releasedetail.cfm?ReleaseID=282418.Retrieved 2009-09-01.

REFERENCES

Agarwal, S., C. Liu, and L. Mielnicki, 2003, "Exemption Laws and Consumer Delinquency and Bankruptcy Behavior: An Empirical Analysis of Credit Card Data", *Quarterly Review of Economics and Finance*, Vol. 43(2), Pp. 273-289.

Benmelech, E. and T. Moskowitz, 2009. "The Political Economy of Financial Regulation: Evidence from U.S. State Usury Laws in the 19th Century." *The Journal of Finance,* Vol. 65, 3, p. 1029-1073.

Campbell, D., Martinez-Jerez, F., and Tufano, P. 2008. "Bouncing Out of the Banking System: An Empirical Analysis of Involuntary Bank Account Closures." working paper.

Carrell, S. and Z. Zinman. 2009. "In Harm's Way? Payday Loan Access and Military Personnel Performance," working paper.

Cerillo, P. 2004. "Payday Advance Customer Satisfaction Survey," Cypress Research.

Commission. 2006. "Federal Trade Commission Annual Report to Congress on Fair Debt Collection Practices Act."

Glaeser, Edward L., and Jose Scheinkman, 1998. "Neither a Borrower Nor a lender Be: An Economic Analysis of Interest Restrictions and Usury laws," J*ournal of Law and Economics*. 1-35.

Damar .2009. "Why do Payday Lenders Enter Local Markets: Evidence from Oregon, *Review of Industrial Organization,* vol. 34, issue 2: 173-191.

Dawsey, Amanda and Lawrence Ausubel. 2004. "Informal Bankruptcy," working paper.

Dawsey, A., Hynes R., and L. Ausubel. 2008. "The Regulation of Non-Judicial Debt Collection and the Consumer's Choice Among Repayment, Bankruptcy and Informal Bankruptcy," working paper.

DeYoung, R and Phillips R. 2009. "Payday Loan Pricing." working paper

Elliehausen, G and Lawrence, E.C. 2001. "Payday Advance Credit in America: An Analysis of Consumer Demand," Monograph No. 35. Georgetown University, Consumer Credit Research Center.

FDIC. 2008. "Study of Bank Overdraft Programs."

Fox J. and E. Mierzwinski. 2001. “Rent-a-Bank Payday Lending: How Banks Help Payday Lenders Evade State Consumer Protections,” Consumer Federation of America and U. S. PIRG

Hynes, R. 2010, “Does Payday Lending Catch Vulnerable Consumers in a Debt Trap,” working paper.

Lawrence, E. and Elliehausen, G. 2008. “A Comparative Analysis of Payday Loan Customers,” *Contemporary Economic Policy.*

Melzer, Brian T. 2009. The Real Costs of Credit Access: Evidence from the Payday Lending Market. working paper.

___________ and Morgan, D. 2009. “Competition and Adverse Selection in a Consumer Loan Market: The Curious Case of Overdraft vs. Payday Credit,” working paper

Morgan, D. 2007. “Defining and Detecting Predatory Lending,” working paper.

Morgan, D. and Strain, M. 2008. “Payday Holiday: How Households Fare After Payday Credit Bans,” working paper.

Morse, Adair. 2009. “Payday Lenders: Heroes or Villains,” working paper.

Prager, R. 2009. “Determinants of the Locations of Payday Lenders, Pawnshops, and Check-Cashing Outlets,” working paper.

Skiba, P. and Tobacman, J. 2008a. “Do Payday Loans Cause Bankruptcy,” working paper.

_____________________ . 2008b. “Payday Loans, Uncertainty, And Discounting: Explaining Patterns of Borrowing, Repayment, and Default, working paper.

Stegman, M. and R. Faris. 2003. “Payday Lending: A Business Model that Encourages Chronic Borrowing,” *Economic Development Quarterly,* 17(1): 8-32.

Stephens Inc. 2008. “Payday Lending Industry Report,” Little Rock, Arkansas.

Stoianovici, P. and Maloney, M. 2008. “Restrictions on Credit: A Public Policy Analysis of Payday Lending, working paper.

Strahan, P. and R. Krozner. 1999. “What Drives Deregulation: Economics and Politics of the Relaxation of Bank Branching Restrictions,” *Quarterly Journal of Economics,* 114(4): 1437-67.

Wilson, B., Findlay, D., Meehan, J. Wellford, Charissa P. and K. Schurter. 2008. "An Experimental Analysis of the Demand for Payday Loans," working paper.

Zinman, J. 2010. "Restricting Consumer Credit Access: Household Survey Evidence on Effects Around the Oregon Rate Cap" *Journal of Banking and Finance*, 34(3): 546-556.