Empirical Essays on Entrepreneurship and Corporate Social Responsibility

by

Aaron Kumar Chatterji

B.A. (Cornell University) 2000
M.S. (UC Berkeley) 2003

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Business Administration

in the

GRADUATE DIVISION

of the

UNIVERSITY of CALIFORNIA, BERKELEY

Committee in charge:
Professor David C. Mowery, Chair
Professor David I. Levine
Professor Bronwyn H. Hall

Fall 2006

Empirical Essays on Entrepreneurship and Corporate Social Responsibility

by

Aaron Kumar Chatterji

Executive Summary

Empirical Essays on Entrepreneurship and Corporate Social Responsibility

by

Aaron Kumar Chatterji

Doctor of Philosophy in Business Administration

University of California, Berkeley

Professor David C. Mowery, Chair

This dissertation is comprised of three empirical essays on entrepreneurship and corporate social responsibility. In the first essay, I investigate the impact of prior employment on entrepreneurial performance and innovation in the medical device sector. For the entrepreneurs in my sample, I find that prior employment at an incumbent medical device firm is consistent with faster time to venture capital funding, higher pre-money valuations, and quicker product approvals from the Food and Drug Administration (FDA). Interestingly, I find that entrepreneurs do not inherit much technical knowledge from their parent firm, but instead acquire valuable non-technical knowledge relating to marketing and regulation. These types of knowledge allow their ventures to perform better than other new entrants in the medical device sector. In the second essay, I examine the validity of corporate social responsibility ratings. These ratings are created

by socially responsible investing companies to grade companies on social performance. Through an analysis of five of the most prominent sets of ratings, I find little convergent validity (low correlations) across ratings, implying that there is considerable uncertainty about the underlying construct of social responsibility and possibly significant measurement error. I also find that the ratings have weak predictive validity in forecasting future scandals, suggesting that these social ratings may not be accurate in separating responsible firms from irresponsible firms. The third essay studies the impact of local contracting set aside programs on minority entrepreneurship. The set asides programs reserved a portion of city contracts for minority owned firms during the 1980s. Utilizing data from the Current Population Survey (CPS), I employ a differences in differences approach using the white entrepreneurship rate as a benchmark. I find that minority entrepreneurship increased between 2.5%-5% after the introduction of these set asides along with a modest increase in minority employment.

Contents

List of Figures

List of Tables

Dedication

To my family, without whom I would have never started, continued, or finished my graduate studies.

Acknowledgements

I would like to thank my parents, Dr. Manas Chatterji and Pradipta Chatterji and my sisters Minki Chatterji and Pinka Chatterji for their consistent love and support during my graduate studies and my entire life. I would also like to thank Stephen Len, Kaushik Bagchi, Milani Len, Kasey Len, Sabina Len, Rohan Bagchi, and Avi Bagchi for enriching my life. I would like to especially thank David Mowery, David Levine, Bronwyn Hall, Kenneth Chay, Robert Fairlie, David Teece, Catherine Wolfram, and the other fantastic faculty members at the University of California. Finally, thank you to all of my friends from Vestal, Cornell, New York City, Berkeley, North Carolina, and beyond who made this possible.

Chapter 1

1. Introduction

This dissertation is comprised of three empirical essays on entrepreneurship and corporate social responsibility. At the heart of each chapter are the questions of 1) Where do new entrants in an industry come from? and 2) How well do they perform? In the first essay, I consider the phenomenon of "spawning", former employees of incumbent firms starting new ventures in the same industry. Using data from the medical device sector, I find that spawned firms perform better than physician founded ventures or ventures founded by industry outsiders. Moreover, I find that spawns do not appear to inherit much technical knowledge from the parent firm, but instead utilize non-technical knowledge relating to regulation and marketing to gain competitive advantage. The results of this essay challenge assertions made in the previous literature about knowledge spillovers from parent firms to their spawns, and suggests further research on the importance of non-technical knowledge in the spawning process. The second essay examines the motives and performance of firms in the nascent socially responsible investing sector. These firms assign social ratings to corporations to aid investors and other stakeholders in appraising non-financial performance. I review extant theory from economics, sociology, and finance relating to convergence among financial raters. I find that these social ratings have very low correlations (low convergent validity) across raters and limited predictive validity with regards to future scandals. These results suggest significant measurement error among raters and imply that further research into the validity of social ratings in required. The final essay examines the impact of a controversial affirmative action program in city contracting. During the 1980s, cities across the U.S. reserved percentages of city contracts for minority owned businesses. The aim of these programs was to increase minority entrepreneurship and employment in American cities. I find that these programs increased minority entrepreneurship rates by 2.5%-5% and also had positive impacts on minority employment, suggesting that these new entrepreneurs were successful enough to hire workers. These results imply that these programs had a very large impact in a short time, suggesting future research on whether other programs may have had an effect during the same period.

Chapter 2

2. Spawned With A Silver Spoon?: Entrepreneurial Performance and Innovation In The Medical Device Industry

2.1 Introduction

Entrepreneurship in high technology industries often occurs through "spawning" (Gompers et al. 2005, Klepper, 2001, Brittain and Freeman, 1986). This is the process by which former employees of incumbent firms found entrepreneurial ventures in the same industry.[1] Employment at an incumbent firm (the parent) frequently provides a springboard for the incipient entrepreneur to launch a new venture (the spawn). Prospective entrepreneurs can use their experience at an incumbent firm to acquire financial resources, accumulate social capital, augment their technical skills, and identify entrepreneurial opportunities. Once these nascent entrepreneurs found a new venture, their progress may be heavily influenced by the characteristics of their former employer, as they raise capital, establish organizational routines, develop and manage intellectual property, and choose the appropriate business strategy.[2]

Despite the considerable influence of prior employment on entrepreneurship, few studies have explicitly related attributes of the parent firm to the strategy and performance of the spawned entrepreneurial venture. How then does prior employment impact the entrepreneurial process? Specifically, how does knowledge gained at the parent firm impact the performance and innovative activities of the spawned venture? In this paper, I

[1] Agarwal et al; 2005 refer to these firms as spin-outs. I use the terms interchangeably in this paper.
[2] Of course, individual characteristics matter greatly as well, and there is a significant literature exploring which individual characteristics relate positively to entrepreneurship

will investigate whether entrepreneurial spawns perform better than their competitors and assess to what extent spawns incorporate knowledge from the parent firm. A key issue in this literature concerns the type of knowledge that the spawn gains from the parent, and I present some preliminary results on this point as well. In the next section, I will discuss the relevant literature from management, economics, and organizational sociology. I use this prior work to develop 4 testable hypotheses, which I propose in Section 3. Section 4 provides a brief overview of the medical device industry and explains why it is well suited for this study. Section 5 describes the dataset and how it was constructed. Section 6 presents the major results on performance and knowledge inheritance. I offer my conclusions and outline the major caveats to this analysis in Section 7.

2.2 Literature Review and Theoretical Background

I draw on recent work from management, economics, and organizational sociology on entrepreneurship, and in particular entrepreneurial spawning and spin-offs to develop 4 testable hypotheses. The key findings of prior work suggest that 1) Organizations differ in their propensity to spawn entrepreneurial ventures; 2) Spawns incorporate technical and other types of knowledge from the parent firm, but also differentiate themselves in their business strategies; and 3) The knowledge inherited from the parent firm, along with other characteristics of the parent firm, will impact the performance of spawns.

2.2.1 Prior Employment and Entrepreneurial Opportunities

While the romantic (and popular) notions of entrepreneurship conjure up images of a college drop-out working out of his parents' garage on the next big thing, most entrepreneurs have significant prior employment experience (Cooper, 1985; Robinson and Sexton, 1994), and many prospective entrepreneurs first identify entrepreneurial opportunities at their previous job (Cooper et al. 1990). In fact, Bhide (1994) found that 71 percent of the founders he studied exploited ideas they came across at their previous employer. Once an employee invents a new technology or identifies a new entrepreneurial opportunity, he may choose to disclose it to his employer or not. Anton and Yao (1995) model this decision, and also theorize about when the innovation is developed at the parent firm or when a spin-off results. In their view, spin-offs occur because of contracting problems between firm and employee, and these new ventures will target niche markets with significant innovations. Thus, how the parent firm designs its internal incentive structure, as modeled by Hellmann (2002), will influence the employee's decision to become an entrepreneur.

2.2.2 Entrepreneurial Spawning

There is also an emerging literature on why some firms spawn more new ventures than others. Gompers et al. (2005) tests two theories of entrepreneurial spawning. In the first case, employees learn valuable skills and gain access to important social and financial networks through the parent firm's suppliers and customers. Furthermore, employees may even have the opportunity to engage in the entrepreneurial process within the firm through corporate venture capital units and other internal initiatives. In theory, this type of organizational environment should attract and produce more entrepreneurially minded employees.

In the second case, entrepreneurship is instead a response to a rigid firm bureaucracy that discourages the pursuit of opportunities outside of the core business and penalizes entrepreneurially minded employees for deviating from an assigned task. For many reasons, industry incumbents may be unable to capitalize on disruptive innovations (Christensen, 1997) and internal capital budgeting may favor ongoing projects instead of innovative ideas. Thus, entrepreneurial employees, frustrated with their employer's reluctance to prioritize innovation and exploit new opportunities, will leave the firm to start their own venture.

Gompers and his coauthors find indirect support for the first case in which employees learn valuable skills and gain access to useful social and financial networks at the parent firm. The authors also suggest that parent firms that were once VC backed themselves may establish organizational routines and practices that encourage innovation and entrepreneurial thinking. Still, they do not test explicitly for knowledge flows (technical or otherwise) from parent to spawn.

Klepper and Sleeper (2002) delve into this relationship further, and argue that spin-offs inherit knowledge from their parent firm that shapes their organization, strategy, and performance. In their model, the employee at the parent firm gets access to key information about new innovations and technical developments as well as market opportunities. The employee may then decide to exploit this opportunity himself. Using evidence from the laser industry, the authors find that spin-offs exploit knowledge from their parents but also tend to differentiate themselves from the parent firm. However, it is important to note that the authors use product characteristics, not measures of knowledge flows, to reach this result.

One unresolved question from Gompers et al. (2005) is posed at the end of their paper. The authors ask,

> *"Finally, how do the characteristics of the spawning firms affect the success of the new ventures? For example, are entrepreneurs from more successful spawning firms more or less likely to be successful themselves?"*

> Similarly, Burton et al. (2002) also ponder whether "*ventures spawned from prominent employers may be more likely to go public successfully, or {whether} they may be more likely to be acquired…"*.

Whether spawns perform better than other entrants is clearly an open question, but it is also important to understand where these supposed performance advantages come from. In particular, to what extent do spawns actually incorporate knowledge or other capabilities from the parent firm that impact performance? For example, spawns may perform better because their prior experience at a prestigious firm is a credible signal of their quality to potential investors, partners, and suppliers. Alternatively, performance advantages may derive from specific knowledge inherited from the parent firm. The question then is, if we demonstrate that spawns perform better, can we also ascertain the difference between the "signaling" component of prior experience that attracts outside financiers, partners, etc. versus the "employee learning" component (discussed both by Klepper and Sleeper (2002) and Gompers et al. (2005)) where spawns use knowledge cultivated at the parent firm to exploit entrepreneurial opportunities?

Furthermore, it need not be technological knowledge alone that is passed from parent to spawn. Agarwal et al. (2004) assert that both technological and marketing know-how can be passed from parent to spawn (they use the term spin-out), using evidence from the disk drive sector. In fact, Agarwal and her co-authors argue that when one type of know-how is not matched by its complement (e.g. a firm with technical know-how but insufficient marketing know-how), employees are more likely to leave the firm to start their own ventures. Thus, the level of technological or marketing know-how in the parent firm is positively related to the spawning rate. Moreover, spawns inherit technical and marketing know-how from their parents, and these capabilities will enhance spawn survival.

2.2.3 Uncertainty and Inter-Organizational Affiliations

While Agarwal et al. (2004) posit that the technological and marketing know-how of the parent firm will impact the performance of the spawned venture, there is other work suggesting that other parent attributes may be relevant to understanding spawning. Burton et al. (2002) focus on the relative "entrepreneurial prominence" of the parent firm and hypothesize that if a particular parent firm is (a) prominent in the industry and is (b) known for spawning entrepreneurial ventures, then their spawns will be more likely to pursue an innovative strategy and obtain external financing. The authors argue that the prominence of the parent company can reduce the uncertainty for investors deciding whether to finance a new venture.

The work on entrepreneurial prominence is related to the concept of inter-organizational affiliation proposed in Stuart et al. (1999). Since there is considerable uncertainty surrounding new high technology firms, potential exchange partners may rely on affiliates to signal the quality of a new venture (Stuart et al. 1999). In this manner, new organizations establish credibility through their associations with existing organizations (Higgins and Gulati, 2003). Often lacking a finished product, customers, or positive revenue, young companies communicate their credibility to outsiders in a variety of specific ways (Stuart et al. 1999). As Higgins and Gulati (2003) describe in the biotech industry, young firms will establish their legitimacy through associations with prestigious scientists, hiring away top managers from incumbent firms, or by partnering with pharmaceutical companies downstream. As discussed in Stuart et al. (1999), these inter-organizational endorsements will reduce the uncertainty surrounding young firms and allows third parties to make more informed judgments about new ventures.

To summarize then, the prior work on entrepreneurial spawning and spin-offs has emphasized the relationship between the parent firm and the spawn, and in particular has suggested that the parent's technical knowledge, marketing know-how, social networks, and prominence in the industry will impact the performance of the spawn. There has been indirect empirical evidence that spawns inherit various types of knowledge from the parent firm, which subsequently impact its performance. Still, large gaps remain in our understanding of the nature of knowledge inherited by spawns and demonstrating that these knowledge spillovers are actually taking place.

2.3 Hypotheses

The previous literature offers some theoretical insights into why spawns might perform better than other entrants. In short, the argument rests on valuable knowledge (usually technical) being inherited by the spawn from the parent, but this has been rarely tested. Moreover, with the exception of Agarwal et al. (2004), the distinction between technical knowledge and other types of knowledge is rarely addressed. In this paper, I seek to test whether spawns actually inherit knowledge from the parent, and if so, how this knowledge impacts performance. I use the prior literature to develop 4 hypotheses on spawn performance and knowledge inheritance. First, as discussed in detail above, prior work in management, economics, and organizational sociology suggests that spawns of large, prominent, firms in the industry will perform better than their competitors. These performance advantages are derived from the lower costs of access to valuable knowledge, skills and experience gained through employment at the parent firm, privileged access to valuable social networks, and the benefits of prominent inter-organizational affiliation for former employees of incumbent firms. These advantages can manifest themselves in several different ways. Timely access to capital is a critical factor to success in entrepreneurship, as several nascent and incumbent ventures are often competing to develop a product at the same time.[3] Hsu (2004) finds the serial entrepreneurs obtain outside funding faster than first time entrepreneurs because their prior experience reduces the uncertainty venture capitalists face in making the funding decision. Prior experience at an industry incumbent will impact time to funding in a similar way, as venture capitalists will face less uncertainty in deciding to fund a spawn rather than a physician with no business experience. In the case of a spawned employee,

---

[3] Hsu, 2004

the affiliation with the parent firm, industry expertise, and valuable social networks may all make it easier to obtain funding more quickly. As implied by Stuart et al. (1999) and Burton et al. (2002), spawns from large incumbent firms in the industry should have an easier time obtaining funding than their competitors.

Thus, I propose hypotheses 1:

Hypothesis 1

Entrepreneurial ventures that have been spawned by incumbent firms in the industry will secure funding more quickly than other entrants, controlling for all observable characteristics

Next, conditional on getting funded, the implications of Klepper and Sleeper (2002) and Gompers et al. (2005) is that spawns benefit from the knowledge and skills gained at the parent firm and should experience better performance in the long run. Agarwal et al. (2004) finds that spawns in the disk drive industry last longer than other entrants and Phillips (2002) also finds that spawned ventures survive longer than other entrants in his study of law firms. Agarwal et al. (2004) argue that spawns will possess superior industry specific knowledge compared to their competitors, which manifests itself into better performance. However, it should be noted that it is quite difficult to measure performance in private entrepreneurial firms. Thus, I propose hypotheses 2A and 2B, using pre-money (private) valuations at the last round of venture capital financing and the successful commercialization of a product as proxies for performance.

Hypothesis 2A

Entrepreneurial ventures that have been spawned by incumbent firms in the industry will have higher valuations at the last round of outside (non-public) financing than other entrants, controlling for all observable characteristics

Hypothesis 2B

Entrepreneurial ventures that have been spawned by incumbent firms in the industry will be more likely to commercialize a product than other entrants, controlling for all observable characteristics

Finally, as suggested by Klepper and Sleeper (2002) and Agarwal et al. (2004), spawns inherit valuable technical knowledge from their parent firm that impacts their own organizational routines, business strategy, and performance. In particular, Agarwal et al. (2004) find that the more technical know-how the spawned venture has, the longer it survives. This work implies that the greater the extent to which spawns incorporate technical knowledge from the parent firm (without directly imitating the parent), the more successful they will be.[4] However, it is difficult to trace the inheritance of this technological know-how across firms. Preliminarily, I examine the relationships between parent and spawn patent portfolios. Therefore, I propose hypothesis 3.

Hypothesis 3

Spawned ventures that incorporate technical knowledge from the parent firm will perform better than other spawns, and other competitors.

[4] Klepper and Sleeper (2002) make an interesting point that this relationship may be non-monotonic

It is important to note that I can only test for one type of parental knowledge here—technological know-how—and assess its impact on the performance of the spawned firm. After the empirical results are presented, I will discuss the alternate forms of knowledge inherited by the spawned firm from the parent in Section 6.

2.4 The Medical Device Industry

The November 1976 meeting of the American Heart Association had a poster session that would revolutionize medicine. [5] Dr. Andreas Gruentzig, dressed in sandals and a leather jacket, presented his research to an audience of skeptical colleagues in Miami, Florida.[6] A young Stanford University trained doctor named John Simpson was enamored with Gruentzig's offbeat idea to use a balloon catheter to clear blocked coronary arteries delivering oxygenated blood to the heart, a procedure called angioplasty.

After tinkering with Gruentzig's design at home, Simpson would eventually found Advanced Cardiovascular Systems (ACS) in 1978, a successful medical device company that helped to create an entirely new specialty within medicine and spawned numerous other medical device firms through its former employees. The entrepreneurial culture at ACS produced the next generation of medical device start-ups that would later commercialize several important medical technologies. [7]

---

[5] http://www.ptca.org/archive/interviews/myler1.html Last Accessed April 4th, 2005
[6] Cath Lab Digest http://www.cathlabdigest.com/cld/displayArticle.cfm?articleID=./200106/cld_200106f6&type=A Last accessed September 10th, 2005
[7] One interview subject I spoke with described ACS as the "university of medical devices" for its reputation as a training ground for future medical device entrepreneurs.

Today, medical technologies are being touted as a potential way to reduce health care costs and improve the quality of care. With industry giants like Medtronic and Johnson & Johnson largely focusing on incremental innovations to their existing products, disruptive innovation has been left largely to physician-entrepreneurs like John Simpson, former employees of industry incumbents like the ACS alums, serial entrepreneurs who found multiple companies, and individuals from outside the industry who develop promising ideas.

The medical device industry has continued to introduce innovative products, especially in the area of interventional cardiology. The development of bare metal stents and drug eluting stents are the most well known. During the first angioplasty procedures, 30 percent of patients eventually developed restenosis, or re-blockage of the arteries.[8] Cardiac stents were then introduced to remedy this problem. A stent is a tiny steel mesh tube that is used to keep arteries open during procedures like angioplasty and atherectomy. During these procedures, a balloon is inflated and the stent will expand to the size of the artery, staying in place permanently after the procedure is over. Stents were first approved for use in 1994 and are now common in cardiac procedures. Bare metal stents eventually reduced restenosis by 20-50%, a significant improvement.[9]

---

8 The Cleveland Clinic Website http://www.clevelandclinic.org/heartcenter/pub/history/future/intervention.asp Last accessed April 12th, 2005

9 The Cleveland Clinic Website http://www.clevelandclinic.org/heartcenter/pub/history/future/intervention.asp Last accessed April 12th, 2005

Still, the simple stents resulted in complications, with scar tissue developing around the stents still sometimes causing re-blockage of the arteries. Many companies, in both biotechnology and medical devices, were actively trying to improve stent technology during the late 1990s and early 2000s. Advances in research led to the development of anti-restenosis drugs, which were quickly seized upon by top stent makers, Guidant, Johnson & Johnson, and Boston Scientific, who tried to combine the new drugs with stents. Johnson & Johnson was arguably most successful in this endeavor, gaining FDA approval for its Cypher stent before its two competitors. The stent market was estimated to be worth $2.6 billion dollars in 2003. The new generation of stents cost $3,000 each, and market analysts expect sales of nearly $5 billion dollars in 2005 for use in 1 million patients (patients often require more than 1 stent).[10]

Today, medical device firms develop and commercialize many kinds of innovative products used by physicians during cardiovascular, neurological, and various other types of medical procedures. These products include diagnostic equipment like CAT Scan machines, therapeutic devices such as pacemakers and surgical instruments like endoscopes. Products are usually marketed and sold directly to healthcare professionals and to supply companies. Overall, medical devices are a growing industry with a market size of $75 billion in 2002 and $1.5 billion venture capital invested in 2003.[11] [12]

---

10 The Motley Fool Website
http://www.fool.com/news/commentary/2003/commentary030602am.htm Last accessed April 12th, 2005

11 AdvaMed2004 Report
12 PriceWaterhouseCoopers Money Tree (www.pwcmoneytree.com) Last accessed March 1st, 2005

The medical device industry is ideal for an empirical investigation of entrepreneurial spawning, since there are a number of large players; Johnson & Johnson, Boston Scientific, Guidant (recently acquired by Boston Scientific), Abbott, Medtronic, St. Jude, and others, surrounded by numerous innovative start-ups. In addition, many employees of the aforementioned large firms have spawned new ventures with varying levels of success. In addition, other medical device entrepreneurs have entered the industry directly from clinical practice or academia. This variation in prior employment is central to the empirical study. The industry is also conspicuously clustered in certain areas of the country, including The San Francisco Bay Area, Minneapolis, Orange County, CA, and Boston, allowing for considerable employee mobility since individuals can easily change jobs without relocating.[13]

Additionally, patenting is crucial in the medical device industry with over 9000 patents issued by 2003. Patenting of medical devices is usually considered a crucial part of firm strategy in this sector.[14] In addition, academic research is a key component of product development. In many cases, advances in the academic literature spur product development and company formation.[15] Moreover, doctors are not only the primary customers for medical devices, they are often innovators as well. These user innovations are an important source of ideas for incumbent firms and new entrants. Finally, there are many active venture capitalists investing in the industry. In the first 2 quarters of 2004,

---

[13] Almeida and Kogut (1999) discuss the implications of regional mobility among engineers in the semiconductor industry

[14] This fact emerged in my interviews with medical device entrepreneurs and from secondary industry sources.

[15] The birth of "Hypothermia companies" between 1996-1998 was partly based on advances in research on cooling the human body.

over $700 million dollars of venture capital was invested in the medical device and equipment sector. Taken together with biotechnology, the life sciences space took in 25 percent of all venture capital in the first 2 quarters of 2004. [16] (See figures 1-4 in Appendix B for general trends regarding venture capital investments in medical device)

2.5 Data and Methods

2.5.1 Data Sources

To answer the research questions posed above, I utilize four major data sources:

1. VentureSource's Venture One database
2. Thomson Financial Venture Economics's Venture Xpert database
3. NBER Patent Database
4. Semi-Structured Interviews with 10 Medical Device CEOs

VentureOne and Venture Xpert have been used by many scholars to investigate issues surrounding venture capital and private companies (Kaplan et al. 2002). VentureOne was established in 1987 and tracks firms that have received venture capital financing. The firms are identified through trade press, company websites, and personal contacts with investors. VentureOne surveys the firms and the investors, and updates and verifies the data monthly. Some of the variables include the names and previous employers of the company founders, industry sector, business strategy, and some limited financial information about the new venture (Gompers et al. 2005).

[16] Venture One

Venture Xpert has similar information on venture capital firms and their investments in private companies. While there are some differences between the databases (see Kaplan et al. 2002), I use both databases together to cross check information. The important variable of prior employment of founders is found in the Venture One dataset.

The NBER patent database has data on nearly 3 million U.S. patents granted between 1963 and 1999 (Hall et al. 2001). I can use medical device patents and citations to analyze innovation among entrepreneurial ventures and incumbents in the industry.

Through industry contacts, I conducted semi-structured interviews with CEO/Founders of privately held medical device companies in my data. The purpose of the interviews was to gain insights on the reasons medical device employees leave large firms to start new ventures and isolate the mechanisms by which past experience impacts the entrepreneurial process and performance.

The interviews were semi-structured, conducted both on the phone and in-person, ranging from about thirty minutes to three hours. If the interviewee agreed, the conversations were also recorded. To develop the interview protocol, I rely heavily on Phillips and Fernandes(2003), where the authors conducted interviews of entrepreneurs in the professional services area.

2.5.2 Building the Dataset

I use Venture Source and Venture Xpert to obtain data on private, venture capital backed medical device companies. Searching under the category "medical device", my original dataset included approximately 1000 firms. For each firm, I also obtained data on executives and their career histories, financing rounds and valuations, and liquidity events. If the executive career history listed a public firm, I used the Compustat database and the NBER patent database to obtain information about the parent company. Finally, I also used the NBER patent database to identify patents for technology invented by and/or assigned to the founders, the entrepreneurial firms, and the parent firms. The final dataset was thus a sample of approximately 650 medical device entrepreneurs from 191 firms, their career histories, their patents and citations, financing information, liquidity events, and information on their parent company.[17]

2.5.3 Constructing Measures

The time to funding variable is calculated by taking the difference (in days) between the first reported round of outside funding in Venture One and the founding date for the company. While seemingly straightforward, there are several possible problems with this measure. First, the founding dates are reported to Venture One by surveys and may not always be accurate. For example, a few companies list their founding date and their first round of funding on the same day, which is feasible, but may not indicate the true age of the firm. Next, neither Venture One nor Venture Xpert has reliable data on individual or "angel" investments that could have been made prior to the first round of venture capital

[17] This sample may be biased towards more successful firms, where the career histories of founders were easier to obtain.

funding. Without this data, we may observe a lengthy "time to funding" in cases where the start-up is actually being financing by angel investors. In practice however, most successful medical device companies receive venture capital funding, and data on non-VC backed medical device companies is not readily available.[18]

Measuring performance in small, privately held, companies is a difficult task. The conventional measures of performance for large companies, like profits, revenues, and sales, do not always apply and financial information about private firms is closely held. However, since each firm in my data received at least one round of venture capital funding, there exists some common metrics of performance across each of the 191 firms.

As some scholars have asserted, entrepreneurship is concerned with "value creation"[19], and the valuation of the private firm, as assessed by outside investors, is an appropriate measure of the value the venture has created, and thus a comparable performance metric. I thus use "pre-money" valuation, or the valuation of the company (as determined by investors at rounds of non-public financing) as one measure of performance in this study.

Since the pre-money valuation is determined by outside investors through a rigorous analysis of the company's management team, market opportunities, milestones, and intellectual property, it is an appropriate measure for comparison across private firms. In practice, valuations can be determined in a few different ways, either by identifying comparable companies in the industry, using a discounted cash flow model where future

[18] Even serial entrepreneurs, who might be the most likely to "bootstrap" their own ventures or secure angel funding, often receive traditional venture capital funding in the medical device industry.
[19] Sarasvathy and Wicks (2005)

earnings are estimated and discounted to present value, or using the replacement value of assets.[20] Hsu (2004) provides some evidence that entrepreneurs may settle for lower valuations to work with prestigious venture capitalists. I will address this point and its potential impact on my results in Section 7.

Since there is considerable variation in the number of rounds of venture capital each firm in my sample has, I only look at the last round of financing and control for the round number and other important variables. This method allows me to compare ventures at the same round of funding, controlling for other observable characteristics. In doing so, I can determine whether the characteristics of the founder impact the firm valuation at a particular round.

As an alternative measure of performance, I also analyze data from the Food and Drug Administration (FDA) on medical device approval. FDA approval is an important milestone for medical device firms and is critical to future rounds of financing. I use the successful approval of a product as a measure of performance for an entrepreneurial venture, controlling for relevant observable characteristics. (It may also be useful to calculate a time to approval measure, but this will depend heavily on unobserved and product specific characteristics.)

These 2 measures of performance are hardly ideal, but are suitable in the context of medical device start-ups, where financial data is difficult to obtain and product approval is a common goal for most start-ups with innovative technologies.

[20] Levine (2001)

In the previous literature, a "spawn" or spin-off is usually defined by the most recent employer of the entrepreneur. In my discussions with medical device entrepreneurs, engineers, and venture capitalists, I found that the career path in medical devices is often characterized by "big company" experience early on (at firms like ACS, Medtronic, Baxter, Johnson & Johnson) followed by multiple start-up experiences. Thus, there were few entrepreneurs in my data whose last job had been at a large medical device firm, while many founders had worked at a large medical device firm earlier in their career.

Therefore, I define a spawn as an individual who at some point in their career worked at a large medical device (or life sciences) firm before becoming an entrepreneur. Life sciences, which includes pharmaceutical and biotechnology firms, is included because of the similar managerial, regulatory, and clinical challenges involved in product development. I have defined a spawned venture broadly here, but I also ran alternative specifications with the strictest possible definition of a spawn (an entrepreneur whose last job was at an incumbent medical device firm) and the performance results did not change significantly.

Since my spawning variable is at the firm level, entrepreneurial founding teams who have at least one individual who worked at any point in their career at a publicly traded medical device or pharmaceutical/biotechnology firm are considered "spawns" for the purpose of this study. A simple dummy variable, 1 if at least one team member has worked at a public firm in the industry, 0 otherwise, is used to denote a spawn.[21]

Several caveats must be made about defining a "spawn" in this manner. First, I do not have detailed information on how long each individual spent at the large medical device company that is labeled their "parent" for the purposes of this study. Next, Venture One does not have data on what date the individual joined the entrepreneurial venture, so I conducted my own research (interviews, Lexis-Nexis, company website searches) to determine whether the individual is actually the entrepreneur, or an employee who joined later on. In almost all cases, I am able to precisely determine who the founder of the company is and where they worked previously, but it is possible that a co-founder has been omitted in some cases.

Through the information in Venture One and my own research, I classify each of the non-spawn founders into 3 groups: 1) serial entrepreneurs, 2) physicians, or 3) industry outsiders. If an entrepreneurial team has one member who has previously founded a venture in medical device (but is not a spawn), then I code the venture as being founded by a serial entrepreneur.[22] Similarly, if the venture is founded by physicians or researchers coming directly from universities and medical schools, I code the venture as physician founded.(This category includes university researchers so this category could

[21] Alternatively, I could use the proportion of founding team members who are spawns of publicly traded firms in the industry as another measure. Since founding team size was difficult to determine reliably, I chose not to use this measure

[22] It is important to note that some spawns actually have prior founding experience. Although I code their ventures as spawns for now, I later exploit this variation between "serial spawns" and traditional spawns to measure the impact of non-technical knowledge on performance.

alternatively be called “academic/non-profit researcher”) Finally, if the founder has previously worked only outside the industry, I code the venture as having an outsider founder. With these 3 additional categories and the previously discussed spawn category, I have classified all of the founders in my sample into 4 broad categories.

Once again, there are several issues to consider with this approach. First, those ventures coded as physician-founded may have unobserved (to the researcher) involvement from a serial entrepreneur or a former employee of a large medical device firm. I have searched company websites and several other sources to mitigate this concern of the “phantom” founder. There may also be a bias against listing failed ventures or brief periods of work experience in self reported surveys. Furthermore, my analysis largely relies on the Venture One and Venture Xpert databases and the underlying survey instrument used to collect the data.

Since valuations will differ depending on which market segment the firm is in (cardiovascular, spine, etc.), I add controls for the 14 different market segments identified by Venture One. In addition, since valuations of private companies are heavily influenced by macroeconomic factors, I also control for the year in which the valuation was determined. I also add dummy variables for each round of financing. In addition, I have some limited data on firm characteristics, so I include dummy variables for whether a firm works with an Original Equipment Manufacturer (OEM), has received corporate venture capital, has an international distribution channel, a co-marketing agreement, or is shipping a product to an existing customer.[23]

2.6 Results

2.6.1 Empirical Strategy

I first use the data from Venture One and Venture Xpert to investigate whether spawns perform better than other entrants, controlling for other important firm characteristics. I then analyze the patents of start-ups and their parents to measure the extent to which spawns incorporate the knowledge of their parent and estimate the impact on performance. Finally, I discuss the crucial role of non-technical knowledge in spawning.

2.6.2 Some Descriptive Statistics

Descriptive statistics are presented in Table 2.1. The mean pre-money valuation in the first round is $6.4 million and $40.4 million in the last round. The standard deviation of last round valuation is $62 million. The average firm has been in business for nearly 7 years. Thus, the sample is composed of small, young, medical device firms, and the distribution of the dependent variable is highly skewed.[24]

Each entrepreneurial venture has been classified as having a founder who was spawned from a large medical device company (36%), a serial entrepreneur with no experience at an incumbent firm (25%), a physician (29%) with no past entrepreneurial experience, or

[23] Two of the 69 spawned ventures received corporate venture capital from their parent firm

[24] I do not have reliable data on the number of employees each firm has at each round. In some cases, I can observe the number of employees at the last round. In these cases, the mean number of employees is 57. 90% of firms in the sample have less than 50 employees.

an industry outsider (9%). These categories are mutually exclusive. I combine outsiders and physicians into one group and they are used as the excluded group in the regressions.[25]

Table 2.2 presents the top spawning firms in the medical device sector. The familiar incumbent firms like Medtronic, Boston Scientific, Johnson & Johnson, and Baxter are present. Guidant has spawned 9 ventures in my sample, going back to its days as Advanced Cardiovascular Systems.

2.6.3 Regression Results

In most of the regressions below, I use an OLS specification that regresses my dependent variable on founder characteristics, year, industry segments, round type, and other controls.

Time to Funding=F (Spawn, Serial, Year, Industry Segment) (1)

Valuation=F (Spawn, Serial, Year, Industry Segment, Round, Firm Controls) (2)

In testing each hypothesis, I first regress the dependent variable on the categorical variables that captures the type of firm (spawn, serial, outsider, physician). I then add in round dummy variables, year effects, industry effects, and finally several other firm

[25] This coding scheme allows the results to be interpreted more easily. I also ran the same regressions using only physicians as the excluded group and the results do not change significantly.

specific controls mentioned above. When the dependent variable is highly skewed, I use a log specification.

In Table 2.3, I find evidence that spawns and serial entrepreneurs get funded faster than other entrants.[26] The dependent variable in each of the three specifications is the log of time to funding. In all three specifications, we see that spawns and serials obtain venture capital funding more quickly than other firms, although the result is weaker when industry segment controls are introduced. The size of the effect is economically significant, approximately equal to a 1 year difference in the time to funding between spawns, serials, and other entrants. This result could reflect the fact that spawns and serials choose certain industry segments that attract heightened venture capital interest, a sign of the market knowledge that such entrepreneurs acquire. As a result, when we control for the industry segment, the variation in time to funding is less pronounced because I am essentially comparing firms in the same sub-market. I will discuss the implications of this result further in Section 6.

In Table 2.4, my results demonstrate that spawns achieve higher valuations at the last round of financing. The dependent variable is the log of the pre-money valuation at the last observed round of financing. The results suggest a positive and significant effect for spawns over other entrants.[27] The results are consistently significant at the 5% level until firm level control variables are introduced in the last specification. At that point, the

[26] For 17 firms, I do not have reliable data on time to funding so they are excluded from this analysis
[27] Although the coefficient on serial is no significant, we cannot completely rule out that they perform as well as spawns. Since the coefficient on spawn is significant in each specification, the results suggest that spawns outperform serials.

results are still significant at the 10% level. With only 191 firms and controlling for founder characteristics, round, industry segment, and financing year, this is not surprising.[28] The effect size is economically significant, representing valuations of over $10 million higher for spawns and serials than other entrants.

These results imply that spawns achieve better performance, but there are several alternative explanations. First, there may be considerable unobserved heterogeneity in the sample. That is, high-ability individuals may select into large industry incumbents and later start successful entrepreneurial ventures because of their inherent skills, rather than anything they learned at the large company. In this case, we would still observe the predicted effect even if the large company experience itself did not shape aspiring entrepreneurs, help them to identify entrepreneurial opportunities, or provide them access to valuable knowledge.

The central question that remains is: What part of spawn performance comes from the individual himself (embodied with skills possessed before the parent firm hired him) and what part comes from knowledge gained at parent firm? (In addition, what types of knowledge gained at the parent firm matter the most?) While it is impossible to completely separate these two components, I now try to estimate what part of spawn performance is driven by technical knowledge gained at the parent firm by using patent data from parent and spawns.

---

[28] I also ran the regressions omitting valuations that were outliers (above the 95% level) and the results remained robust

2.6.4 Patent Analysis

The empirical results suggest that spawns perform better than other entrants. What accounts for the superior performance of spawns? The prior literature has conjectured that knowledge (primarily technical) spillovers from the parent to the spawn account for these advantages, but has never, to my knowledge, provided evidence to support these claims. Furthermore, theoretical models have predicted that spawns will “look” like their parents in terms of the technologies they develop. This proposition has usually been tested using product characteristics, not the underlying technology. Thus, I first investigate whether spawns are entering into technology areas in which its parent firm is active. Next, I measure technical knowledge spillovers from the parent to the spawn.

I am interested in 2 broad questions:

1) Do spawns patent in the same technology classes as their parent firm and how does this strategic decision impact firm performance?

2) Do spawns inherit technical knowledge from their parent and how do these spillovers impact firm performance?

I select data from the NBER Patent database on all patents assigned to the spawned firms up until 2002. I then construct patent portfolios for each of the parent firms as well. As predicted by several models of spawning, we expect that spawns will patent in the same area (or in a very similar area) to the parent firm.[29]

69 of the 191 firms with financing information available are classified as spawns. Examining these 69 firms, I first calculate which patent class (IPC classification) these firms are most active in. Since many of these firms are quite small and new, they often have several patents in one primary class and a few more patents distributed across the rest of the patent classes. I then identify the three primary patent classes for the parent firms.[30] Overall, the parent firms have 78% of their patents in three classes.

6 out of the 69 spawned ventures (8%) have the statistical majority of their patents in their parent firm's primary patent class. 23 of the 69 spawned firms (33%) have the majority of their patents in one of the top 3 patent classes for their parent. These descriptive statistics imply that not all spawns patent in the same patent classes as the parent firm.[31] In fact, most spawns patent in classes that account for less than 25 percent of their parent's patent portfolio, or do not have an issued patent in the sample period.[32]

I also analyze the citations file from the NBER Patent Database to trace technological spillovers from parent to spawn more precisely. For each spawn patent, I can identify which previous patents it has cited, and begin to understand whether the patented

[29] In the Klepper and Sleeper model framework, we might view the knowledge that former employees retain from their parent firm to be most useful in technologically-related areas.
[30] To do so, I divide the parent firm's patent portfolio by class and identify the 3 classes with the most patents.
[31] This fact does not preclude the spawn from working in the same product area as the parent, but makes it less likely.
[32] 22 spawned ventures do not have issued patents as of 1999

technology draws on work done at the parent firm. There are some reasons why patent citations might not be an appropriate measure of technological spillovers. Spillovers between technologies may occur, even when there is no citation, and alternatively citations might be made when no technological spillover has occurred. For my purposes then, I may be underestimating the amount of technological spillovers from parent to spawn by focusing on patent citations, but alternative measures suffer from the same limitations. I am less concerned about "gratuitous" citing, where spawns cite parents even in the absence of true spillovers. Jaffe et al. (1993) provides a longer discussion about the merits of using patent citations as a measure of technical spillovers.

12 of the 69 spawned firms in my data cite the parent firm and the average citation rate (number of cites to parent/number of total cites) for citing firms is 3%. Of course, the overall citation rate to the parent firm is very low for spawns, around 1% of all citations, which suggests that not all spawns draw on technical knowledge from the parent firm. But some spawns do cite their parents and appear to build on their work, with firms like Genyx Medical and Arterial Vascular Engineering(not in the financing sample) devoting more than 10% of their cites to the parent firm.

There are a few reasons why spawns might not cite their parents even if they are working in a similar area. Perhaps non-compete clauses and other covenants to prevent former employees from taking valuable knowledge outside the firm compel spawns to omit citations to their old firm's patent portfolio. Another reason may be that the spawned employees themselves are non-technical and do not carry with them valuable technological expertise per se, but rather managerial skill.

To account for this, I collect educational, occupational, and patent data on each of the spawned employees in my sample to classify them as technical or non-technical founders. If the founders have an M.D., a science graduate degree, or any patents, I characterize them as a technical founder. 46 out of the 69 spawns have technical founders and I include a dummy variable for this the regressions that follow and my results changed very little.

The results from the patent analysis reveal that spawns devote 1% of their total overall cites to their parent firm. Without a proper benchmark, we cannot know whether this is significant. Furthermore, a benchmark would also help us to understand whether the citation rates to the parent firm are truly due to the relationship between spawn and parent, instead of the technological position of the parent firm in the space. That is, a spawn may cite their parent firm because the parent firm owns the most important patents in the technological area, not because they are building on the parent's technology.[33]

To address some of these concerns, I create a control set of patents for the spawn patents based on application year, IPC 4 digit class, and grant year.[34] For each of the 792 spawn patents I identify a control patent that has the same application year, same IPC classification, and closest grant date. This method was used in Jaffe et al. (1993). The

[33] Alternatively, a spawn might also cite its parent often simply because it is most familiar with its parent firm's patent portfolio.

[34] Of course, we could generate control samples based on other criteria, including patents in the same class and year from other firms in my sample.

control patent set is 712 patents, since some control patents match to more than one spawn patent. Intuitively, the difference in citation rates between spawns and controls will be the extent to which the parent-spawn relationship matters in terms of technical spillovers. In other words, the control patent citation rate represents the percentage of citations that would go to the parent firm, if no parent-spawn link existed.[35]

Table 2.9 presents the top ten cited organizations by spawn patents and the control set. The lists are similar, with prominent medical device companies receiving most of the citations from spawn patents and the control patents.[36] This fact suggests that the control patents are an appropriate benchmark.

For each spawn, the average citation rate (total citations to parent/total citations made) is 1.04%. Assigning each spawn patent a corresponding control patent, I create a control patent portfolio for each spawn. The next step is to calculate what percentage of total cites for the control patent portfolio are to the parent firm of the matched spawn firm. The average citation rate to the parent for the control patent portfolio is .00906 or 0.906%.

For each spawn, I calculate the difference between their parent citation rate and corresponding control rate. The average difference is 0.00135 or 0.135%. This difference cannot be shown to be different than zero in terms of statistical significance.

[35] A concrete example would be to compare two start-up companies that were both producing a new type of stent. If one of the firms was a spawn of Johnson & Johnson and other was not, we could compare the percentage of citations to Johnson & Johnson in their patent portfolios to discern whether the parent-spawn relationship matters.

[36] Though unreported in this version, an analysis of the citations of the non-spawn firms in the original sample of 191 firms reveals that these firms are also citing many of the same medical device firms, including Medtronic, Baxter, and Advanced Cardiovascular Systems.

Thus, we have strong evidence that most spawns do not inherit technical knowledge from their parent firm in the medical device industry. This result is quite surprising, given the prior literature on spawning and the technical nature of the industry.

Many of the citations spawns do make to the parent firm are self-citations. In these cases, the founder is citing his own previous work at the parent firm. In this section, I investigate the self-citations in my data set and discuss the implications for understanding technological knowledge inheritance in the medical device industry.

I read each citation to a parent firm patent and coded it as a self-citation in either of the following 2 cases.

1) The founder is the inventor on the citing patent (the spawn patent) and listed on the cited patent (the parent patent).
2) The founder is not listed on the citing patent (the spawn patent) but is listed on the cited patent (the parent patent)

The first case is clearly a self-citation. In this case, a former employee of a large medical device firm founds his own venture and cites his prior patents at the parent firm on the new patents he files. The second case is less straightforward. Looking at the data, I find many well-known technical founders who had several patents at incumbent firms sometimes did not appear on all patents filed by their start-up. Still, in some cases, these patents still cite the founder's work at the parent firm. In these instances, it may be that the start-up is building on the founder's knowledge as in case 1.

Thus, I define self-citation broadly, including case 1 and case 2 citations to ensure that I account for all types of technological spillovers that arise from the founder citing his own previous work. Out of the 262 citations made by spawn patents to parent patents, 38 citations are self-citations. 18 citations are "case 1" types where the founder is on both the citing and the cited patent, and 20 citation are "case 2 "types. The differences between case 1 and case 2 self-citations might have important implications for understanding knowledge inheritance from parent to spawn. Case 1 self-citations reflect the most extreme form of individual-embodied knowledge, since the entrepreneur is citing his own work at his prior employer. The technology (insofar as it can be accounted for by patents) of the medical device start-up is primarily based on individual knowledge in this case. In Case 2, employees of the new venture cite the founder's work at the parent firm, implying that the knowledge has now been shared among a larger group of employees. While in both cases, the start-up is building on knowledge from the parent firm, the second case might be considered a better example of a technological spillover since the technical knowledge is being transferred among individuals.

Doing the same calculations as above, the average parent cite rate (now excluding self-cites) is 0.862%, which is not statistically different from the control rate. These results confirm that most spawns are not citing their parent firm, but those that do sometimes use self-citations. In these cases, the individual founder is the conduit for technological spillovers from parent to spawn. This pattern suggests that a significant portion of technical knowledge inheritance in the medical device industry is being driven by entrepreneurs who cite their own prior work. Do these spawned ventures perform better than other entrants because of the technical knowledge they bring from the parent firm?

After demonstrating that most spawns do not inherit technical knowledge from their parents, we would like to know if the few spawns that do build on their parent firm's technology perform better. To differentiate between spawns who inherit technological knowledge from their parent and those who do not, I run the same performance regressions as above, except with new variables indicating how related the spawn's patents are to the parent. I create 4 dummy variables, Spawn Unrelated (for those spawned firms that did not have the majority of their patents in their parent's top three patent classes), Spawn Related (for those spawned firms that do patent predominantly in one of their parent's top three classes), Spawn Simple (for those spawned firms that do not cite their parent), and Spawn Tech Inheritor (for those spawned firms that do cite their parent).

The spawn variable in the earlier analysis is now replaced by Spawn Related and Spawn Unrelated in one model, and Spawn Tech Inheritor and Spawn Simple in another model.[37] The purpose of these regressions is to estimate what part of superior

[37] In the first model, the "Spawn" variable in equation (2) is replaced by "Spawn Unrelated" and "Spawn Related". Each of the 69 firms previously coded as spawn are recoded according to relatedness criteria described above. In this case, 23 of the 69 firms are coded as "Spawn Related" and the rest are coded as "Spawn Unrelated". The other categorical variables remain the same. In the second model, the process is the same, except 12 of the 69 spawned firms are coded as "Spawn Tech Inheritor" and the rest as "Spawn Simple".

performance for spawns is derived from spawns that "look like" their parents in terms of innovative activities or those spawns who benefit from technological spillovers from the parent firm.

The results are presented first in Table 2.5. As before, the dependent variable is the log of the pre-money valuation at the last round of non-public financing. Here, we see some limited evidence that spawns that patent in related areas perform better, but once the industry segments are introduced, we cannot claim that this group of spawns outperforms other spawns.

Table 2.6 shows the results of the same model, except with Spawn Tech Inheritor and Spawn Simple as the main explanatory variables. Here, we find no definitive evidence that spawns that cite their parent's patent portfolio perform better.[38] Thus, I find no support for the claim that spawns that inherit technical knowledge from the parent firm perform better than other entrants.

Table 2.7 presents the results from an alternative probit model using product approval as the binary dependent variable for performance.

Pr (Product Approved=1)=F (Spawn Unrelated, Spawn Related, Serial, Year, Industry Segment, Round, Firm Controls) (3)

[38] If anything, the evidence suggests that spawns who do not cite their parent have higher valuations

I find that spawns that are more related to their parents are more likely to get FDA approval. Interestingly, the main effect for spawns is not significant, which means that hypothesis 2b is not confirmed. Rather, only certain spawns (those related to their parent) seem to outperform their competitors in terms of obtaining FDA approval. Table 2.8 presents results of a hazard specification that allows for censoring. We see that spawns who are working in related areas to their parents get products approved by the FDA faster.

These results indicate that spawns who work in technology areas closely related to their parents are more likely to get a product approved by the FDA and also navigate the process more quickly. This result is particularly interesting since I also find that technical spillovers between parent and spawn are not extensive and they do not impact performance. How then, can these results be reconciled?

These spawns who patent in the same area as their parents may also have other complementary knowledge that is useful to obtaining approval for products in therapeutic segments. If a former employee of Medtronic starts a new venture around cardiac rhythm management (CRM) (an important area for Medtronic) for example, he may possess regulatory and marketing knowledge that helps his new venture to get a product approved. While the new venture may not cite Medtronic on its patents or build directly on an existing Medtronic technology, these other types of knowledge gained at the firm could be useful as well. Furthermore, it could be that the employee's prior experience at

Medtronic alerted him to entrepreneurial opportunities in CRM in the first place. I explore the importance of non-technical knowledge below.

2.6.5. Implications

What do the empirical results suggest? After establishing first that spawns perform better than other new ventures, we see some evidence that this superior performance is being driven by firms that patent in related areas to the parent firm. Still, I do not find evidence that spawns who cite their parent firm's patent portfolio perform better. Most importantly, I find that spawns do not cite their parent firm very much (compared to a control group), which belies the conjecture that spawns inherit technical knowledge from the parent firm.

If not technical knowledge inheritance, what else could be driving spawn performance advantages? In my interviews, several explanations were given. First, many spawned employees inherit other types of non-technical knowledge from the parent firm, such as regulatory knowledge (understanding both the FDA approval process and the Medicare reimbursement process), marketing knowledge (especially regarding how to market to physicians), and how to identify new market opportunities in the medical devices. In most cases, these types of knowledge were suggested as being more important than technical knowledge inherited by the spawn from the parent.

For example, the management team of Acorn Cardiovascular, also included in this dataset, believed they were close to FDA approval in 2002 for their device that helps to shrink enlarged hearts, but the FDA instead recommended a much larger clinical trial that ended up taking three more years and costing the company $30 million.[39] Acorn acknowledged that the slow enrollment in their earlier trials (which was part of the reason for the original delay) could have been predicted since Medtronic and Guidant had also had similar problems enrolling patients in their trials for similar devices.[40] Former employees of large medical device firms who have managed clinical trials often have crucial knowledge that can speed up the approval process.

Importantly, interviewees stressed the importance of prior experience in identifying entrepreneurial opportunities. Employees at large device firm often participated in strategic analysis of their competitors, and were well aware of which segments were being filled up by new start-ups and where new opportunities existed. Similarly, marketing products to physicians provided valuable insight into what new devices might look like and who might buy them.

I find some evidence that spawns target segments where entrepreneurial opportunities are most likely to exist. When I introduce segment controls into my specifications, the impact of spawn on performance is decreased, suggesting that between segment variation accounts for a significant part of the performance advantages for spawns. This trend implies that spawns target particular segments where market opportunities are large, and as a consequence, venture capitalists are active in.

---

[39] McCartney (2005)

[40] IN VIVO Business and Medicine Report (Summary) February 2004, Vol. 18, No. 2. Windhover Information Inc.
http://www.windhover.com/contents/monthly/exex/e_2004800038.htm; Last accessed September 11th, 2005

In addition, the finding that spawns who work in areas related to their parent perform better also suggests that identifying profitable opportunities is an important component of success for these firms. After all, it is most likely that employees will gain knowledge in those technological areas where the firm is active. I find that the most successful spawns enter these related areas, but do not directly build on the technology of the parent firm.

While it might seem surprising that some medical device entrepreneurs leave their parent firm without specific technical ideas, my research indicates that this model is fairly common. For example, Mike Hooven, the founder of Atricure, began his job at Johnson & Johnson with the expressed intention of starting his own medical device firm as quickly as possible.

According to Hooven:

> *"I told my superiors right from the start that I wanted to work here for about five years, get the experience, and make the contacts so I could start my own business and sell products back to the J&J's of the world,"*[41]

Interestingly, when Hooven left in 1994 to start his own company, he did not know exactly what clinical area he would specialize in. Instead he met with doctors to figure out what clinical needs were not being addressed, and eventually founded Atricure to

[41] INVIVO Business and Medicine Report, March 2002, Windhover Information Inc.

focus on atrial fibrillation, a major cause of stroke and congestive heart failure.[42] His experience at Johnson & Johnson had helped him identify new entrepreneurial opportunities, rather than a specific technical idea.

In sum, my results suggest that spawns do perform better than other entrants, but that technological knowledge inheritance is not the major reason. Instead, non-technical types of knowledge help spawns in the regulatory process, marketing to physicians, and identifying profitable market opportunities to pursue. Even in a highly technical industry such as medical device, it would seem that these types of knowledge are more important.

It is difficult to measure these non-technical forms of knowledge as precisely as I have measured technical knowledge inheritance. Using patent data, I was able to identify which spawns had inherited technical knowledge from the parent firm. I could not measure however, whether these spawns possessed other types of non-technical knowledge. One coarse method to differentiate between spawns would be to examine which spawns also have prior founding experience. As mentioned above, many spawns had several jobs in between their experience at a large firm and the start-up listed in my data, including founding previous ventures.[43] Prior founding experience could allow an entrepreneur to acquire many of the non-technical types of knowledge listed above, especially identifying entrepreneurial opportunities and management skills. Thus, we might think that those spawns with prior founding experience possess important non-technical knowledge that my interviewees have suggested are crucial to success.

---

[42] INVIVO Business and Medicine Report, March 2002, Windhover Information Inc.

[43] To be classified as a serial entrepreneur in the previous specifications, the entrepreneur was required to have no prior experience at an incumbent medical device firm. Thus, spawned entrepreneurs with prior founding experience were coded as spawns.

I divide spawns into 2 groups based on those that have prior founding experience and those who do not. Using these new categories, "serial spawn" and "spawn", I run the same performance regressions as above. The results are presented in Table 2.10. We can see some evidence that serial spawns (spawned employees who have already started at least one venture) perform better than other spawns. This result implies that prior experience at industry incumbents coupled with founding experience will result in superior entrepreneurial performance compared to prior experience at industry incumbent alone. These results also support the importance of non-technical knowledge in influencing entrepreneurial performance.

The results suggest that technical knowledge inheritance is less important in impacting the capabilities of start-ups than previously documented. However, if the technologies for new medical device entrants are not coming from incumbent firms, where are new medical devices conceived? My interviewees described a spawning process that is slightly different from the Gompers et al. (2005) and Klepper and Sleeper (2002) perspectives. Rather than ideas coming from ex-employees of large companies, the innovative ideas often come from the users themselves, in this case physicians conducting research (a fine example of user innovation). When a physician discovers a new way to address a clinical need and/or builds a new device, a venture capitalist is usually a key intermediary is matching a manager from a large firm with the physician/inventor. These contacts are often made at medical conferences or through industry associations. This model of spawning is more nuanced than the previous literature and deserves closer inspection.

The start-up Velocimed, founded in 2001 and included in this study, is a typical example of this variant of spawning. The idea for the firm came from Dr. Dennis Wahr, formerly director of interventional cardiology at St. Joseph Mercy Hospital in Ann Arbor, Michigan, who invented an embolic protection device.[44] The device worked in a novel way to remove embolic debris during interventional procedures by blocking and reversing the flow of the treated vessel. Wahr had previously participated in clinical trials to test other firms' devices, including the popular Guidewire device produced by Medtronic.[45] Ironically, he teamed up with 2 former executives from Boston Scientific, another incumbent medical device firm, to form Velocimed around his embolic protection device. Interestingly, Wahr was also CEO despite being the "scientific" founder.[46] The market for embolic protection is estimated to be in the hundreds of millions of dollars, and the combination of an unmet clinical need, huge market potential, and experienced executive team guided Velocimed through 4 rounds of venture capital funding.[47] The company was eventually acquired by St. Jude Medical in 2005 for $74 million. [48]

[44] Reilly (2003)
[45] Reilly (2003)
[46] Riverwest Venture Partners, http://www.rivervest.com/newsletter/newsletter3.pdf, Last Accessed September 6th, 2005
[47] Reilly (2003)
[48] Velocimed Corporate Website, www.velocimed.com/StJude.pdf Last Accessed September 6th, 2005

If this model of spawning is prevalent in other industries as well, we must refine our current models of the process accordingly. After all, the theoretical models as they stand now assume that all valuable knowledge resides in the universe of all existing firms and their employees, while in reality this may not be the case. Employees may leave large firms to be matched with physician-innovators and exploit an entrepreneurial opportunity that was never available to their prior employer at all, which challenges many assumptions of the current literature. For example, some of theoretical literature models an employee deciding whether or not to reveal his invention to the firm. In reality, the employee may be weighing an offer to found a start-up based on technology outside of the firm, which would change the contracting problem considerably. In Anton and Yao (1995) for example, one possible outcome is duopoly profits shared between the employee-inventor and the firm. In the case where the employee is working with someone outside the firm, this outcome might not exist. Future research may seek to incorporate this variant of spawning by incorporating it into theoretical models and empirical testing.

2.7 Discussion

While we know that many high technology ventures are founded by former employees of incumbent firms in the same industry, we know little about how well these spawns perform and to what extent they incorporate knowledge from the parent firm. In this paper, I analyze performance and technical knowledge inheritance for spawns in the medical device industry. I find that spawns do obtain funding more quickly than their competitors (excluding serial entrepreneurs). I also find that spawns receive higher valuations at round, compared to other entrants. These results imply that spawns perform better than other entrants, a finding that has been documented in the prior literature.

Most interestingly though, I find little evidence that technical knowledge gained at the parent firm is a large component of the superior performance of spawns in the medical device sector. Rather, preliminary results and interviews suggest that non-technical types of knowledge acquired at the parent firm, such as the ability to identify entrepreneurial opportunities, seem to drive spawn performance. This is an important contribution to the existing literature, which has hypothesized that spawns inherit technical knowledge from the parent but has rarely demonstrated it empirically.

I examine spawn and parent patent portfolios in detail to measure technological relatedness between firms. I find that not all spawns patent in the same technological areas as the parent. I also construct a set of control patents and find that spawned firms in the medical device sector do not cite their parents more than expected. After concluding that technical knowledge inheritance is probably not the source of superior spawn performance, I discuss other types of non-technical knowledge that impact performance of spawned ventures. I preliminarily test the importance of non-technical knowledge and find that it may account for some of the advantages spawns have over other entrants.

In addition, I describe a more nuanced view of spawning than the previous literature, whereby university researchers or practicing physicians join together with spawned employees from incumbent firms to form new ventures. This variant of spawning deserves more attention, as the technical knowledge possessed by the start-up may not be derived from the parent firm, which might have different implications for performance. In these cases, it may be the managerial, regulatory, and industry specific knowledge that spawned employees bring to the new enterprise that are most important, rather than technical knowledge. Investigating the different types of knowledge gained through prior employment is an area for future research.

Furthermore, I have constructed a unique dataset which has yielded several important descriptive contributions as well. The medical device industry has been understudied in the academic literature, and not studied at all in the context of spawning. As Klepper (2001) points out, the small existing literature on spawning has only focused on a few industries, so this evidence from the medical device sector is important for making improvements to existing theory and guiding the direction of future empirical work.

I find that spawns are common in the medical device industry and that a handful of serial entrepreneurs account for a large part of new entrants. With the exception of Hsu (2005), few scholars have empirically examined serial founders. Through my fieldwork, I came across the names of the same 5-7 serial entrepreneurs who have started multiple successful companies in the medical device industry. These founders were described as having the "Midas touch" for attracting capital to new companies, recognizing entrepreneurial opportunities in unmet clinical needs, and helping to develop nascent technologies. I plan to investigate these founders and their companies in future work.

Despite these results, there are several significant alternative explanations that cannot be definitively dismissed. It is possible that the large medical device firms simply recruit the best and brightest employees, who, regardless of their work experience, would later have become successful entrepreneurs. In this case, the employee learning component of spawning would be a red herring; the screening process of the parent firms would actually be driving success for spawned entrepreneurs. While I cannot rule out this possibility entirely, the interviews point to a stronger impact of employee learning rather than parent screening. My interview subjects frequently discussed the product development process at the parent firm and the valuable lessons they learned from it.

Furthermore, this alternative explanation that "good people work for good firms" does not explain why spawns that "look like" their parents perform better. If spawn performance were driven completely by screening, how can we account for the fact that those spawns who rely most on the parent firm do better than those spawns who work in entirely different areas?

Next, the coding of the founding teams is heavily dependent on the data from Venture One and Venture Xpert and my own research, which may be biased by unobserved trends. To mitigate this, I have cross checked my coding and run alternate specifications to ensure the robustness of the results. In addition, patents are not necessarily the best way to track technology flows in the medical device industry and I could be missing knowledge transfer that is not codified in patents. Judging from the perceived importance of patents in the industry and the interviews, I am less worried about this issue than the other caveats I have described above.

Also, there may be additional unobserved spawning of medical device employees into other industries, like biotechnology, which I would not observe. Thus, this study only applies to medical device employees who start ventures in the same industry, and cannot speak to what valuable knowledge these individuals may carry with them to other fields. It is likely that the knowledge gained at the parent firm is most important for spawns in the same industry, so the impact of prior experience on spawn performance is likely biased upwards in this study since I do not consider spawns in different industries.

It is also possible, as Hsu (2004) suggests, that entrepreneurs accept lower valuations to work with more prestigious venture capitalists. This fact would actually seem to strengthen my results. If spawns are more likely to be working with prestigious venture capitalists, we would observe lower valuations during rounds of financing. My results imply that spawns receive higher valuations at their last observed round than other entrants, so these estimates might be biased downwards if spawns are not accepting their best offer. In other words, I may be underestimating the performance advantages of spawns since these firms may be worth more than their valuation.

Furthermore, pre-money valuation is not an ideal metric for performance for reasons discussed above. In particular, without detailed knowledge of capital equipment, licensing agreements, and detailed financial information, comparisons between pre-money valuations should be considered carefully. As a robustness check, I use other dependent variables such as product approval to measure performance. Still, measuring performance of private companies, some without products on the market, is a noisy process and always open to criticism.

Finally, I only measure (somewhat indirectly) technical knowledge inherited by the spawn from the parent, and do not empirically account for other types of knowledge, including managerial, regulatory, and marketing knowledge. It is difficult to measure this type of knowledge reliably and my data limitations preclude me from addressing this more comprehensively. Ideally, we would also like to have data on the FDA approval process and Medicare reimbursement experience for all of the medical device entrepreneurs in my sample, but such data might be difficult to obtain. Unless a spawned entrepreneur formally held the title of vice president of regulatory affairs (which almost none of the entrepreneurs in my sample did), it would be difficult to assess and compare the regulatory experience of the individuals in my sample.[49]

There are several interesting research questions to pursue based on the results of this paper and other work on spawning. One of the weaknesses of the literature is the inability to refute the "good people work for good firms" explanation for the seemingly large effects of inter-organizational affiliation and prominence. In particular, it would be useful to understand what variables mediate the impact of inter-organizational affiliation for the spawn. Is the effect stronger in the same industry or for employees who were more senior

[49] Another alternative would be to develop a coarse measure of FDA approval and CMS reimbursement experience at the parent firm level, but we could not know how much experience any particular spawned entrepreneur had with either process at the parent firm.

at the parent firm? That is, does the parent firm's reputation matter more for some spawns than others?

We may also be interested in how spawning impacts the subsequent performance of the parent firm. (Jaffee and McKendrick, 2005) When valuable human capital leaves the firm to start a new venture, what is the impact on the parent firm? Also, in studying technological spillovers, it may be enlightening to investigate whether parent firms eventually cite their spawns as evidence that knowledge can travel from parent to spawn and back again.

Finally, when examining the impact of working at an incumbent firm versus starting an entrepreneurial venture, research should focus on differentiating the types of knowledge that can be acquired in each case, and explore how this knowledge is used by individuals who have done both. The spawned entrepreneurs with entrepreneurial experience are interesting candidates for further research, since they presumably embody much of the valuable technical, industry-specific, regulatory, and managerial knowledge (taken together to embody "the silver spoon") necessary to launch successful entrepreneurial ventures.

References

AdvaMed 2004 Annual Report, Advanced Medical Technology Association

Agarwal, Rajshree, Raj Echambadi, April Franco, and MB Sarkar (2004). "Knowledge Transfer Through Inheritance: Spin-Out Generation, Development, and Survival," *Academy of Management Journal,* Vol 47, No 4, pp 501-522.

Almeida, Paul and Bruce Kogut (1999). “Localization of Knowledge and the Mobility of Engineers in Regional Networks,” *Management Science,* Vol. 45, No. 7, pp. 905-917.

Anton, J.J. and D.A. Yao (1995) “Start-ups, Spin-Offs, and Internal Projects,” *Journal of Law, Economics, and Organization*, 11, 362-378.

Bhide, Amar (1994). “How entrepreneurs craft strategies that work,” *Harvard Business School Review, March-April, 1994, 150-161.*

Brittain, Jack W. and John Freeman (1986). “Entrepreneurship in the Semiconductor Industry,” *Working Paper*

Burton, M. Diane, Jesper Sorensen, and Christine Beckman (2002). “Coming for Good Stock: Career Histories and New Venture Formation,” *Research in the Sociology of Organizations, Forthcoming*

Carroll, Glenn R. and Elaine Mosakowski (1987). “The Career Dynamics of Self-Employment.” *Administrative Science Quarterly, 32, 570-89*

Christensen, Clayton M. (1997) “The Innovator's Dilemma: When New Technologies Cause Great Firms to Fail.” *Harvard Business School Press*

Cooper, A.C. (1985). “The Role of Incubator Organizations in the Founding of Growth Oriented Firms.” *Journal of Business Venturing 1, 75-86.*

Cooper, A. C., Dunkelberg, W. C., Woo, C. Y. & Dennis, W. J., Jr. (1990). “New Business in America: The Firms & Their Owners.” *Washington, D.C.: The NFIB Foundation*

Dobrev and Barnett (2005). “Organizational Roles and Transition to Entrepreneurship.” *Academy of Management Journal,* Vol. 48, No 3, pp 433-449.

Garvin, David A. (1983). “Spin-Offs and the New Firm Formation Process.” *California Management Review, Volume XXV, Number 2*

Gompers, Paul, Josh Lerner, and David Scharfstein (2005). “Entrepreneurial Spawning: Public Corporations and The Genesis of New Ventures, 1986-1999,” *Journal of Finance,* Volume 60: Issue 2, April 2005

Hall, B.H., A.B. Jaffe, and M. Tratjenberg (2001). "The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools." *NBER Working Paper 8498*

Thomas Hellmann (2002). “When Do Employees Become Entrepreneurs?” Stanford GSB *Working Paper 1770*

Higgins, Monica and Ranjay Gulati (2003). Getting Off To a Good Start: The Effects of Upper Echelon Affiliations on Underwriter Prestige," *Organization Science, Vol 14, No. 3, May-June 2003, pp. 244-263*

Hsu, David (2004). "What do entrepreneurs pay for venture capital affiliation?," *Journal of Finance,* Vol. 59: Issue 4, pg. 1805-1844, August 2004

Hsu, David (2005). "Experienced Entrepreneurial Founders, Social Capital, and Venture Capital Funding". Working Paper

Jaffe, A., M. Trajtenberg, and R. Henderson. (1993). "Geographic localization of knowledge spillovers as evidenced by patent citations". *Quarterly Journal of Economics 63: 577-98*

Jaffee, Jonathan, and David McKendrick. (2005) "Innovation and Reputation: The Effect of Spin-outs on the Performance of Progenitors in the Hard Disk Drive Industry, 1963-1998." *Working Paper*

Kaplan, Steven N., Berk A. Sensoy, and Per Stromberg (2002). "How well do venture capital databases reflect actual investments?" *Working Paper*

Kirsner, Scott (2004) "Fantastic Voyage." *Fast Company Magazine, Issue 81, April 2004*

Klepper, S. (2001). "Employee Startups in High-Tech Industries," *Industrial and Corporate Change 10: 639-674*

Klepper, Steven and Sally Sleeper. (2005). "Entry by Spin-offs," *Working Paper*

Levine, Scott D., "Business Valuation Issues Related to Start-Ups", posted at http://www.beysterinstitute.org/onlinemag/mar01/trends.htm
Last Accessed April 20th, 2005

Mannen, Ted. (2004). "*Changing Perspectives: The Reimbursement Roller Coaster."* Medical Device and Diagnostic Industry, August 2004.
http://www.devicelink.com/mddi/archive/04/08/018.html;
Last Accessed September 11th,2005

McCartney, Jim (2005). *"Medical-Device Companies Lack Clinical-Trials Patients*." Saint Pioneer Press, August 4th, 2005.

Phillips, Damon (2002). "A Genealogical Approach to Organizational Life Chances: The Parent-Progeny Transfer among Silicon Valley Law Firms, 1946-1996," *Administrative Science Quarterly,* 47: 474-506.

Phillips and Fernandes (2003) "Network Based Entrepreneurship: An Interview Based Investigation of Professional Service Firm Start-Ups", *Polsky Center for Entrepreneurship Working Paper*

Podolny, Joel M (1993). "A Status-Based Model of Market Competition." *American Journal of Sociology,* 98, 4:829-872

Podolny, Joel M., and Damon J Phillips (1996). "The Dynamics of Organizational Status." *Industrial and Corporate Change,* 5, 2: 453-472

Podolny, Joel M., Toby E Stuart, and Michael T Hannan (1996). "Networks, Knowledge, and Niches: Competition In the Worldwide Semiconductor Industry, 1984-1991." *American Journal of Sociology,* 102, 3: 659-689

Radinsky, Greg (2004). "Reimbursement." *Medical Device and Diagnostic Industry*, *January 2004*

Reilly, Mark (2003). "Medical Start-Up Lands $18 Million." *The Business Journal,* July 18th, 2003.

Robinson, P.B. and Sexton, E.A. (1994). "The Effect of Education and Experience on Self-Employment Success." *Journal of Business Venturing,* 9(2)*, 141-156*

Sarasvathy, Saras and Andrew C. Wicks (2005). "Value Creation through Entrepreneurship: Reconciling the Two Meanings of the Good Life." Working Paper

Stuart, Toby E., Ha Hoang and Ralph Hybels (1999). "Interorganizational Endorsements and the Performance of Entrepreneurial Ventures," *Administrative Science Quarterly,* 44, 2: 315-349.

Websites

Angioplasty.Org
http://www.ptca.org; Last Accessed April 4th, 2005

Cath Lab Digest
http://www.cathlabdigest.com; Last Accessed September 10th, 2005

PriceWaterhouseCoopers Money Tree
http://www.pwcmoneytree.com; Last Accessed September 10th, 2005

Riverwest Venture Partners
http://www.rivervest.com; Last Accessed September 6th, 2005

The Cleveland Clinic
http://www.clevelandclinic.org; Last Accessed April 12th, 2005

The Motley Fool
http://www.fool.com; Last Accessed April 12th, 2005

Windhover Information Inc.

http://www.windhover.com; Last Accessed September 10th, 2005

Table 2.1 Descriptive Statistics

Variable	N	Mean	S.D.	Min	Max
First Round Pre-Money Valuation (MM)	191	6.4	6.6	0	39.3
Last Round Pre-Money Valuation (MM)	191	40.4	62	0.5	499
Spawn	191	0.36	0.48	0	1
Serial	191	0.25	0.43	0	1
Physician	191	0.29	0.46	0	1
Outsider	191	0.09	0.29	0	1
age(yrs)	191	6.7	2.9	0.67	18.9
financingrounds	191	4.7	2.4	1	13
CorporateVC	191	0.14	0.34	0	1
OEM	191	0.07	0.26	0	1
GenDistr	191	0.13	0.34	0	1
CoBrdMkt	191	0.05	0.22	0	1
GlobalDistr	191	0.09	0.29	0	1
JointRD	191	0.06	0.23	0	1
Customer	191	0.1	0.3	0	1

Table 2.2 Top Spawners in Medical Device

Parent	# of Spawns
Medtronic	11
Johnson and Johnson	9
Guidant	9
Baxter	5
Nellcor	4
Pfizer	3
American Hospital Supply	2
BSX	2
Bard	2
CardiacThoracic Systems	2
Intec Systems	2

Table 2.3 Regression Estimates of Determinants of Time To Funding

Variable	(1)	(2)	(3)
Spawn	-0.967***	-0.689**	-0.563*
	(.255)	(0.267)	(0.312)
Serial	-1.08***	-0.740**	-0.522*
	(0.282)	(0.289)	(0.302)
Constant	6.027	6.306	6.426
	(0.146)	(0.000)	(0.587)
Year Effects	N	Y	Y
Segment Effects	N	N	Y
N	174	174	174
R Squared	0.102	0.228	0.307

Robust standard errors in parentheses

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

Table 2.4 Semi-Log Regression Estimates of the Determinants of Last Round Of Outside Financing

Variable	(1)	(2)	(3)	(4)
Spawn	0.358***	0.372***	0.322**	0.283*
	(0.137)	(0.136)	(0.146)	(0.146)
Serial	0.232*	0.258*	0.210	0.175
	(0.135)	(0.140)	(0.142)	(0.149)
Constant	2.855	1.187	1.111	0.852
	(0.135)	(0.404)	(0.387)	(0.516)
Round Dummies	Y	Y	Y	Y
Year Effects	N	Y	Y	Y
Segment Effects	N	N	Y	Y
Firm Controls	N	N	N	Y
R-squared	0.69	0.7	0.72	0.73
Observations	191	191	191	191

Robust standard errors in parentheses
* significant at 10% level ** significant at 5% level; *** significant at 1% level

Table 2.5 Semi-Log Regression Estimates of the Determinants of Last Round Of Outside Financing

Variable	(1)	(2)	(3)	(4)
Spawn-Unrelated	0.323**	0.347**	0.307*	0.264
	(0.152)	(0.151)	(0.163)	(0.163)
Spawn-Related	0.436**	0.428**	0.36*	0.327
	(0.175)	(0.180)	(0.197)	(0.202)
Serial	0.231*	0.258*	0.211	0.175
	(0.136)	(0.141)	(0.143)	(0.149)
Constant	4.696	1.187	1.109	0.855
	(0.136)	(0.408)	(0.390)	(0.52)
Round Dummies	Y	Y	Y	Y
Year Effects	N	Y	Y	Y
Segment Effects	N	N	Y	Y
Firm Controls	N	N	N	Y
R-squared	0.69	0.7	0.72	0.73
Observations	191	191	191	191

Robust standard errors in parentheses
* significant at 10% level; ** significant at 5% level *** significant at 1% level

Table 2.6 Semi-Log Regression Estimates of the Determinants of Last Round of Outside Financing

Variable	(1)	(2)	(3)	(4)
Spawn Simple	0.359**	0.373***	0.322**	0.285*
	(0.142)	(0.141)	(0.150)	(0.152)
Spawn-Tech Inheritor	0.353	0.364	0.323	0.268
	(0.222)	(0.229)	(0.257)	(0.273)
Serial	0.232*	0.258*	0.21	0.175
	(0.136)	(0.141)	(0.143)	(0.149)
Constant	3.924	1.187	1.111	0.849
	(0.222)	(0.407)	(0.389)	(0.526)
Round Dummies	Y	Y	Y	Y
Year Effects	N	Y	Y	Y
Segment Effects	N	N	Y	Y
Firm Controls	N	N	N	Y
R-squared	0.69	0.7	0.72	0.73
Observations	191	191	191	191

Robust standard errors in parentheses

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

Table 2.7 Probit Regression Estimates of the Determinants of FDA Product Approval

Variable	(1)	(2)	(3)
Spawn-Unrelated	-0.015	-0.093	-0.105
	(0.095)	(0.101)	(0.103)
Spawn-Related	0.422***	0.365***	0.361***
	(0.098)	(0.117)	(0.121)
Serial	0.038	-0.029	-0.041
	(0.094)	(0.106)	(0.109)
FinancingRounds	0.035**	0.029*	0.022
	(0.015)	(0.017)	(0.017)
Segment Effects	N	Y	Y
Firm Controls	N	N	Y
Pseudo R Squared	0.08	0.17	0.20
Observations	191	184	184

Robust standard errors in parentheses (Marginal Effects Reported)
* significant at 10% level; ** significant at 5% level *** significant at 1% level

Table 2.8 Hazard Model Time to Product Approval

Variable	(1)	(2)
Spawn Unrelated	-0.069	-0.242
	(0.315)	(0.322)
Spawn Related	0.888***	0.512**
	(0.291)	(0.299)
Serial	0.172	-0.041
	(0.302)	(0.322)
Segment Effects	N	Y
R-squared	0.02	0.69
Observations	191	191

Robust standard errors in parentheses

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

Table 2.9 Patent Analysis

	Spawn	Control
Total Number of Patents	792	712
Top Ten Cited Orgs	Target Therapeutics, Inc. Cordis Corporation Advanced Cardiovascular Systems, Inc. Devices for Vascular Interventions Olympus Optical Co; Ltd Sci-Med Life Systems, Inc. Medtronic, Inc. Danek Medical, Inc. Cardiovascular Imaging Systems, Inc. Cook Inc.	Advanced Cardiovascular Systems, Inc. Medtronic Inc. United States Surgical Corporation Cordis Corporation Olympus Optical Co; Ltd Everest Medical Corporation 3M Baxter International Inc. ValleyLab, Inc. General Electric Company
Citation Rate to Parent	1.040%	0.906%

Table 2.10 Semi-Log Regression Estimates of the Determinants of Last Round of Outside Financing

Variable	(1)	(2)	(3)	(4)
Serial Spawn	0.394**	0.438**	0.392**	0.339*
	(0.166)	(0.170)	(0.189)	(0.188)
Spawn	0.332**	0.324**	0.276	0.248
	(0.159)	(0.156)	(0.169)	(0.174)
Serial	0.232*	0.259*	0.215	0.18
	(0.136)	(0.140)	(0.143)	(0.150)
Constant	2.855	1.162	1.083	0.835
	(0.136)	(0.413)	(0.401)	(0.529)
Round Dummies	Y	Y	Y	Y
Year Effects	N	Y	Y	Y
Segment Effects	N	N	Y	Y
Firm Controls	N	N	N	Y
R-squared	0.69	0.7	0.72	0.73
Observations	191	191	191	191

Robust standard errors in parentheses

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

Figure 2.1 Interview Protocol

(see Phillips and Fernandes, 2003)

Firm History

When was the firm founded and when was it incorporated?
Why did you decide to start this firm?
Did you begin with a written business plan?
How many employees do you currently have and can you describe how your organization is structured?
What are your current revenues and how fast are you growing?

The Entrepreneurial Process

Is this your first entrepreneurial venture?
Tell me about the process of raising capital. What were your biggest strengths? Weaknesses?
Tell me about your experience identifying potential partners and suppliers?
What are the reasons you think that your firm can exploit this entrepreneurial opportunity better than large, established firms or your competitors?
What type of work experience do you value in hiring potential employees?

The Parent Firm

Where did you work immediately prior to this venture? How long were you employed there? Why did you decide to leave?
How would you describe the corporate culture of your previous employer and how does it differ from your current venture?
Does your firm compete directly or indirectly with your old firm?
How does your management of R&D differ from your previous employer?
In your opinion, what are the most crucial factors in identifying entrepreneurial opportunities in medical device?

Intellectual Property and Regulatory Process

How many products does the firm have?
How many patents does the firm currently have and when were they granted?
Were the inventors ever employed by your firm and if so, are they currently employed by your firm?
Where and when were the patented innovations first conceived?
Do your patents cite older patents from your previous employer? If so, why did you choose to commercialize this invention through a new firm rather than an established firm?
Could this innovation be developed in a large company? Why or why not?
How many times have you gone through the FDA approval process at your current firm? How important was your prior experience in managing this process?
Did you outsource the FDA approval process or handle it in-house?
Does your device qualify for Medicare reimbursement?
On a scale of 1-7, could you rate the following sources of innovation in the medical device industry, with 1 being insignificant and 7 being extremely critical?
1) University Research 2) Practicing Physicians 3) Medical device manufacturers 4) Medical device firm research and development employees
Is there anything else you think is important about the entrepreneurial process in the medical device industry that you would like to share?
Are there other entrepreneurs in medical device that you think may be interesting for me to interview?

Figure 2.2 Venture Capital Investments in Medical Device Industry 1995-2004

Source: PWC MoneyTree

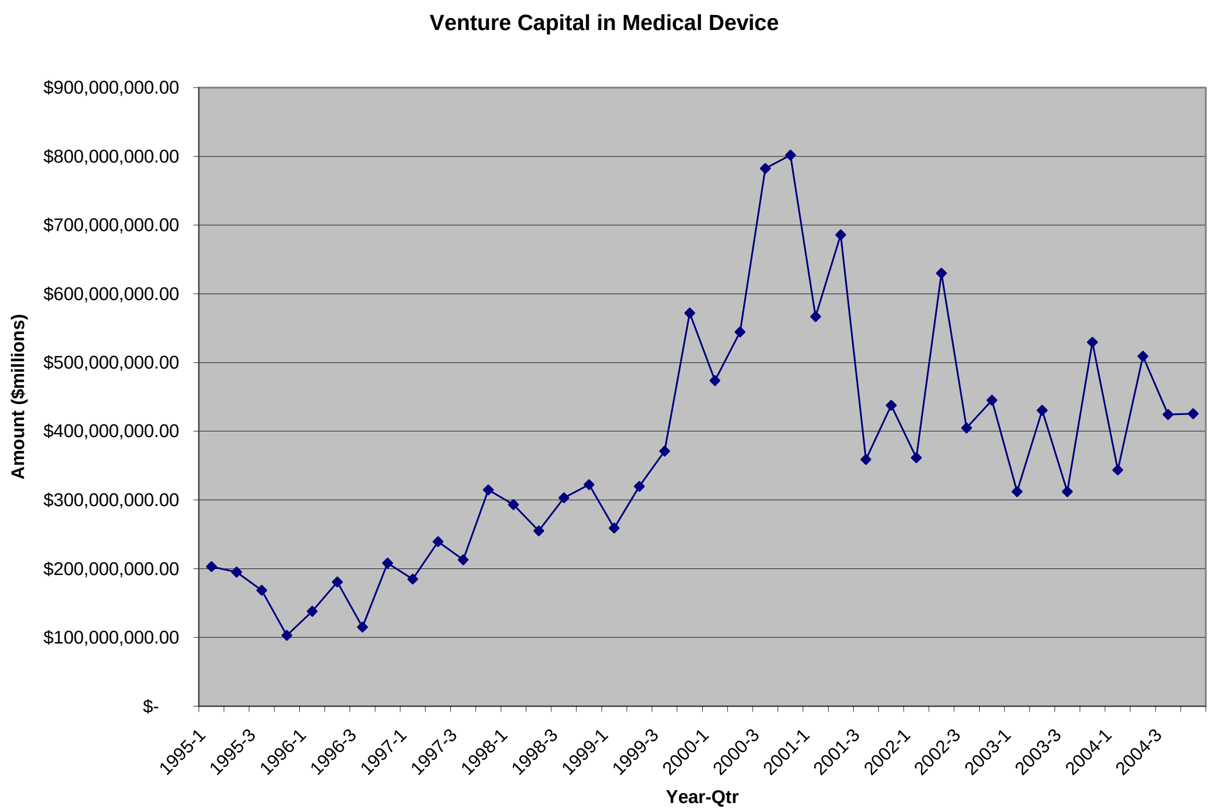


Figure 2.3 Venture Capital Deals in Medical Device Industry 1995-2004
Source: PWC MoneyTree

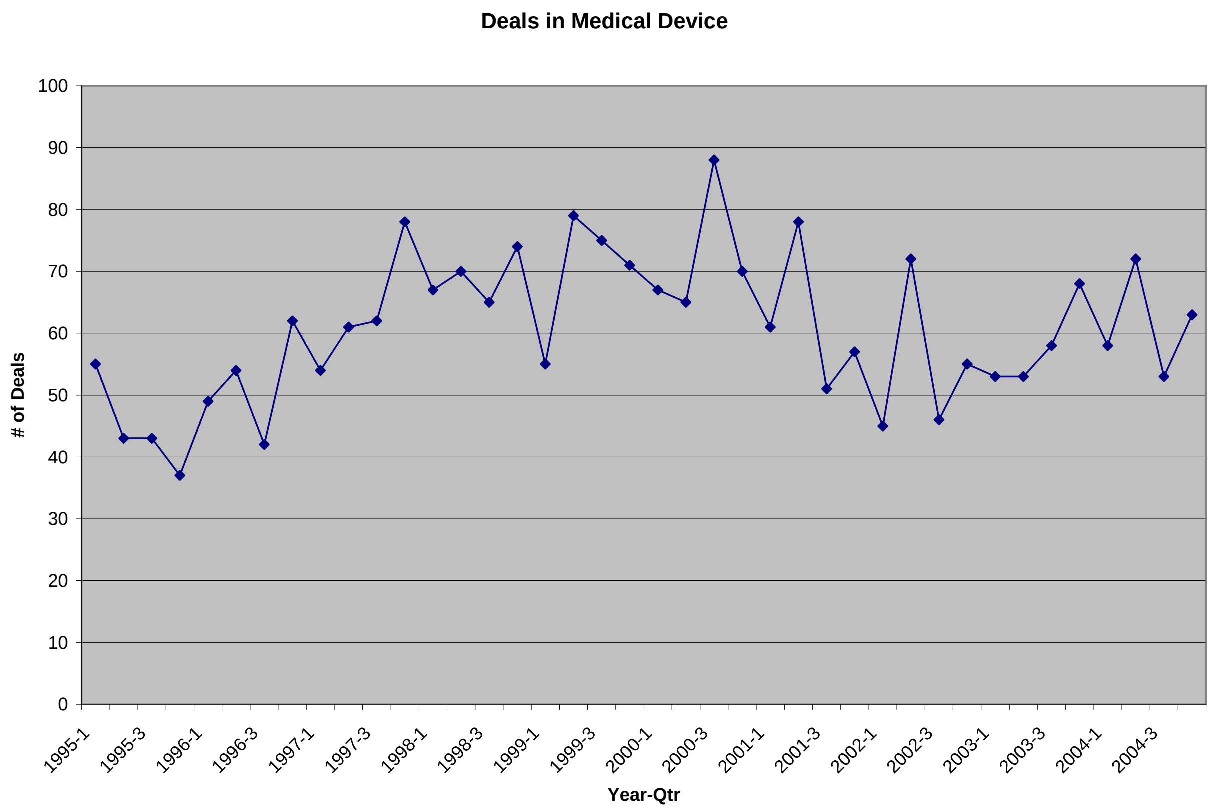


Figure 2.4 Venture Capital Investments Across All Industries (2004-Q2)

Venture Capital Investments Across All Industries (2004-Q2)

Industry	Amount	% of Total	# of Deals
Software	$1187M	20.22%	213
Biotechnology	$1009M	17.19%	89
Telecommunications	$516M	8.79%	61
Networking and Equipment	$513M	8.74%	46
Semiconductors	$505M	8.60%	59
Medical Devices and Equipment	$482M	8.22%	70
Media and Entertainment	$353M	6.02%	38
Computers and Peripherals	$239M	4.06%	36
Industrial/Energy	$206M	3.51%	38
Business Products and Services	$194M	3.30%	33
Electronics/Instrumentation	$185M	3.15%	23
IT Services	$174M	2.96%	36
Healthcare Services	$151M	2.57%	18
Financial Services	$118M	2.01%	17
Retailing/Distribution	$33M	0.56%	10
Consumer Products and Services	$6M	0.11%	7

Source: PWC MoneyTree

Figure 2.5 Medical Device Venture Capital Investments by Region (2004-Q2)

Region	Amount	% of Total	Deals
Silicon Valley	$234M	48.47%	21
San Diego	$82M	16.97%	8
New England	$35M	7.32%	6
North Central	$30M	6.23%	4
Northwest	$29M	6.01%	3
LA/Orange County	$19M	4.01%	3
Midwest	$19M	3.86%	6
Southeast	$14M	3.00%	4
Texas	$10M	2.03%	3
Colorado	$5M	0.94%	2
DC/Metroplex	$2M	0.49%	3
Philadelphia Metro	$2M	0.45%	3
NY Metro	$1M	0.19%	3
Upstate-NY	$0M	0.04%	1

Source: PWC MoneyTree

Chapter 3

3. How Responsible are Measures of Corporate Social Responsibility?

3.1 Introduction

In 2005, of the $24.4 trillion under professional management, at least $2 trillion was invested in socially responsible portfolios.[50] The huge amount of capital invested under the banner of social responsibility has drawn considerable attention from scholars, activists, managers, and policymakers. Socially responsible investing (SRI) has drawn praise from some advocates of corporate social responsibility, who believe that SRI can be effective in directing capital towards the most responsible firms while penalizing firms with poor social performance. Moreover, transparent and accurate social ratings of firms can help consumers and other stakeholders make more informed choices. Finally, public ratings could also encourage improvements in social responsibility, particularly among poor performers. Skeptics argue that the firms that rate the social performance of enterprise, referred to simply as "raters" in our study, cannot truly discern which firms are socially responsible, resulting in metrics that are invalid and misleading to stakeholders.[51] More specifically, critics like Paul Hawken argue that the various methodologies employed by socially responsible raters allow for almost any public firm to be considered for a SRI index. (Hawken, 2004)

For their part, academics have produced dozens of articles on SRI. (See recent review by Orlitsky et al; 2003.) Most recent research has examined whether socially responsible

---

[50] Social Investment Forum 2005 Report
[51] Entine, 2003

investing (SRI) affects investors and/or managers (see Waddock, 2003 for a review). A more fundamental question is whether commonly used indicators of social responsibility are valid measures of the social, environmental, ethical, governance, and sustainability performance of enterprises. If metrics are invalid, none of the hypothesized benefits of socially responsible investing can occur. In the worst-case scenario, if firms expend resources to achieve high scores on invalid metrics, then social welfare can decline as SRI becomes increasingly important. Thus, it is crucial to understand the validity of the metrics used by SRI raters. Unfortunately, no careful validation of SRI metrics has been conducted.

It is challenging to discuss the validity of SRI metrics without an agreed-on goal of "responsibility." When investors have varied social and ethical preferences, we expect socially motivated investors to choose quite different portfolios. Thus, some analysts have claimed SRI researchers should abandon the task of discussing the validity of metrics.[52] We analyze two aspects of validity that do not require a detailed consensus on what is "corporate social responsibility; convergent validity and predictive validity. By *convergent validity* we mean metrics that claim to measure the same construct are highly correlated. By *predictive validity* we mean an uncontroversial minimal level of social responsibility: Whether highly rated firms are less likely to be in a major scandal.

To test for convergent validity and predictive validity, we must first understand the organization of the socially responsible investing sector. After briefly discussing the major actors in the sector, we examine the theoretical motivations for studying

[52] http://oae.sagepub.com/cgi/reprint/16/3/381.pdf p. 387

convergent and predictive validity, followed by our hypotheses. Next, we discuss our datasets and methods, our results, and offer some concluding remarks.

3.2 The Organization of the SRI Sector

3.2.1 Investors

While most investors do not use social screens or concerns in choosing their portfolio, a potentially important minority does. Investors' motives for socially conscious investment can come in various forms and any particular investor can hold more than one motive. Investors might choose socially responsible companies because they associate social responsibility with better financial performance. Investors with consequentialist motives hold socially-screened investments to "invest for their own futures and a better world at the same time" (Entine 2003). It is likely such investors assume that their decision on holding stocks lowers the cost of capital for socially-desirable firms and raises the cost of capital for disfavored firms (in effect "punishing" less responsible firms). For example, marketing materials from socially responsible funds routinely claim that such investments will help improve the world.[53] On the other hand, investors with deontological motives

---

[53] KLD Homepage (http://www.kld.com/about/index.html) Last accessed October 2nd, 2006
"Those who invest in a socially responsible manner attempt to improve the world by investing in companies that function in an ethical manner. SRI is frequently described as the attempt to 'do good while doing well.' " http://www.investorhome.com/sri.htm
Amy Domini, founder and managing principal of Domini Social Investments and a passionate believer that social investing can improve the world.
http://www.gwsae.org/ExecutiveUpdate/2001/February/money.htm also at
http://www.centeronline.org/knowledge/article.cfm?ID=609&ContentProfileID=122789&Action=searching
Modern portfolio theory suggests that SRI will have a small effect on most firm's cost of capital (Beltratti, 2003). An exception is possible if SRI raters identify small niche firms that are otherwise "below the radar screen" of most investors (as in Merton 1987). SRI

consider it unethical to receive profits from sectors they consider evil or harmful. Such investors might make these portfolio choices even knowing that their decisions do not raise the cost of capital for firms the investors dislike.

Finally, the expressive motive for social investment involves investing in firms marked as "responsible" because such choices express the investors own social responsibility to both the investor and others. Expressive motives for SRI arise because, in our culture, many people perceive that "good" people act "socially responsibly." For example, 80% of Americans call themselves environmentalists. In addition, 75-80% say they would pay more for environmentally responsible products (although such products have a far lower share of the market). Thus, it is internally consistent for a good person to invest in a fund that calls itself "socially responsible."

An example of this expressive motive in another sphere would be a successful individual who earns a high salary and decides to volunteer his time to a charitable organization. If he only cared about the ends of the charitable organization, he would instead work a few extra hours at his highly compensated job and donate the money to the charity. However, there is something about the act of volunteering itself that it is important to his identity as the sort of person who volunteers.

---

ratings can have much larger effects if they affect consumers, employees, managers' self-image, regulators or other stakeholders. Allanovo uses KLD data and IdealsWork uses IRRC SRI data to provide information to consumers.

Importantly, variation in investors' social preferences should lead to a variety of social portfolios with different assets.[54] At the same time, when SRI screens and portfolios vary substantially when reflecting the *same* preferences, all or almost all of the portfolios have high measurement error. An important issue recurring in the discussion below is the extent to which variation in SRI portfolios is due to desirable variation in explicit goals or to undesirable variation in measurement.

Because few investors can measure social performance on their own and because there are economies of scale in such measurement, most investors use social rating agencies to gather information on the social performance of enterprises. These rating agencies either run a mutual fund themselves or design and license stock indexes that other financial firms use to market mutual funds to investors. In addition, mutual funds and other investors (such as pensions) purchase research and ratings from the agencies.

The use of raters, however, just pushes back the monitoring problem one level: How can investors measure whether rating agencies are measuring corporate social performance with high validity? Investors' problem monitoring advisors is familiar from the literature on mutual funds. In the case of mutual funds, however, past stock returns give a (noisy) signal of managerial quality (Berk and Green, 2004). For socially responsible funds, investors do not have such indicators of the validity of advice (other than studies such as this and Chatterji, et al.2006) Two related studies found limited evidence that KLD's

[54] Entine (2003) discusses the variation in screens directed at investors with different social preferences as a critique of SRI. We disagree and consider it a feature of market systems that people with different preferences can make different choices.

social ratings were correlated with Fortune's Most Admired Companies list. (Brown and Perry, 1994; Sharfman, 1992)

3.2.2 Raters

The rating agencies have two overlapping sets of goals. First, many firms enter the social rating industry in part to foster their vision of a better society. Thus, they would like to measure corporate social responsibility in ways that engender changes that the raters themselves find desirable. To achieve their own social goals, it is important the rating agencies have valid measures of corporate social performance – where "validity" reflects the correlation of the raters' social preferences and the rating agency's recommendations (perhaps also valuing the scope of investment funds the raters can influence). Second, the raters are competing in a marketplace for customers (mutual funds and investors), described above. The drive for survival and profits provides incentives for social raters to match potential investors' preferences, differentiate their product, and to appear competent to these investors.

In our study, we consider 4 raters that do not manage money, (KLD, FTSE4Good, DJSI, and Innovest) along with one rater that runs its own mutual fund. (Calvert) The first 4 raters sell licenses to fund managers, while investors can invest directly into Calvert's socially screened fund. This distinction may have an impact on the motivations of each rater, especially with regards to the financial performance. After all, Calvert may have stronger incentives to select companies that satisfy particular social criteria and also financially outperform the market. Still, in examining the criteria used by each rater in detail (discussed below), we find little difference in their stated goals. The fund managers that purchase licenses from KLD, FTSE4Good, DJSI, and Innovest presumably also care about financial performance. However, we will revisit these distinctions again when considering the empirical results.

3.2.3 Companies receiving ratings

The companies being rated also play an important role in providing raters with social performance data along access to the company for interviews and surveys. The raters we study examine a wide range of data including information companies report to governments such as data on toxic emissions or product safety, press reports, corporate social and environmental reports, and information received directly from the firm (typically in response to a survey or interview request). Firms have incentives to engage in "greenwashing" by trumping up modest achievements in environmental and social performance (Lyon and Maxwell, 2005). If firms provide incorrect or biased information to raters, ratings could exhibit poor predictive validity (that is, ratings do not predict future performance well). Raters may also exhibit low convergent validity if some raters are better at detecting greenwashing than others.

3.2.4 Perspectives on Convergent Validity

Even after many years of scholarly and practical work, it is not always clear how to measure the performance of an organization. Meyer and Gupta (1994) argue that 1) Performance measures are proliferating along with the performance measurement industry; 2) Performance measures are weakly correlated; and 3) There is frequent change in which performance measures corporate managers consider most salient. While they provide examples of these trends across several different empirical settings, their discussion of performance metrics for private sector firms is especially instructive. First, they note that there are several different performance metrics for firms, including accounting measures such as return on equity (ROE), return on investment (ROI), and return on assets (ROA), along with financial measures like return to shareholders. Some of these measures, among others, have been used as performance metrics in the empirical literature on business strategy. However, in several studies, these measures have been found to be weakly correlated or not correlated at all with each other. Why would this be the case? After all, we might intuitively assume that firms that performed well on one performance metric would also do well on another.

Meyer and Gupta (1994) assert that these results are driven by a complex process by which new performance measures are developed and implemented by corporate managers, and eventually lead to a reduction in variance of performance across organizations. This decrease in variance is due to positive learning or because organizations "game" the system to succeed on the dominant performance metric, while ignoring others. This trend is observed in diverse settings, including performance measures for hospitals, nuclear power plants, and baseball players. New performance metrics must then be introduced, because existing metrics have lost their ability to differentiate between good and bad performers. The new metrics are by design uncorrelated with the old metrics, starting the process all over again.

However, neo-institutional theory offers a different explanation that might be more applicable to a nascent industry such as socially responsible investing. The critical assumption in institutional theory is that organizations seek legitimacy in their environment, rather than strictly efficiency. (Scott, 1995; Staw and Epstein, 2000) The implication is that SRI rating organizations may adopt a practice, routine, or metric because another organization has already adopted it, rather than because it is efficient. This phenomenon is especially acute when the underlying mechanisms are not clear to managers. Staw and Epstein (2000) draw on prior work to make a point especially relevant to measurement of social responsibility:

"As March and Olsen (1976) noted, when technologies are poorly understood and organizations face problems with ambiguous causes and unclear solutions, copying other organizations (and their executives) may simply be a low-cost heuristic for finding useful solutions."

Staw and Epstein (2000) examine the adoption of popular management techniques and find that while the adoption of these techniques increases legitimacy, it does not lead to better economic performance. Other scholars have also modeled imitative behavior as a social force, where early adoption by high status actors may be especially important in creating information cascades. (Rao et al. 2001)

There has been other interesting research related to raters, ratings, and the information cascades that lead to herding behavior, especially in the economics and finance literature. Individuals and organizations are heavily influenced by the decisions of their peers, whether the choices are being made by securities analysts (Rao et al. 2001), physicians making a diagnosis (Bikhchandani et al. 1992), or voters choosing a candidate. (Bartels, 1988) Whether these choices are individually rational (Banerjee, 1992) or a product of a more complicated social process (Rao et al. 2001), convergence can occur even when individuals have limited information. (Bikhchandani et al. 1992).

In the area of finance specifically, Scharfstein and Stein (1990) find that investment managers will exhibit herd-like behavior and even ignore private information that would suggest divergence. This result is driven by the managerial concerns over their own reputations. Hong et al. (2000) find empirical evidence for the impact of career concerns on herding among sell-side securities analysts. Zitzewitz (2001) provides a different perspective on career concerns, arguing that certain analysts may have an incentive to exaggerate their differences from the herd. In particular, when their earnings are noisier, their expected career tenure shorter, and when they are "underrated" by the market, we would expect these analysts to exaggerate more. Thus, we may see anti-herding behavior in some cases as well.

3.2.5 Applications to SRI

The literature discussed above all relates to difficulties in measuring the performance of organizations. While Meyer and Gupta (1994) expect performance metrics to be weakly correlated, other work predicts considerable convergence among ratings due to herding or the social pressure to imitate competitors. Socially responsible investing ratings are an ideal empirical setting to test the convergent validity of ratings. There are several different raters and ratings, but the underlying construct of social performance is still somewhat vague.

While there has been some theoretical and empirical literature on convergence among analysts of firm financial performance, almost no prior work has examined convergence among analysts of firm social performance. The key difference between financial and social performance is that the underlying construct of social performance is more opaque, and there is considerable uncertainty over how to measure social performance. Those creating socially screened funds face severe constraints of limited information about corporate actions, bounded rationality, and missing evidence on the causal links between corporate actions and social outcomes. The social raters described above may have similar career concerns and incentives to exaggerate their ratings, but there is also considerable risk that they will simply measure social performance inaccurately. It is important to note however that much of the finance literature has focused on individual analysts with career concerns, while we examine ratings produced by firms. In this work, we intend to first assess the level of convergent validity among raters, and then separate out the divergence that is due to measurement error versus purposeful differentiation.

In a world where investors, companies and the raters themselves are uncertain about what domains of social responsibility are important and how to measure each domain, socially screened funds have strong incentives to follow the crowd through rational choice and are also vulnerable to social forces that lead to convergence. However, using the insights of Zitzewitz (2001), we might not expect complete agreement across all or perhaps even most firms. After all, there may be incentives to deviate from the crowd to send a signal of superior ability. Thus, social raters may include different firms in their indexes or rate firms much higher or lower than their competitors. In fact, at a very practical level, SRI funds also have incentives to have imperfect convergent validity (that is, to come up with quite different measures) as a differentiation strategy. Meyer and Gupta (1994) also find that prominent performance measures are often uncorrelated. (However, they were examining performance measures constructed by managers not by raters)

Since the socially responsible investing industry is still in an early stage of development, firms are struggling to find the right balance between rigorous measurement and distinctiveness. It is notable that raters make a significant effort to persuade potential investors that their methods and ratings are based on careful analysis and not simply imitating their competitors. For example, we see socially screened funds drawing from the established list of scientific procedures such as the use of multiple research methods: review of company documents; review of press reports; interviews, and often surveys. Their marketing literature stresses the care given to careful analysis of company's social record. Finally, they often describe their own services as similar to traditional financial research firms.

3.2.6 The Relationship to Financial Performance

As discussed above, much of the empirical work on corporate social responsibility has focused on whether socially responsible companies experience better financial performance. In fact, according to Margolis and Walsh (2003), over 100 academic studies have investigated the link between financial performance and social responsibility. Many of the studies have found a positive link, but the direction of causality remains unclear. After all, socially responsible practices could lead to better financial performance, but firms that succeed financially may also have more money to spend on social responsibility and promoting their good works. There are also significant omitted variable bias issues with most empirical studies, since good managers might be causing both superior financial performance and superior social performance. For a more detailed discussion of these issues see Chatterji et al; (2006).

In this paper, we only focus on whether the ratings assigned to firms by socially responsible investing companies are correlated across raters and whether they predict social performance. As discussed in Chatterji et al; (2006), establishing the link between social metrics and social performance is a crucial intermediate step towards gaining insight into whether social metrics will predict financial performance. After all, if the link between social metrics and actual social performance is weak, it is unclear why social metrics would be an appropriate measure in the empirical studies relating social responsibility and financial performance. In other words, if existing social metrics are imperfect proxy variables for actual social responsibility, we should interpret the results of previous empirical studies with caution.

It is important to note that some social investors may believe that corporate social responsibility is related to financial performance (if only because more responsible companies avoid costly scandals), implying that the ratings we consider in this study may be viewed by some as important lead indicators of financial success. Thus, many investors, including institutional investors, are also interested in using SRI ratings to identify firms that will financially outperform the market. The Investor Responsibility Research Center (IRRC) provides a corporate governance rating for firms, based on a methodology developed in Gompers et al. (2003). The score is a simple count (0-24) of the number of anti-takeover provisions that management has in place, such as “poison pills” and “golden parachutes”. (Gompers et al. 2003) Each provision implies greater entrenchment for managers, so higher scores imply worse governance. Gompers et al. (2003) find that better governance is associated with higher firm value and higher stock returns. Importantly, there is no direct evidence that these better governed firms experience better social performance. We will examine this issue empirically in the discussion of predictive validity.

3.2.7 Do Social Investors Expect Convergent Validity?

Socially responsible investors are not a monolithic group and may be looking for different types of validity. Investors with consequentialist motives are concerned about identifying firms with high and with low levels of social responsibility, so their investments can support socially responsible firms. Deontological investors will likely have strong aversions to specific sectors, so their concern with validity will emphasize identifying a subset of firms they consider evil. Finally, expressive investors are concerned with the act of “responsible” investing itself, and will be satisfied with a high level of perceived validity (by themselves and others) – regardless of true validity. Interestingly, after examining the criteria used by social raters, it is not clear that a particular rater is targeting just one type of investor. Rather, the raters appear to be trying to attract each of investors described above.

The several SRI rating agencies we study do not purport to measure precisely the same constructs. At the same time, the basic description of their products is remarkably similar, leading investors to expect fairly strong convergent validity (with exceptions we note below). For example, compare these summaries on their web sites:

1. find companies that are particularly strong models of corporate behavior
2. broad-based, rigorously constructed benchmark for measuring the performance of US-based socially responsible companies
3. leading sustainability-driven companies
4. balance the level of environmentally and socially driven investment risk with the companies' managerial and financial capacity to manage that risk successfully and profitably into the future
5. Identify and invest in companies that meet globally recognized corporate responsibility standards [and] contribute to the development of responsible business practice.

We suspect to most clients, firms that are "strong models of corporate behavior" or "socially responsible companies" are also leading "sustainability-driven companies" and "meet globally recognized corporate responsibility standards." Having spent countless hours reviewing these websites and related materials, we ourselves have trouble identifying the sources (1: KLD; 2: Calvert; 3: Dow Jones Sustainability Indexes; 4: Innovest; and 5: FTSE4Good).[55]

The major social raters also share many common screens. For example, KLD, Calvert, FTSE4Good all screen out “sin” industries such as alcohol, tobacco, military, and gambling. This fact provides more reason to believe that they are attempting to measure similar constructs and that investors are expecting convergent validity among the various ratings.

3.3 Hypotheses

3.3.1 Convergent validity

SRI rating schemes differ along a number of dimensions including what domains of social activity they rate (e.g., environment, human rights) and what activities they rate in each domain, the relative importance they give to different domains, and sampling frames. We carefully adjust for the systematic differences leaving measures that claim to identify similar features of companies.

---

[55] 1. KLD [http://www.kld.com/benchmarks/dsifaq.html accessed July 18, 2005]”
2: Calvert [www.calvert.com]
3. Dow Jones Sustainability Indexes [http://www.sustainability-indexes.com/ accessed July 18, 2005]
4. Innovest's mission statement
5. {FTSE4Good} Key Objectives

If the two measures have the same validity (that is, correlation with the true construct), the correlation between them is also equal to their validity. To understand this intuition, consider the SAT. If the math and reading portion of the SAT do not predict each other well, they cannot both strongly predict success in college. In general, the correlation between two measures purporting to measure the same construct is the upper bound on one measure's validity. The other could be highly valid or even completely invalid (uncorrelated with the true construct). Thus, if we find low correlations between two measures of social responsibility, we know that one or both measure is not valid, but we do not know which one. Even if the correlation is high, if the correlation is due to measurement error shared by the two rating agencies, high convergent validity could also be consistent with low validity. Thus, convergent validity is a necessary but not sufficient test for a set of SRI indices to possess if they are all measuring similar constructs that are all highly valid.

As discussed above, the prior literature has expressed two conflicting views of convergent validity across performance metrics. On one hand, Meyer and Gupta (1994) find low correlations among financial performance metrics. Alternatively, other work in economics and sociology expects herding by raters due to rational career concerns or the social forces that encourage imitation.

> H1: SRI raters will have high convergent validity prior to adjusting for explicit differences in methods and goals.

H1': SRI raters will have low convergent validity prior to adjusting for explicit differences in methods and goals.

After accounting for explicit differences in the rating methods, we might expect ratings to converge unless there is significant measurement error. After all, if designing portfolios for well-informed heterogeneous instrumental (consequentialist + deontological + expressive) investors drove differences in ratings, then these differences would disappear after adjusting for differences. On the other hand, if significant measurement error exists or if the raters are influenced by the desire to be distinctive or some other bias, we might still observe differences even after adjustments. It is important to note that it will be difficult to infer from the data exactly where these differences originate from, but we will offer some preliminary thoughts in the discussion section.

H2: Adjusting for explicit differences in methods and goals should greatly reduce differences in ratings compared to not adjusting for explicit differences in methods and goals.

H2': Large differences in ratings remain even *after* adjusting for explicit differences in methods and goals.

3.3.2 Predictive validity

While various social responsibility ratings may be correlated or not, it is also crucial to consider whether the ratings actually capture the underlying construct of social responsibility. Simply put, do social metrics predict future social performance?

Around the year 2000, Enron, WorldCom, and Tyco went from major companies few in the public had heard of to household names. It is difficult to know exactly what investors want when they shop for a "socially responsible" company. Nevertheless, it is safe to assume few socially minded investors do want to invest in enterprises that commit massive fraud against investors (e.g., Enron and Tyco), illegally exploit consumers (e.g., Enron in California's electricity crisis), kill thousands of nearby residents (Union Carbide in Bhopal), destroy a local ecosystem (Exxon Valdez), or discriminate massively against female employees (State Farm's anti-discrimination settlement of over $100 million). While no investor wants to invest in such companies if the resulting scandal destroys stock market wealth, presumably most socially-minded investors also place value on not harming other stakeholders, which is encompassed by most definitions of social responsibility.

One recent analysis of the growth of SRI has claimed that corporate scandals and the resulting erosion of trust in top management is a major driver of SRI. "Many investors are attracted to an investment process based on research that goes deeper and considers qualitative information designed to identify corporate character."[56] At the same time,

---

- [56] Steven J. Schueth, "SRI in the US", http://www.firstaffirmative.com/news/sriArticle.html. Similar claims are made at http://www.novonordisk.com/sustainability/socially_responsible_investment/socially_responsible_investment.asp
- "As Markets Reel From Corporate Scandals, Screened Mutual Funds Stand Strong and Push For Greater Corporate Responsibility" http://www.greenmoneyjournal.com/index.mpl.
- The recent crisis in confidence elicited by the string of corporate governance scandals from Enron to WorldCom may inspire investors to seek investments that have demonstrable integrity. http://www.msu.edu/~divest/faqinvest.html
- "The Ethical Investor (1995) changed my views on what could done to harness investment to social responsibility in the UK context. This book now moves the agenda to the world stage and is

critics of SRI rating firms such as John Entine have noted, “Many of the recently disgraced companies, including Anderson, Enron, WorldCom, Adelphia, Tyco, and Tenet Healthcare were favorites of social funds.” (Entine 2003: 357)

If social raters are simply “jumping on the bandwagon” described in Staw and Epstein (2000), we would expect the ratings to have very little predictive validity in terms of actual social performance. Similar to adopting popular management techniques that do not yield actual economic benefits, raters may be imitating each others’ ratings even though the ratings themselves do not accurately capture the social responsibility of the rated firm. In this case, we would expect low predictive validity for social metrics. Meyer and Gupta (1994) argue that measures because less effective over time, so we might also expect the predictive validity of ratings to decline over time but we do not have the data to test this secondary hypothesis.

H3: SRI ratings will have low predictive validity.

3.4 Data

We use data from several of the major raters. Each rater uses some of the same sources, so we will focus on the differences between them below. For the most part, all firms use information from companies, government, and other stakeholders in determining their social ratings.

---

essential reading for all those who can see the need to harness capitalism to SRI objectives in the post Sept. 11th world, and that after the Enron and Worldcom scandals, SRI can deliver the better world we need" Tony Colman MP, House of Commons International Development Select Committee http://www.bookworkz.com/construction/finance/0471499536.html

3.4.1Social Indices

KLD

We first examine KLD social ratings. KLD is one of the oldest (1988) and most influential social rating agencies with $8 billion invested in funds based on its index.[57] KLD's objective is "to provide global research and index products to facilitate the integration of environmental, social and governance factors into the investment process. "KLD evaluates seven domains of the social performance of companies: community relations, corporate governance, diversity, employee relations, environment, human rights, and product quality and safety. Within each domain, KLD assigns a positive score (Strength) or a negative score (Concern) based on specific rating criteria. The strengths and concerns do not cancel each other out but are rather reported together. KLD has explicit screens for alcohol, tobacco, nuclear, firearm, military, and gambling involvement. KLD reveals no specific weights on its various sub-scores, although we are able to glean some insights into their weights in our analysis. KLD does not rank firms relative to their industry averages.

KLD researchers study company, government, media and NGO reports to rate over 3000 companies each year. In this study, we will use two of KLD's indexes—Domini 400 and the Large Cap Social Index (LCS). To construct the Domini 400, KLD begins with the S&P 500 and eventually chooses the top 250 firms from the index along with 150 other socially responsible firms to comprise its Domini 400 index. To construct LCS, KLD

[57] Kinder (2005)

begins with the Russell 1000 universe, and selects 662 firms using their ratings system for inclusion into the index.

Many have argued that (e.g., Waddock 1993) KLD creates the highest quality metrics of social responsibility. The data has been used by several academic research articles and considered as the standard for measuring corporate social performance in the academic literature We examine both membership in KLD's indexes and the detailed sub-scores KLD provides for each firm.

Calvert Social Index

Calvert is the only rater in our study that manages its own socially responsible fund. Calvert manages $12 billion in assets for over 400,000 investors. Calvert was founded in 1976 and offers over 30 different funds, including the Calvert Social Index. Calvert describes its social index as a "broad-based, rigorously constructed benchmark for measuring the performance of US-based socially responsible companies." Selected from the Russell 1000, Calvert selects approximately 600 firms for inclusion into its social index. These firms are evaluated on the following criteria: Products, Environment, Workplace, and Integrity. Calvert also reports ratings for the top 100 largest companies, rating the firms on a 1 to 5 scale across 5 categories, the Environment, Workplace, Business Practices, Human Rights, and Community Relations. Calvert uses many of the same explicit screens that KLD does and does not report explicit weights on the various social criteria. Calvert also rates firms according to average performance in their industry. Calvert keeps information on nearly 7000 companies, using Lexis Nexis, trade publications, government and NGO reports, and various other sources to formulate its decisions.

FTSE4Good

FTSE4Good was launched in 2001 by The Financial Times and the London Stock Exchange, and donates all of its license revenue ($1.6 million by 2006) to the UNICEF. One of FTSE4Good's primary objectives is "to provide a tool for responsible investors to identify and invest in companies that meet globally recognized corporate responsibility standards." Companies are judged on the following criteria: environment, stakeholder relationships, human rights, supply chain management, and "countering bribery". FTSE4Good also includes screens for tobacco, nuclear, and military concerns.

FTSE4Good divides industries into high, medium, and low impact sectors, and employs different criteria for each. For each criterion, FTSE4Good evaluates companies on policy, management, and reporting activities. Depending on the impact of their sector, firms have to meet a fraction of the recommended criteria to be included in one of FTSE4Good's indexes. In this study, we use FTSE4Good's US 100, which includes the 100-US based firms in the FTSE4Good index. FTSE4Good works with Ethical Investment Research Services to conduct its research. They collect information from annual reports, company websites, and other publicly available information.

Dow Jones Social Index (DJSI)

DJSI was launched in 1999 by Dow Jones Indexes, STOXX Limited, and SAM group and has sold 56 licenses in 14 nations. In sum, these licensees managed over $4 billon Euro in 2006.[58] DJSI's goal was to create the "world's first equity benchmark to track the financial performance of sustainability leaders on a global scale."

DJSI begins with a universe of 2,500 firms from the Dow Jones Global Index and aims to select the top 10% for the Sustainability World Index. DJSI divides firms into 58 industry sectors, and places 60% weight on general criteria and 40% on industry-specific criteria. They place equal weight across three broad categories, Economic, Environmental, and Social. Importantly, DJSI uses relative rankings by industry, so they are seeking to identify the top 10% in each industry. Thus, DJSI does not use the same screens as other raters, while instead trying to identify the best in class, even among "sin" industries like tobacco.

DJSI uses information from 4 sources: company questionnaire, company reports, media and stakeholders, and direct company engagement. DJSI only rates firms that respond to its questionnaire. DJSI works with Sustainable Asset Management (SAM), which employs 20 analysts who spend on average 2 days per company.

For the Dow Jones Social Index (DJSI) we examine over 80 US-based firms with explicit rankings ratings, as well as the other firms below that level. We rescaled the DJSI measures to a percentile of their industry. Recall that DJSI ranks the top 10 percent of

---

[58] Dow Jones Sustainability Indexes Homepage, (http://www.sustainability-index.com/), Last Accessed October 3rd, 2006)

each industry plus the highest runner-up. For example, if Dow Jones ranked 3 of 30 firms in an industry we would have relative information on 4 firms (the 3 ranked firms plus the runner up). The other 26 firms would be classed collectively as "not ranked,"

Innovest

Innovest was founded in 1995 and has licensees with $1.1 billion under management in 20 nations. Its web site explains: "At the heart of Innovest's analytical model is the attempt to balance the level of environmentally and socially driven investment risk with the companies' managerial and financial capacity to manage that risk successfully and profitably into the future."[59]

Innovest tracks 1750 companies world-wide on 120 individual factors. They focus on 4 areas : EcoValue (environmental issues), Human Capital, Stakeholder Capital, and Sustainable Governance. Innovest analyzes these factors in the context of how they affect financial performance. For the current study, we use the 18 US-based firms in Innovest's Top 100 leaders in sustainability. Innovest claims to have the largest number of analysts in the world and over 90% hold advanced degrees. Rather than using questionnaires, they interview executives.

3.4.2 Other Measures

IRRC index of weak governance

[59] Innovest Webpage, (www.innovestgroup.com) Last accessed October 2nd, 2006

In our predictive validity analysis, we use an additional explanatory variable—The governance index produced by IRRC. The governance index was created by Gompers, Ishii, and Metrick (2003) ranges from 0 to 24, with higher numbers being associated with more protection for managers and poorer governance. As the number of provisions entrenching managers increases, the G index increases. The Investor Responsibility Research Center (IRRC) provides this rating for 1990, 1993, 1995, 1998, and 2000.[60] The IRRC was founded in 1972 and provides information to more than 500 institutional investors on corporate governance and social responsibility.[61]

The governance score is different than the other ratings discussed above because it focuses directly on corporate governance and is much more transparent in its methodology. To examine convergent validity with other raters, we would need KLD, FTSE, DJSI, or Innovest sub-scores on corporate governance for the years covered by IRRC. Unfortunately, these ratings do not exist. However, we can still use the governance index in the predictive validity analysis. By comparing the KLD scores and IRRC governance scores across firms involved in major scandals, we gain insight into whether social ratings have the ability to separate responsible companies from irresponsible ones.

It is important to note that the governance score is provided by IRRC to institutional investors primarily as a tool to identify well managed companies that will presumably

[60] http://finance.wharton.upenn.edu/~metrick/governance.xls

[61] The Investor Responsibility Research Center Webpage, (http://irrc.com/index.html) Last accessed November 15th. 2006

deliver above average financial returns. Thus, the validity of the governance index will be an interesting complement to our results for KLD, which includes an explicit social vision in computing their ratings. If the governance index has high predictive validity, and the KLD score does not, it may because IRRC measures corporate governance in a direct and effective manner, and good corporate governance is associated with fewer scandals. (Or there could be an omitted variable driving both) If the KLD score performs better, it may imply that overall social responsibility (community giving, environmental impact, product safety, etc.) is a better predictor of future scandals than corporate governance alone.

3.4.3 Scandals

We first identified firms that underwent “major” scandals. We identified such scandals using several methods. Our main data source was the Corporate Crime Reporter’s list of the largest corporate crimes/scandals of 1990-1999. To identify more recent scandals we used web searches and the Equal Employment and Opportunity Commission (EEOC) website (looked through annual reports for lawsuits filed by EEOC. We did a web search with the terms “largest fine” and “largest penalty” coupled with phrases such as “environmental” and “discrimination.” Our second source was CEPD list of the 100 largest environmental spills and accidents from 1991 to 2003.

These methods are likely to identify major scandals as measured by the size of the fine and the publicity of the scandal. Thus, these methods emphasize scandals at large companies. Thus, in the analysis below we carefully control for company size, as $100 million fines are much more likely to arise at a company with 50,000 employees than with 500.

In addition we also used data from the General Accounting Office on 919 earnings restatements between 1997-2003.[62] The database collected restatements data from Lexis-Nexis searches, and focuses on restatements associated with "accounting irregularities". The restatements are classified into 13 categories.

3.4.4 Concerns

Our measures of scandals have several limitations. We falsely classify some firms as not having had scandals when unethical behavior is legal. For example, consider employment discrimination in the 1950s or massive campaign contributions to political parties in recent presidential elections that is intended to corrupt the political process. Similarly, we miss companies that just have not (yet) been caught.

In the other direction, we falsely classify some firms as having had scandals when their behavior was socially responsible. For example, not all restatements or accusations of criminal conduct indicate a lack of social responsibility. We do not know why such failures to be caught or such innocent restatements of earnings should be correlated with our measures of social responsibility or weak governance. Thus, the presence of these types of measurement error reduces the statistical power of our tests, but needs not create bias.

[62] http://www.gao.gov/new.items/d03395r.pdf

3.4.5 Summarizing explicit differences in method

While the raters are generally consistent in their philosophy and the firms they choose to rate, this section summarizes several important explicit differences. First, all of the social raters except KLD rank firms within industry or sector. KLD is also distinctive in that it uses S&P 500 membership as a significant factor in gaining membership.

The use of screens varies across the raters. KLD, Calvert and FTSE4Good use explicit screens, but not DJSI or Innovest. All three screen out firms with substantial military and tobacco interests, although the precise definitions of "substantial" vary and FTSE4Good only screens out nuclear weapons makers. KLD and Calvert screen out alcohol. KLD and FTSE4Good screen out firms with revenue from nuclear power

All of the indices cover similar high-level topics such as Environment, Workplace, Business Practices, Human Rights, and Community Relations (the five Calvert sub-scores). The precise items within each major sub-score vary across raters. Standard investors would not be able to easily see such items or their decision weights for most of the raters. (DJSI and FTSE4Good however, have their full weighting scheme on their websites). We had to negotiate and/or pay for access to company reports for some of the rating agencies.

The weighting schemes for the top-level sub-scores also vary. KLD has no explicit weights on sub-scores, but a committee reviews all the evidence and sub-scores and decides which companies to include. Calvert's funds are chosen similarly, with 7 different categories and no explicit weights detailed on their website; DJSI places explicit weights of 1/3 each on economic, environmental, and social sub-scores. FTSE4Good has weights that vary by industry (for example, with more weight on the environment in environmentally sensitive sectors), while Innovest scores firms in 36 different categories and uses different weighting schemes in each industry.

Ideally, explicit differences in goals and methods allow consumers to pick the index that suits their own preferences. Unfortunately, because many of the differences in methodology are subtle and difficult to understand from simply reading the websites, we find this doubtful. For a more complete treatment of the differences between the indexes, see Chatterji and Levine (2006)

3.5 Methods

We discuss methods to measure two forms of validity: how closely ratings correlate with each other (convergent validity) and how well they predict future scandals (predictive validity).

Issues in measuring convergent validity

There are two sets of problems in measuring the degree of convergent validity. First, there is no natural definition of “high” or “low” agreement between two ratings. This is a standard problem in examining correlations and validity metrics in other spheres.

We will use several objective benchmarks to describe convergent validity as high or low compared to objective measures. Unfortunately, none of these metrics is fully satisfactory. To get a feel for what our measures mean substantively, a correlation below 0.5 shows that when one rater finds a firm two standard deviations above average, another rater finds it less than one standard deviation above average. If they are rating the same construct, it is difficult to consider that level of agreement as very high. Nevertheless, each reader will need to decide if the level of convergent validity is high or low in economic terms.

A second (but related) problem is that there are numerous measures of similarity among the discrete and continuous measures we analyze. Each of these measures, in turn, has problems in capturing the substantive meaning of convergent validity. For example, we examine the extent to which memberships in the several SRI indices overlap. A measure of the share of overlapping membership can be misleading if one index covers only a few firms. For example, if one index includes 500 firms from a universe of 1000 and second index includes only 10 of that universe, it would be surprising if almost all of the second index's members were *not* in the top half of the first index. Substantively, it would indicate that the second index's top 1% (roughly 3 standard deviation outliers) did not all fall in the top half of the first index. Thus, we present both the raw figures on overlap and also measures that are invariant to the number of members in each index

We also perform several tests of the statistical significance of the overlap of memberships and correlation across measures. Statistical significance can be misleadingly encouraging about the economic importance of a relationship, in that it tests a null hypothesis of *zero* relation between two measures of social responsibly. Convergent validity requires a *strong* relation, not just one different from zero. At the same time, statistical significance can be misleadingly discouraging about the strength of convergent validity for indices that cover only a few firms. Specifically, two indices that have a substantively large correlation may fail to be statistically significantly related if one or both has too few members.

Several of our correlations have a natural measure of what we mean as “very strong” convergent validity. Specifically, we have multiple sub-scores for KLD firms and for a subset of the Calvert firms. We can use the KLD sub-scores to predict membership in KLD’s LCS index and the Calvert sub-scores to predict membership in the Calvert fund. We cannot expect the KLD sub-scores to do a better job predicting membership in other indices than they do in KLD’s own LCS index. Thus, the predictive power of KLD sub-scores on LCS membership and of Calvert sub-scores on Calvert membership provides a meaningful upper bound on what we can hope for as convergent validity when KLD or Calvert sub-scores predict membership in other companies’ indices.

3.5.1 Membership

We first examine the overlap of the several indices. A problem with this method is that the overlap is sensitive to the size of the various indices. In addition, overlap is not presented in units familiar to most social scientists. We can get a better feel for the quantitative magnitude by switching to correlations adjusted for the dichotomous nature of the data. We begin by assuming a standard measurement model:

$$R_{ij} = b\ T_i + e_{ij} \qquad (1)$$

where:

T_i is the unobserved (latent) true level of social responsibility of firm *i*;
R_{ij} is the unobserved continuous rating given by an SRI firm *j* of firm *i*'s true level of responsibility;
b is a regression coefficient; and
e_{ij} is rater *j*'s measurement error and idiosyncratic definitions of "social responsibility."

We assume that e_{ij} is normally distributed and independent across raters and firms and that the true level of social responsibility of firms is also distributed normally. Important realistic extensions address correlated measurement error; discussed below. We assume that measurement error of different raters has identical variance, which we normalize to unity. Without loss of generality we normalize the mean true responsibility level $T_i = 0$.

For most of our raters, we observe only the discrete measure of whether SRI rater *j* has firm *i* in its membership: M_{ij}. We assume that the discrete membership rating M_{ij} equals one when the unobserved continuous rating R_{ij} is above SRI rating agency *j*'s cutoff for membership ($Cutoff_j$):

$$M_{ij} = 1 \text{ if } R_{ij} > Cutoff_j\text{, and 0 otherwise.}$$

In this model variation in $Cutoff_j$ may be driven by each rating agency's desired size for its membership. (It could also be driven by a fund manager wanting to manage a larger fund and take in a higher management fee) We can then use maximum likelihood techniques to estimate the correlation of two rating agencies' unobserved continuous ratings (that is, the squared coefficient, *b*). These estimated correlations are known as tetrachoric correlations.

3.5.2 Continuous Scores

We combined KLD's domain-specific sub-scores into an index of predicted KLD LCS membership and use Calvert's similar 5 scores to create an index of predicted membership in the Calvert Social Index. To identify the weights to use for this index, we estimate a logit regression:

$$\text{Membership in relevant Index} = F(\Sigma_i \, \beta_i \text{ sub-score}_{fi})$$

Here F(.) is the logit function and the β_i variables capture the importance of each sub-score on predicting membership in KLD's LCS or the Calvert Index. We refer to the predicted probability of membership as the KLD Score or Calvert Score of a firm.

That is, the KLD or Calvert Score is a weighted average of the KLD or Calvert sub-scores transformed to range from zero to unity. The weights are chosen to best predict membership in the relevant index, and the units of each Score are the predicted probability of index membership. .

3.5.3Adjusting for explicit differences

Because KLD include several explicit screens, we create 4 "styles" of KLD scores to adjust for explicit differences among the indices. The first, "KLD style" is produced using the logit of the KLD sub-scores where the dependent variable is membership in KLD's LCS index. For those few sub-scores that perfectly predicted success or failure, we assigned a coefficient 1.5 times as large as the next positive or negative (whichever appropriate) coefficient in the logit estimation.

The next 3 KLD scores, "Calvert style", "FTSE style", "DJSI/Innovest style" are computed the same as above except we consider the screens that are common with KLD and the focal index. Companies that are screened out are assigned a Score of zero. Furthermore, we include industry dummies when calculating these 3 scores, because KLD is the only index in our sample that does not norm their ratings by industry.

3.5.4 Predictive Validity

For each scandal firm we matched a firm in the same 2-digit SIC industry code that was closest in employment 2 years prior to the scandal. We balanced the sample so matches were as likely to be larger or smaller than scandal firms. Our tests of predictive validity, then, involve asking whether scandal firms were systematically less likely to be in KLD's Domini 400 and had systematically higher indices of weak governance. We use the Domini 400 because it is more selective index than the LCS. The Domini 400 criteria should allow KLD to better differentiate between responsible and irresponsible firms.

3.6 Results

3.6.1 Convergent Validity

We first discuss overlap among the memberships of the various SRI indices and correlations among scores and sub-scores. In all cases, recall that small indices' overlap with large indices can be misleadingly high, while their statistical significance can be misleadingly low. In the next section we examine correlations among the several continuous measures we construct as well as the relations between discrete measures of index membership and the several continuous measures. We conclude this section by examining the extent to which correlations are low due to purposeful differences in index measurement that we are able to adjust for versus other sources of divergence (either purposeful or erroneous).

3.6.2 Overlaps of membership

In this section we describe how well the memberships of the different indices correlate. By comparing which firms are included in various indexes, we will test for a minimum level of convergent validity. In the next section, we will turn to the actual score that the rater assigns the firm. For each pair we first describe the overlap of memberships and then the tetrachoric correlation in implied underlying social ratings (Table 3.2 and Appendix 3.1). The correlations' track the pattern of membership overlap (from which they are derived), although the correlations are designed not to respond to changes in relative index size.

KLD LCS membership correlated fairly well with Calvert, but not the other indexes. Specifically, 89% of the 493 Calvert members, but only 47% of the 490 Calvert nonmembers, are in the LCS. 79% of the FTSE4Good top 78 are in LCS, and 65% of the 513 FTSE nonmembers. 65% of DJSI's 80 members and waitlist companies are in LCS, as opposed to 60% of the DJSI "others." 47% of the 18 Innovest elite and an even greater 67% of the 542 non-elite are in LCS. The tetrachoric correlations of LCS membership and the other indices follow a similar pattern of 0.66 with Calvert, 0.01(not significant) with DJSI, 0.22 with FTSE4Good, and -0.22(not significant) with Innovest. With the exception of Calvert, KLD has little overlap with the other major social indexes.

Calvert had a weaker relationship with most of the other indices than did KLD. Specifically, 61% of the FTSE4Good's top 77 firms and 51% of the 501 FTSE "others" are members of Calvert, corresponding to a tetrachoric correlations of 0.13 (not significant). A higher 57% of the 77 DJSI members and waitlist firms, and 53% of the 834 others are in Calvert, with a tetrachoric correlation of 0.07 (not significant). Finally, 59% of the 17 Innovest elite and 51% of the 485 non-elite are in Calvert, implying an underlying correlation of ratings of 0.09 (not significant).

Innovest's elite are not statistically significantly related to FTSE4Good or DJSI. Five of Innovest's 14 elite are in the 75 FTSE4Good, while 9 are in the 317 nonmembers. This relationship is not statistically significant in part due to the small sample size of Innovest elite – as the rate of FTSE membership is more than twice as high for Innovest elite as for others (7% vs. 3%). At the same time, when Innovest is only selecting 14 firms from FTSE4Good's universe, it is not impressive that less than a third are in FTSE4Good's top quarter. Eleven percent of DJSI's 79 members and waitlist are in Innovest's elite 18. This share is small, but higher than the 2% of DJSI "others." Finally, 40% of the 67 DJSI members and waitlist are in FTSE4Good top 100, as opposed 10% of the 503 DJSI "others". DJSI has statistically significant tetracorrelations with both FTSE4Good (0.60) and with Innovest (0.71).

To summarize these results, overall, almost all the relations are essentially all positive (other than the LCS-Innovest correlation), and about half are statistically significantly. The mean of the ten distinct tetrachoric correlations is 0.23. There is not strong evidence that all the rating agencies are measuring a single construct with high accuracy. This result does not support Hypothesis 1 in that there is low convergent validity between memberships in various indexes.

These modest correlations are consistent with explicit differences among the methods used by raters – the hypothesis we turn to next. If one or two rating agencies had a distinctive approach to rating then we would expect them to be less strongly correlated with the other indices than they are with each other. In fact, there is also no evidence in the data that one or two rating agencies has a distinctive approach to rating. We find each rating agency's membership has a mean tetrachoric correlation with the other agencies of 0.17 to 0.29. Finally, there is no evidence that a one or two raters has a distinctive competence in rating a single agreed-on definition of "responsibility." Such a rater would have ratings consistently more highly correlated with other raters than average.

3.6.3 Using continuous scores

In this section we examine the relationship among rating agencies using continuous scores computed from sub-scores for KLD and Calvert and based on explicit within-industry rankings for DJSI (although for DJSI rankings are only available for the top 10% plus one firm per industry).

3.6.4 Estimating the KLD & Calvert Scores

We begin with logit regressions using KLD and Calvert sub-scores to predict membership in the KLD LCS and Calvert indexes. Because the KLD Score and Calvert Score are predicted probabilities, their potential range is from zero to unity. In the baseline specification with 63 sub-scores, the mean KLD Score is .204 and the standard deviation is .231

3.6.5 Correlations among continuous scores

The KLD Score estimated based on sub-scores is correlated 0.49 with the Calvert Score (Table 3.3, col. 1). Results for the DJSI rating are less encouraging. DJSI ratings are available only for the top 10% of each industry plus one runner-up per industry. As such, any correlations will be reduced by the restriction of range. Within that upper strata, the correlation with KLD Scores is only 0.03 (n.s.) (Table 3.3, col. 1). Adjusting for screens and industry increases the correlation (as theory suggests it should). These findings provide no evidence that within the upper strata, DJSI and KLD can distinguish which are the best among the best.

Finally, the correlation between Calvert Scores and DJSI Rankings is a tiny 0.0073 (n=32). The sample is small and there is restriction of range; nevertheless, this correlation is surprisingly close to zero. In short, the continuous measures of social responsibility correlate well for KLD and Calvert, but poorly for either with DJSI. Below we examine whether the correlations rise if we standardize the Scores to adjust for explicit differences in each raters' goals and methods.

3.6.6 Are differences in rankings due to explicit differences in method?

In this section we examine whether explicit differences in method, actual use of screens, and implicit differences in weights on sub-scores account for divergences in membership and ratings across the several indices. This calculation is most straightforward if we examine the relation between KLD Scores and the other indices, as KLD has the most distinctive set of criteria. First, KLD is the only index that rates firms across industries; the others all claim to be set relative to industry benchmarks. Second, KLD has more explicit screens than the other funds.

The results reported in Table 3.3 show that there is no systematic increase in the correlation of KLD scores with membership as we adjust the KLD scores to more closely resemble the methods used by the other indices. Across Calvert, FTSE4Good, DJSI, and Innovest, we see no significant improvement when we adjust scores to reflect explicit differences in method. For example, if the average KLD style Score for the firms that make the Calvert index is 0.30 higher than those firms that do not, we would expect a higher gap when we use the Calvert style score. (because the maximum possible gaps are equal) As seen in Table 3.4, the gap goes to 0.29 when we adjust for Calvert methods and screens. Looking at the changes in the absolute gaps and maximum gaps when we adjust for the methods and screens, the same general pattern holds, leading us to conclude that significant measurement error likely exists in the social ratings industry.

The fit of FTSE4Good and DJSI go up slightly, the fit of Calvert and Innovest decline, and none of the changes are statistically significant. In short, the fairly good fit with

Calvert and relatively low fit with the other indices is not improved if we follow the other indices in the screens they use (or do not use) of if we follow the other indices in norming scores within each industry.

3.6.7 Predictive Validity

Now that we have considered convergent validity, we evaluate the predictive validity of the ratings discussed above. In this section we test whether KLD membership and ratings or the IRRC index of weak governance "G" predicts scandals two years later.

3.6.8 Does Domini 400 membership or KLD sub-scores predict fewer scandals?

The results on predictive validity are fairly clear. We identified 126 scandals in the universe of firms rated by KLD at least two years prior to the scandal. For each scandal firm we identified a comparison firm with similar employment two years previous and the same 2-digit industry that was also in KLD's universe. We then measured membership in the Domini 400 two years prior to the scandal for the scandal firm and the comparison firm.

If Domini membership strongly predicts fewer scandals, we expect the pool of 126 scandal firms to have fewer Domini members than the pool of 126 matched non-scandal firms. In fact, the scandal set is 52% Domini firms and the comparison set if 55% Domini firms. (Table 3.5A). This difference of 3 percentage points is not statistically significant and the odds ratio of 1.375 is not statistically significantly different from 1. At the same time, the confidence interval on the odds ratio is very wide: from 0.52 to 1.49.

Thus, our test does not have power to rule out economically meaningful effects -- in either direction.

When we examine KLD sub-scores we find a shred of evidence that KLD has predictive power, but it is coupled with a shred of evidence in the opposite direction. Specifically, we summed the number of concerns on Community, Diversity, Employee Relations, Environment, and Product each firm had. The mean scandal firm had 1.88 of these five concerns, while the mean non-scandal firm had only 1.39 concerns. The half-point difference is statistically significant at the 5% level ($t = 2.24$, matched t test; results for KLD sub-scores and predictive validity are in Appendix 4).

While encouraging for KLD sub-scores to have some predictive validity, this correlation is not convincing for two reasons. First, we also examined whether companies that KLD rated as having many *strengths* had fewer scandals. Of the five possible strengths, the average scandal firm had 2.62 while the average non-scandal control had 2.10 -- that is, that is, scandal firms averaged half a strength *more* than did controls without a scandal. While the difference is only marginally significant ($t = 1.75$), the magnitude of this gap in strengths (that is inconsistent with theory) is as large as for the gap in concerns (that was consistent with theory). Second, we have run at least 15 t-tests on subsets of the KLD scores. Thus, we do not want to over-emphasize a single statistically significant finding. The statistical significance would not be present if we adjusted for the number of permutations we examined. Finally, when we use all KLD sub-scores or just the KLD

Score (the index of KLD sub-scores that best predicts KLD membership, estimated in Appendix 1) to predict scandals, there is not statistically significant relationship.

3.6.9 Does the Gompers et al. IRRC index of weak governance predict fewer scandals?

Our tests of predictive validity for the Gompers, et al., index of weak governance "G" follows the same method as our test using Domini 400 membership. For each scandal firm that had Gompers, et al., index of weak governance 2 or 3 years prior, we found a control firm of the most similar employment and the same 2-digit industry.

If the index of weak governance strongly predicts fewer scandals, we expect the scandal firms to have a higher mean index (that is, worse governance) than the non-scandal firms. We find the scandal firms had a mean index of weak governance of 9.39; that is, a bit over 9 of the 24 dummy variables measuring aspects of weak governance, were true two years prior to the scandal (Table 3.5B). At the matched comparison firms that had no scandal, the mean index of weak governance was 9.95. The weaker governance (by 0.66 of 24 items, about a fifth of standard deviation) of the control firms is not statistically significant (t = 1.1) and is in the opposite direction predicted by theory. Thus, we find no evidence that the index of weak governance predicts scandals.

Our statistical power is modest, in that the 95% confidence interval ranges from scandal firms having roughly ½ point stronger governance to 1½ points weaker governance than comparison firms. Thus, even with this modest sample we can rule out a strong relationship between measures of weak governance used by Gompers, et al. and scandals two to three years later.

3.7 Conclusion

3.7.1 Summary

For convergent validity we find evidence that these firms are rating related constructs in the sense that the underlying correlations in ratings seem to correlate around 0.23. This figure is below unity due to purposeful differences and due to measurement error. If one of the raters measured the underlying construct of social responsibility with more precision than the others, we would observe consistently higher levels of correlation for that rater. Our current results indicate no such pattern. In our own view of high convergent validity, we find no support for Hypothesis 1 but for each reader it will depend on the interpretation of a 0.23 correlation between ratings. We also find no support for Hypothesis 2 since our differences in measurement remain even after adjusting for explicit distinctions in measurement. None of our adjustments for explicit differences in rater goals or methods substantially increased the correlations. Thus, we are left with the strong suspicion that measurement error accounts for a significant share of the variance in raters' true ratings of corporations' social performance. However, future research will be necessary to estimate the exact amount of measurement error in these social ratings.

Our results on predictive validity are even less encouraging than our results on convergent validity. Our data on scandals is imperfect in that misses some scandals and it includes some indicators of suspicious behavior that, in fact, are innocent. Nevertheless, our results provide support for Hypothesis 3 that Domini 400 membership and the Gompers, et al., index of weak corporate governance have weak predictive validity. Neither a narrow focus on governance or broad measure of social responsibility (including charitable giving, environment impact, product safety, etc.) seem to distinguish firms that will have major scandals from those who will not.

While we find little support for the theoretical predictions of strong convergence through imitation, we cannot yet assert that Meyer and Gupta (1994) explanations for weak correlations between performance measures is driving all of the divergence here. After all, significant measurement error likely exists, and future research can advance our understanding of how much impact this error has on the social ratings in this study.

Finally, the distinction between “independent” raters like KLD, FTSE4Good, DJSI, and Innovest, and raters that also manage their own mutual fund like Calvert, did not seem to impact our results. Calvert ratings did not systematically converge or diverge from the other ratings, and tracked KLD scores most closely. It would be interesting to know whether Calvert’s socially responsible funds outperform the other 4 raters consistently and this can be addressed in future research. Furthermore, we found no evidence that, despite their differences, either IRRC or KLD ratings had better predictive validity than the other. With a small sample of scandals however, we should not dismiss the hypothesis that these differences between raters might have substantial impact. Future research should address this question if larger datasets become available.

3.7.2 Discussion

The key insight of this paper is that social science methods can shed light on the morality of SRI without defining morality, as in our convergent validity test or with an uncontroversial minimal level of morality -- not being in a major scandal. The results on predictive validity are hardly definitive in that the standard errors are large. Thus, defenders of SRI metrics need not lose their faith. Similarly, critics of SRI metrics (e.g., Entine, 2003) need not back down from the idea that the metrics of high *perceived* social responsibility might even predict *more* scandals.

Thus, the paucity of data to examine predictive validity has two related implications; one about management incentives in the SRI rating business and one about validation studies. First, the lack of ability to test predictive validity with any precision implies that SRI agencies must convince investors of their validity largely by using procedurally rational means. As in astronomy or evolutionary biology, SRI agencies can take advantage of many of the strengths of the scientific method, but are unable to perform experiments. Even worse, because many of the scandals are rare events, models to predict the scandals are hard to test. That fact is not a critique of SRI agencies constructing such models, but is a limitation in evaluating the quality of the SRI ratings. As the SRI industry matures and more years of data exist, the hypothesis of predictive validity should become testable with more precision. Until that time, the industry will remain open to fads, neoinstitutionalist stories, dueling anecdotes, and other weak forms of evidence.

Due to the high level of uncertainty, these failures can arise even in settings where the analysts at socially screened funds are personally hard-working and dedicated to measuring corporate social responsibility. That is, nobody knows what metrics are valid predictors of future corporate behavior. Thus, even if their analysts have high effort, skill, and good will, socially screened funds may be able to produce highly valid metrics. This paper is not a critique the effort or goodwill of current practitioners, but does lay a cautionary tale about the current state of knowledge. While the evidence does not support the neo-institutional predictions about convergence, we cannot be sure that all of the differences are do to measurement error. There could be several reasons, among the ones suggested by Meyer and Gupta (1994), for these measures to be weakly correlated.

There are several potential extensions that could lead to further insights into the topics discussed in this paper. First, we could include more ratings from additional raters in the socially responsible investing industry. In addition, along with considering a particular rater's screens, we could compute the ratings using our own definition of screens, though we doubt that this would improve the fit or the correlations. In terms of predictive validity, we could add different types of scandals, including the recent option back-dating controversies. With enough scandals, we could potentially separately test for environmental and governance scandals. We could also test whether IRRC governance scores predict membership into KLD and the other indexes. The broader results on predictive validity could inform the results of our companion study (Chatterji, et al; 2006), which found low predictive validity for green metrics. We find low predictive validity using a broader set of scandals in this paper, but it would also be worthwhile to measure how various CSR metrics (governance, environment, etc.) predict scandals in different domains. (i.e. Do poor governance scores predict environmental disasters?) After all, if there is truly a “culture of responsibility” in particular firms, perhaps domain specific predictive validity is not the correct expectation. Furthermore, we could also address the construct validity issue by unpacking the governance scores (e.g. exploiting variation by state) and rater scores, and looking at the measurement properties of the original surveys.

The two portions of this paper complement each other. First, high convergent validity may arise through shared error. Thus, the tests of predictive validity are crucial to test for such shared errors. Second, our tests of predictive validity suffer from modest sample size. The lack of precision for these results is motivation for studying convergent validity. Furthermore, prior work on financial ratings implies that the lack of predictive validity (or the fact that raters “overvalue” companies) is actually a consequence of the limited herding behavior that we observe in the convergent validity results.

Our findings are consistent with several interpretations, each with implications for SRI raters and for investors. To the extent the current low convergent validity of social screens reflects different preferences that investors know about and use to make decisions, that diversity is a desirable outcome of market forces and should persist. Thus, we expect both funds that avoid contraception providers (on religious grounds) and funds that that search out contraception providers (so as to empower women) to continue to thrive. Our results are also relevant for other areas of finance where ratings are crucial, whether in the recommendations of sell-side analysts or the ratings of corporate bonds. Our study is among the first to test for convergent validity among social ratings, and it would be interesting to examine the parallels between the early stages of the financial industry and the socially responsible investing industry. Our sense is that the current state of SRI is similar to other ratings systems in their early stages of development.

As we have argued, much of the current diversity in screens reflects inconsistent definitions and measures of social responsibility coupled with measurement error – not marketing to distinct niches. Divergences that reflect experimentation and learning can be a strength of a new industry; the key is to build in learning and a continued stream of validation studies that inform social screening funds and socially conscious investors of the validity of various metrics It is important that convergent validity rise over time due to increased knowledge, not due to herding or institutionalist forces that promote conformity to norms and "best practices" – even if those norms have low validity. Finally, if there is not a steady stream of additional research validating the numerous measures, diversity of measures may continue as both old and new funds market new approaches to measurement.

References

Banerjee, Abhijit, 1992: "A Simple Model of Herd Behavior." *Quarterly Journal of Economics*, Volume 107, Issue 3: 797-818.

Bartels, Larry M. *Presidential Primaries and the Dynamics of Public Choice.* Princeton, N.J.: Princeton Univ. Press, 1988.

Beltratti, Andrea: "Socially Responsible Investment in General Equilibrium." FEEM Working Paper, October 2003, No. 93: http://www.feem.it/web/activ/_activ.html

Berk, J.B. and R.C. Green. 2004. "Mutual Fund Flows and Performance in Rational Markets." *Journal of Political Economy,* vol. 112, no.6, pages 1269-1295

Bikhchandani, S., D. Hirshleifer, and I. Welch: "A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades." *Journal of Political Economy*, 1992, Vol. 100: 992-1026.

Brown, B; and S. Perry. 1994. "Removing the Financial Performance Halo from Fortune's "Most Admired" Companies." The Academy of Management Review, Vol. 37, No. 5. pp. 1347-1359.

Chatterji, A.; D. Levine, and M. Toffel: 2006. "Do Corporate Social Responsibility Ratings Predict Corporate Social Performance?" Working Paper

Corporate Crime Reporter: List of the largest corporate crimes/scandals of 1990-1999. (www.corporatecrimereporter.com) Last Accessed November 14th, 2006

DiMaggio, P., and W. Powell: "The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields." *American Sociological Review*, 48: 147-160.

Entine, John: "The Myth of Social Investing: A Critique of its Practice and Consequences for Corporate Social Performance Research." *Organization and Environment*, September 2003, Vol. 16, No. 3: 352-368 http://oae.sagepub.com/cgi/rapidpdf/16/3/352.pdf

Gompers, Paul A., Joy L. Ishii, and Andrew Metrick: "Corporate Governance and Equity Prices." *Quarterly Journal of Economics,* February 2003, Volume 118, No. 1: 107-155 http://finance.wharton.upenn.edu/~metrick/gov.pdf

Hawken, Paul. 2004. "Socially Responsible Investing." The Natural Capital Institute, October 2004.

Hong, H., J. Kubik, and A. Solomon: "Security Analysts' Career Concerns and the Herding of Earnings Forecasts." *Rand Journal of Economics*, Spring 2000, Vol. 31: 121-144.

Kinder, Peter: "A Note from Peter Kinder and the KLD Press Release." GreenMoneyJournal.com, Fall 2005. (Last Accessed September 14, 2005)

Lyon, T. P., and J. W. Maxwell. 2005. "Greenwash: Corporate Environmental Disclosure under the Threat of Audit." Working Paper, University of Michigan.

Margolis, J., and J. Walsh. 2003. "Misery Loves Company: Rethinking Social Initiatives by Business. Administrative Sciences Quarterly, 48: 268–305.

Merton, Robert: "A Simple Model of Capital Market Equilibrium with Incomplete Information." *Journal of Finance*, July 1987, Vol. XLII, No. 3.

Meyer, J. W., and B. Rowan: "The Structure of Educational Organizations," in *Environments and Organizations*, M. W. Meyer (ed.), San Francisco: Jossey-Bass: 78-109.

Meyer, Marshall W., and V. Gupta: "The Performance Paradox," in *Research in Organizational Behavior*, L. L. Cummings and B. M. Staw (eds.), Greenwich, CT, JAI Press, Volume 16: 309-369

Rosenbaum, Virginia: Governance Index ("G") for the full sample of Investor Responsibility Research Center (IRRC) firms for each publication of Corporate Takeover Defenses, by Virginia Rosenbaum
http://finance.wharton.upenn.edu/~metrick/governance.xls

Scharfstein, David S., and Stein, Jeremy C. 1990. "Herd Behavior and Investment." American Economic Review 80, pg. 465-479

Scott, W.R. 1995. "*Institutions and Organizations"* Thousand Oaks, CA; Sage

Sharfman, Mark. 1996. "The Construct Validity of the Kinder, Lydenberg & Domini Social Performance Ratings Data." Journal of Business Ethics 15: 287-296

Social Investment Forum: *2003 Report of Socially Responsible Investing Trends in the United States*, 2003.

Stata, "Tetrachoric Correlations for Binary Variables," in *Stata Reference R-Z*, 2006, Plano TX: 426-434.

Statman, Meir: "Socially Responsible Mutual Funds." *Financial Analysts Journal*, May/June 2000.

Statman, Mier, S. Hamilton, and H. Jo: "Doing Well While Doing Good?: The Investment Performance of Socially Responsible Mutual Funds." *Financial Analysts Journal*, November/December 1993.

Staw, Barry M. and Epstein, Lisa D. 2000. "What Bandwagons Bring: Effects of Popular Management Techniques on Corporate Performance, Reputation, and CEO Pay." *Administrative Science Quarterly,* 45: 523-556

The Financial Statement Restatement Database, United States General Accounting Office, 2003 [Annex 2 of http://www.gao.gov/new.items/d03395r.pdf].

Waddock, Sandra: "Myths and Realities of Social Investing." *Organization and Environment,* 2003, Volume 16, Number 3: 369-380.

Zitzewitz, Eric. 2001. Opinion-producing agents: career concerns and exaggeration“, Working Paper

Table 3.1 Summary Statistics

Membership in Social Indices (2005xx)	In		Out	Universe (N)	Universe name
KLD LCS membership	670		313	983	Russell 1000
Calvert Membership	493		490	983	Russell 1000
FTSE4Good Membership	77		501	578	FTSE All World USA
DJSI Membership (including waitlist)	77		834	911	Dow Jones World Index
Innovest Membership-(Top 18)	17		485	502	"Innovest Universe"
Continuous Measures of Responsibility	*Mean*		*SD*	*Min*	*Max*
KLD Score = Predicted probability of KLD LCS membership; from Appendix 1					
Using 63 subscores (0 or 1) and KLD screens	0.204		0.231	0	0.999
With common KLD and Calvert screens	0.22		0.241	0	0.999
With common KLD and FTSE screens	0.151		0.205	0	0.999
With common KLD and DJSI/Innovest Screens	0.164		0.195	0	0.999
Calvert Score = predicted probability of Calvert Index membership based on 5 subscores (each 1 to 5), from Appendix 1. Calvert subscores are available only for the 100 largest firms	0.53		0.395	0.0000863	0.9997
DJSI rank within 22 industries. Ranks are available for DJSI's top 10% of each industry (so range = 90-100 percentile) and 22 waitlisted firms (where rank = 88%xx)	0.93		0.029	0.895	0.9857
G Index of Weak Governance (# of 24 items Gompers et al. coded as weak corporate governance)	9.1		2.77	2	19
Scandal measures 1993 to 2006	at least once	never			
Largest 100 spills 1993-2006					
Earnings Restatements 1993-2006	69		635	704	
other large scandals (see text)					

^ Obtained through logit post estimation, where subscores the perfectly predict success and failure are multiplied by 1.5 times the largest coefficient (positive or negative) and subtracted out

^^ Same as above, except without SP 500 as a control in the logit equation

^^^ Same as above, except with explicit screens, like alchol, tobacco, etc., included in the logit equation and explictly screened out companies set to zero

^^^^ Same as above, except with industry controls in the logit equation

Table 3.2 How much do the memberships of top Social Ratings funds intersect?

Tetrachoric correlations

Index and [maximum N in index]	**Calvert**	**DJSI**	**FTSE4Good Top 100**	**Innovest Elite 18**	**Mean correlation of this index with the other 4**
KLD's LCS	TC=0.6856 *; Chi2=198.83 *, LO=9.12 N=983;	TC=.01; Chi2=0; LO=1.02; N=911;	TC=0.2196 *; Chi2=5.89 *; LO=2.03; N=578;	TC=-.22; Chi2=2.93; LO=.438; N=560;	0.1738
Calvert Social Index [607]	X	TC=0.07; Chi2=0.77; LO=1.24; N=911;	TC=0.13; Chi2=2.64*; LO=1.50; N=578;	TC=0.09; Chi2=0.41; LO=1.38; N=507;	0.2439
DJSI top 10% plus 1 in each industry [88]	X	X	TC=0.53 *; Chi2=45.53*; LO=5.98; N=570;	TC=0.54 *; Chi2=26.06*; LO=8.52; N=534;	0.2875
FTSE4Good Top 100 [101]	X	X	X	TC=0.23; Chi2=2.58; LO=2.44; N=392;	0.2774
Innovest US firms in Innovest's global top 100 [18]	X	X	X	X	0.16

Notes: In each cell, TC means tetrachoric correlation, Chi2 is Chi Squared Statistic, LO is the log odds ratio,. and N is the intersection of the universes of the two rating agencies
As described in the text, tetrachoric correlations are similar to standard correlations, but are adjusted for the dichotomous nature of the data.

Membership is from 2005.

* Significant at the 5% level

Table 3.3 How correlated are Social Ratings? (Calvert scores & DJSI ratings & KLD scores)

	1	2	3	4
	KLD_kld Score	**KLD_calvert score**	**KLD_FTSE score**	**KLD_DJSI score**
Calvert Score = predicted probability of Calvert Index membership based on 5 subscores	0.4861 (N=100)	0.4922 (N=100)	0.5196(N=100)	0.4869 (N=100)
DJSI rank within 22 industries. Available for the top 10% of each industry (coded with rank = 90-100 percentile) and 22 waitlisted firms (rank coded 88%xx)	0.0311 (N=88)	0.0264 (N=88)	0.0952 (N=88)	0.0863 (N=88)

Notes: The correlation between Calvert Score and DJSI Ranking (among DJSI's top-ranked firms in each industry) is 0.0073 (N=32)

N is the overlap of the two indices' universes.

Table 3.4 Gap between index members and non-members

Gap in KLD Score (= Predicted probability of KLD LCS membership) between index members and non-members
SE in parentheses, maximal gap possible with membership of this size in {braces}.

	KLD LCS	**Calvert**	**FTSE4Good Top 100**	**DJSI**	**Innovest**
1. KLD style: Screened-out firms get Score = 0. Others get probability predicted based on 63 subscores (0 or 1) .	0.55 (0.02) * {0.73}	0.30 (0.02)* {0.52}	0.05 (0.05) {0.40}	0.02 (.04) {0.37}	0.15 (.09) {0.40}
2. Calvert style: Screened-out firms (usingKLD measure of Calvert screens) get Score = 0. Others get probability predicted based on 63 subscores (0 or 1). Logit, but not prediction, also has industry dummies.	0.53 (0.02)* {0.70}	**0.29** **(0.02)*** **{0.52}**	0.02 (0.04) {0.40}	0.02 (0.04) {0.40}	0.11 (.09) {0.42}
3. FTSE4Good style: Screened-out firms (using KLD measure of FTSE screens) get Score = 0. Others get probability predicted based on 63 subscores (0 or 1). Logit, but not prediction, also has industry dummies.	0.37 (0.02)* {0.43}	0.24 (0.02)* {0.50}	**0.08** **(0.04)*** **{0.60}**	0.11 (.04)* {0.56}	0.02 (.08) {0.39}
4. DJSI and Innovest style: All firms get probability predicted based on 63 subscores (0 or 1). Logit, but not prediction, also has industry dummies.	0.29 (0.02)* {0.58}	0.22 (0.02)* {0.44}	0.02 (0.03) {0.51}	**0.05** **(.03)** **{0.52}**	**0.03** **(.07)** **{0.40}**
Change in gap between KLD style and "Actual" (In Bold)		.01 (.03)	-0.03 (.05)	-0.03 (.05)	.12 (.11)
N =overlap of universes of KLD and other index	983 *	983	578	911	502

* The universe for KLD LCS is the Russell 1000, but we have social ratings for 983 of the 1000 firms in the index

Notes:
A firm's Score is the predicted probability a firm would be included in the KLD LCS index. We vary the formula for the Score to approximate the methods used by indices other than KLD.

If the rater in question would screen out the firm, the firm is given a Score of zero.

Screens for the other indices are proxied with KLD's measures of that screen. For example, FTSE4Good has a screen on gambling, but due to data limitations we use KLD's, not FTSE's definition of "screened out for gambling."
If not screened out, the company's score is the predicted probability of membership in the KLD LCS based on its 64 KLD subscores. See appendix for an example of this regression.

Industry norming of KLD score is carried out by including industry dummies in the logit equations predicting KLD LCS membership, but dropping the dummies when predicting membership.

The {maximal gap possible} for an index of size n in a universe of size N compares the mean score of the n highest Scores and the (N-n) lowest scores; that is, the gap in Scores that would arise if KLD's subscores were all that were used in choosing the index in question.

Table 3.5 Predictive Validity Analyses

Table A: Does membership in the Domini 400 Index predict fewer scandals?

	Domini 400 membership status in Year t-2			
Scandal in Year t	*Member*	*Non-Member*	*Column Total*	N (row)
Scandal firms[a]	52%	48%	100%	126
Non-scandal comparison firms	55%	45%	100%	126
N(column)	134	118		252
McNemar chi2=1.42 testing whether column shares are equal (n.s.); Log Odds Ratio=1.375				

a: Firms with scandals are all firms rated by KLD that two years later had earnings restatements, one of the largest environmental spills, or were members of a list of major scandals collected by the authors.

Table B: Do IRRC Governance Scores predict fewer scandals?

G Index of weak governance ranges from 1-16, with higher numbers associated with weaker corporate governance

Correlation between G Index of Weak Governance and Scandals	-0.0953	N=134		
Mean G Index of Weak Governance for Scandal Firms	*9.39*	*N=69*		
	(sd=2.91)			
Mean G Index of Weak Governance for Matched Non-Scandal Firms	*9.95*	N=69[b]		
in same 2-digit industry and closest # of employees in year t-2	sd=(3.01)			
Gap	0.16			
t statistic on gap (n.s.)	1.1			
b: 7 of the controls matched to more than 1 Scandal firm				

Appendix 3.1 How much do the memberships of top Social Ratings funds intersect? *

Index and [maximum N in index]	Calvert	FTSE4Good Top 100	DJSI	Innovest Elite 18
KLD's LCS (670)	89% of the 493 Calvert members, but only 47% of the 490 Calvert nonmembers, are in the LCS.	79% of the 78 FTSE4Good top 100 are in LCS, and 65% the 513 FTSE nonmembers	DJSI's 60 members and 20 waitlist are both about 65% in LCS, as opposed 60% of the DJSI "others"	47% of the 18 Innovest elite and 67% of the nonelite are in LCS
Calvert Social Index [607]	X	61% of the 77 FTSE4Good top 100 and 51% of the 501 FTSE "others" are members of Calvert	60% of the 57 DJSI members, 55% of the 20 waitlist, and 53% of the 834 others are in Calvert	59% of the 17 Innovest elite and 51% of the 485 non-elite are in Calvert
DJSI top 10% plus 1 in each industry [88]	X	X	42% of the 52 DJSI members are in FTSE4Good top 100, as opposed to 33% of the 15 waitlist firms and 10% of the 503 DJSI "others".	5 of Innovest's 14 elite are in the 75 FTSE4Good, while 11 are in the 317 nonmembers.
FTSE4Good Top 100 [101]	X	X	X	15% of DJSI's 59 members are in Innovest's elite 18. This share is small, but it is higher than the 6% of the 20 DJSI "waitlist" firms and higher yet again than the 2% of DJSI "others".
Innovest global top 100 [18 in U.S.]	X	X	X	X

Chapter 4

4. The Impact of City Contracting Set-Asides on Minority Self-Employment

4.1 Introduction

Set-aside programs that target government contracts for disadvantaged and minority-owned firms are controversial. During the late 1970s and 1980s there was substantial growth in the value of federal, state, and local government contracts reserved for minority-owned businesses. The purpose of these set-aside programs was to develop minority enterprise, counter the effects of past discrimination, and reduce unemployment among urban minorities.

For the last fifteen years, however, the state and local programs established in the 1980s have been both judicially and legislatively challenged and dismantled. The *City of Richmond v Croson Co.* Supreme Court decision in January 1989 invalidated the use of such programs unless they were used as narrowly tailored remedies for identified discrimination. The 1995 *Adarand Constructors, Inc. v. Peña* Supreme Court decision and state referendums passed in California (Proposition 209 in 1996) and Washington (1998) further brought into question the constitutionality and political viability of government set-asides.

Given the debate over the constitutionality of these affirmative action programs, remarkably little is known about the actual effects of the original programs. The set-aside programs implemented in the late 1970s and 1980s are among the most substantive pieces of antidiscrimination legislation passed since the Civil Rights Act of 1964 and

Executive Order 11246 in 1965 which outlawed discrimination by federal contractors. In this study, we evaluate the impact of the programs passed in large cities during the 1980s on the self-employment rates of black men.

Black self-employment is a natural initial outcome of interest since, historically, self-employment has been a route of economic advancement for disadvantaged groups. Also, some analysts suggest that stimulating black self-employment in sectors with high growth potential (e.g., construction, wholesale trade, and business services) is necessary public policy for promoting economic development in urban minority communities and alleviating poverty (Bates 1993).

Using comprehensive Current Population Survey data from 1979-1989, we document the evolution of black self-employment rates during the 1980s. A striking empirical regularity emerges from the analysis. We find that self-employment rates for black men rose substantially in the mid-1980s in cities in which set-aside programs were implemented. In contrast, the self-employment rates of white men were stable in these same cities during this period.

Next, we utilize the staggered timing of set-aside programs across U.S. cities during this period to estimate their impact on black self-employment rates. To test for endogeneity bias induced by the potential nonrandom location and timing of these programs we: 1) use an "event study" or double difference methodology, based on the exact dates of program implementation, to examine trends in black self-employment rates before program implementation; and 2) contrast these with the self-employment rates of white men before and after program implementation.

Remarkably, there is little consensus in the literature on the exact dates that city programs were implemented. To address this, we compile date information from four sources, including court cases and city records, which has not been previously done. The results suggest large increases in black self-employment rates soon after program implementation, with the black-white self-employment gap narrowing 2.5 to 5 percentage points within five years of program initiation. There is little evidence of a pre-existing trend in black self-employment. Further, the results imply a reduction in the black-white gap in the employment-to-population ratio of about 1.5 percentage points for every percentage point decrease in the self-employment gap.

The timing of the gains and their location suggest that city-level minority business set-aside programs may be an important explanatory factor underlying these differential changes. The purpose of these programs is to develop minority enterprise, counter the effects of past discrimination, and reduce unemployment among minorities. In the future, we hope to be able to contrast these benefits with the costs associated with such programs and to infer whether the programs had "crowd out" effects on the self-employment of white men.

4.2 Background on Business Set-Aside Programs

The purpose of minority business set-aside programs is to develop minority enterprise, counter the effects of past discrimination, and reduce unemployment among minorities in urban communities. These programs originated from government policies that attempted to strengthen the viability of small businesses. Initially, set-asides were focused on increasing the number of minority-owned firms during the late 1960s and early 1970s. During the next fifteen years, however, set-asides were increasingly targeted to businesses that had greater future growth potential (Bates 1985).

In general, there are two types of set-aside programs. In one type, a specified percentage of the number or total dollar value of government contracts is allotted to minority-owned businesses. In the other type, prime contractors are required to allot a specified percentage of the total amount of government contracts to minority-owned subcontractors and/or suppliers (Rice 1991 and Myers 1997).[63] The percentage goals vary across programs and sometimes within programs for different purchases, such as construction contracts, procurement of goods and services, and professional services. Data on local set-aside programs listed in a 1988 report by the Minority Business Enterprise Legal Defense and Education Fund (MBELDEF) indicate that these goals range from 1 to 50 percent, with most programs having goals of 5 to 15 percent. A large proportion of set aside programs appear to be targeted towards the construction sector (MBELDEF 1988).

---

[63] The constitutionality of this type of set-aside was challenged in the 1995 *Adarand v. Peña* Supreme Court case.

Set-aside programs are also often complemented with procurement officials who aid minority-owned businesses in obtaining assistance (Bates and Williams 1993).

Set-aside programs exist at the federal, state, city, county, and special district (e.g. airport, water, sanitary, park, and school) level. A well-known program at the federal level is the Small Business Administration's (SBA) 8(a) program, established in 1968 (and still in operation today) as an amendment to section 8 of the Small Business Administration Act of 1953. In this program, the SBA serves as the prime contractor for goods and services to various federal agencies. The SBA then provides subcontracts to firms that are owned by individuals who are socially and economically disadvantaged.[64] These SBA 8(a) contracts totaled $2.3 billion in 1983 (Bates 1985).

Another important federal program is the 1977 Public Works Employment Act, which required that 10 percent of all federal public works contracts be given to minority-owned businesses. This program earmarked $400 million of local public works to minority-owned firms (Bates 1985). The constitutionality of this program was soon challenged leading to the U.S. Supreme Court's ruling in *Fullilove v. Klutznick*, in 1980, which upheld the federal government's use of these programs. *Fullilove v. Klutznick* sparked the creation of set-aside provisions among other federal agencies, and state and local governments. Minority business set-asides were mandated for federal transportation and highway construction, national defense, NASA contracts, international development grants, and for the development, construction and operation of the super collider (Myers

[64] The SBA considers blacks, Hispanics, Native Americans, and Asian Pacific Americans as socially disadvantaged. In 1978, 96 percent of 8(a) firms were owned by minorities (Bates, 1985).

1997). The federal government reported $4.4 billion in contract awards to minority and disadvantaged firms in FY 1986 (Rice 1991).[65]

Most states also created set-aside programs for minority-owned businesses.[66] This was a direct response to requirements that state departments of transportation administering federal highway grants and contracts oversee implementation of the federal set-aside provisions (Myers 1997).

Another response to the *Fullilove v. Klutznick* ruling was the creation of minority business set-aside programs by more than 200 local governments (Myers 1997). Most of these programs were created in the early to mid-1980s (MBELDEF 1988), and many of them, especially in large central cities, were quite substantial (Bates 1985). For example, minority- and white female-owned businesses received $191 million between 1979 and 1989 through Atlanta's set-aside program (Boston 1998).[67]

The existence of overlapping federal, state, and local programs will complicate our empirical work and is very difficult to control for without detailed contract data at each level. Many of the federal and state programs have evolved, along with the local programs, into initiatives targeting disadvantaged businesses. Further research will be required to ascertain whether these reconstituted programs are having a similar impact to

---

[65] Of the total, three billion dollars in contract awards were through the 8(a) set-aside program (Rice 1991).
[66] Rice (1991) reports that 36 states had set-aside programs in place by the late 1980s. To provide an example of the size of these programs, Myers and Chan (1996) report that the state of New Jersey awarded $93 million (or 3.2 percent of the total amount awarded) of public procurement and construction prime contracts to minority-owned firms in 1988.
[67] Procurement in Washington, D.C. to minority-owned firms was $170 million in 1985 (Rice, 1991).

the minority set aside programs discussed in this paper. In interpreting our empirical results, we must be careful not to ascribe all of the effect on black self-employment to local set asides, since other programs on the state and federal level could have also had an impact. Furthermore, most of the anecdotal evidence on set asides has come from the construction industry, so our ability to interpret our empirical results in other sectors is limited.

Minority business set-asides represent a multi-billion dollar annual governmental expenditure and have recently become very controversial both politically (e.g., Proposition 209 in California) and judicially (e.g., the 1995 *Adarand Constructors, Inc. v. Penã* Supreme Court decision). Remarkably, little is known about their actual effectiveness in promoting growth in the number of minority-owned businesses and in alleviating unemployment among blacks in the inner city. In particular, only a handful of studies have attempted to analyze whether these programs have met their goals.

The first question is whether set-aside programs actually increased the number and/or total dollar amount of government contracts received by minority-owned businesses. Myers and Chan (1996) examine the award of public procurement and construction contracts to minority- and non-minority-owned firms before, during, and after the implementation of the state of New Jersey's set-aside program.[68] They find that the average number of contract awards going to black-owned firms submitting bids remained unchanged from the period before set-asides (1980-84) to the period during set-asides

[68] New Jersey's set-aside program started in 1985 and was suspended in 1989 due to the *City of Richmond v. Croson* decision. The authors define the pre-, during, and post-periods as 1980-84, 1985-88, and 1989-90, respectively.

(1985-88) and decreased from the period during set-asides to the period after set-asides (1989-90). In contrast, average contract awards for white male-owned firms increased from 1980-84 to 1985-88 and decreased markedly from 1985-88 to 1989-90.[69] The authors conclude that New Jersey's set-aside program did not have a substantial impact on the average number of contracts awarded to black-owned firms submitting bids on state contracts.

Additional evidence on the "first-stage" relationship between set-aside programs and contract awards is provided in a recent review of 58 disparity studies conducted in response to the *Richmond v. Croson* decision by the Urban Institute (Enchautegui, et al., 1996). Disparity is defined as the ratio of the percentage of total contract dollars awarded to minority-owned firms to the percentage of all available firms that are minority-owned. The study finds evidence of greater disparity in contract awards (i.e., lower disparity ratios) in jurisdictions without affirmative action programs, suggesting that such programs positively affect the amount of government contracts received by minority-owned firms.

The next natural question is whether set-aside programs have had an effect on the growth and viability of minority-owned firms. There is little evidence on this question. Boston (1998) uses published data from the Survey of Minority-Owned Business Enterprises (SMOBE) to examine the growth rate in the number of black-owned businesses in cities

[69] They also find that the ratio of awards to bids decreased for black-owned firms from 1980-84 to 1985-88, whereas the ratio increased for white males during the same period. The authors argue that this decline in black bid success rates was therefore due to an increase in bids without a corresponding increase in awards.

that implemented affirmative action programs in the 1980s relative to cities that did not. The data on which cities installed affirmative action programs and their dates come from MBELDEF (1988). He finds that the average growth rate from 1982 to 1992 was 65 percent in cities with programs and 61 percent in cities without programs and that this difference is not statistically significant.

Bates and Williams (1993) find that from 1982 to 1987 total sales by black businesses and the number of black firms increased more in cities with black mayors than in cities without black mayors. Citing evidence from case studies suggesting that black mayors place a high priority on contracting with minority-owned businesses, Bates and Williams argue that the positive effect of black mayors on black business outcomes is partly due to their support of minority business set-aside programs.

Bates and Williams (1996) use data from the U.S. Census Bureau's Characteristics of Business Owners (CBO) survey to examine the survival rates of minority-owned enterprises that sell to state and local government relative to minority-owned firms that do not. Controlling for a number of firm and owner characteristics, they find that minority firms with local government sales are no more likely to survive than minority firms with no local government sales from 1987 to the end of 1991. They also find that minority firms that derive at least 25 percent of their sales from state and local government are less likely to survive than minority enterprises that are less reliant on state and local government. They note that these findings are consistent with the practice of minority firms serving as front companies and small minority firms being awarded large procurements that they are not equipped to handle.[70] Alternatively these results could also reflect as underlying instability in the set aside programs that negatively impact the survival of new firms that are especially dependent on government contracts.

Finally, Bates and Williams (1995) explore whether the characteristics of preferential procrurement programs have an effect on survival among minority-owned businesses. The authors and the Joint Center for Political and Economic Studies (JCPES) collected detailed profiles on minority-business set-aside programs in 28 large cities in the United States (JCPES 1994). The profiles include information on program dates, program assistance staffing, provision of capital assistance, bonding, downsizing of large procurement contracts, certification of minority business enterprises, penalties for violation of certification or program regulations, and treatment of brokers. They find higher survival rates among minority-owned businesses that derive 1-24 percent of their sales from state and local governments in cities with affirmative action programs that have a rigorous certification process, a staff assigned to assist minority firms, routinely waive bonding requirements or provide bonding, and/or provide working capital assistance to minority firms receiving contracts.

In sum, the set aside programs were designed to remedy past discrimination against minority entrepreneurs with regards to access to finance and government contracts. The architects of set aside programs viewed low rates of entrepreneurship in minority communities as driven at least in part by these factors and often saw increased

[70]Bates and Willams (1995) note that many state and city programs do not make a serious attempt to verify that minority firms receiving contracts are actually minority-owned and operated, and that in some programs it is not even illegal to act as a front company. See Bates and Williams (1995) for more details.

entrepreneurship as a first step towards increasing minority employment as well. By setting aside a percentage of contracts that could only go to minority owned firms, local governments could provide strong incentives for new business formation. For example, the local set aside programs in construction were designed to provide incentives for minority employees in existing construction companies to start small firms that could be supported by city contracts early on and eventually gain other clients to generate a sustainable business. As these businesses grew, they could eventually hire more employees from the local community, hopefully having an impact on city employment rates. Since set aside programs also targeted other sectors, like wholesale trade and professional services, we will also analyze data from other sectors in this study. However, as discussed below, much of the qualitative evidence on set aside programs comes from the construction sector.

4.3 Data on City-Level Set-Aside Programs

We conclude that little is known about whether set-aside programs have met the goal of generating successful minority-owned businesses. In this study, we examine whether the set-aside programs established in many of the largest U.S. cities during the 1980s had an impact on self-employment rates among black men relative to white men.[71]

---

[71] Communications with Thomas Boston and Timothy Bates suggested that state-level set-aside programs are much less substantive than city-level programs and that most county-level programs follow the city programs since targeted minorities generally live in central cities. According to a communication with Timothy Bates, construction, business services, and wholesale trade are the sectors that were impacted the most by set-aside programs. Construction is the largest in terms of dollar amounts, and business services is the largest in terms of number of firms. Examining the MBELDEF (1988) document, it appears that the bulk of city-level program coverage was targeted at minority construction firms.

As mentioned above, the vast majority of city-level set-aside programs were implemented during the early and mid-1980s. Our data on the years in which these programs were enacted come from the two sources previously used in the literature and two new sources that have not been used. The first source is the 1988 Report on the Minority Business Enterprise Programs of State and Local Governments by the Minority Business Enterprise Legal Defense and Education Fund. This report was intended to list all local affirmative action programs in the United States as of 1988.[72] The report contains information on program initiation dates, authority, coverage, and percentage goals for most programs. These data were previously used in Boston (1998).

A second source of data on program dates is a report to the U.S. Department of Commerce Minority Business Development Agency entitled, Assessment of Minority Business Development Programs, by the Joint Center for Political and Economic Studies (JCPES) in 1994. This report contains detailed profiles on minority-business set-aside programs in 28 large cities in the United States. For our analysis the most important information contained in the profiles is the program initiation dates, although the profiles contain much more information as noted above in our review of Bates and Williams (1995).

The first set of columns of Appendix Table 4.1 presents the program dates provided in MBELDEF (1988) and JCPES (1994) in the 44 cities identifiable in the Current

[72] A personal communication with Franklin M. Lee, Chief Counsel of MBELDEF, revealed that the report is probably not an exhaustive list. Information from other sources also revealed that a few cities with programs existing in 1988 are not listed in the report and that the listed starting date for Atlanta's program is incorrect.

Population Survey. Altogether, these two sources provide program date information for 33 of the 44 MSAs. The 11 cities without listed set-aside dates may still have had programs since Bates and Boston suggested that both data sources may be incomplete.[73] According to both sources, most city-level programs were passed in the early to mid-1980s, particularly from 1983-85. Seventy percent (14 of 21) of the dates from MBELDEF are in the 1983-85 period. Although the mode of the JCPES dates is also 1983-84, the JCPES distribution is more spread out during the 1980s.

Both MBELDEF and JCPES contain information on 12 of the cities. Remarkably, it appears that there is little consensus on even the dates at which cities passed set-aside programs. The two sources agree on only 4 of the 12 cities (Los Angeles, San Jose, San Francisco, and Boston). The two sources disagree by at least 2 years for 7 cities and by at least 5 years for 4 cities. Consequently, we are very concerned about the reliability of these dates, and are concerned that all subsequent analyses based on these dates may be suspect.[74]

Because obtaining the correct dates is crucial to the analysis, we obtained program dates from two additional sources. The first source is based on a search of court cases brought against the cities. Using Westlaw, we searched federal and state court cases that involved

[73] For example, Bates informed us that the original goal of JCPES (1994) was to get data for 50 cities, but some cities refused to answer while others did not have programs.
[74] Additional complications exist with respect to the timing of the programs listed in the two sources. For the MBELDEF data, the date of the administrative order or resolution can often be an inaccurate measure of the date that the program started. In many cases, the actual program did not start for several years after the order/resolution. Boston found that the MBELDEF dates were wrong for Atlanta. Bates suggested similar problems with the JCPES data. His guess is that the actual start date for several programs was abut 2-3 years after the passing of the resolution/ordinance. He also found that city contacts often did not know the start dates, and that different people would sometimes give him different start dates.

minority business programs and the 44 cities identified in our sample. For 14 cities, we found information on program start dates from court cases, which typically involved a lawsuit by a construction firm against the city's minority contracting program following *Croson*. Unfortunately, the availability of information on additional aspects of the programs was limited and inconsistent across cities.

To better understand how city level set aside programs worked in practice, we discussed the programs with key administrators in more than 30 different cities either by phone or email. In most cities, the relevant city employees were easily identified since set asides programs have been reconstituted into small business or disadvantaged business outreach programs. In other cases, when calling the city administrative offices, we simply asked to speak to someone familiar with the minority business outreach programs in that city. In each case, we spoke with current city employees or affiliates of an auxiliary business outreach program. During these conversations, the dates of the city set asides programs from other sources were cross-checked and additional documentation was requested when applicable. It is important to note that most individuals provided details on programs that are now defunct or radically restructured, so the traditional concerns about retrospective bias will apply. Still, given our supporting documentation, we feel that this bias is not especially severe in our current study.

Some of the city level programs date back as early as 1977 (Portland, OR), while most others started during the mid 1980s. In most cases, the programs began as loosely defined goals for minority business utilization and later were formally enacted by city ordinance. At this point, the amount of time necessary to set up agencies to register minority owned businesses, track statistics on utilization, and conduct outreach in the community varied between cities.

The construction industry was the focus of the set-aside programs in most cities because of the low costs of entry, presumed minority experience in the sector, and the considerable opportunities to move from worker to manager to owner (City of Greensboro, North Carolina Disparity Study, 1992). Still, the construction set asides were not ideal for every city. The Portland program administrators recognized early on that minority participation in construction was low, and that more opportunities were needed in the goods and services sector. In Seattle, the administrators found that construction was still a difficult business for many minority entrepreneurs, even with the set asides, because the programs lacked substantial assistance for developing business skills. According to one source, with regards to almost every aspect of the construction business, from financing to bond insurance to relationship management, minority owned firms in Seattle continued to struggle because the set aside programs only guaranteed a steady flow of contracts without requiring or supporting the necessary upgrade in business management skills.

The qualitative question of "Did these programs actually achieve their goals?" was posed to many programs administrators, and the responses were mixed. Most administrators concluded that while the programs seemed to be having tangible results, outreach (identifying and contacting minority owned businesses) was sometimes difficult, the list of “certified” firms did not include every minority owned firm in the city, and enforcement of the targets or goals was often difficult. There were a few concerns about fraud, (i.e. a husband transferring his business into his wife’s name to take advantage of the set-aside), but some cities had measures to protect against this type of behavior.

Most administrators identified other economic and social factors, not directly addressed by the set-aside programs, which often kept the initiatives from succeeding as well as expected. The availability of performance bonds for minority owned firms, the role of unions in the local economy, and the support of the program by key officials, like the Mayor, were mentioned often. Despite these obstacles, most of the program administrators we spoke to felt that minority participation in public contracting had declined after the programs ended, suggesting that the programs had some positive effect.

Many program administrators identified the obtainment of performance bonds as a main challenge for minority owned businesses. An administrator who was involved in Portland’s minority business program remarked that since bond insurance was likely required for prime contractors, minority owned firms often filled the subcontractor role so they would not have to post a costly bond. The City of Charlotte’s (North Carolina) 1993 MWBE Disparity report also found that minority owned firms were less likely to have performance bonds than white-owned firms. Similar insights were included in the 1993 study from the city of Ft. Worth, Texas and in comments from an administrator in Seattle, Washington.

Unions also played a large role in limiting opportunities for minority entrepreneurs, according to those involved in many city programs. In Albany, NY, one source familiar with the program remarked that unions often "adapted" to meet the goals for minority participation without really meeting them. The city of Greensboro, N.C. also noted, in a 1992 disparity study, that the lack of union projects in North Carolina was a positive benefit for minority entrepreneurs, since "union shops and apprenticeship programs" were often subject to "discriminatory practices". Since working with unions was often the most effective way to reach potential minority entrepreneurs, especially in construction, many city employees complained that unions could have done more to include minorities in their membership and expand opportunities.

Finally, in cities where public officials strongly supported the program, like Atlanta, Georgia, or where minority political and economic power were high, the set aside programs were perceived to have greater success and tended to survive legal and political challenges. Since prime contractors were responsible for meeting the goals of the programs in their subcontracting behavior, the political power and enforcement capabilities of the relevant city administrators had an appreciable impact on compliance. In some cities, program officials complained that primes did not make good faith efforts to meet the goals of the programs, but there was little recourse.

In sum, while the city set aside programs were far from comprehensive in addressing the obstacles to minority self-employment, most of the city employees we spoke with asserted that the programs did achieve at least some of their intended goals. When the programs were challenged, and later discontinued or restructured, it was not clear whether most minority firms survived, since many had been dependent on government contracts given out through the preferential bidding process, and had not expanded their client base.

Appendix Table 4.1 shows the dates from all four sources. For many cities there is agreement with the MBELDEF and JCPES dates, for quite a few there is disagreement. Appendix Table 4.2 shows the 13 cities in which the dates are verified in at least 2 of the sources. Note that many of these cities are large. Most had set-aside dates in the early-to-mid-1980s, while Washington DC had a program established before the period of interest and New York City did not have a program until after the 1980s. Appendix Table 4.3 shows the additional 15 cities that had a date from only one source, thus it did not conflict across sources.

The below analysis will utilize two different estimation samples – the 13 cities with cross-validated dates and the 28 cities with either cross-validated dates or dates from only 1 source. Thus, we only consider cities where we feel certain that a program was instituted at a particular date. Thus, we drop the remaining 16 cities that we could not get program data from or that had a conflicting date in multiple sources. Note that the 13 cities with non-conflicting dates comprise the majority of the population in the 28-city sample. All analyses are weighted using the CPS population weights.

4.4 CPS Data and Trends in Self-Employment

The primary data source on self-employment comes from the 1979-1989 Current Population Survey (CPS) Merged Outgoing Rotation Group (ORG) files which contain information for one-quarter of the individuals in each monthly CPS. Combining the observations from the monthly surveys results in annual samples which are roughly three times larger than a monthly CPS. The large sample sizes are important since the group of interest, self-employed black men, is small relative to the population. In addition, it allows for a more informative, disaggregated city-level analysis. On the other hand, the ORGs only include data on work characteristics in the survey week and do not contain information on self-employment income. As a result, we also analyze the 1979-1989 CPS March Annual Demographic Files (ADF), which contain information on work characteristics and income received during the previous calendar year. Due to its smaller sample sizes, a precise analysis with the ADF files is not possible, and we do not present the results here (but they are presented in Chay and Fairlie 1998).

From these data we construct annual information on minority and non-minority patterns of self-employment from 1979 to 1989. The period after 1989 is not included since the 1989 *Croson* decision led to the suspension or dismantling of set-aside programs in several cities. For now, we focus on examining the impact of the original set-aside programs and leave an examination of the effects of *Croson* to future research. Although the CPS only allow for an analysis of self-employment rates (defined as the ratio of the self-employed to the population), it is a natural initial outcome to examine. Historically, self-employment has been a route of economic advancement for disadvantaged groups (Glazer and Moynihan 1970). In addition, high self-employment rates for racial and ethnic groups are strongly associated with high average incomes (Fairlie and Meyer 1996).

Identification of self-employed workers is based on the unedited class of worker question in the CPS and is defined to be those individuals who self-report being self-employed in their own incorporated or "not incorporated" business.[75] This question refers to the job with the most hours during the reference week in the ORG files.[76] We include only non-Hispanic white and black men aged 20 to 64 in this analysis. All group specific self-employment rates are taken relative to the entire race-specific populations.

Since the analysis is focused on city-level changes in racial self-employment rates, consistently matching cities over time in the CPS is imperative. In the ORG files, the 44 largest Statistical Metropolitan Statistical Areas (SMSAs) can be identified from 1979 to September 1985 based on their 1970 Census population size ranking (57 SMSA identifiers). After 1985, the city coding scheme changed to a more detailed system with 252 Consolidated MSA (CMSA) ranking identifiers, some subdivided into as many as 12 Primary MSA (PMSA) ranking codes. Matches for the 1986-89 cities to the 1979-85 cities in the ORG were based on making the CMSA and PMSA rankings compatible with the SMSA rankings. Population totals were examined to gauge the quality of the match. City identifiers are missing for all records during the last three months of 1985. The final

[75] Unpaid family workers are not counted as self-employed.
[76] In the future, we will redo the analysis dropping individuals who are not employed in the reference week and individuals who report being part-time status.

analysis is based on the 44 MSAs that can be consistently matched over the entire decade.[77]

Table 4.1 presents some summary information on the characteristics of the self-employed, by race and gender, for the 1979-89 ORG data. There are several points of interest. First, blacks men are much less likely to be self-employed than white men. Similarly, black men who are self-employed are two times less likely to be incorporated than self-employed white men. The 13-city and 28-city samples used below account for 28 and 39 percent of all black men in the U.S and 14 and 22 percent of all white men. The unemployment rate and the employment-population ratio are much higher and lower, respectively, for black men relative to white men in the MSAs with set-aside information. Finally, the self-employed are generally older and better-educated than those not self-employed.

Table 4.2 shows the distribution of self-employment across industries by race. The industries are roughly ordered from top to bottom by sectors that are more and less likely to be impacted by minority business set-asides. Following the results from Bates (1993) and conversations with city officials, we identify construction as the most affected industry followed by wholesale trade, business and repair services, and transportation,

[77] It is worth noting that the matched MSAs are the largest cities and contain a substantial fraction of the black population. In addition, San Francisco and Oakland and Tampa and St. Petersburg are not separately identified from each other before 1986. Consequently, although these areas had different affirmative action dates, San Francisco-Oakland was assigned San Francisco's date and Tampa-St. Petersburg was assigned Tampa's date.

communication, and other utilities.[78] 22 percent of all self-employed black men and 19 percent of self-employed white men are in construction.

Figure 1A plots the self-employment rate (relative to the population) for white and black men in the entire U.S. and in the 13-city and 28-city MSA samples. Black self-employment rates in the entire U.S. increased from about 4.2 percent to nearly 5 percent (nearly 20 percent) in the mid-1980s. The mid-decade jump for black self-employment is even larger in the MSAs with reliable set-aside program dates. In the 13-city sample, black self-employment increased by 1.5 percentage points (over 30 percent) in the mid-1980s. For the 28-city sample, the black self-employment increase is nearly 25 percent. Interestingly, self-employment rates for white men in the 13-city and 28-city samples grew more from 1979-1983 than after 1983.[79]

Figure 1B shows the patterns in self-employment for cities during the 1980s that did not have a program until after 1989 and the cities that had a program preceding 1979. (Programs after 1989 were often fashioned as disadvantaged business programs as discussed above) Consistent with a program effect, while the trends in black and white self-employment within the two groups are similar during the 1980s, the black-white self-employment gap is significantly smaller in cities that had a pre-existing program relative to cities that did not implement a program before 1990.

[78] Bates (1993) suggests that black-owned businesses historically have been heavily concentrated in the small-scale service and retail trade sectors, which had limited growth potential and where most owners had very low incomes. During the 1980s and 1990s black firms have emerged in construction, wholesale trade, and business services; sectors with large-scale firms that have higher survival rates and that are more likely to create jobs for minorities.

[79] It is noteworthy that the fraction of men in incorporated self-employment also rose for blacks after 1983 and leveled off for whites. Fairlie and Meyer (1998) use the decennial censuses to document that while the relative black-white self-employment rate ratio declined from 1970-1980, it increased from 1980-1990.

Figure 2 graphs employment-population ratios by race in the U.S. and in contributing MSAs to provide a window to the changing economic conditions during this period. The employment experiences of black and white men in contributing MSAs were similar to those of the U.S. male population. Employment rates declined sharply for black men in contributing MSAs from 1979-82, but rebounded after 1983 to above its original level by the end of the decade. White employment rate patterns appear to be less cyclically sensitive. Consequently, the striking jump in self-employment rates for black men in the early to mid-1980s does not seem to be attributable to worsening economic conditions for blacks. In addition, although the trends are only suggestive, the rebound in overall black employment rates coincided with the growth in black self-employment.[80] This point is especially notable, because a secondary goal of the set aside programs was to increase black employment in urban areas.

These empirical regularities were previously undocumented and are suggestive of a potential impact of city-level set-aside programs. The natural next step is to try to measure the factors that underlie the large changes in black self-employment that occurred in the early to mid 1980s. The timing of the gains and their geographic and industrial location suggest that one important explanatory factor may be city-level minority business set-aside programs.

[80] Trends in average education and age during the 1980s are very similar for self-employed black and white men and black and white workers.

4.5 Econometric Models

First we estimate the unrestricted "program effects" using an "event study" or double difference analysis that derives the set-aside "response function" while regression adjusting for unrestricted year effects, city fixed effects, and the observable characteristics of black and white men.

(1) $$y_{ijt} = \alpha_j + \lambda_t^r + X_{ijt}'\beta + \sum_{t-s=-8}^{8} \theta_{t-s}^r \cdot 1(t_{ij} - s_j = t - s) + u_{ijt} ,$$

where y_{ijt} is an indicator variable equal to one if person i in city j is self-employed in survey year t, α_j are city-level fixed effects, λ_t^r are survey year fixed effects that are allowed to vary by race r, and X_{ijt} is a vector of personal characteristics (education, age, age-squared, and education interacted with age).

The next set of indicators, $1(t_{ij} - s_{ij} = t - s)$, measure the program response function – that is, each is equal to one if the survey year for an individual in city j minus the year of set-aside program initiation in city j is equal to some value between -8 and +8. The specification allows the program effects to vary by race, and will be used to plot the unrestricted response functions of black and white self-employment rates, as well as their difference. Below, we present plots of the response functions from specifications that include: 1) no controls (i.e., the raw patterns); 2) controls for city fixed effects and black and white year effects; and 3) controls for the city fixed effects, race-specific year effects, and individual characteristics.

Next, we fit more restrictive regression models to the data to derive estimates of the magnitude and statistical significance of the program effects. In particular, we estimate: 1) difference-in-differences models of the black-white relative gain in self-employment after set-aside initiation date; and 2) the post-program black-white differences in self-employment deviated from a pre-program trend that is allowed to vary by race.

(2)

$$y_{ijt} = \alpha_j + \lambda_t + \lambda_t^{Black} + X_{ijt}'\beta + constant + \delta \cdot Black_{ijt} + \gamma \cdot 1(t > s) + \theta\left(Black_{ijt} \cdot 1(t > s)\right) + u_{ijt},$$

and

(3)

$$y_{ijt} = \alpha_j + \lambda_t + \lambda_t^{Black} + X_{ijt}'\beta + \delta_1 \cdot (t - s) + \delta_2 \cdot Black_{ijt} \cdot (t - s) + \sum_{t-s=0}^{8} \gamma_{t-s} \cdot 1\left(t_{ij} - s_j = t - s\right) + \sum_{t-s=0}^{8} \theta_{t-s} \cdot Black_{ijt} \cdot 1\left(t_{ij} - s_j = t - s\right) + u_{ijt}.$$

Here, the policy effects of interest are θ in equation (2) – as it measures the gain in black self employment relative to whites after program initiation – and θ_{t-s} in equation (3), which measures the black-white differences in self-employment after program initiation deviated from race-specific pre-program trends.

The analysis below allows the error term, u_{ijt}, to be heteroskedastic and to be correlated over time within cities conditional on the city fixed effects (i.e., time-series clustering at the city-level). We estimate all 3 questions separately for the 13-city and 28-city samples.

We also estimate these equations when the outcome is an indicator for whether an individual is employed – thus, measuring changes in the employment-to-population rates relative to the timing of the set-aside programs.

4.6 Double Difference Results on the Impact of Set-Asides

Figure 3 shows the patterns from estimating equation (1) for the 13-city sample. In Panel A, we include controls for MSA fixed effects and race specific year effects. In Panel B, we add additional controls for education and age. The broad pattern illustrated in these figures is consistent with a large gain in black self-employment rates of approximately 4% after program initiation. Interestingly, after controlling for education and age, the gap between black and white self-employment narrows.

In Table 4.3, we report regression results for the 13 city sample as outlined in equations 2 and 3. Similarly to the pattern in Figure 3, we find significant convergence in the black-white gap in self-employment after the programs are initiated, especially when we adjust for race-specific year effects, individual characteristics, and MSA fixed effects. Columns 4, 5, and 6 report the differences in differences results. In sum, the program effects are highly statistically significant even after we adjust for many potential forms of clustering and heteroskedasticity.

In Figure 4, we find the same patterns for the 28 city sample. Table 4.4 reports the differences in differences regression results in Columns 4, 5, and 6. With a larger sample, our results are even more precise and Figure 4 demonstrated an even more precisely timed improvement in black relative self-employment rates. For more detailed trends by city, see Appendix Figure 1 and Figure 2.

In Figure 5, we analyze the program effects by industry. In Panel A, we look at the 13 city sample and consider the 28 city sample in Panel B. We once again find evidence of relative improvement, but the effect is less striking than in previous figures. Tables 4.5 and 6 present the regression results to complement Figure 5. The effect sizes are again significant, but it is important to note that construction accounts for only 10% of the overall improvement in black self employment rates after program initiation, contrary to anecdotal claims that set asides only impacted the construction industry. The other sectors that were most affected, including wholesale trade and professional services, are also important drivers of the growth in self-employment. As mentioned above, we have little qualitative evidence of how set asides worked in these sectors so further research is required to properly interpret these results.

Next, we turn to the impact on overall employment. In Table 4.6, we estimate equation 1, but the dependent variable is now whether or not the individual is employed. We find some evidence that relative employment for blacks increases after the programs are initiated. In Table 4.7, we estimate equations 2 and 3 using employment as the dependent variable. We find a significant increase in black relative employment rates after the programs are enacted. In fact, for every 1% increase in black self-employment, the black employment rates increases 1.5%. Thus, we find evidence that set asides programs may have had a small impact on black employment as well.

4.7 Conclusion and Future Directions

Based on the available data and careful econometric approach, we find that minority set-asides programs had statistically and economically significant effects on black self-employment and employment rates during the 1980s. Our findings represent a major contribution to the policy debate on affirmative action and the most comprehensive evaluation of contracting set-asides to date. However, we note several caveats with our analysis. First, our approach is a descriptive first pass at the question and reduced-form in nature since more detailed data on the "first-stage" importance of these programs (e.g., contract awards and amounts) are not readily available. Second, since the goals of set-asides are to enhance the long-run success of minority-owned firms and alleviate unemployment among blacks in cities, a definitive analysis would entail examining outcomes such as business revenues, profits, and failure rates and the racial composition of those employed in minority firms. In the current paper, we treat all of the programs as identical. However, since the set aside percentage differed by city, it may be interesting to look at the varying impact across cities in a more robust manner. Using insights from prior work, we could use mayoral characteristics, voting patterns, and demographics to proxy for the level of enforcement that the programs may have received. Finally, we conclude below that the dates provided by MBELDEF (1988) and JCPES (1994) are often inconsistent with each other and may not be accurate. Remarkably, it appears that there is no consensus in the literature on even the dates in which cities passed set-aside programs. Consequently, the modest goal of this study is to establish some empirical facts on the evolution of self-employment rates of black men during the 1980s. We will discuss the above issues more below.

Additionally, the effects of these relatively targeted set aside programs seem quite large based on our empirical results. We can do some very rough calculations to gain perspective on whether our results are within a reasonable range in terms of the expected impact of these programs. For example, the city of Atlanta spent $191 million on minority (including women) owned businesses during the 1980s. The approximate population of blacks in Atlanta during this period was 500,000, so a 2.5% to 5% increase in self employment would imply 12,500 to 25,000 new black entrepreneurs in the 5 years after the program was enacted. This is a large increase, even considering the fact that many construction and professional service firms can be very small. After all, this would mean that the average yearly local contract would be worth between $3800 and $7600.

There are several possible explanations for this. First, other local, state, and federal programs may have increased the self-employment rate among blacks during this period. If the typical amount spent on city contracts could not reasonably create a 2.5% to 5% increase in the black self-employment rate, perhaps these other programs had a large impact as well. Future research should establish a credible and comprehensive list of these programs and redo the analysis in this paper. Next, there may be an unobserved reason why blacks began to list themselves as self-employed in the years after the programs were enacted. Anecdotal evidence suggests that white business owners may have begun to list black employees as owners to take advantage of set asides, but no robust empirical study has demonstrated this. The significant increase in black employment during this period also makes this type of widespread fraud seem less likely because dramatic increases in black self employment due to fraud could actually decrease the black employment rate.

Future research should also further consider the crucial role of networks among minority entrepreneurs that may magnify the impact of programs like minority set asides. For example, Kilby (1983) argues that minority entrepreneurs develop valuable networks that allow them to minimize risk, obtain capital, and identify new opportunities. Thus, programs that initially create a critical mass of African-American entrepreneurs may result in creating a far greater number of African-American entrepreneurs over time, as networks are developed to provide capital, risk sharing, and mentorship. Since networks of entrepreneurs can often lower transaction costs through mechanisms like the effective enforcement of contracts (Greif, 1993; Ramachandran and Shah, 1999), an emerging network of African-American entrepreneurs could reduce the costs of entry for prospective African-American entrepreneurs, thus multiplying the impact of a particular set aside program. In addition, other scholars have discussed the potential importance of a “demonstration” effect that can result from entrepreneurship programs. (Nolan, 2003) While difficult to establish empirically, successful entrepreneurs in minority communities not traditionally involved in entrepreneurship can act as an inspiration to other members of the group to pursue the same goal. Nolan (2003) points out that Scotland introduced a program to promote entrepreneurship that showcased successful Scottish entrepreneurs, intending to reverse cultural biases against entrepreneurship. The cultural barrier to entrepreneurship is yet another “cost” of entry, and policies that create successful minority entrepreneurs can lower this cost for other members of the same group. While these explanations may be part of the large effect observed in this study, it is unlikely that they can account for the entire effect, so future work must investigate the effect size more comprehensively.

Furthermore, it is surprising that these significant increases in self-employment occurred so soon after the programs were enacted. While our research design utilizes the staggered introduction of these laws in each city, perhaps these set asides were announced publicly some time before they were actually passed as law, allowing incipient entrepreneurs ample time to start a new business. While a new construction firm can be started quickly by an experienced employee, we might also expect the impact of set asides on self-employment to grow sharply over time while prospective entrepreneurs assess the structure and stability of the programs. Our results do not indicate this pattern, so further investigation into the timing of business formation in response to these types of government programs is required.

Noting the caveats above, this analysis represents the first step towards understanding the impact of local set aside programs. We hope that the data used for this study will spur the collection of new data to address the concerns and limitations above. For future research, we hope to obtain access to the Census-like microdata in the Characteristics of Business Owners (CBO) surveys and the Surveys of Minority-Owned Business Enterprises (SMOBE) at a Census Data Center secure site (e.g., the Berkeley-UCLA center). In addition to containing very detailed data on the characteristics of the business owners, their businesses, and the workers in the firm, these surveys have information on the fraction of sales coming from government contracts (Bates 1990 and Boston 1998). In addition, the surveys are conveniently timed every five years (e.g., 1982, 1987, and 1992), and can be complemented with the 1980 and 1990 5-percent Censuses of the Population.

Finally, we will examine other, potentially more convincing, research designs for measuring the impact of set-aside programs. As a result of the 1989 Croson decision, which effectively nullified race-based local contracting programs that were not based on proven past discrimination, many state and city governments suspended or dismantled their set-aside programs. A potential "quasi-experiment" would use the differential changes in minority contract awards across localities induced by the Croson decision to identify the set-aside program effects. Such an analysis could be more robust since the differential effects of the court decision may be unrelated to underlying local economic and political conditions.

In addition to getting the right dates, it will be important to find annual city-level contract awards information, by industrial sector. This data would also lend further insight into which industries may have been most impacted. The ideal analysis of the CPS data would entail relating black-white relative self-employment rates to variation in the timing, proportion, and amount of government contracting dollars set-aside for minority businesses across cities and industrial sectors over time. We aim to address these extensions in future research and will validate or reconsider the results found in this study based on our findings.

References

Bates, Timothy. 1993. Banking on Black Enterprise. Washington, D.C.: Joint Center for Political and Economic Studies.

Bates, Timothy. 1990. "New Data Bases in Human Resources: The Characteristics of Business Owners Data Base," Journal of Human Resources, 25(Fall): 752-56.

Bates, Timothy. 1988. "Do Black-Owned Businesses Employ Minority Workers? New Evidence," Review of Black Political Economy, Spring: 51-64.

Bates, Timothy. 1985. "Impact of Preferential Procurement Policies on Minority-Owned Businesses," Review of Black Political Economy, Summer: 51-65.

Bates, Timothy, and Darrell Williams. 1996. "Do Preferential Procurement Programs Benefit Minority Business?" American Economic Review, 86(2): 294-97.

Bates, Timothy, and Darrell Williams. 1995. "Preferential Procurement Programs and Minority-Owned Businesses," Journal of Urban Affairs, 17(1): 1-17.

Bates, Timothy, and Darrell L. Williams. 1993. "Racial Politics: Does It Pay?" Social Science Quarterly, 74(3): 507-22.

Boston, Thomas D. 1998. "Trends in Minority-Owned Businesses," Paper presented at the National Research Council Research Conference on Racial Trends in the United States.

Chay, Kenneth Y. and Robert W. Fairlie. 1998. "Minority Business Set-Asides and Black Self-Employment," Paper presented at the 1999 American Association Meetings.

Enchautegui, Maria E., Michael Fix, Pamela Loprest, Sarah von der Lippe, and Douglas Wissoker. 1996. Do Minority-Owned Businesses Get a Fair Share of Government Contracts. Washington, D.C.: Urban Institute.

Fairlie, Robert W., and Bruce D. Meyer. 1998. "Trends in Self-Employment Among White and Black Men: 1910-1990," mimeograph, University of California, Santa Cruz.

Fairlie, Robert W., and Bruce D. Meyer. 1996. "Ethnic and Racial Self-Employment Differences and Possible Explanations," Journal of Human Resources 31(4): 757-93.

Glazer, Nathan and Daniel P. Moynihan. 1970. Beyond the Melting Pot: the Negroes, Puerto Ricans, Jews, Italians, and Irish of New York City, 2nd Edition. Cambridge: MIT Press.

Greif, Avner, "Contract Enforceability and Economic Institutions in Early Trade: The Maghribi Traders' Coalition," *American Economic Review*, June 1993

Joint Center for Political and Economic Studies. 1994. Assessment of Minority Business Development Programs, Report to the U.S. Department of Commerce Minority Business Development Agency, Washington, D.C.

Kilby, Peter, "The Role of Alien Entrepreneurs in Economic Development," *American Economic Review* Papers and Proceedings, Vol. 73, pp. 107-111, May 1983.

LaNoue, George R. 1992. "Split Visions: Minority Business Set-Asides," Annals of the American Academy, AAPSS, 523(September): 104-16.

Minority Business Enterprise Legal Defense and Education Fund. 1988. Report on the Minority Business Enterprise Programs of State and Local Governments.

Myers, Samuel L., Jr. 1997. "Minority Business Set-Asides," Entry in Encyclopedia of African-American Business History. Greenwood Press.

Myers, Samuel L. Jr., and Tsze Chan. 1996. "Who Benefits from Minority Business Set-Asides? The Case of New Jersey," Journal of Policy Analysis and Management, 15(2): 202-26.

Nolan, Alistair. 2003 ""Entrepreneurship and Local Economic Development: Policy Innovations in Industrialized Countries-Discussion." Kansas City Federal Reserve Bank

Ramachandran, V. and Shah, M.K. (1999) "Minority Entrepreneurs and Firm Performance in Sub-Saharan Africa." Regional Program on Enterprise Development Discussion Paper 86

Rice, Mitchell F. 1991. "Government Set-Asides, Minority Business Enterprises, and the Supreme Court," Public Administration Review, 51(2): 114-22.

Table 4.1 Summary Statistics for Black and White men aged 20-64 from 1979 to 1989, CPS-ORG Data

	Entire United States		13 cities with verified program dates		28 cities with no conflict in program dates	
	Black	**White**	**Black**	**White**	**Black**	**White**
All men						
Age	**37.3**	**39.1**	**37.4**	**39.0**	**37.3**	**38.9**
Education	**11.7**	**13.0**	**12.2**	**13.8**	**12.1**	**13.6**
Percent unemployed	**10.9**	**4.7**	**10.6**	**3.9**	**11.1**	**4.3**
Percent out of labor force	**16.7**	**10.3**	**17.8**	**9.8**	**17.5**	**9.8**
Percent self-employed	**4.6**	**14.8**	**5.2**	**14.0**	**4.9**	**13.4**
Percent in MSA	**71.7**	**59.4**	**96.6**	**97.2**	**97.6**	**98.3**
Self-employed men						
Age	**42.8**	**43.1**	**42.1**	**42.7**	**42.0**	**42.6**
Education	**12.0**	**13.4**	**12.7**	**14.3**	**12.6**	**14.1**
Percent incorporated	**13.2**	**27.0**	**15.8**	**33.5**	**15.2**	**33.5**
Percent in MSA	**72.1**	**53.4**	**96.6**	**97.7**	**97.5**	**98.5**
Sample Size	**112,111**	**1,132,459**	**31,904**	**154,271**	**43,678**	**251,756**

Notes: Data come from the 1979 to 1989 Merged Outgoing Rotation Groups files of the Current Population Survey, and are limited to observations that are aged 20-64 and are black or non-Hispanic white men. The statistics are weighted by the CPS sample weights. The samples underlying each set of columns are explained in the text and in the Appendix tables.

Table 4.2 Distribution of Self-Employment across Industries for Black and White Men (in percent), 1979 to 1989 CPS-ORG Data

	Entire United States		13 cities with verified program dates		28 cities with no conflict in program dates	
	Black	**White**	**Black**	**White**	**Black**	**White**
Most affected industries						
Construction	**21.8**	**18.5**	**18.5**	**16.4**	**18.3**	**18.0**
Professional services	**7.8**	**12.0**	**9.3**	**18.0**	**9.1**	**16.9**
Wholesale trade	**2.6**	**5.4**	**3.1**	**6.7**	**3.3**	**6.7**
Business services	**6.5**	**4.7**	**8.1**	**7.6**	**8.7**	**6.9**
Repair services	**9.3**	**5.8**	**9.2**	**5.0**	**9.3**	**5.3**
Less affected industries						
Transport, communicat, and utilities	**9.9**	**4.0**	**11.4**	**3.6**	**11.3**	**3.8**
Retail trade	**11.9**	**14.4**	**14.3**	**14.4**	**13.6**	**14.2**
Finance, insurance, and real estate	**3.4**	**6.1**	**4.4**	**8.1**	**4.1**	**7.6**
Personal services	**5.0**	**2.6**	**5.5**	**3.1**	**5.2**	**2.9**
Entertainment and HH Services	**4.3**	**2.1**	**6.6**	**3.9**	**6.0**	**3.5**
Least affected industries						
Manufacturing	**3.6**	**5.8**	**2.1**	**7.1**	**2.3**	**7.1**
Agriculture	**8.2**	**14.3**	**2.8**	**2.8**	**3.8**	**3.5**
Mining, forestry, fishery	**0.4**	**1.0**	**0.2**	**0.3**	**0.2**	**0.4**
Public administration	**0.0**	**0.0**	**0.0**	**0.1**	**0.0**	**0.1**
Other	**5.4**	**3.5**	**4.5**	**3.1**	**4.8**	**3.1**
Sample Size	**5,327**	**177,765**	**1,701**	**21,561**	**2,212**	**33,790**

Notes: Data come from the 1979 to 1989 Merged Outgoing Rotation Groups files of the Current Population Survey, and are limited to observations that are aged 20-64 and are black or non-Hispanic white men. The statistics are weighted by the CPS sample weights. The samples underlying each set of columns are explained in the text and in the Appendix tables.

Table 4.3 Estimated changes in black relative self-employment rates after set-aside program,
For the 13 cities with program dates verified in multiple sources

	Difference-in-Differences			Post-Program Effects deviated from Pre-Program Trend		
Regression coeff (x100)	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Pre-program black diff	**-7.45*****	**-7.57*****	**-5.57*****			
	(0.70)	**(0.79)**	**(0.77)**			
Pre-program trend diff				**-0.11**	**-0.07**	**-0.13**
				(0.14)	**(0.16)**	**(0.17)**
Post-program b-w gain	**2.28****	**2.62*****	**2.59*****			
	(0.88)	**(0.96)**	**(1.03)**			
Post-program b-w diff (relative to pre-program trend)						
Year 0				**1.42****	**1.33***	**1.58****
				(0.66)	**(0.75)**	**(0.74)**
Year 1				**2.45**	**2.46****	**2.61****
				(1.45)	**(1.28)**	**(1.19)**
Year 2				**3.45****	**3.43*****	**3.78*****
				(1.29)	**(1.32)**	**(1.52)**
Year 3				**2.99***	**3.07****	**3.43****
				(1.52)	**(1.42)**	**(1.51)**
Year 4				**2.38**	**2.54***	**3.08****
				(1.58)	**(1.41)**	**(1.57)**
Year 5				**3.53****	**3.46****	**4.25*****
				(1.53)	**(1.56)**	**(1.72)**
Year 6				**4.74*****	**4.67*****	**5.40*****
				(1.36)	**(1.60)**	**(1.88)**
Year 7				**5.91****	**5.82****	**6.48****
				(2.25)	**(2.57)**	**(2.93)**
Year 8				**8.58*****	**8.38*****	**8.63*****
				(1.83)	**(2.11)**	**(2.29)**
Post-program white gain	**-1.45**	**-0.85****	**-0.84****			
	(0.87)	**(0.43)**	**(0.43)**			
Pre-program white trend				**0.12**	**0.34*****	**0.24****
				(0.22)	**(0.10)**	**(0.12)**
City fixed effects	**N**	**Y**	**Y**	**N**	**Y**	**Y**
Black, white year effects	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**
Education, age controls	**N**	**N**	**Y**	**N**	**N**	**Y**

Notes: ***, **, and * indicate statistical significance at the 1, 5, and 10 percent level, respectively. The dependent variable is equal to one if the person reports being self-employed and zero otherwise. The sample consists of 163,812 black and non-Hispanic white men, aged 20-64, in 13 MSAs. Data come from the 1979 to 1989 Merged Outgoing Rotation Groups files of the Current Population Survey, and are limited to observations that are no more than eight years away from the set-aside program date in the city. The estimates come from linear probability models, where the standard errors have been corrected for heteroskedasticity and cross-sectional and time-series clustering in the residuals at the MSA level. The education, age controls include linear terms for education and age, age-squared, and the interaction of education and age. The regressions are weighted by the CPS sample weights.

Table 4.4 Estimated changes in black relative self-employment rates after set-aside program,
For the 28 cities with program dates that do not conflict with other sources

	Difference-in-Differences			Post-Program Effects deviated from Pre-Program Trend		
Regression coeff (x100)	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Pre-program black diff	**-7.47*****	**-7.58*****	**-5.80*****			
	(0.59)	**(0.66)**	**(0.64)**			
Pre-program trend diff				**0.03**	**0.05**	**-0.01**
				(0.13)	**(0.15)**	**(0.15)**
Post-program b-w gain	**2.94*****	**3.11*****	**3.07*****			
	(0.70)	**(0.74)**	**(0.79)**			
Post-program b-w diff (relative to pre-program trend)						
Year 0				**2.89*****	**2.79*****	**2.89*****
				(0.73)	**(0.75)**	**(0.70)**
Year 1				**2.45****	**2.37****	**2.57*****
				(1.04)	**(1.01)**	**(0.94)**
Year 2				**2.27****	**2.13****	**2.45****
				(0.90)	**(0.95)**	**(0.99)**
Year 3				**3.29*****	**3.14*****	**3.41*****
				(0.97)	**(1.00)**	**(1.02)**
Year 4				**2.79****	**2.73****	**3.13****
				(1.10)	**(1.10)**	**(1.23)**
Year 5				**2.86*****	**2.76****	**3.44*****
				(1.00)	**(1.09)**	**(1.22)**
Year 6				**3.69*****	**3.65****	**4.01****
				(1.35)	**(1.47)**	**(1.65)**
Year 7				**4.21****	**4.21****	**4.69****
				(1.64)	**(1.80)**	**(2.01)**
Year 8				**5.18*****	**5.09*****	**5.13****
				(1.87)	**(1.92)**	**(2.17)**
Post-program white gain	**-2.02*****	**-0.86****	**-0.86****			
	(0.60)	**(0.36)**	**(0.36)**			
Pre-program white trend				**-0.07**	**0.33*****	**0.23****
				(0.16)	**(0.08)**	**(0.09)**
City fixed effects	**N**	**Y**	**Y**	**N**	**Y**	**Y**
Black, white year effects	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**
Education, age controls	**N**	**N**	**Y**	**N**	**N**	**Y**

Notes: ***, **, and * indicate statistical significance at the 1, 5, and 10 percent level, respectively. The dependent variable is equal to one if the person reports being self-employed and zero otherwise. The sample consists of 268,987 black and non-Hispanic white men, aged 20-64, in 28 MSAs. Data come from the 1979 to 1989 Merged Outgoing Rotation Groups files of the Current Population Survey, and are limited to observations that are no more than eight years away from the set-aside program date in the city. The estimates come from linear probability models, where the standard errors have been corrected for heteroskedasticity and cross-sectional and time-series clustering in the residuals at the MSA level. The education, age controls include linear terms for education and age, age-squared, and the interaction of education and age. The regressions are weighted by the CPS sample weights.

Table 4.5 Estimated changes in black relative self-employment after implementation of a set-aside program by industry group,
13- and 28-city samples

	Difference-in-Differences Model								
	Most affected industries			Less affected industries			Least affected industries		
Regression coeff (x100)	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**
A. 13-city sample results									
Pre-program black diff	-4.41***	-4.47***	-2.91***	-1.78***	-1.86***	-1.54***	-0.92***	-0.89***	-0.81***
	(0.47)	(0.53)	(0.52)	(0.33)	(0.35)	(0.35)	(0.23)	(0.25)	(0.23)
Post-program b-w gain	1.32***	1.41***	1.42***	0.59	0.83	0.81	0.23	0.23	0.23
	(0.36)	(0.31)	(0.32)	(0.45)	(0.50)	(0.49)	(0.32)	(0.35)	(0.35)
Post-program wht gain	-0.50	-0.40	-0.41	-0.85**	-0.32*	-0.30*	-0.04	-0.08	-0.07
	(0.37)	(0.32)	(0.32)	(0.37)	(0.17)	(0.17)	(0.25)	(0.14)	(0.14)
B. 28-city sample results									
Pre-program black diff	-4.31***	-4.37***	-2.96***	-1.75***	-1.84***	-1.55***	-1.07***	-1.04***	-0.98***
	(0.39)	(0.43)	(0.42)	(0.26)	(0.29)	(0.28)	(0.18)	(0.19)	(0.18)
Post-program b-w gain	1.76***	1.68***	1.66***	0.64*	0.94**	0.93**	0.45*	0.37	0.37
	(0.29)	(0.26)	(0.25)	(0.37)	(0.38)	(0.38)	(0.26)	(0.32)	(0.32)
Post-program wht gain	-0.90***	-0.35	-0.36	-0.85***	-0.25*	-0.24*	-0.20	-0.21*	-0.20*
	(0.30)	(0.23)	(0.22)	(0.27)	(0.14)	(0.14)	(0.22)	(0.12)	(0.12)
City fixed effects	N	Y	Y	N	Y	Y	N	Y	Y
Black, white year effects	Y	Y	Y	Y	Y	Y	Y	Y	Y
Education, age controls	N	N	Y	N	N	Y	N	N	Y

Notes: ***, **, and * indicate statistical significance at the 1, 5, and 10 percent level, respectively. The dependent variable is equal to one if the person reports being self-employed in the industry group and zero otherwise. See Table 2 for the listing of industries in each of the three industry groups. See notes to Tables 3 and 4 for descriptions of the samples. Data come from the 1979 to 1989 Merged Outgoing Rotation Groups files of the Current Population Survey, and are limited to observations that are no more than eight years away from the set-aside program date in the city. The estimates come from linear probability models, where the standard errors have been corrected for heteroskedasticity and cross-sectional and time-series clustering in the residuals at the MSA level. The education, age controls include linear terms for education and age, age-squared, and the interaction of education and age. The regressions are weighted by the CPS sample weights.

Table 4.6 Estimated changes in black relative employment rates
after implementation of a set-aside program, 13 cities and 28 cities

	Post-Program Effects deviated from Pre-Program Trend					
	13 Cities with verified dates			28 Cities with non-conflicting dates		
Regression coeff (x100)	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Pre-program trend diff	**-0.476***	**-0.492***	**-0.492****	**-0.401**	**-0.461**	**-0.431**
	(0.227)	**(0.253)**	**(0.207)**	**(0.360)**	**(0.340)**	**(0.345)**
Post-program b-w diff (relative to pre-program trend)						
Year 0	**-0.08**	**0.19**	**-0.22**	**0.18**	**0.15**	**-0.17**
	(1.41)	**(1.41)**	**(1.45)**	**(1.53)**	**(1.54)**	**(1.60)**
Year 1	**1.59**	**1.56**	**1.13**	**1.59**	**1.26**	**1.03**
	(1.58)	**(1.55)**	**(1.63)**	**(1.52)**	**(1.37)**	**(1.54)**
Year 2	**3.62**	**3.40**	**3.80**	**3.60**	**2.67**	**2.78**
	(2.46)	**(2.44)**	**(2.26)**	**(2.41)**	**(2.20)**	**(2.29)**
Year 3	**5.05***	**4.71**	**4.48**	**4.92**	**3.87**	**3.40**
	(2.74)	**(2.72)**	**(2.85)**	**(2.95)**	**(2.66)**	**(2.91)**
Year 4	**1.70**	**1.38**	**2.12**	**3.78**	**2.88**	**2.88**
	(3.31)	**(3.30)**	**(3.37)**	**(3.38)**	**(3.14)**	**(3.33)**
Year 5	**8.90****	**9.03****	**9.24*****	**6.79***	**6.18***	**6.18***
	(3.26)	**(3.31)**	**(3.20)**	**(3.45)**	**(3.31)**	**(3.48)**
Year 6	**7.44****	**7.53****	**8.30*****	**4.72**	**4.07**	**4.33**
	(3.24)	**(3.29)**	**(2.89)**	**(3.65)**	**(3.46)**	**(3.58)**
Year 7	**8.06***	**7.69**	**8.65***	**9.94***	**8.30**	**8.68**
	(4.34)	**(4.54)**	**(4.28)**	**(5.52)**	**(5.07)**	**(5.34)**
Year 8	**18.90*****	**18.43*****	**19.23*****	**14.76****	**12.57****	**12.74****
	(3.84)	**(4.14)**	**(3.12)**	**(6.27)**	**(5.96)**	**(6.23)**
Pre-program white trend	**0.407*****	**0.196**	**-0.108**	**0.367****	**0.250***	**-0.080**
	(0.139)	**(0.152)**	**(0.143)**	**(0.165)**	**(0.145)**	**(0.138)**
City fixed effects	**N**	**Y**	**Y**	**N**	**Y**	**Y**
Black, white year effects	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**
Education, age controls	**N**	**N**	**Y**	**N**	**N**	**Y**

Notes: ***, **, and * indicate statistical significance at the 1, 5, and 10 percent level, respectively. The dependent variable is equal to one if the person reports being employed and zero otherwise. See notes to above tables for sample construction and sizes. Data come from the 1979 to 1989 Merged Outgoing Rotation Groups files of the Current Population Survey, and are limited to observations that are no more than eight years away from the set-aside program date in the city. The estimates come from linear probability models, where the standard errors have been corrected for heteroskedasticity and cross-sectional and time-series clustering in the residuals at the MSA level. The education, age controls include linear terms for education and age, age-squared, and the interaction of education and age. The regressions are weighted by the CPS sample weights.

Table 4.7 Estimated changes in black relative employment rates after set-aside program implementation by industry group

	Difference-in-Differences Model								
	Most affected industries			Less affected industries			Least affected industries		
Regression coeff (x100)	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**
A. 13-city sample results									
Pre-program black diff	-6.80^{***}	-6.99^{***}	-3.39^{*}	-2.50	-3.42^{**}	-4.89^{***}	-5.31^{***}	-3.57^{***}	-4.62^{***}
	(2.11)	(1.93)	(1.78)	(1.60)	(1.35)	(1.28)	(1.25)	(0.90)	(0.85)
Post-program b-w gain	1.04	0.56	0.54	1.72	2.79^{*}	2.93^{**}	-1.47	-1.61	-1.53
	(0.96)	(0.77)	(1.04)	(1.44)	(1.49)	(1.36)	(1.19)	(1.53)	(1.39)
Post-program wht gain	0.46	-0.12	-0.17	-3.74^{**}	-0.44	-0.44	1.39	-0.75	-0.71
	(0.86)	(0.73)	(0.76)	(1.49)	(0.81)	(0.81)	(4.98)	(0.53)	(0.50)
B. 28-city sample results									
Pre-program black diff	-7.41^{***}	-7.49^{***}	-4.19^{***}	-2.81^{**}	-3.37^{***}	-4.46^{***}	-4.42^{***}	-3.14^{***}	-4.25^{***}
	(1.53)	(1.31)	(1.23)	(1.19)	(0.99)	(0.96)	(1.22)	(0.75)	(0.74)
Post-program b-w gain	0.75	0.04	-0.05	0.82	2.30^{*}	2.41^{**}	-1.47	-1.53	-1.42
	(1.12)	(0.66)	(0.87)	(1.57)	(1.26)	(1.19)	(1.19)	(1.16)	(1.03)
Post-program wht gain	-0.51	-0.03	-0.10	-3.34^{**}	-0.62	-0.60	2.87	0.12	0.19
	(0.78)	(0.48)	(0.49)	(1.34)	(0.53)	(0.52)	(3.42)	(0.59)	(0.58)
City fixed effects	N	Y	Y	N	Y	Y	N	Y	Y
Black, white year effects	Y	Y	Y	Y	Y	Y	Y	Y	Y
Education, age controls	N	N	Y	N	N	Y	N	N	Y

Notes: ***, **, and * indicate statistical significance at the 1, 5, and 10 percent level, respectively. The dependent variable is equal to one if the person reports being employed in the industry group and zero otherwise. See Table 2 for the listing of industries in each of the three industry groups. See notes to Tables 3 and 4 for descriptions of the samples. Data come from the 1979 to 1989 Merged Outgoing Rotation Groups files of the Current Population Survey, and are limited to observations that are no more than eight years away from the set-aside program date in the city. The estimates come from linear probability models, where the standard errors have been corrected for heteroskedasticity and cross-sectional and time-series clustering in the residuals at the MSA level. The education, age controls include linear terms for education and age, age-squared, and the interaction of education and age. The regressions are weighted by the CPS sample weights.

Appendix Table 4.1: Dates of City Set-Aside Programs from 4 different sources

MSA	City	MBELDEF(1988)	JCPES(1993) Ordinance	JCPES Enacted	City Record	Court Case
80	Akron	1984				
160	Albany	1984			1984	
360	Anaheim	1985			No program	
520	Atlanta	1982	1970s	1991	1975	1982
720	Baltimore	1982	1987	1988		1986
1000	Birmingham	1980				1977
1120	Boston	1987	1987	1987	1987	
1280	Buffalo					
1600	Chicago		1985	1985	1985	1985
1640	Cincinnatti	1983	1978	1978		
1680	Cleveland	1984	1982	1982	1994	
1840	Columbus	1983	1980	1981		1981
1920	Dallas		1984	1984		
2080	Denver		1983	1983	1983	1983
2160	Detroit		1983	1983		1983
2800	Ft. Worth		1986	1986	1988	
2960	Gary					
3120	Greensboro				1985	
3360	Houston		1981	1981	1984	
3480	Indianapolis		1984	1984		1987
3760	Kansas City		1981	1981		
4480	Los Angeles	1983	1983	1983	1987	1983
5000	Miami	1985			1985	
5080	Milwaukee	1987	1989	1989		
5120	Minneapolis		1980	1980		
5380	Nassau, NY				No program	
5560	New Orleans		1984	1984		
5600	New York		1991	1992	1992	1992
5640	Newark	1984				
5720	Norfolk				No program	
6040	Passaic, NJ				No program	
6160	Philadelphia	1984	1982	1983	1982	1982
6280	Pittsburg				1980	
6440	Portland				1977	
6840	Rochester				1980	
6920	Sacramento	1985				
7280	San Bernadino					
7320	San Diego		1986	1986	1985	
7360	San Francisco	1984	1984	1984	1984	1984
7400	San Jose	1983	1983	1983	1984	1983
7600	Seattle	1986	1980	1980	Pre-1984	
7040	St. Louis					
8280	Tampa	1985				
8840	Washington DC	1980	1975	1975		1977

Appendix Table 4.2: Cities and Dates of Set-Aside Programs verified in multiple sources

1. Albany (46) – 1984 (MBELDEF, City Record)

2. Boston (8) – 1987 (MBELDEF, JCPES, City Record)

3. Chicago (3) – 1985 (JCPES, City Record, Court Case)

4. Columbus (36) – 1981 (JCPES, Court Case)

5. Denver (28) – 1983 (JCPES, Court Case, City Record)

6. Los Angeles (2) – 1983 (MBELDEF, JCPES, Court Case)

7. Miami (26) – 1985 (MBELDEF, City Record)

8. New York (1) – 1992 (JCPES, City Record, Court Case)

9. Philadelphia (4) – 1982 (JCPES, City Record, Court Case)

10. San Diego (24) – 1985 (City Record, JCPES)

11. San Francisco (6) – 1984 (MBELDEF, JCPES, City Record, Court Case)

12. San Jose (31) – 1983 (MBELDEF, JCPES, Court Case)

13. Washington DC (7) – 1977 (JCPES, Court Case)

Appendix Table 4.3 Cities and Dates of Set-Aside Programs with only one source (but not divergent from other sources)

Most weight put on JCPES, Court Case, or City Record

14. Akron (49) – 1984 (MBELDEF)

15. Dallas (17) – 1984 (JCPES)

16. Detroit (5) – 1983 (JCPES)

17. Fort Worth (44) – 1986 (JCPES)

18. Greensboro (57) – 1985 (City Record)

19. Kansas City (27) – 1981 (JCPES)

20. Minneapolis (16) – 1980 (JCPES)

21. New Orleans (32) – 1984 (JCPES)

22. Newark (15) – 1984 (MBELDEF)

23. Pittsburgh (10) – 1980 (City Record)

24. Portland (34) – 1977 (City Record)

25. Rochester (38) – 1980 (City Record)

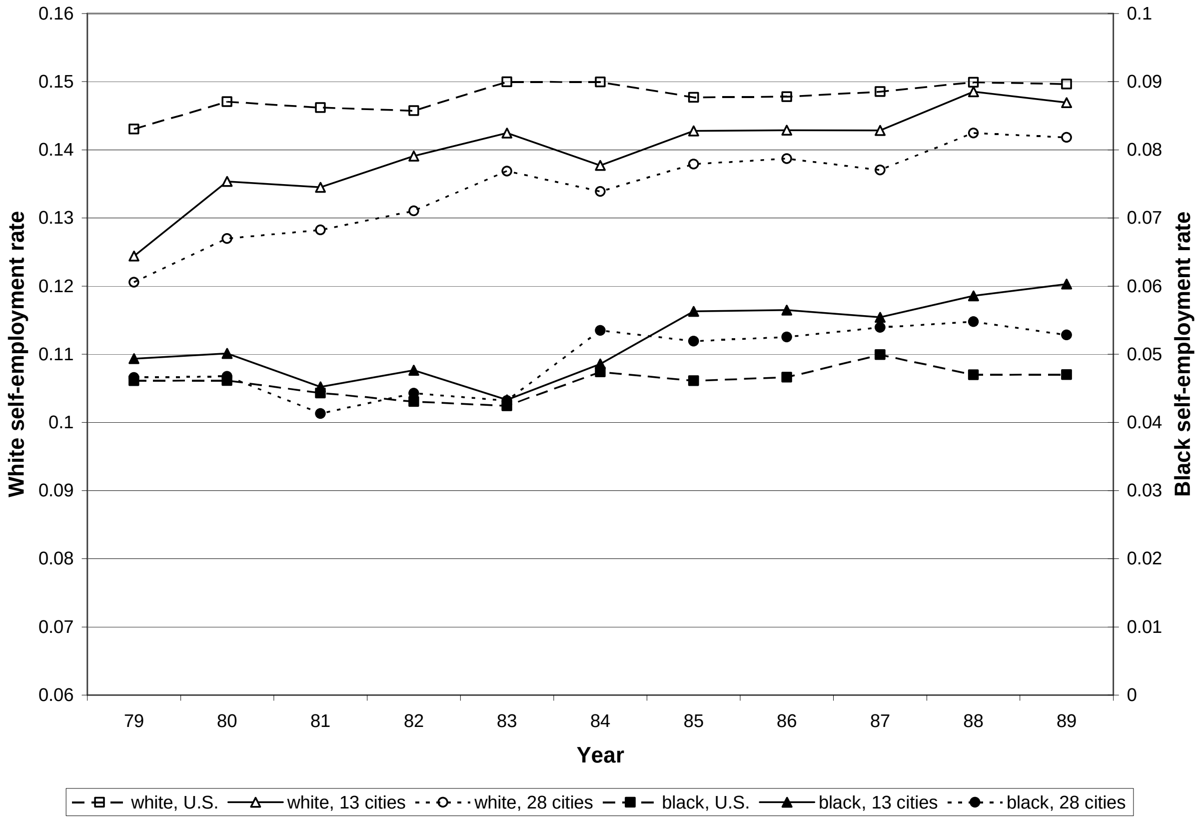
26. Sacramento (42) – 1985 (MBELDEF)

27. Seattle (18) – 1980 (JCPES)

28. Tampa (33) – 1985 (MBELDEF)

Figure 4.1 Black and White Self-Employment Rates from 1979 to 1989

A. Entire United States and 13 (28) cities with program dates that do not conflict across sources



B. Cities with and without a set-asides program for the entire 1979 to 1989 period

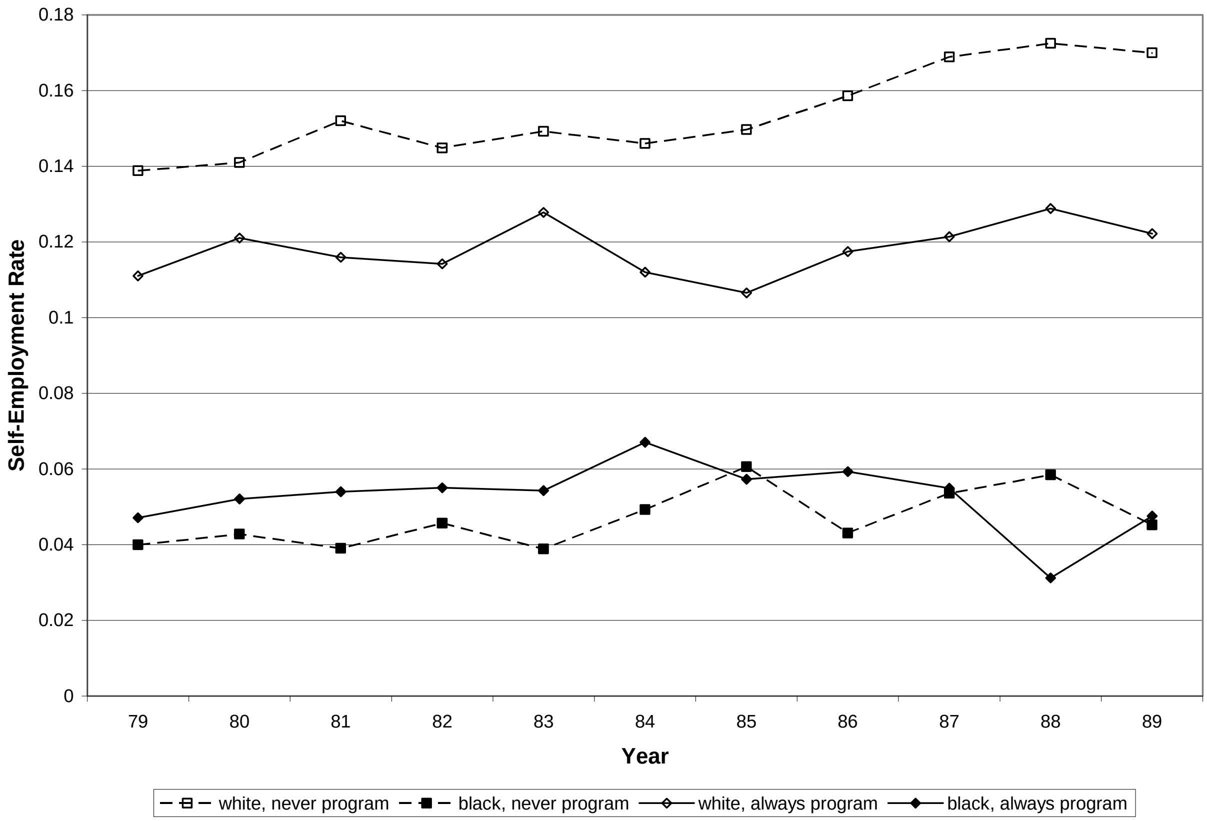


Figure 4.2 Black and White Employment rates for the three samples

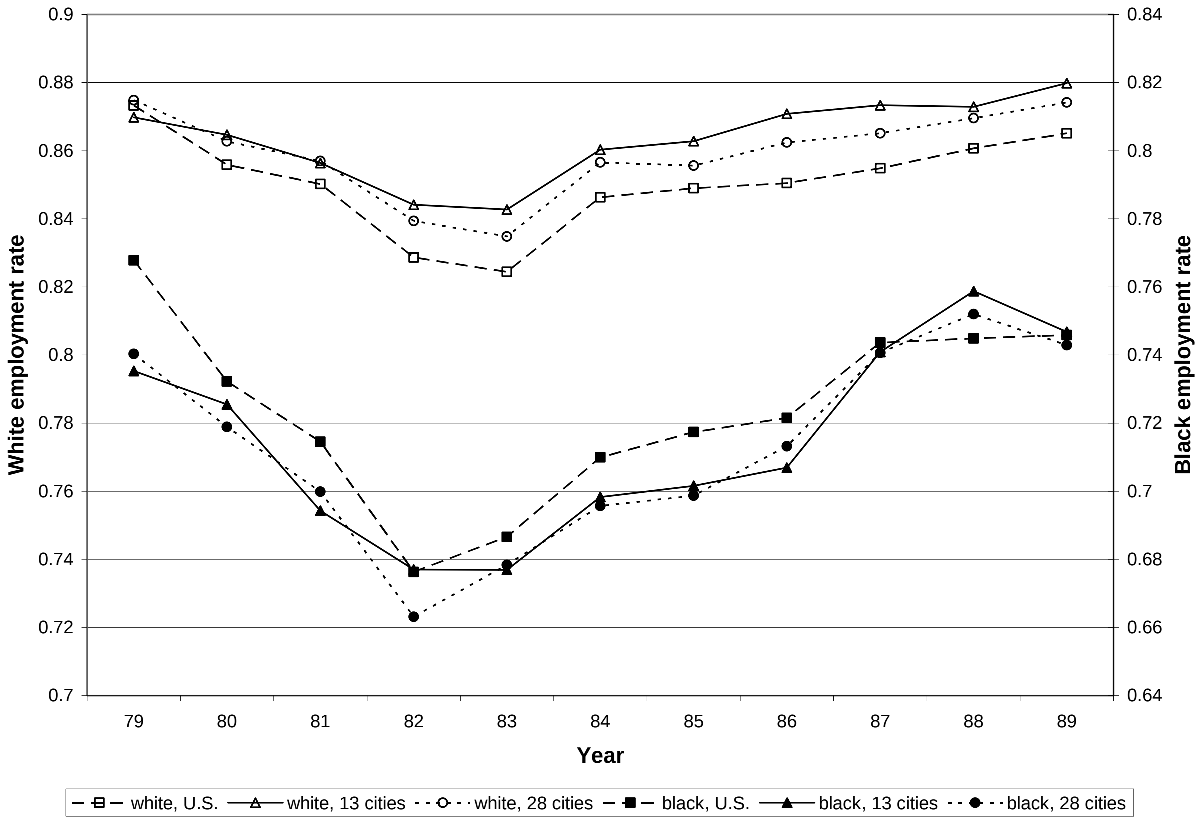
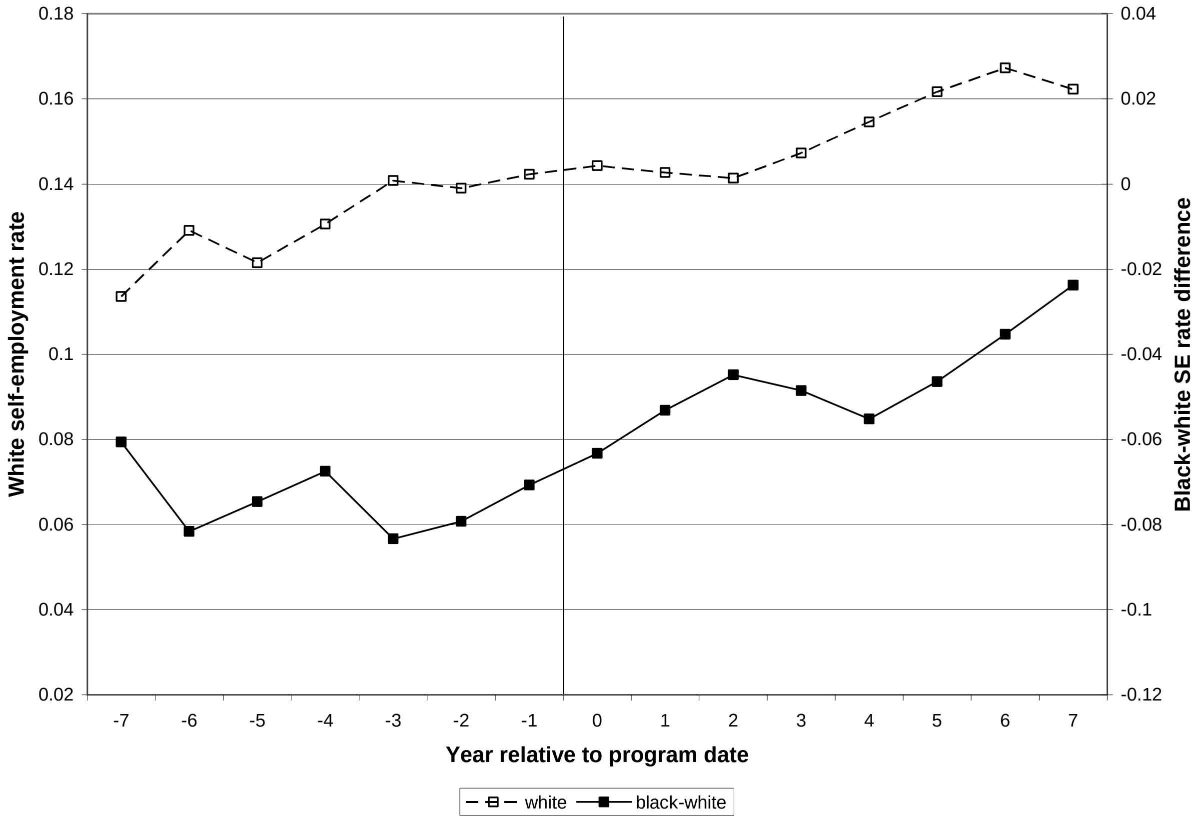


Figure 4.3 Event Study Figures for Self-Employment Rate, 13 cities with verified program dates

A. Controls for MSA fixed effects and race-specific year effects



B. Controls for MSA effects, race-specific year effects, education, and age

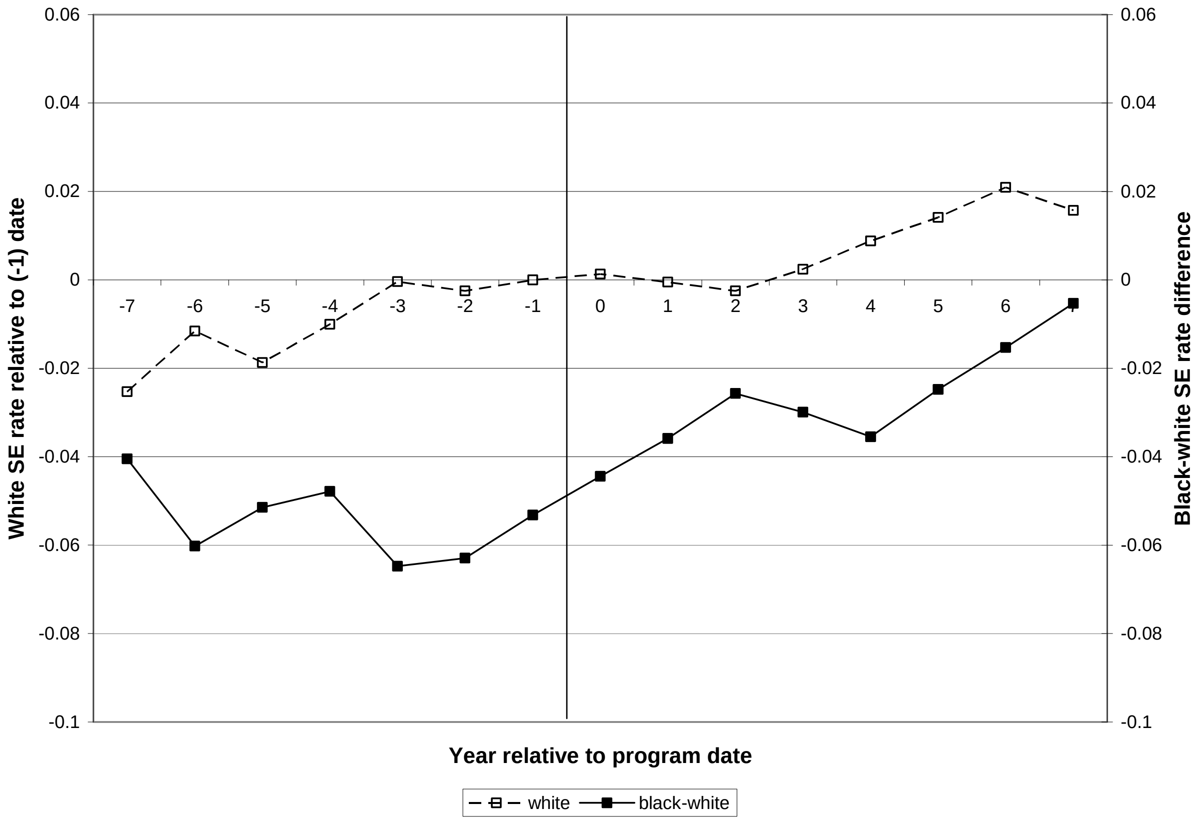
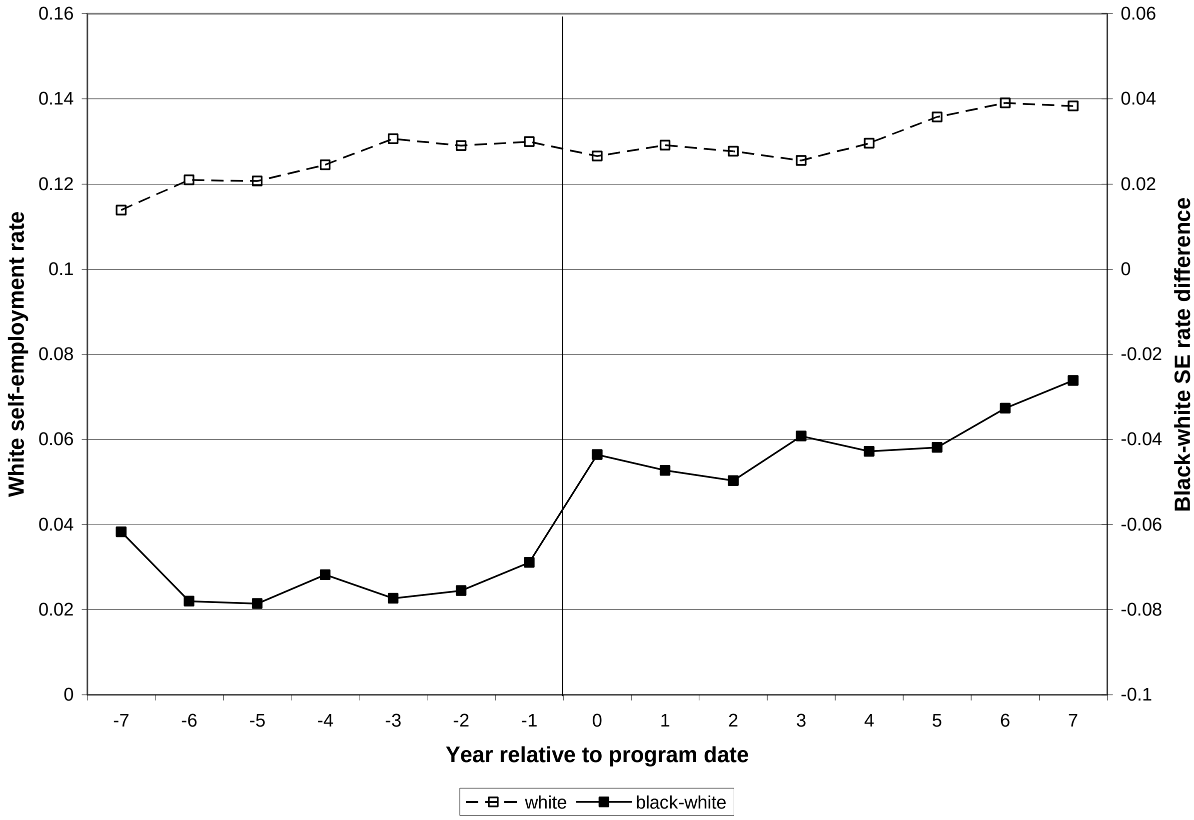


Figure 4.4 Event Study Figures for SE Rate, 28 cities with program dates that do not conflict

A. Controls for MSA fixed effects and race-specific year effects



B. Controls for MSA effects, race-specific year effects, education, and age

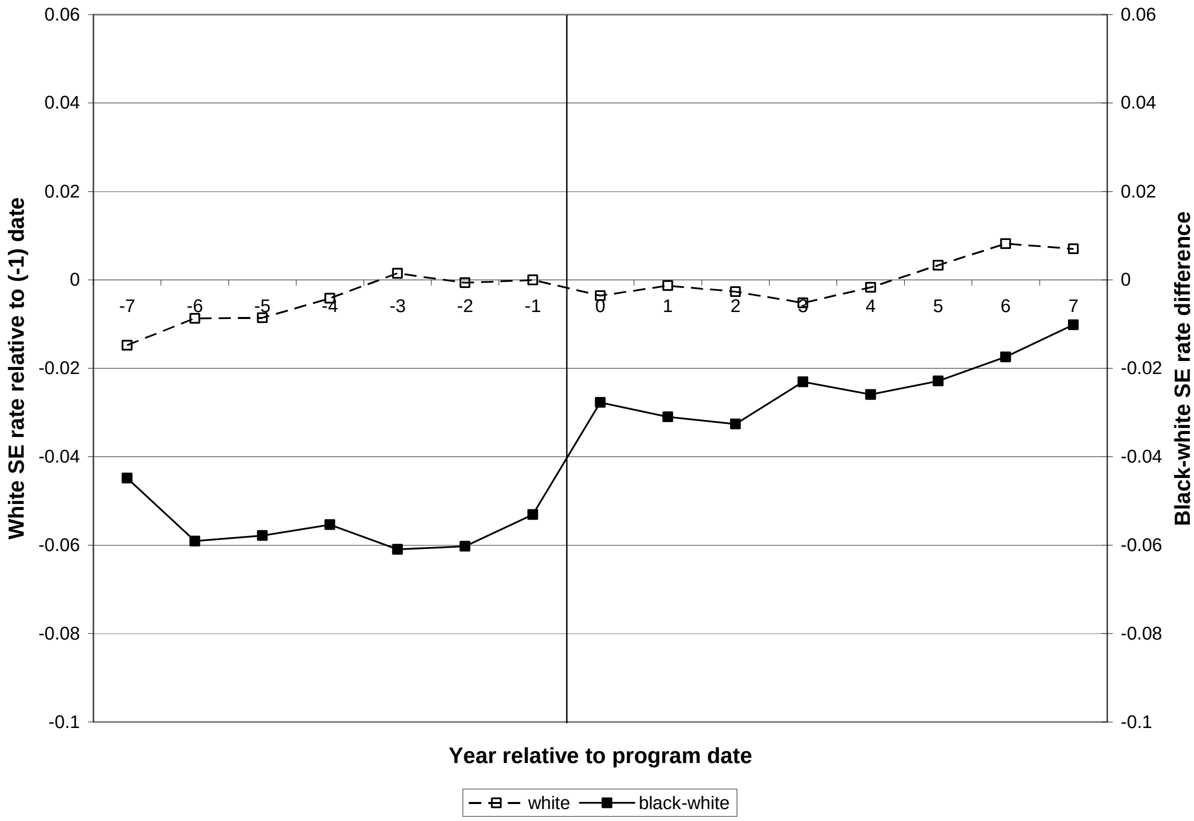
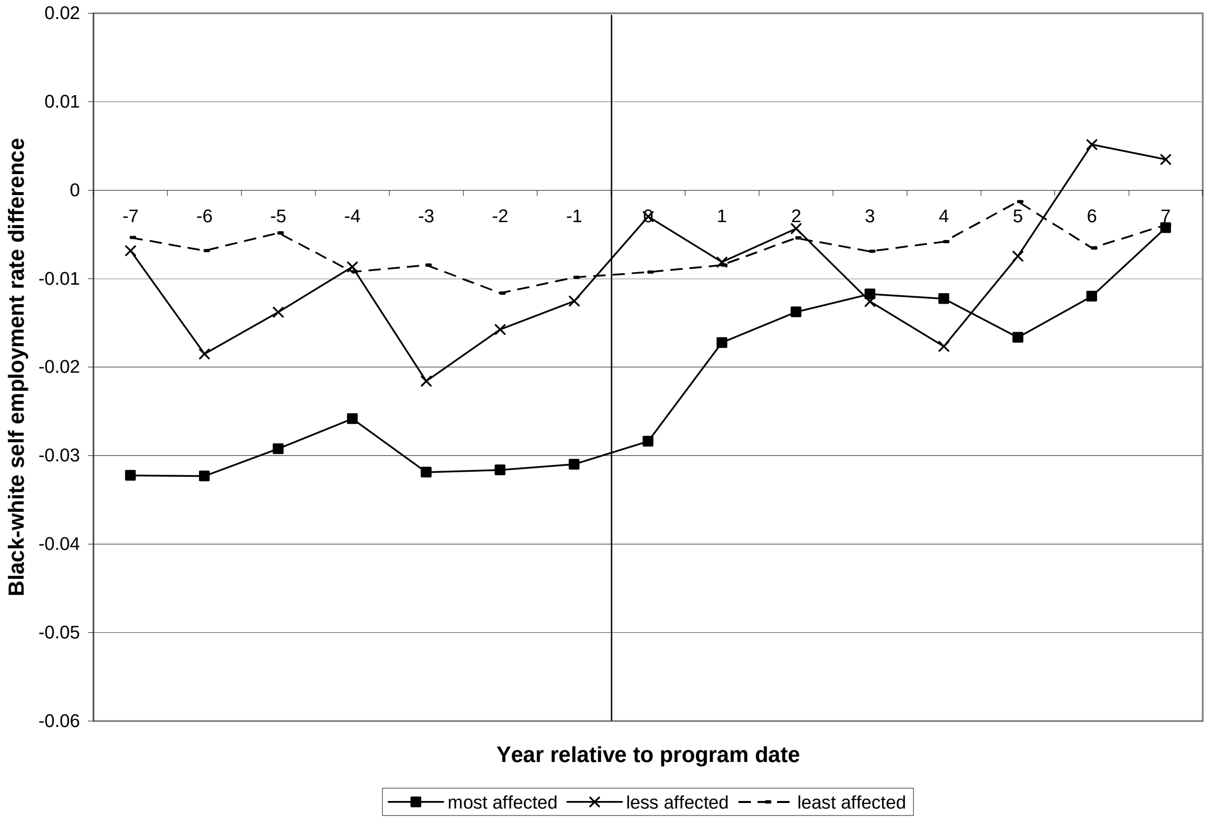


Figure 4.5 Black-white Self-employment rate differences by industry group

A. 13 cities, controls for year and city effects and control for age, education



B. 28 cities, controls for year and city effects and control for age, education

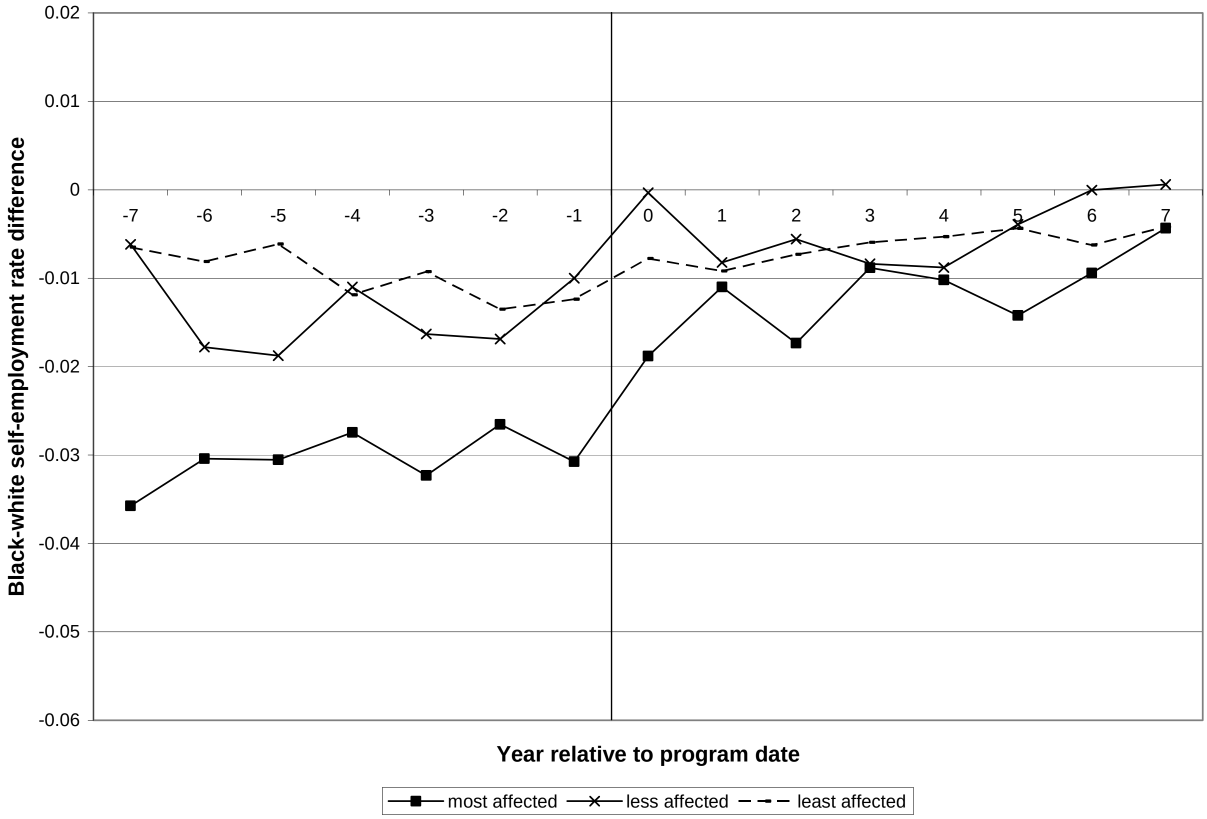
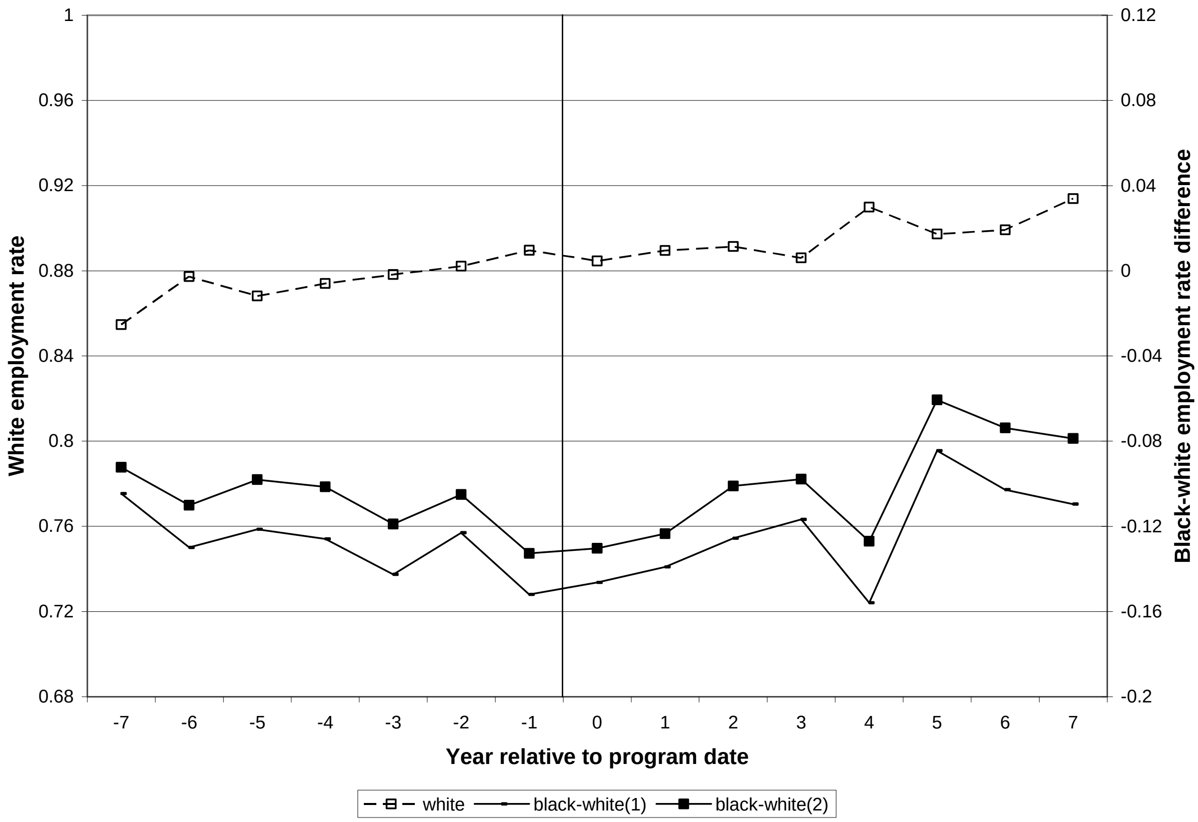
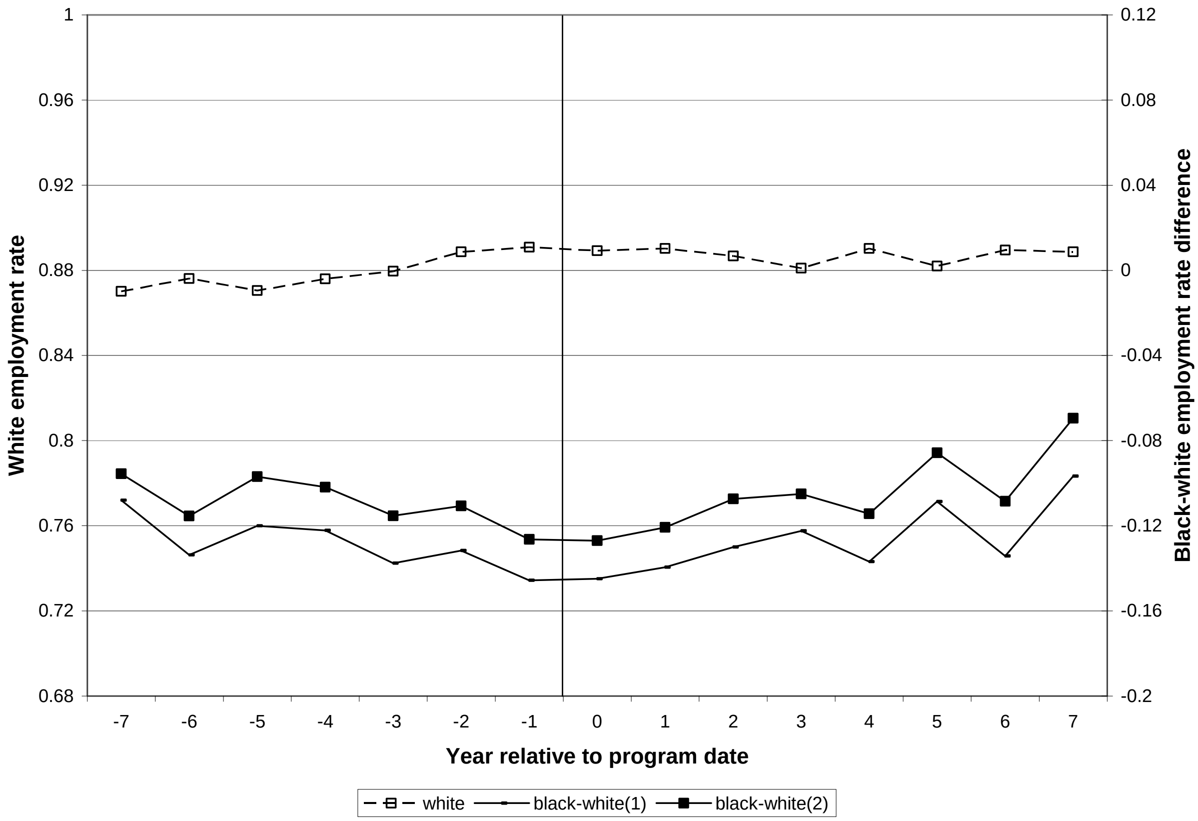


Figure 4.6 Employment rates of black and white men with respect to program date

A. 13 cities



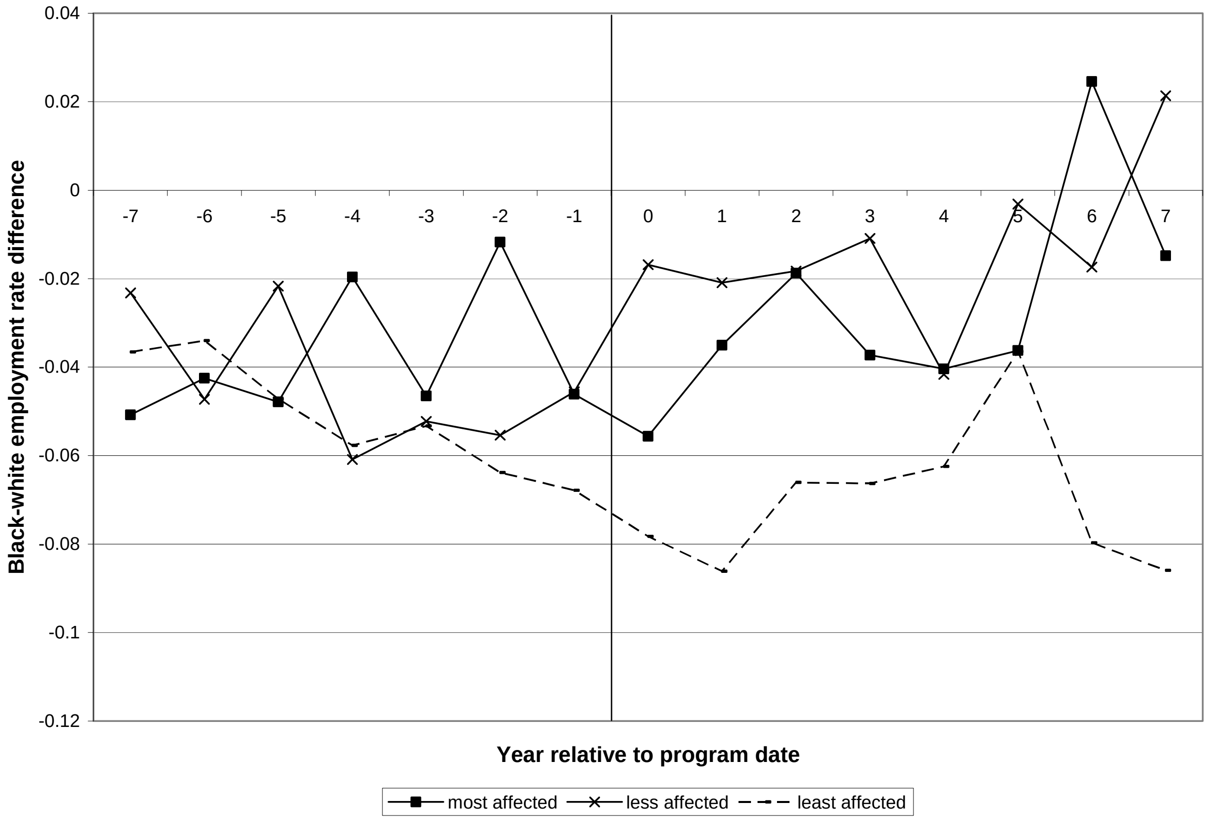
B. 28 cities



White employment rate
Black-white employment rate difference
1
0.96
0.92
0.88
0.84
0.8
0.76
0.72
0.68
0.12
0.08
0.04
0
-0.04
-0.08
-0.12
-0.16
-0.2
-7
-6
-5
-4
-3
-2
-1
0
1
2
3
4
5
6
7
Year relative to program date
white
black-white(1)
black-white(2)

Figure 4.7 Black-white Employment rate differences by industry group

A. 13 cities, controls for year and city effects and control for age, education



B. 28 cities, controls for year and city effects and control for age, education

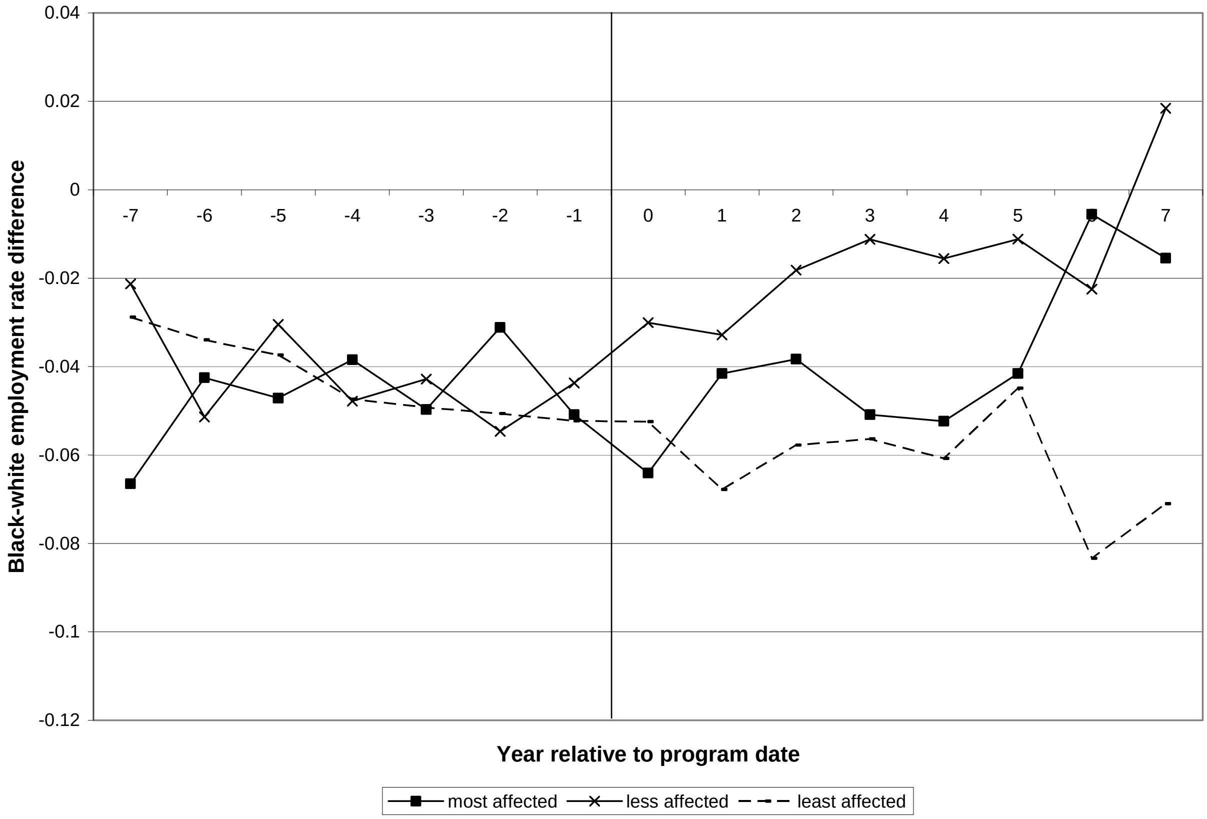
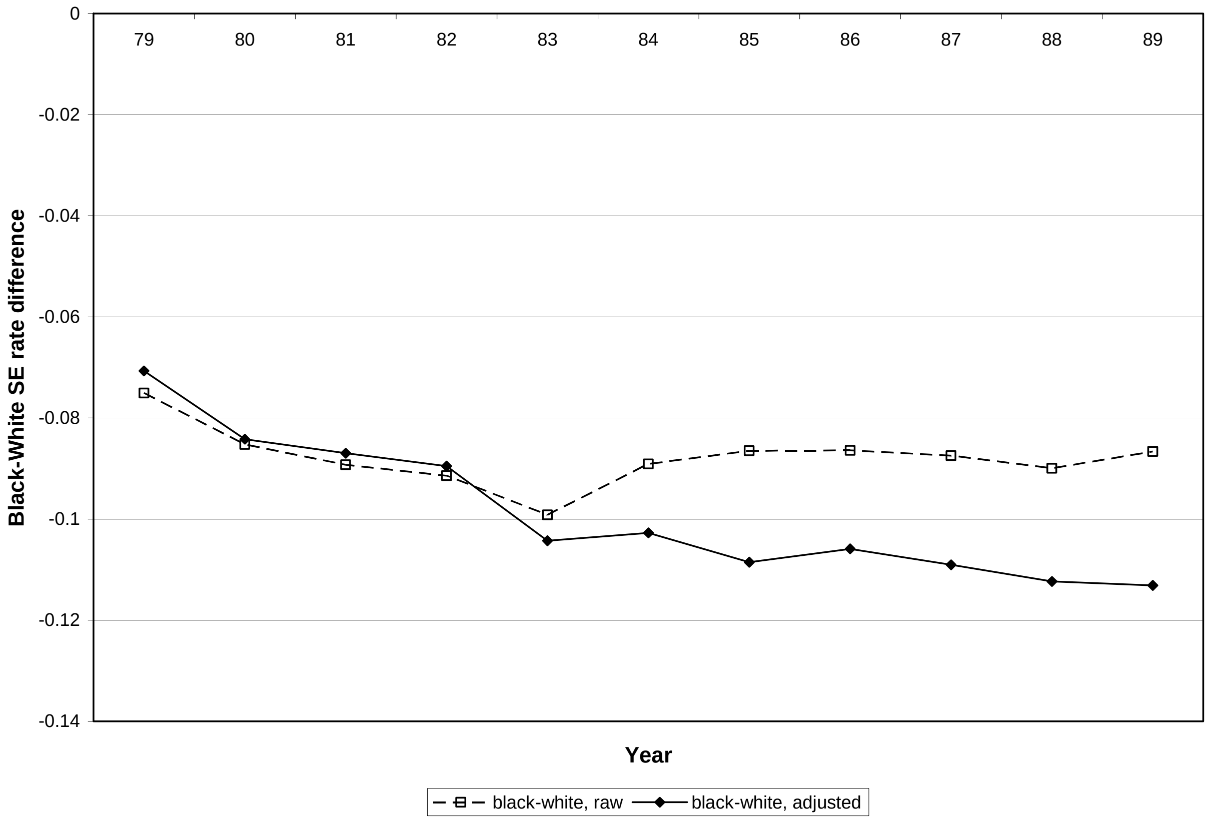
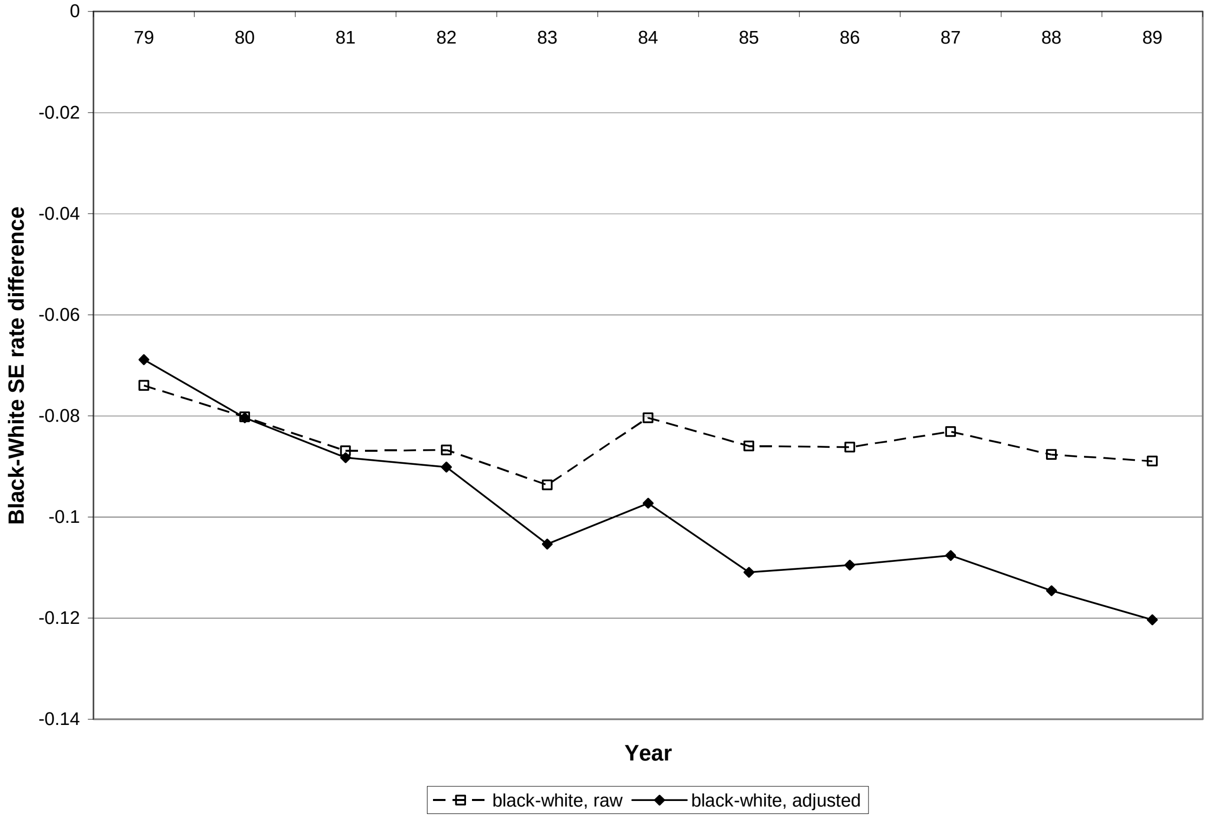


Figure 4.8 Black-White difference in self-employment year effects netting out effects of program dates

A. 13 cities



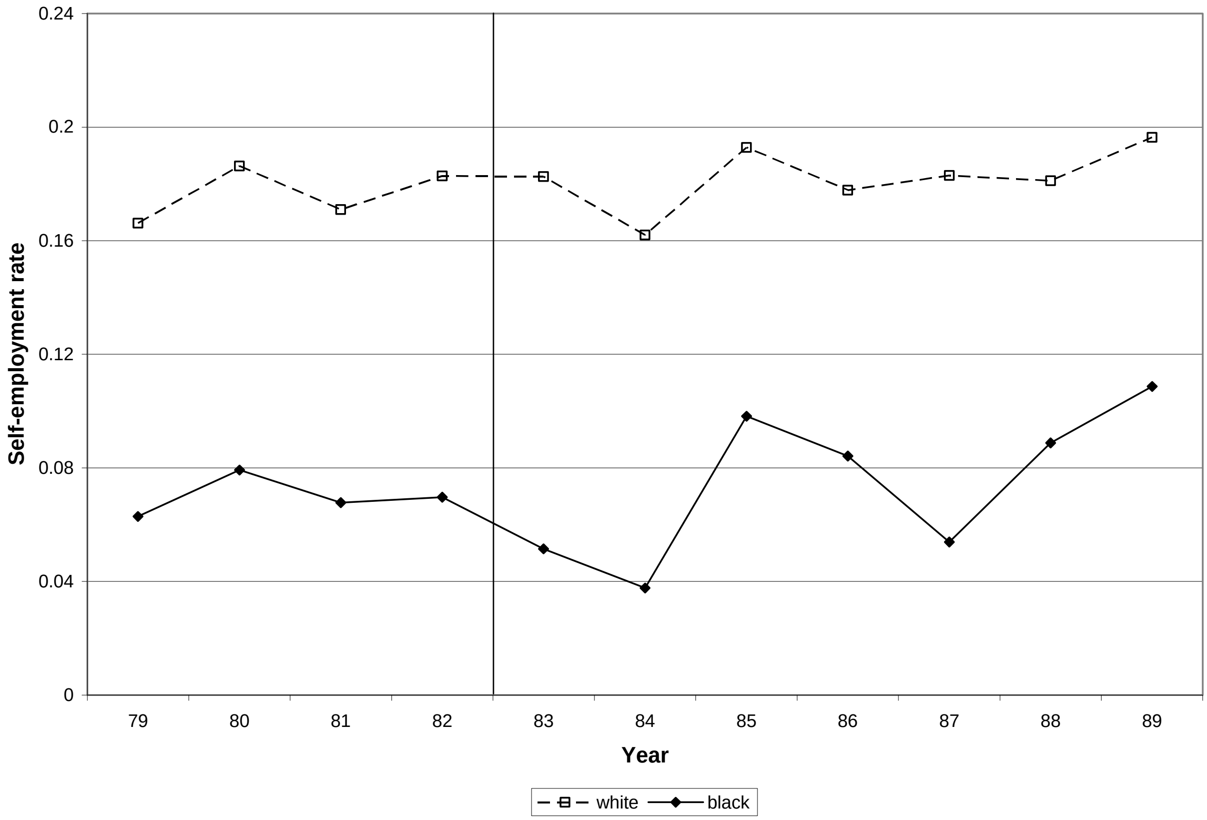
B. 28 cities



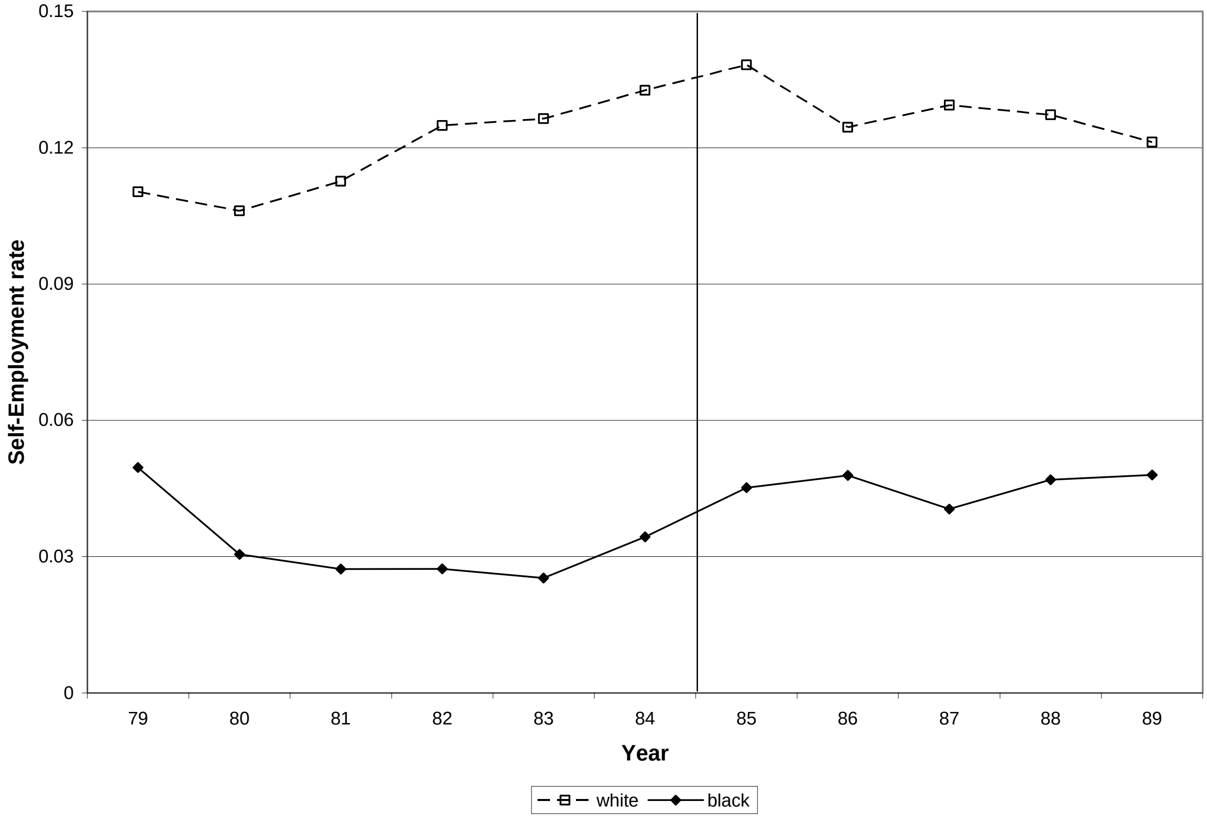
0
-0.02
-0.04
-0.06
-0.08
-0.1
-0.12
-0.14
79
80
81
82
83
84
85
86
87
88
89
Black-White SE rate difference
Year
black-white, raw
black-white, adjusted

Appendix Figure 4.1 Self-employment rates and program dates for cities with verified program dates

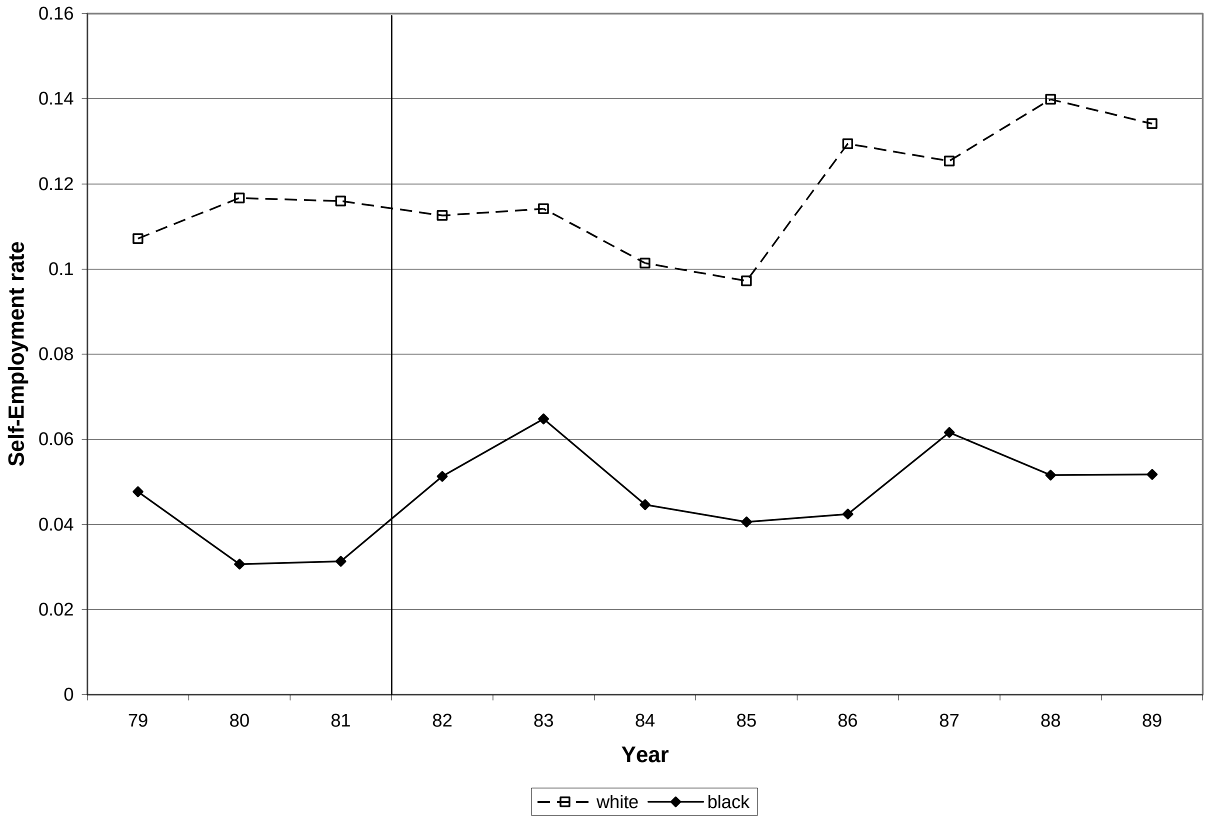
A. Los Angeles – 1983 program date



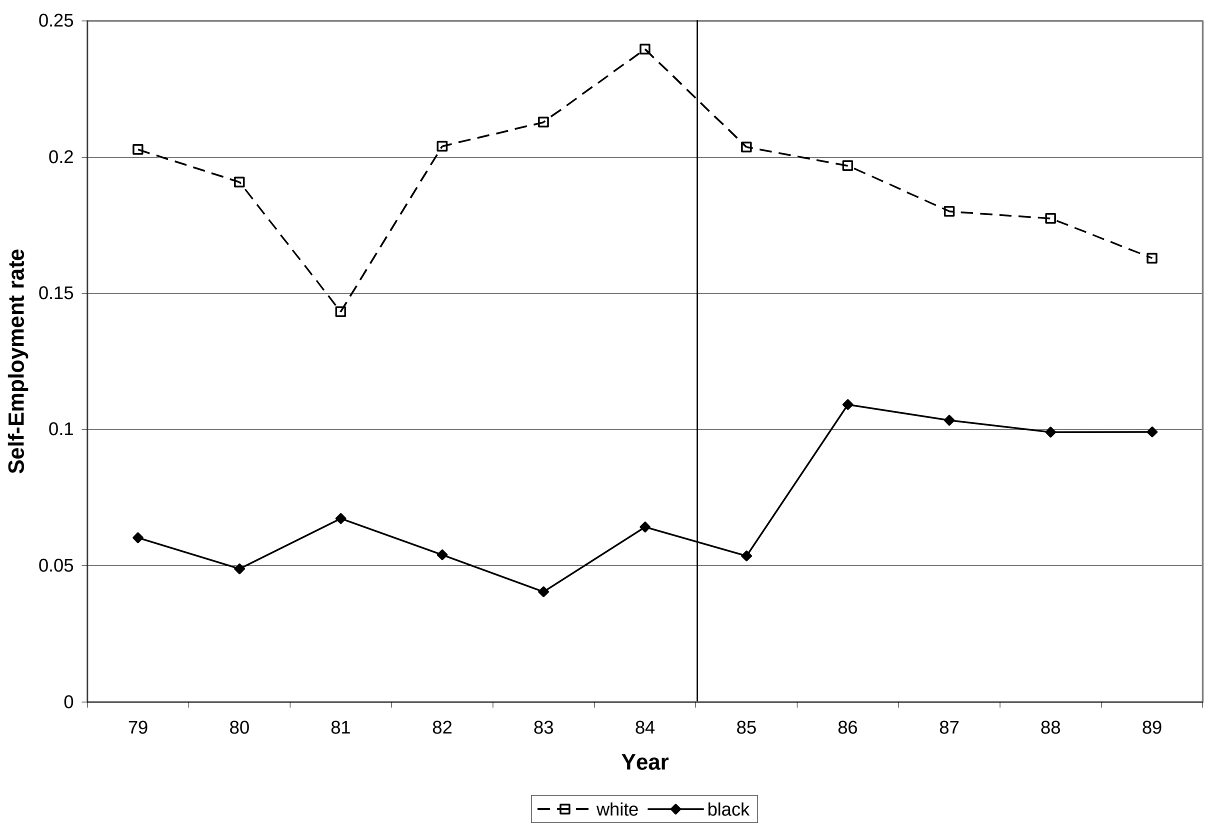
B. Chicago – 1985 program date



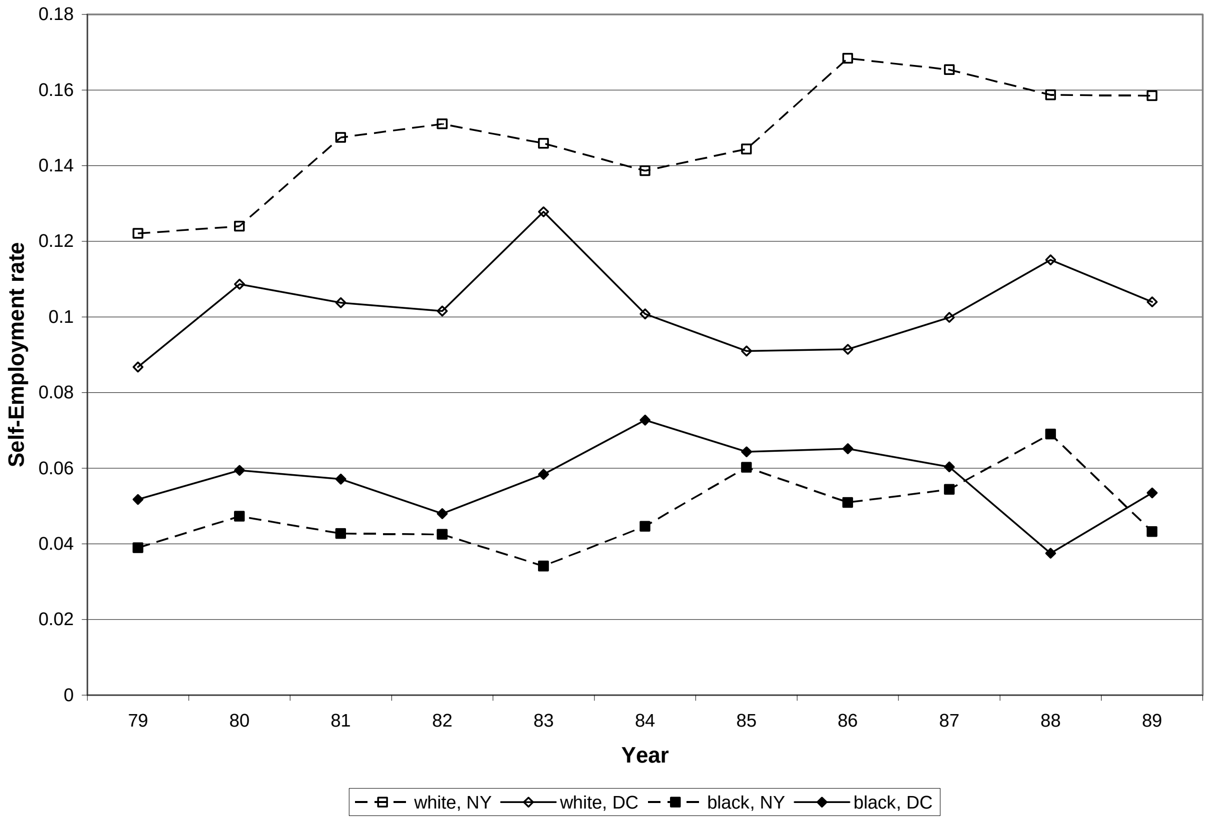
C. Philadelphia – 1982 program date



D. Miami – 1985 program date

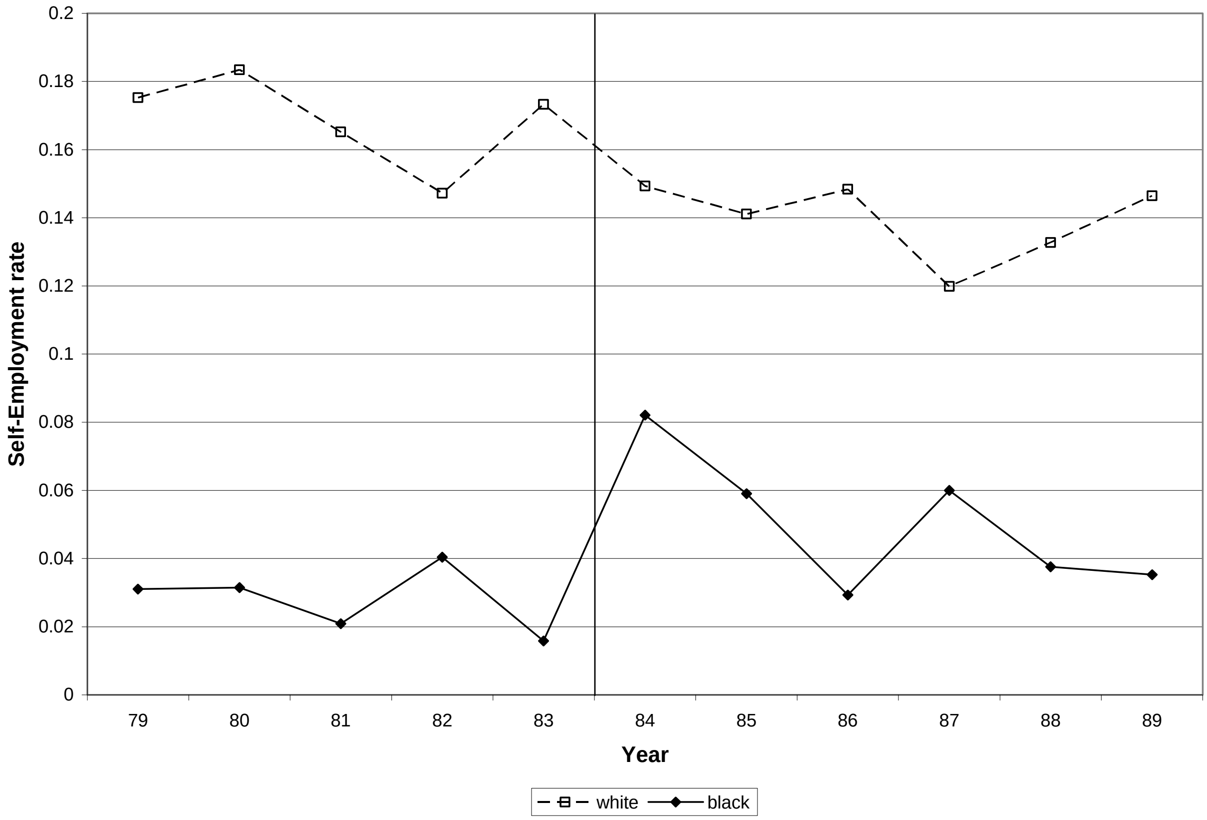


E. New York (1992) and Washington DC (1977)

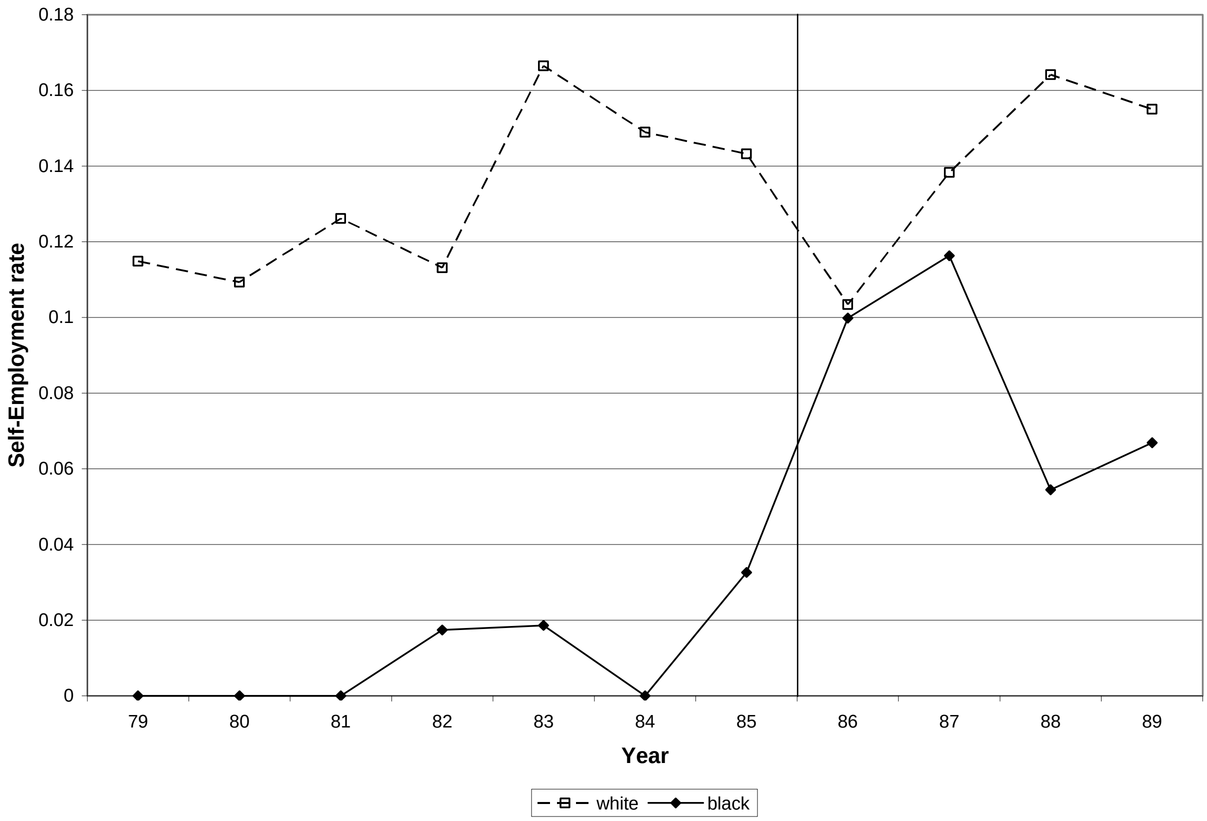


Appendix Figure 4.2 Self-employment rates and program dates for cities with non-conflicting program dates

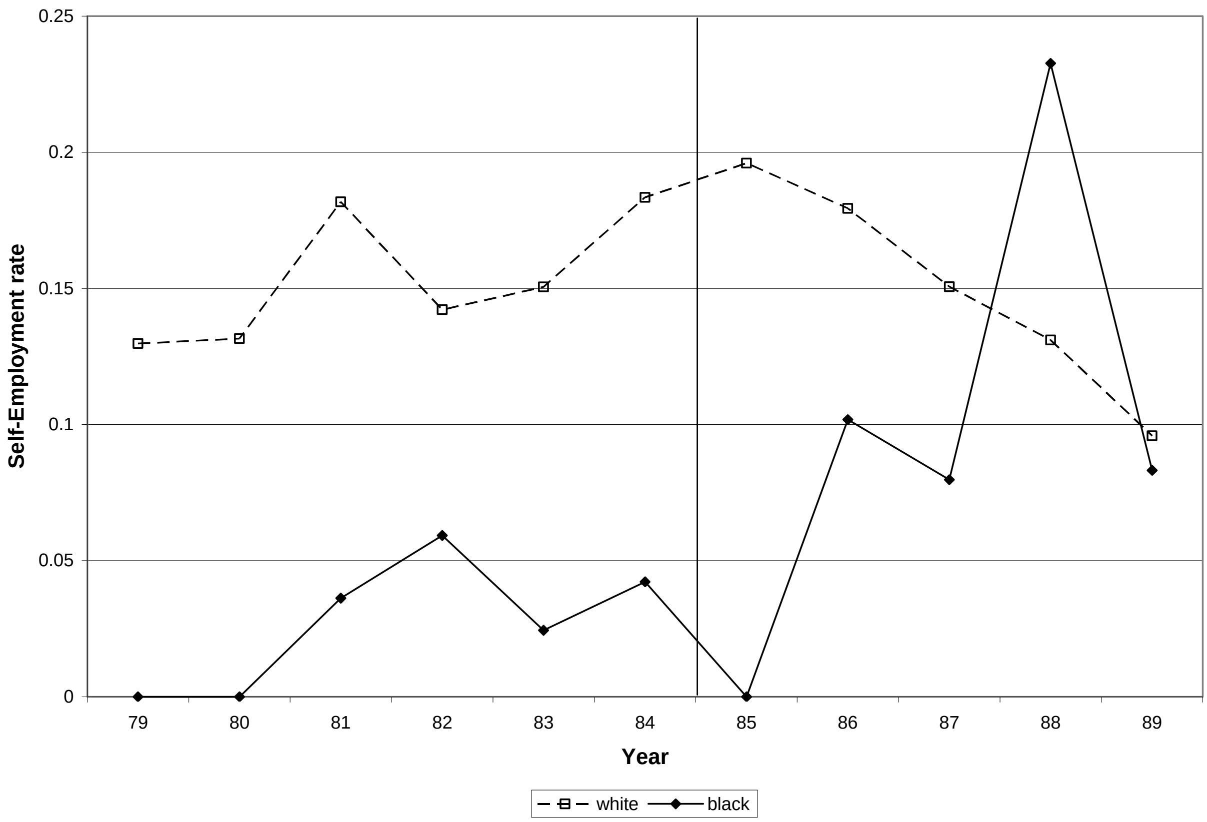
A. Dallas (1984)



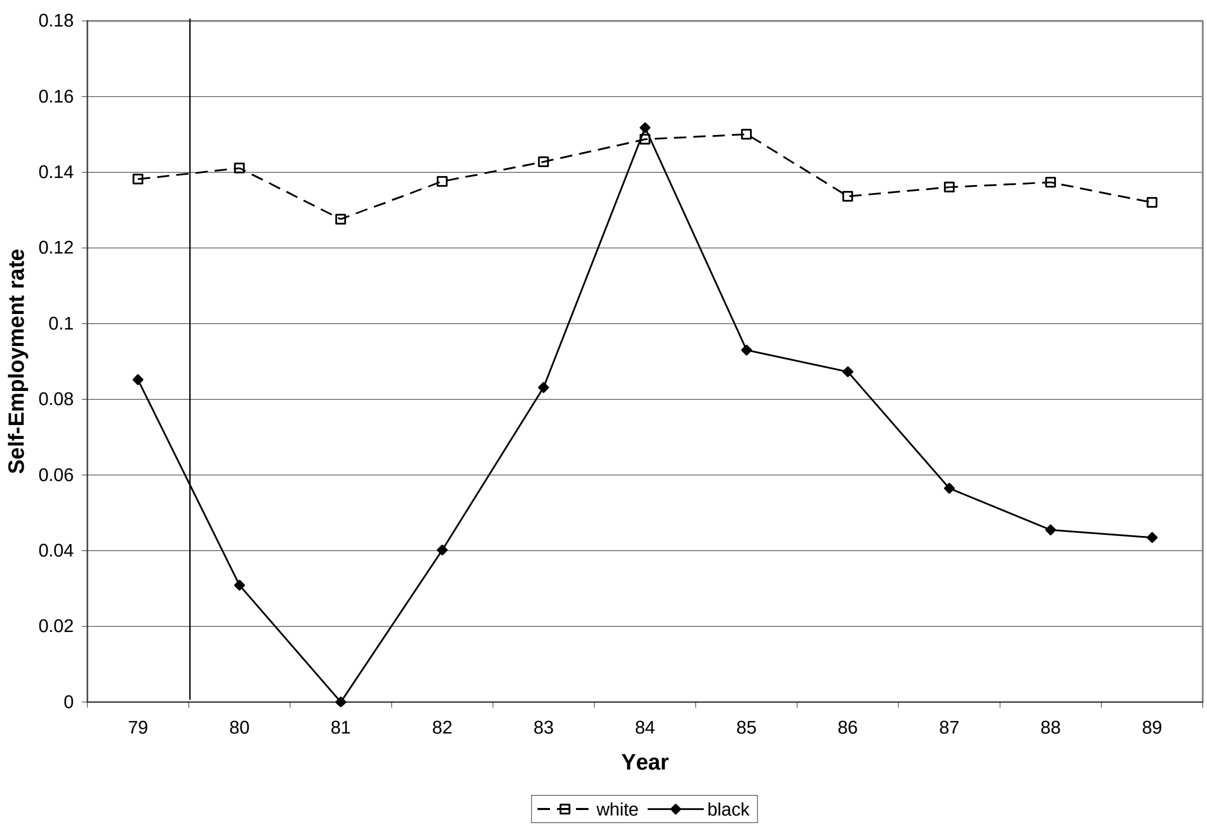
B. Fort Worth (1986)



C. Sacramento (1985)



D. Seattle (1980)



Chapter 5

5. Conclusion

Taken together, each of the preceding chapters has advanced our understanding of entrepreneurship and corporate social responsibility. In Chapter 2, I found that entrepreneurs with prior experience at incumbent firms perform better than other entrants and inherit valuable non-technical knowledge. Clearly, prior knowledge impacts entrepreneurial performance and future research should further analyze the inheritance of non-technical knowledge from parent to spawn. In Chapter 3, I found that raters in the nascent socially responsible investing sector produce ratings that have low convergent validity and low predictive validity. This result calls for further research on the underlying construct of social responsibility and more rigorous testing of existing ratings. In Chapter 4, I find that controversial set aside programs in city contracting had a large impact on minority self-employment and minority employment in the targeted cities. Further research is necessary to identify other programs that may have had an impact during the same period and how well these new firms performed over time.