INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

ProQuest Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

EMPIRICAL ESSAYS ON NETWORK EFFECTS IN MARKETS

by

Supriya Sarnikar

A Dissertation submitted to the Faculty of the

DEPARTMENT OF ECONOMICS

In Partial Fulfillment of the Requirements
For the Degree of

DOCTOR OF PHILOSOPHY

In the Graduate College
THE UNIVERSITY OF ARIZONA

2 0 0 2

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

THE UNIVERSITY OF ARIZONA ®
GRADUATE COLLEGE

As members of the Final Examination Committee, we certify that we have read the dissertation prepared by Supriya Sarnikar

entitled Empirical Essays on Network Effects in Markets

and recommend that it be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy

Price V. Fishback	7/31/02. Date
Stanley S. Reynolds	7/31/02 Date
Ronald L. Oaxaca	7/31/02 Date
Shawn E. Kantor	7/31/02 Date
	Date

Final approval and acceptance of this dissertation is contingent upon the candidate's submission of the final copy of the dissertation to the Graduate College.

I hereby certify that I have read this dissertation prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.

Dissertation Director 7/31/02 Date
Stanley S. Reynolds / Price V. Fishback

STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at The University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgment of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the head of the major department or the Dean of the Graduate College when in his or her judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

SIGNED: Supriya

ACKNOWLEDGEMENTS

I am grateful to my committee, Prof. Stan Reynolds, Prof. Price Fishback, Prof. Ronald Oaxaca, Prof. John Wooders, Prof. Dan Houser, Prof. Shawn Kantor and Prof. William Horrace for their support and guidance. I owe special thanks to Stan Reynolds for helping me improve my thoughts and ideas on network effects in computer software markets and helping me express them in a mathematical form. I am thankful to Ronald Oaxaca for his invaluable guidance and help in improving the empirical analysis presented in this dissertation and for always finding the time to help me despite his busy schedule. I am deeply indebted to Price Fishback for not only his guidance in improving this dissertation but also for providing constant encouragement and support throughout my career as a graduate student and for believing in me even when I did not believe in myself.

My understanding of computer software markets has benefited greatly from conversations with Surendra Sarnikar and Ram Pudupet. I am thankful for their insight and support. I thank all my colleagues particularly, Zeynep, Francesco and Bill, for sharing their ideas and helping me improve my own.

Finally, I am thankful for the loving support of my family.

In the memory of my grandfather,

Shri. Dattatreya Bhanudas Sarnikar

who taught me by example

the twin virtues of patience and perseverance.

TABLE OF CONTENTS

TABLE OF CONTENTS - CONTINUED

LIST OF TABLES

LIST OF TABLES - CONTINUED

ABSTRACT

This dissertation examines the impact of network effects in two settings - the computer software markets and self-employment decisions by individuals. Although there have been strong developments on the theory of network effects, relatively little empirical work has been done to examine their importance.

The first part of this dissertation focuses on network effects in the market for computer software. It has been hypothesized that the presence of network effects in this market might often lead to lock-in of an inferior technology. An indirect test of this hypothesis is devised by taking advantage of a natural experiment afforded by the introduction of the programming language, Java. Java made it possible for programmers to write a single program that would run on any operating system. It therefore had the potential to eliminate the indirect network externalities in the operating systems market. Hedonic price regressions with fixed time and firm effects are estimated to test for the effect of Java on the extent of competition in the software market. Results using data compiled from magazine

reviews of graphics applications programs indicate that Java was successful in creating more competition in the market for software applications.

The second part of this dissertation examines whether social networks might explain the persistent racial gap in Self-Employment (SE) rates in the United States. Self-employment rates in the United States fell dramatically for most of the twentieth century before starting to increase in the 1970's. The racial gap in self-employment rates however, remained constant throughout this period. Many theories have been proposed in the literature but none of them successfully explains the persistence of the gap. A multinomial logit specification is used to model individual decisions to become self-employed. The average SE rate in the neighborhood is used as a measure of the network effect. Results indicate that social networks played an important role in promoting self-employment among blacks since 1950. Given the initial conditions of lower SE rates among blacks, the role of social networks in promoting SE might be able to explain the persistence of the racial gap in SE rates.

I. INTRODUCTION

Network effects are said to exist whenever an individual's utility from a given choice depends on the number of other individuals that have made the same choice. The individuals that make the same choice may be thought of as forming a 'network'. Since the network influences an individual's choice, this phenomenon is called a network effect. This thesis examines the impact of network effects in two settings - computer software markets and labor markets. Networks are generally believed to be important determinants of individual choices in the computer software markets and also in the labor markets. Although the reasons for their importance might be different in the two markets, it is generally believed that network effects lead to conformity of individual behavior.

Economists have devoted much attention in the last twenty years to this phenomenon. The idea that 'networks' or 'social interactions' influence individual choices has been used to explain such diverse phenomena as the adoption of technology to residential segregation and the persistence of ghettoes. Also, the impact of network effects on social

welfare has been studied quite deeply in a theoretical setting. Despite all the attention that the phenomenon of network effects has drawn from the academic community, very few studies are of an empirical nature[1]. In the absence of empirical evidence to test and support the implications of the theoretical models, our understanding of the phenomenon remains limited. This dissertation contributes towards this gap in the literature by making use of a unique data set compiled by its author and by empirically examining, for the first time, whether economic theories of network effects might help to explain the persistent racial gap in self-employment rates in the United States.

Many reasons have been proposed as to why individual choices might be interlinked with their social interactions. When individual choices are interlinked in such a manner, it has been shown to induce conformity of behavior across individuals. The conformity of behavior may be due to either direct/strategic dependence of the individual utilities/payoffs or due to informational asymmetries.

---

[1] Some empirical studies in the industrial organization literature are Saloner and Shepard (1995), Gandal (1995)

When the choices under consideration pertain to product markets, such as the choice between several different products or technologies, then individual choices are thought to be interlinked due to a direct/strategic dependence of their payoffs resulting in a positive externality. Such interdependencies have been termed "network externalities" in the industrial organization literature and the concept has been used to explain the process of technological diffusion and to study the impact of the presence of such externalities on social welfare[2].

When the choices under consideration pertain to urban living choices in general, the interdependencies have been described as "neighborhood externalities" which might be either positive or negative. This concept has been used in the urban economics/macroeconomics literature to explain such phenomena as residential segregation, patterns of crime rates in cities, geographic concentration of urban unemployment rates and income distribution patterns.

Economides and Himmelberg(1995) etc.

[2] Farrell and Saloner (1986, 1992), Katz and Shapiro (1985,1986), Church and Gandal (1992, 1993), Matutes and Regibeau (1992) are some studies that examine the impact on social welfare in the presence of network effects in the adoption of technology.

Another strand of mostly theoretical literature has sought to understand the reasons for conformity of individual behavior when there are no obvious reasons for interdependence of individual payoffs. In the absence of any such externalities, it has been shown that individual choices might be interlinked due to informational asymmetries that lead individuals to base their decisions at least in part on their neighbor's experiences. This idea of 'social learning' occurring due to informational asymmetries is used to test whether social networks might have played a role in the persistence of the racial gap in self-employment rates in the United States.

1.1. Economic Theories of Conformity

Much of the earlier literature on urban macroeconomic issues as well as the more recent literature in industrial organization have focused on understanding the impact of externalities imposed by social interactions on individual choices. Bernheim (1994) analyzes a model of social interaction in which individuals care about their social status as well as the "intrinsic utility" of their choice. Status is assumed to depend on public perceptions of an individual's predisposition rather than their actions.

Since predispositions are unobservable, actions signal predispositions and therefore affect status. When popularity is sufficiently important relative to intrinsic utility, many individuals conform to a homogenous standard even with heterogeneous underlying preferences. Social interactions may also explain cross-group variations in behavior if the various groups conform to different types of behavior.

Miyao (1978) shows that in the presence of negative inter-group externalities and/or positive intra-group externalities, the mixed city equilibrium is not necessarily unstable; i.e., there is no tendency for mixed cities to become homogeneous with respect to the characteristics of its inhabitants. Spatial segregation of inhabitants that differ along ethnic/racial/economic or other criteria within a city is shown to be a stable equilibrium.

Benabou (1993) models the links between residential choice, educational investment and production in a city composed of several communities. Identical agents choose whether to become high-skill workers, low-skill workers or remain

outside the labor force while simultaneously choosing which part of the city to live. Education is a local public good and therefore in each community, the more agents invest in acquiring high-skills, the easier it is to do so. These local complementarities in human capital investment result in occupational segregation even though the social optimum is to have identical communities. Individual incentives to pursue perfect segregation, while feasible, are self-defeating and can lead to a collapse of the productive sector. Benabou (1996) uses a general model of community formation and human capital accumulation with social spillovers to demonstrate how even small differences in education technologies, preferences or wealth lead to a high degree of economic segregation.

Each of the applications discussed above assumes a spillover effect or externality effect from social interactions. A different approach to explaining conformity of behavior is found in Bikhchandani, Herschleifer and Welch (1992). They show that localized conformity of behavior and fragility of mass behavior can be explained by informational cascades. An informational cascade occurs when it is optimal for an individual having observed the

actions of those ahead of him, to follow the behavior of the preceding individual without regard to his own information.

In Ellison and Fudenberg (1993, 1995), economic agents make decisions without knowing the costs and benefits of the possible choices and therefore rely on word-of-mouth communication. Two models of technology choice are considered. In both models, there is no direct or strategic dependency of choices. The only reason why choices might be linked is the way information is transmitted. First, the two competing technologies are of unequal quality. The structure of communication is important in determining whether the superior product is chosen overall. Second, both products are equally good. Word-of-mouth communication might allow efficient social learning that leads to adoption of technology that is superior on average. Thus conformity of behavior is shown to result simply from information asymmetry and 'social learning' occurring through social interactions. No explicit or direct dependence of individual payoffs on social choices is assumed in these models.

More generalized models of individual behavior enhanced by social interactions are found in Brock and Durlauf (2000), Blume and Durlauf (2000) and also Banerjee (1992). Banerjee (1992) uses a sequential decision model in which each decision maker looks at the decisions made by previous decision makers in making their own decision. He concludes that the decision rules chosen by rational, optimizing individuals are characterized by herd behavior that is shown to be inefficient. No information asymmetry or strategic dependence of payoffs is assumed in this model.

Brock and Durlauf (2000), and Blume and Durlauf (2000) both use a discrete choice model where individual utilities from a given choice include a social component that characterizes the social interaction effects on individual choices. The models do not make any explicit assumptions of informational asymmetry or a positive externality. Equilibrium conditions are derived under the assumption of rational expectations of agents regarding the level of social choice.

Besides developing theories of the mechanisms by which networks might influence individual behavior, economists have also directed much attention towards understanding the impact of the presence of network effects on social welfare. The theoretical literature leads to several different conclusions about the welfare impact of the market outcomes in the presence of network effects. It has been shown that under certain conditions the presence of network effects might lead to inefficient outcomes such as "excess inertia" and a socially undesirable 'de-facto standardization' (Katz and Shapiro 1986).

On the other hand it has also been shown that in product markets with network effects, when consumers prefer variety, the market outcome is socially undesirable as it leads to multiple 'networks' when only one 'network' is optimal. In general, the theoretical conclusions depend heavily on the simplifying assumptions made by an otherwise extremely complex model. Consequently, very little is known about how and why networks influence people's choices. One way to resolve this question is to conduct empirical investigations to see if the conclusions of any one model hold uniformly across different markets or if different

markets exhibit characteristics suggestive of different mechanisms. A few empirical studies have been conducted to measure the role of network effects in explaining several socio-economic phenomena and these studies are the subject of discussion in the following section.

1.2. Applications of Economic Theories of Conformity

The idea that individuals tend to conform to the behavior of their reference groups has been used to explain and understand a wide range of social and economic phenomena. This section discusses the various empirical studies that attempt to measure the role of network effects in explaining several social and economic phenomena and to infer their impact on social welfare. The results and implications are mixed as described below.

Ellison and Glaeser (1997) develop a test for whether observed levels of geographic concentration of U.S. manufacturing industries are greater than would be expected to arise randomly and explore the nature of agglomerative forces such as industry spillovers and natural advantages, in describing patterns of concentration. Their results suggest a high level of geographic concentration of many

industries that cannot be explained simply by random occurrences or by the natural advantages of specific locations for some industries. Glaeser, Sacerdote and Scheinkman (1996) show that social interactions might be an important factor for explaining the incidence of crime in neighborhoods.

Evans, Oates and Schwab (1992) examine the influence of peer groups on teenage pregnancy and school performance. The neighborhood characteristics such as unemployment rate, median income level in metropolitan area can be considered as exogenous determinants of the choice of peer group but do not directly affect the probability of getting pregnant or dropping out of school. They find that when the endogeneity of peer group choice is accounted for in a simultaneous equation bivariate probit model, the peer group effects are insignificant.

Isolde and Kapteyn (1998) investigate the influence of habit formation and preference interdependence on labor supply behavior of females. They incorporate survey information on reference groups of individuals. Their results suggest that preference interdependence and habit

formation contribute significantly to female labor supply decisions. Allen (2000) examines the role of social networks in self-employment decisions and argues that it might explain gender differences in self-employment rates.

The idea of network effects has also found many applications in the industrial organization literature on diffusion of technology. Several studies have attempted to measure the strength of these network effects in the diffusion of technologies and products such as Automated Teller Machines (Saloner and Shepard, 1995) computer spreadsheet software (Gandal 1995), compact disc players (Gandal, Kende and Rob, 2000) and fax machines (Economides and Himmelberg, 1995).

In summary, even though the existing literature does not provide sufficient clues as to how or why social interactions affect individual choices, the essential idea that the presence of such interactions leads to conformity of behavior has found a great deal of support among economists and social scientists in general.

In Chapter 2 of this thesis, I examine the impact of network effects in computer software markets. Network effects are thought to exist in the computer software markets mainly due to a user's desire to swap files with other users and the desire for a large number and variety of complementary products and services. It has often been hypothesized that the presence of network effects in these markets leads to socially sub-optimal outcomes[3]. Specifically, it has been argued that the need for compatibility often leads to standardization and subsequent lock-in of an inferior technology[4], as the network effects offer a strong barrier to entry. This was indeed one of the key arguments in the recent antitrust case against computer software producer, Microsoft Corporation. It was argued that an "applications barrier to entry" protected Microsoft's dominant position in the operating systems market. That the reluctance of firms to write software applications that would be compatible with the less popular operating platforms presented a significant barrier to

---

[3] Katz and Shapiro (1985 & 1986), Farrell and Saloner (1986), Church and Gandal (1993)

[4] Arthur (1989), David (1985), Cusumano et al (1992)

entry into the operating systems market and thereby ensured the continued dominance of Microsoft's operating systems.

The argument that a new operating platform would not be adopted even if it were superior in technology to Microsoft's operating systems is mainly based on a few inadequately tested theoretical models. Despite the absence of reliable empirical evidence in support of the theory of inefficient standardization, its use in a major antitrust case further emphasizes the urgency of subjecting it to an empirical test. The data required for a direct test of the hypothesis however, is mostly unavailable to researchers. I therefore, devised an indirect test that takes advantage of a natural experiment that was afforded by the introduction of the new software programming language, Java. Java made it possible for software programmers to write a single program that would run on any operating system enabled with a Java Virtual Machine. It therefore had the potential to eliminate the disadvantage faced by competing operating systems developers in obtaining the "critical mass" required for successful adoption by consumers. In chapter 2, I employ a hedonic price regression model to examine the impact of Java on the level of competition in the software

applications market. The results indicate that Java was successful in increasing the competition in the market for software applications. This result not only has important implications for the ongoing antitrust trial of Microsoft but also for the broader question of whether or not a superior technology would always be successful in the presence of network effects.

In Chapters 3 and 4, I examine the role of social networks in determining labor market outcomes. Self-employment rates in the United States fell dramatically for most of the twentieth century before starting to increase in the 1970's. The racial gap in self-employment rates however, remained almost constant throughout this period. Given that self-employment has long been considered as an important means of closing the racial gap in wealth, studying the reasons for the persistence of the gap in self-employment rates is important for policy formation. Even though many theories have been proposed in the sociology and economics literature, none of them has been successful in fully explaining the persistence of the gap and the observed patterns of ethnic and racial self-employment. A possible explanation for the persistence of the gap could be that an

individual's decision to become self-employed depends, among other things, on the number of individuals in his/her social network that are self-employed. This dependence may be due to several different reasons such as better access to information, better access to credit etc. Past empirical studies in economics have mainly focused on the impact of labor market disadvantages and consumer discrimination on individual self-employment decisions.

In Chapter 3, I replicate and compare the results from earlier studies of minority self-employment decisions. I employ and enhance the empirical model developed in Borjas and Bronars (1989) to study the impact of consumer discrimination on minority self-employment rates. A comparison of the results indicates that this model of discrimination is incomplete and inadequate in explaining the racial differences in self-employment decisions. A comparison of the results helps to identify some limitations and inaccuracies in the empirical analysis of Borjas and Bronars (1989). The main result in Borjas and Bronars (1989) is that when Black entrepreneurs face discrimination from consumers, the more able Blacks tend to choose wage/salary employment instead of self-employment.

This result did not hold up when the study was replicated. The source of the discrepancy appears to be the difference in the construction of the samples. In the replicated study, I attempted to follow the construction of the stratified random samples exactly as described in Borjas and Bronars (1989). However, adhering to the exact rules for sample selection resulted in much larger samples than was reported in Borjas and Bronars (1989). As a result, the weights used in the probit regression in Borjas and Bronars (1989) may have been inappropriate and may account for the difference in the results. In addition, the model of employer and consumer discrimination may be incomplete if Black entrepreneurs also faced discrimination from workers. The model also omitted controls for industry, geographic location and social network measures that might affect self-employment decisions.

In chapter 4, I extend this model to include neighborhood characteristics, social network measures and industry dummies. A likelihood ratio test conducted leads to the rejection of the restricted model of Chapter 3. Inclusion of several different network measures allows us to reasonably identify and examine the role of networks in

self-employment decisions of individuals. The results indicate that social networks indeed played an important role in determining an individual's self-employment status. Given the past inexperience of Blacks in business, social networks therefore might explain the persistence of the racial gap in self-employment rates.

This dissertation thus explores the role of network effects in two markets. In the computer software markets, where network effects have received a good deal of attention, the results of the study indicate that network effects may not be as persistent or as damaging as previously thought. In labor markets and in self-employment decisions in particular, where the presence of network effects had received much less attention, the results indicate that social networks might have a larger impact on self-employment decisions than previously thought.

II. THE EFFECT OF JAVA ON NETWORK BENEFITS IN COMPUTER SOFTWARE MARKETS

2.1. Introduction

A network benefit is said to exist if the value of a product to an individual consumer is higher when a large number of other consumers also consume the same product. Some products that may exhibit network benefits are videocassette recorders, compact disc players, telecommunication equipment, personal computers (PCs), and computer software. In computer software markets for example, individual consumers benefit from a bigger 'network' of consumers because they can exchange files with a greater number of other users and also because a greater number of complementary software and services are available for the bigger network.

The presence of network benefits is often believed to lead to inefficient market outcomes and therefore has attracted a great deal of attention from economists. However, little empirical evidence is available to support or refute such a belief. Given its policy implications however, it is

increasingly important to subject the theory to empirical testing.

This chapter focuses on the network effects in the market for PC Operating Systems (OS). Specifically the goal is to empirically examine whether or not, the programming language Java® helped to reduce the extent of the network effects in this market.

2.1.1. Background Literature

The theoretical literature on network externalities has seen rapid growth in the last decade. The predictions of the various models depend heavily on the assumptions made about the underlying preferences and expectations of consumers.

Katz and Shapiro (1985) show that the presence of network externalities and the need for compatibility lead to multiple equilibria and that the equilibrium that actually obtains is determined by consumer expectations and the firms' actions. Farrell and Saloner (1986) use a dynamic

model, to show that a new and 'superior' technology may not be adopted due to 'excess inertia' caused by the presence of an installed base. Katz and Shapiro (1986) show that the presence of network benefits can lead to 'excessive' standardization. Church and Gandal (1992 &1993) study the adoption of (hardware) technology when complementary (software) products are produced by different firms and show that (i) when consumers place a high value on software variety, there is a sub-optimal amount of standardization by the market (ii) when software firms are Bertrand competitors, a hardware technology with lower software development costs is adopted when it is socially optimal to adopt the other technology.

Liebowitz and Margolis (1994, 1995a & b), Witt (1997) present a competing opinion and argue that even though the presence of network benefits might lead to standardization, it does not necessarily lead to inefficient lock-in. The effect of the presence of network benefits on social welfare therefore remains debatable.

The empirical literature in this area is mostly limited to studies that show that significant network benefits exist

in various markets such as ATM machines (Saloner and Shepard 1995), computer spreadsheets (Gandal 1995) etc. Gandal, Kende and Rob (2000) estimate the speed of adoption of compact disc players. The degree of standardization on Microsoft's operating systems and the degree of substitutability between niche operating systems and upgrades of the mainstream OS is documented by Kretschmer (2001).

There is little empirical evidence however, to support the phenomenon of sub-optimal standardization and technological lock-in. David (1985), Cusumano et al. (1992), Liebowitz and Margolis (1990) are some studies that examine the hypothesis of premature and inefficient standardization. Bresnahan and Greenstein (1999) study the level of technological competition in the computer industry.

These studies however mostly provide anecdotal evidence. Empirical demonstration of inefficient standardization has proved to be a tricky and difficult task due to non-availability of data and due to the absence of a clear consensus on what can be considered as evidence of lock-in.

See Foray (1997) for a detailed discussion of the empirical issues involved in demonstrating inefficient lock-in.

Although it was not possible to conduct a direct test of this hypothesis, an indirect test is devised in this chapter using the data generated by the natural experiment afforded by the introduction of Java®. The following sub-sections provide a brief background of the PC operating systems market and the introduction of Java®.

2.1.2. The PC Operating Systems Market

A computer operating system provides an interface between the computer's Central Processing Unit (CPU) and the various pieces of hardware such as the monitor, printer etc. Operating systems also control the interaction between various applications software such as word-processing or spreadsheet programs, and the CPU. Because of the complex interactions between operating system software, applications software, and the hardware attached to the PC, in general, application software written for a particular operating system cannot run on a different operating system without extensive and costly modifications and add-ons. As

a result, a user's ability to exchange programs and files increases with the number of users of the same operating system. Such benefits are termed direct network externalities[5] since these benefits are an increasing function of the number of consumers.

In addition, vendors of software applications tend to provide a greater number and variety of programs and complementary services for the most popular system. There might be two reasons why firms prefer to write application software for the dominant OS. One is that the potential market size is larger for applications written for the dominant OS. Since a huge component of software development costs is fixed, it might be argued that there are increasing returns to scale in software development that make smaller networks less attractive for firms writing applications software. Second, firms might write software for the dominant OS if they expect to be able to extract a premium for compatibility with the dominant OS. Therefore as the network size grows, the number of complementary

---

[5] See Gandal (1995) for a definition and explanation of direct and indirect network externalities.

software programs increases. This effect has been termed as indirect network externalities[6].

Microsoft Corporation currently dominates the market for PC operating systems with its family of Windows® OS and its predecessor MS-DOS. It has been argued that the availability of a greater number of complementary software products was crucial for the widespread adoption of Microsoft's operating systems and for its continued dominance in this market (Gandal, Greenstein and Salant 1999). Firms prefer to write applications software for Microsoft's dominant OS which in turn causes its market share to grow more thereby initiating a "positive feedback cycle" that is believed to be very difficult if not impossible to break. Despite scant empirical evidence, it is commonly believed that such technological lock-in is a very common occurrence in industries with significant network externalities.

Direct tests of this hypothesis however are very difficult to conduct, as the data required for such a test are not

---

[6] Ibid.

easily available. This chapter takes an indirect approach to testing the strength of indirect network effects in the operating systems market by exploiting the 'natural experiment' afforded by the introduction of Java. The following sub-section contains a background history of the introduction of Java and explains why it was an important 'natural' experiment for the purpose of hypothesis testing.

2.1.3. The introduction of Java and its impact on the PC-OS market

The growing popularity of the Internet in the early 1990's and the introduction of the programming language Java by Sun Microsystems Corp. in 1995, posed a serious potential threat to the dominance of Windows in the operating systems market.[7] Java made it possible, through its Java Virtual Machine (JVM), for programmers to write applications that would run on any operating system. Thus it became possible for developers of application software to write a single program that would run on any operating system enabled with

---

[7] InfoWorld (Nov.13, 1995) "Java OS; Sun focuses on eclipsing Windows."

JVM[8] thereby greatly reducing the costs of software development. (In the absence of Java, software developers would have to write separate programs in order to ensure compatibility with each of the different operating platforms.)

Netscape Navigator®, which was the leading Internet browser at the time, was the major distribution channel for JVM[9]. Even though the need for an operating system could not be eliminated, Netscape and Java together presented an alternative platform to which applications could be written thereby creating greater potential for competition both in the market for operating systems and the market for applications software. This meant that consumers using non-Windows OS could expect to have a greater number and variety of application software available to them as a

---

[8] Java World (May 1996) " Java jumps to operating systems: All major operating systems to build in support for Java"

[9] InfoWorld May 29, 1995 "Netscape Inks Pact with Sun" - '"Netscape 2.0 came with a built in support for Java. This version became available in late 1995. Until then, Sun's Hot Java browser was the only java-compatible browser."'

result of this new technology.[10] Thus Java possessed the potential to eliminate or at least, reduce the network effects in the market for operating systems.

However, this promise did not materialize as programs written in Java tended to be much slower in executing than programs written in other languages such as C. In addition, there were security concerns being voiced in the industry regarding Java 'applets'[11]. Despite these initial concerns, Java enjoyed a good deal of support and a steady rate of adoption in all of 1996 and most of 1997[12]. In the second

---

10 See InfoWorld (Oct. 30, 1995) "Smell of Java lures scores of vendors."

11 "Despite growing support, some security concerns are being raised about Java applets…although Sun maintains that Java includes facilities for screening out viruses, Java applets might open up undetectable security holes on a client computer." InfoWorld Oct.30 1995: 'Smell of Java lures scores of vendors'.

"Lord knows, interpreters are slow. Pre-compiling Java source code into portable byte codes saves the time needed for translating syntax, but interpreting byte codes in the Java virtual machine (JVM) is still exceedingly slow compared to the native code produced by a compiler. Thus, Java performance is generally deemed acceptable for small applets but not for any sizable application." Java World, March 1998.

12 "Software Development '96: Java is this year's hot ticket" Java World, April 1996.

half of 1997 however, Microsoft introduced its own "impure" version of Java that ran better with its own browser and OS than with the other browsers or operating systems. These actions along with other alleged anti-competitive behavior led to the antitrust trial of Microsoft.

Microsoft is accused of using its dominant position in the PC operating systems market to push its own Internet browsing technology. (It is beyond the scope of this chapter to state and evaluate the various anti-competitive actions that Microsoft was accused of. See Gilbert and Katz (2001), Klein (2001) Whinston (2001), Gilbert (1995), Lopatka and Page (1995), Hall (1999) for a history and analysis of the various arguments in the Microsoft antitrust case).

These actions coupled with Sun Microsystems's inability to get its application approved for turning Java into an international standard, was a setback to the further adoption and development of Java as a cross-platform

---

"Educators embrace Java: High marks in college signal long-term, real world success of new programming language", Java World, January 1997.

language. As a consequence, there are only a few applications written in Java that are commercially available to date. Most of the applications tend to be custom-written for specific customers. However Sun Microsystems continues to try to improve its technology in order to encourage its adoption.[13]

In summary, it may be argued that even though Java had not yet demonstrated itself to be a superior technology it did have the potential to create more competition in the market for application software and the market for PC operating systems. Therefore, it had the potential to raise the indirect network effects for non-windows operating sytems, which would reduce the gap in indirect network effects between Windows and non-Windows operating systems.

If it can therefore be shown that Java succeeded in creating more competition and reducing the extent of network effects in this market despite its technical shortcomings it would be consistent with no inefficient

---

13 "Sun's Hotspot technology promises to deliver interpreted byte codes that run faster than a compiled program" 'Hotspot: A new breed of virtual machine', Java World, March 1998.

lock-in. If Java were shown to have not been effective in creating the expected competition however, we would need further evidence to identify whether this was evidence of inefficient technological lock-in or of Java's inefficient technology.

2.2. Model

A random utility model is used to show that the firms whose software products are compatible with the dominant operating system will command a higher price than firms whose software products are not compatible with the dominant operating system. It is argued that the effect of the introduction of Java was (1) to reduce the difference in the extent of network benefits of the Windows vs. non-Windows operating system by increasing the value of the non-Windows network and (2) create more competition both in the application software markets and the OS market.

Consider the choice of application software by consumers who have already invested in one of the two operating systems, A or B. Operating system A has a bigger network.

Assumptions:

1) There are two firms 1 and 2 that produce application software compatible with the dominant operating system A and firm 3 produces software compatible with the operating system B.

2) The software compatible with operating system A is incompatible with the operating system B and therefore cannot be substituted by software that is compatible with the operating system B and vice versa.

3) The greater the variety of complementary products compatible with any product, the higher the network benefit. The benefit from compatibility with the dominant OS of network of size S_A is $W(S_A)$ and the benefit from compatibility with the less popular operating system B of network size S_B is $W(S_B)$. By assumption, therefore $S_A > S_B$ and $W(S_A) > W(S_B)$.

The consumers on the operating system A-network have a choice either to buy nothing or to buy application software from firm 1 or from firm 2. The corresponding utilities of each choice are as follows:

$$U_0^A = V^A + \varepsilon_0^A$$

$$U_1^A = W(S_A) + c_1^A - p_1^A + \varepsilon_1^A$$

$$U_2^A = W(S_A) + c_2^A - p_2^A + \varepsilon_2^A$$

Whereas consumers on the operating system B-network choose between buying nothing and buying from firm3. The corresponding utilities of each choice are as follows:

$$U_0^B = V^B + \varepsilon_0^B$$

$$U_3^B = W(S_B) + c_3^B - p_3^B + \varepsilon_3^B$$

c_1^A, c_2^A and c_3^B denote the value of the different product characteristics offered by firms 1, 2 and 3 respectively. V^A and V^B denote the value of waiting another period(s) before buying any software. p_1^A, p_2^A, p_3^B denote the prices charged by firms 1, 2 and 3 respectively.

The ε's represent consumers' idiosyncratic tastes and are assumed to have a double exponential distribution therefore the difference between the ε's will be distributed according to the logit. The ε^A's are independent of the ε^B's. The probability that an individual consumer chooses firm 1's product is then given by the probability that

$U_1^A > U_2^A and\ U_1^A > U_0^A$ which is equal to the probability that the difference between the random components of the utility is greater than the difference in the non-random components. The total number of consumers is normalized to one so that the demand for each product is given by the probability that the product is chosen.

Given our assumptions, the demand function for each product is given by:

$$D_1^A = \frac{\exp((W(S_A) + c_1^A - p_1^A))}{\exp(V^A) + \exp((W(S_A) + c_1^A - p_1^A)) + \exp((W(S_A) + c_2^A - p_2^A))}$$

$$D_2^A = \frac{\exp((W(S_A) + c_2^A - p_2^A))}{\exp(V^A) + \exp((W(S_A) + c_1^A - p_1^A)) + \exp((W(S_A) + c_2^A - p_2^A))}$$

$$D_3^B = \frac{\exp((W(S_B) + c_3^B - p_3^B))}{\exp(V^B) + \exp((W(S_B) + c_3^B - p_3^B))}$$

Given the above consumer demand functions, the firms choose prices simultaneously to maximize their individual profit functions, which are denoted by π below. 'F' denotes the fixed costs of developing software. The subscripts refer to the firm and the superscript refers to the OS network.

$$\pi_1^A = p_1^A D_1^A - F_1^A$$
$$\pi_2^A = p_2^A D_2^A - F_2^A$$
$$\pi_3^B = p_3^B D_3^B - F_3^B$$

The first order conditions yield the following implicit relationships for the prices.

$$p_1^A = \left[1 + \frac{\exp((W(S_A) + c_1^A - p_1^A))}{\exp(V^A) + \exp((W(S_A) + c_2^A - p_2^A))}\right]$$

$$p_2^A = \left[1 + \frac{\exp((W(S_A) + c_2^A - p_2^A))}{\exp(V^A) + \exp((W(S_A) + c_1^A - p_1^A))}\right]$$

$$p_3^B = \left[1 + \frac{\exp((W(S_B) + c_3^B - p_3^B))}{\exp(V^B)}\right]$$

Predictions from the above model are as given below.

(1) $\frac{\partial p_1^A}{\partial c_1^A} > 0, \quad \frac{\partial p_2^A}{\partial c_2^A} > 0, \quad \frac{\partial p_3^B}{\partial c_3^B} > 0$

The prices are increasing in the respective quality characteristics.

(2) $\frac{\partial p_1^A}{\partial W(S_A)} > 0, \quad \frac{\partial p_2^A}{\partial W(S_A)} > 0, \quad \frac{\partial p_3^B}{\partial W(S_B)} > 0$

The prices increase as the size and the value of the networks increases.

(3) $\frac{\partial p_1^A}{\partial V^A} < 0, \frac{\partial p_2^A}{\partial V^A} < 0, \frac{\partial p_3^A}{\partial V^B} < 0,$

As the value of waiting increases, the prices fall.

The introduction of Java (and the possibility of cross-platform applications becoming available in the future) should have increased the value of waiting for consumers on both networks. Therefore the potential for increased competition in the future should have depressed prices on both networks[14].

---

[14] Even though the value of waiting another period may be higher for consumers on network B, it is not clear whether this would result in the prices of B-compatible software falling more than A-compatible software.

2.3. Empirical Analysis - Hedonic Price Regressions

The empirical analysis presented here is similar to Gandal (1994), which provides evidence that the PC spreadsheet market exhibits network externalities and that consumers are willing to pay a premium for spreadsheets that are compatible with the LOTUS platform[15]. Gandal (1995) tests for indirect network externalities arising from common file compatibility standards among spreadsheet software and database management systems. To control for compatibility with hardware, Gandal (1994 & 1995) focuses on programs written for Intel-based PCs.

In this chapter, I focus on the market for graphics software written for Intel-based PCs[16]. Graphics software applications have recently begun to be integrated into office productivity suites. But during the time period under consideration in this chapter (1993-97), they were

---

[15] Lotus was the dominant spreadsheet in the time period considered in Gandal's (1994) paper.

[16] Ideally I would have liked to focus on spreadsheet or database software applications for comparison with earlier studies. But during the time frame that I was interested in, the spreadsheet software became part of office productivity suites and it would have been difficult to

largely sold as stand-alone products. Another advantage of using data on graphics software was that there was no clear leader or dominant product among the various graphics software programs that were available during this time frame.

Using hedonic price regressions, I attempt to measure the extent of the network externality in the graphics software market and the effectiveness of Java in reducing it. Even though estimation of a structural model would have been ideal, lack of data on sales and market shares of individual firms, made it imperative to use hedonic regressions.

The regressions included the log of the real price of the graphics application as the dependent variable and a number of independent variables that measured the quality and performance characteristics of the applications. In addition, there were TIME dummies to control for any time effects. Dummies representing each firm in the dataset that

---

control for the changes in quality, as standalone prices were not available for the spreadsheet applications.

developed graphics software application programs, control for any firm fixed effects.

The installed base of the operating system that the software application program is primarily written for, served as a measure of the size of the OS network. The coefficient on the installed base variable, if significant, would be an indicator of whether the application providers were able to capture any premium for compatibility with the dominant operating system.

The introduction of Java increased the *potential* for competition in both the operating systems market and the software applications market. In the applications market it would have increased the value of the option to wait for cross-platform applications. Therefore, as shown in the previous section, prediction 3 of the model indicates that it would result in a fall in the prices of platform-specific applications.

It is conceivable that the waiting option becomes more attractive for the OS/2 users than for the Windows users but it is impossible to tell from our model whether this

would lead to a bigger reduction in prices of OS/2 - compatible software in comparison to Windows-compatible software.

2.3.1. Data

Data on the price and quality characteristics of graphics application software was compiled from product reviews published in PC magazine, InfoWorld, PC World, PC Computing, Compute!, Windows magazine and OS/2 e-zine during the years 1993-1997. Graphics software comes in several different packaged forms - some are written for very specific purposes, such as illustration, image viewing and editing, desktop publishing (DTP), charting etc., while others include two or more of these functions. The data correspond to all of these programs and every effort was made to control for all the features that described these software packages. The resulting sample is an unbalanced panel with a total of 107 observations out of which 13 were compatible with OS/2, and the rest with the Windows operating system.

There were 42 distinct software products. Whenever multiple versions (such as standard and professional) were available for any product in the same year, each version was treated as a separate product i.e. each version is a separate observation within the same year. Version upgrades (such as version 1.0 and version 2.0) are treated as the same product i.e., whenever two versions (old and new) were available in the same year, only the new version appears in the sample in that year. If no new version was available in the subsequent year(s) but the old version continued to be sold and was reviewed in any of the above mentioned magazines, then the old version appeared in the sample in the subsequent year(s) also as it was a choice available to consumers in the subsequent years. Such observations however were few and the results presented are not sensitive to their inclusion.

The following is a list of all the variables in the empirical model.

Dependent Variable

RPRICE: is the list price of the product in real dollars. As the street price was not available for all the products

in the sample, I had to use the list price even though it does not represent the ultimate cost to consumers. However, this should not be of great concern since the list price is found to be very highly (over 99 percent) correlated with the street price and the discount, which was typically found to be about 30-60 percent, is found to be uncorrelated with the characteristics of the software.[17] The natural log of the real list price is the dependent variable in my model.

The independent variables include a list of quality characteristics, a measure of network size, a variable that is constructed to measure the impact of Java and a list of variables that control for fixed time and firm effects.

Independent Variables - Measures of Quality

CLIPART: This dummy variable takes on a value of 1 if the package comes with a library of pre-selected pictures that may be useful in constructing illustrations. This feature is most useful for users who are not professional graphics designers. Ideally I would have liked to control for the

[17] See Appendix A for the relevant data and correlation

actual number of images supplied with the software, as there is a huge amount of variability in the number of images that are supplied. Some packages, for instance come with a library of 15 images whereas others had a library of over 5000 images. However as the actual number of images was not available for all observations, I could only control for whether or not the software provided a clipart library.

MULTIMEDIA: The variable MULTIMEDIA takes on a value of 1 if the package offers support for sound, video and/or animation, zero otherwise.

TEXT: The variable TEXT takes on a value of 1 if the software is capable of importing either formatted or unformatted (ASCII) text. All software had some capability of inserting text onto an image but not all could import text. This feature is most useful for desktop publishing tasks.

DATA: This variable takes on a value of 1 if the program can import data from at least one of the popular data storage and manipulation programs such as Excel, Lotus, and DBase.

---

statistics.

RAM: This is a continuous variable representing the minimum amount of Random Access Memory (RAM) that was required to run the program. This was included to control for the hardware requirement standards that might influence consumer's demand for any software. Also generally speaking, it is observed especially for graphics applications that the more features a program has, the more RAM it requires to run smoothly. Therefore, this may also serve as a proxy for the many quality characteristics that cannot be easily measured.

CHARTS: This variable takes on a value of 1 if the program is capable of creating at least basic charts, such as bar, pie and line graphs etc. A better control would have been the number of different chart types that a program could create, but once again I am constrained by non-availability of data. I could only control for whether or not the software was able to create charts but since there was considerable variation among the products even in this respect, it may be expected to suffice.

OLE: This variable takes on a value of 1 if the program supports Object Linking and Embedding (OLE). This feature enables active links between different programs- thus if the data underlying a chart is changed in the spreadsheet

program, it is automatically reflected in the linked chart. This feature is only available for Windows compatible software. However since not all Windows compatible software included this feature, it is important and possible to control for the feature.

DDE: This variable takes on a value of 1 if the program supports Dynamic Data Exchange (DDE). This also is a feature that is available only for Windows compatible software however all not Windows compatible software provide it.

The above list of characteristics may not however be a complete list of the features that may be important for evaluation of graphics software. Other features such as the number and types of drawing tools, the kind of color output generated, the types and number of different input and output devices supported, presumably are also important, but it was not possible to control for all of these features due to lack of consistent data for all products in the sample. I could however, control for general characteristics such as whether the applications were more suited for specialized functions such as image editing, illustration, desktop publishing or presentations.

DRAW: is a dummy variable that takes on the value of 1 if the software is primarily meant for illustration purposes and therefore provides specialized and sophisticated drawing tools that are generally unavailable in other graphics applications.

EDIT: is a dummy variable that takes on a value of 1 if the software is primarily meant for viewing and editing photos/images.

DTP: (Desktop Publishing) is a dummy variable that takes on a value of 1 if the software consists of specialized tools for publishing activities.

PRES: is a dummy variable that takes on the value of 1 if the software provides special presentation tools.

Independent Variables - Measures of Compatibility

File format standards: The dummy variables BMP, EPS, GIF, JPEG and TIFF take on the value of 1 if the package offers full/partial compatibility with the respective formats. Compatibility with a particular format allows the user to exchange files with other software applications that support the format. Graphics image formats fall in two broad categories - bitmap and vector. Bitmap images store

information in the form of dots and pixels whereas vector images are collections of vectors and shapes. Of all the file formats listed above, .EPS is a vector format whereas the others are bitmaps. The .GIF and .JPG (JPEG) formats are most used for web publishing. An alternative variable (BITMAP) was also constructed, which took on the value of one if the application provided compatibility with all of the four bitmap formats listed above and zero otherwise.

WIN95: This variable indicates compatibility with Windows95 operating platform. Microsoft released Windows95 in August/September of 1995 even though it had pre-announced this new improved upgraded version much earlier. Windows95 was a 32-bit operating system that would have made incorporation of many additional quality features possible for software applications. This variable was therefore included to account for the improved quality features that it may have made possible. The coefficient on this variable therefore is expected to be positive.

INST_BASE: is a continuous variable that is a measure of the installed base (in millions of units) of the operating system in any given year. For example, if the observation

corresponds to a program that is compatible with the Windows operating system in the year 1993, then the variable takes on the value of the installed base of Microsoft's operating systems in 1993. The sample considered consists of software written for the OS/2 and both Windows3.1 and Windows95 platforms. The installed base data for 1992 was obtained from extracts of International Data Corporation's (IDC) reports published in various software magazines. The installed base data for the rest of the years was constructed from the market share and shipments data in the now publicly available trial exhibits from the Microsoft Antitrust case. As this data only gives the aggregate market share and shipments of all of the operating systems (including MS-DOS) sold by Microsoft, it was not possible to separate the shares of the various versions of the Windows operating systems and MS_DOS[18].

Fixed Effects

TIME: The time dummies YR_93, YR_94, YR_95, YR_96, YR_97 take on a value of 1 if the data correspond to that year.

---

[18] See Appendix B for details on construction of installed base estimates.

These dummies serve as a control for any market-wide or economy-wide time-specific exogenous factors that might have impacted the prices of the software under consideration.

FIRM: The data correspond to graphics application software that was developed by either of 22 firms. The firm dummies serve as a control for any number of unobserved characteristics that are unique to a firm and/or its product (e.g. the product interface, ease of use, organizational culture etc.). In addition, the dummies also serve as controls for the firm's market share and resultant network effects within the graphics applications market. As there was no single firm or product that was a clear leader in all the segments of this market during the time period under consideration, it is reasonable to assume that this is an adequate control for the direct network effects, if any, within the graphics software market.

The Effect of Java

JAVA: is a dummy variable that takes on a value of 1 only if the software was introduced after October 1995. It is

important to note that this variable is *not* simply equal to the sum of the 95, 96 and 97-time dummies. This was because (1) this variable takes on a value of one for only a subset of the observations in year 1995 and (2) because I have observations for versions of software applications that continued to be sold after October 1995 but were actually introduced before October 1995 and so these observations had a value of zero for the JAVA dummy. Typically firms introduced new versions of their product on a cycle of one to two years. However there were some applications that did not see new updated versions for the entire period under consideration. The sample includes observations on all software that was available for sale and that was reviewed during the time period under consideration even if the versions were old. To check for robustness of results the regressions were estimated without the old software as well. The results were not sensitive to including the old software.

2.4. Results

The descriptive statistics are shown in the Tables 2-1 through Table 2-3 below. Figure 2-1 shows the distribution of software prices by operating platform. It may be observed that there was a general downward trend in software prices during the period 1993-1997. Also the prices of Windows based applications were higher on average than prices of applications written for OS/2.

Table 2-1. Real Price by Operating Platform

OS	Observations	Mean	Std Dev	Minimum	Maximum
OS/2	13	226.33	154.48	43.53	555.56
Windows 3.X	64	377.35	154.89	84.59	625.44
Windows95	30	336.71	133.168	127.81	583.06

Table 2-2. Real Price by Year

YEAR	Observations	Mean	Std Dev	Minimum	Maximum
93	21	402.13	135.49	172.68	625.44
94	20	371.54	153.35	87.16	612.18
95	21	339.01	150.33	84.59	595.48
96	22	321.82	163.41	63.05	579.66
97	23	309.54	166.97	43.53	558.68

Figure 2-1. Relative Frequency Histogram of Real Prices by Operating System

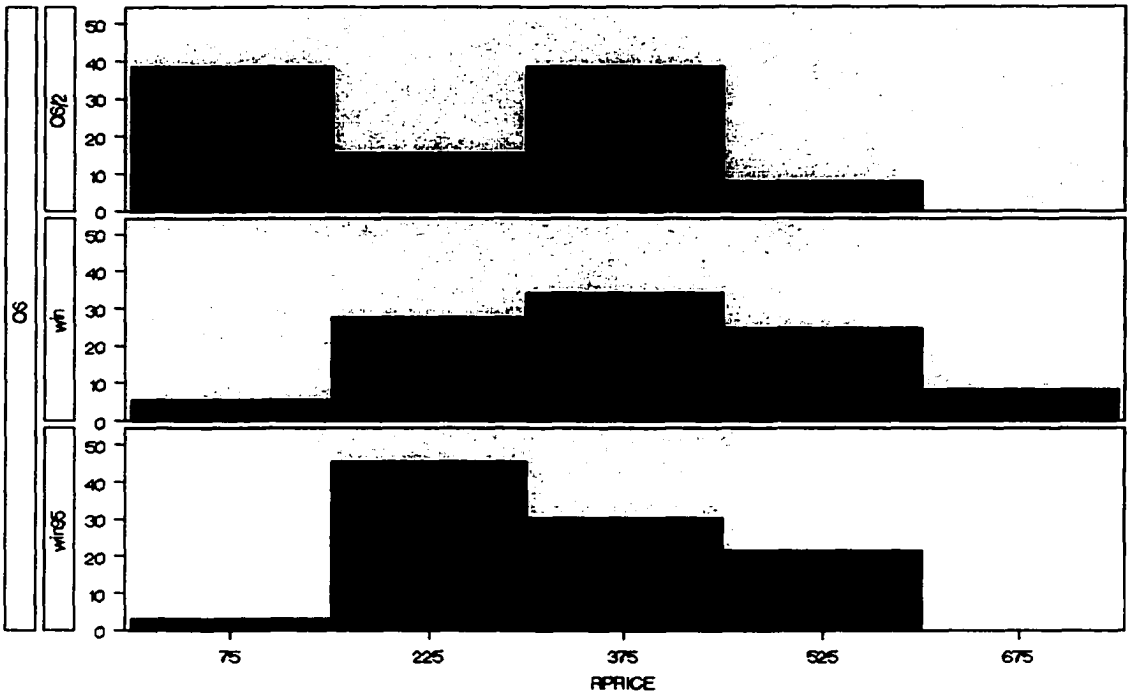


Table 2-3. Real Prices by Year and Operating System

OS	Observations	Mean	Std. Dev	Minimum	Maximum
YEAR=93					
OS/2	2	450.73	148.24	345.91	555.56
Windows 3.X	19	397.01	137.43	172.68	625.44
YEAR=94					
OS/2	1	331.33	.	331.33	331.33
Windows 3.X	19	373.66	157.25	87.16	612.18
YEAR=95					
OS/2	2	323.43	1.64	322.27	324.59
Windows 3.X	13	333.77	168.95	84.59	595.48
Windows 95	6	355.56	146.26	220.85	583.06
YEAR=96					
OS/2	4	182.03	110.94	63.05	316.09
Windows 3.X	8	371.20	174.27	160.57	579.66
Windows 95	10	338.24	153.07	127.81	572.62
YEAR=97					
OS/2	4	83.64	50.27	43.53	155.47
Windows 3.X	5	439.83	172.56	180.40	558.68
Windows 95	14	327.55	121.87	210.82	558.68

Table 2-4. Hedonic Price Regression Estimates
(Dependent variable is the natural logarithm of the software's list price in real dollars)
Number of observations = 107

Variable Name	Parameter Estimate	Standard Error	Parameter Estimate	Standard Error	Parameter Estimate	Standard Error	Parameter Estimate	Standard Error
INTERCEPT	5.4354	0.3078	5.5796	0.2689	5.5377	0.2023	5.55098	0.2043
Installed Base	-0.0009	0.0030	-0.0004	0.0028	-0.0007	0.0027	0.000189	0.00302
JAVA	-0.2553	0.0974	-0.2486	0.0924	-0.2543	0.0903	-0.31075	0.12729
DRAW	0.3160	0.1158	0.3272	0.1080	0.3124	0.1003	0.30623	0.10119
EDIT	0.3696	0.1221	0.3586	0.1144	0.3610	0.1127	0.34695	0.11536
DTP	0.2363	0.1857	0.2191	0.1722	0.2581	0.1447	0.24528	0.14675
PRES	-0.1152	0.2239	-0.2433	0.1917	-0.2403	0.1885	-0.23822	0.18937
BMP	0.2971	0.2895	----	----	----	---	----	----
EPS	0.0903	0.1961	-0.0115	0.1513	----	----	----	----
GIF	0.1092	0.1629	----	----	----	----	----	----
JPEG	0.1516	0.1438	----	----	----	----	----	----
TIFF	-0.1333	0.1997	----	----	----	----	----	----
Clipart	-0.0944	0.1137	-0.0465	0.1045	----	----	----	----
Multmedia	-0.2121	0.1301	-0.1877	0.1126	-0.2016	0.1062	-0.22126	0.11106
TEXT	0.3928	0.1126	0.4181	0.1066	0.4203	0.1035	0.41898	0.10394
DATA	0.2015	0.1791	0.2986	0.1382	0.3004	0.1356	0.29576	0.13636
RAM	0.0145	0.0106	0.0146	0.0100	0.0138	0.0097	0.01096	0.01075

Table 2-4 continued on next page

Table 2-4 - continued

Charts	-0.0902	0.1328	-0.1227	0.1209	-0.1182	0.1186	-0.11695	0.1191
OLE	0.0731	0.0923	0.1239	0.0805	0.1153	0.0765	0.11707	0.07689
DDE	-0.0936	0.0985	-0.1043	0.0950	-0.1064	0.0933	-0.10538	0.09375
YR_94	-0.0886	0.0829	-0.1105	0.0770	-0.1106	0.0744	-0.12077	0.07643
YR_95	-0.1630	0.1116	-0.1870	0.1041	-0.1902	0.1007	-0.22132	0.11254
YR_96	-0.1261	0.1640	-0.1706	0.1476	-0.1659	0.1450	-0.19772	0.15408
YR_97	-0.1121	0.2460	-0.1786	0.2210	-0.1655	0.2161	-0.22497	0.23657
BITMAP	----	----	0.2595	0.1109	0.2656	0.1084	0.26998	0.10911
WIN95	----	----	---	----	----	----	0.09021	0.14269
Condition Index	76.5500		57.6500		44.4500		45.485	
R-Square		0.9204		0.9206		0.9204		0.9208
Adjusted R- Square		0.8639		0.8705		0.8740		0.8729

Notes: The Regressions also included firm dummies that are not reported here in order to save space. I rejected the pooled OLS model in favor of the firm fixed effects at the one percent significance level based on results from an F-test.

The first column of Table 2-4 shows the parameter estimates for the full model. The quality and performance characteristics captured by the measures of text and data handling capabilities are positive and statistically significant. On average, the software that provided text capability was priced 42 percent higher than software that did not provide such capability. Similarly data handling capabilities increased prices by almost 30 percent.

The results also show that the memory requirement RAM serves as a proxy for performance features that are hard to measure and on average every additional megabyte of RAM required resulted in a price increase of about 1.1 - 1.4 percent. The negative coefficient on multimedia is baffling. This variable was highly correlated with RAM and therefore the coefficient on RAM should be appropriately interpreted in conjunction with the coefficient on Multimedia capability.

The coefficient on installed base is statistically insignificant suggesting that the applications providers are unable to charge any premium for

compatibility with the Windows platform. This result is not surprising given that the installed base measure includes all versions of Microsoft's Windows and DOS operating platforms. This measurement error might very well result in an underestimated coefficient as the installed base measure is overstated. It is also possible that the application providers are unable to charge a higher premium for the bigger network since they also face greater competition in the bigger market.

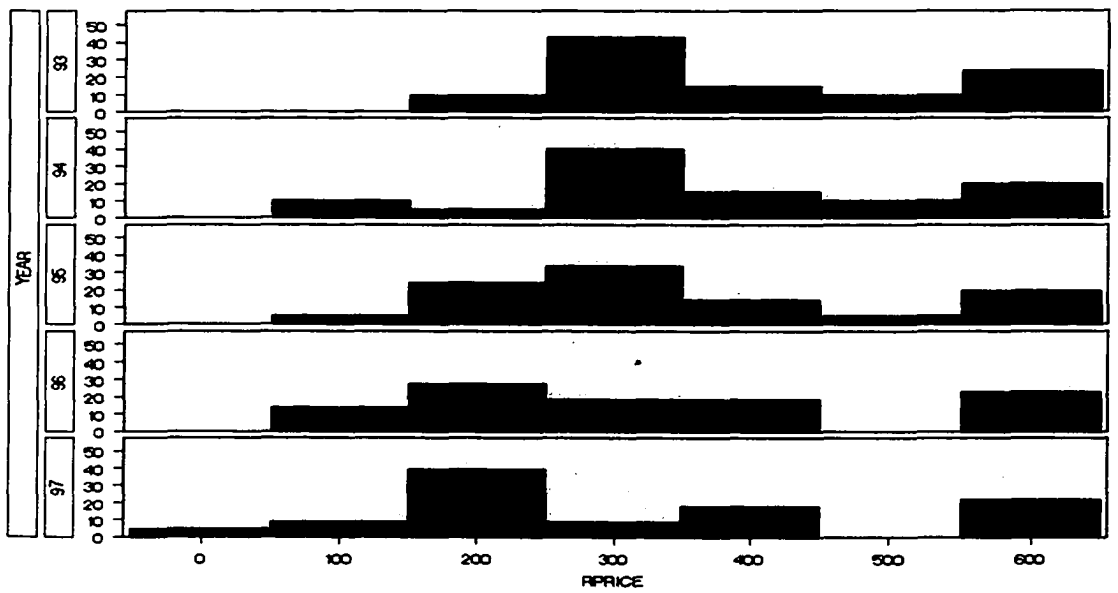
The coefficient on the JAVA term is negative and statistically significant at the one percent significance level. This suggests that even though software providers were already unable to charge any premium for compatibility with the dominant operating system, the introduction of Java further eroded their price-markups by 25-30 percent as a result of an increasing possibility of more competition in the future. The increase in competition is also evident in the number of software programs written exclusively for OS/2 after 1995. Even though these applications were not cross-platform capable, there was a rise in

the number of software programs written exclusively for OS/2 as is evident from our data. See Table 2-5 and Figure 2-2 below for evidence of this.

Table 2-5. Real Prices by Operating Platform and Year

YEAR	Observations	Mean	Std Dev	Minimum	Maximum
		OS=	OS/2		
1993	2	450.73	148.24	345.91	555.56
1994	1	331.33	.	331.33	331.33
1995	2	323.43	1.64	322.27	324.59
1996	4	182.03	110.94	63.05	316.09
1997	4	83.64	50.27	43.53	155.47
		OS=	Windows 3.X		
1993	19	397.01	137.42	172.68	625.44
1994	19	373.66	157.25	87.16	612.18
1995	13	333.77	168.95	84.59	595.48
1996	8	371.20	174.27	160.57	579.66
1997	5	439.83	172.56	180.40	558.68
		OS=	Windows95		
1995	6	355.55	146.26	220.85	583.06
1996	10	338.24	153.07	127.81	572.62
1997	14	327.55	121.88	210.82	558.68

Figure 2-2. Relative Frequency Histogram of Real Prices



The column 4 of Table 2-4 shows regression estimates with the variable WIN95 included. The inclusion of this variable increases the overall condition number index and lowers the efficiency of estimates. The variable itself however, is statistically insignificant. Therefore, we may conclude that the negative coefficient on JAVA is indeed a result of greater anticipated competition rather than a coincidental effect of the almost simultaneous introduction of Windows95. Moreover, the effect of the introduction of Windows95 on prices was expected to be positive due to the improved quality features. Most of the firms in the sample introduced newer versions for Windows95 simultaneously with the introduction of Windows95. Therefore, any drop in prices due to consumers waiting for a Windows95 version would have occurred prior to the introduction of software compatible with Windows95 i.e. before August 1995. As a result, the drop in prices observed after the introduction of Windows95-compatible software may be reasonably attributed to the introduction of Java.

Robustness checks

The least squares dummy variable approach taken here to estimate firm fixed effects, leads to a high level of multi-collinearity. The Belsley-Kuh-Welch condition index was found to be between 40-80. Such collinearity among the regressors is not unusual in hedonic regressions where most of the regressors are dummy variables. The problem is aggravated due to the additional dummies that control for firm fixed effects. It was however, not possible to estimate the usual within estimator as some of the software applications appeared only once in the data. Dropping these observations from the estimation procedure would have thrown away valuable information.

It was not feasible to use the random effects estimator due to the three-way error components (firm, product and time components in the residual error term) and the unbalanced nature of our panel data. As the data collected corresponds to almost the entire population of graphics applications, it was not possible to further enhance it.

Another remedy for multi-collinearity is to drop the variables that are highly collinear and/or most insignificant from the model. The results show that the file format standards in general are highly collinear and not an important determinant of the product's price. But instead of dropping them from the model, they were combined into a single measure of compatibility. In column 2 of Table 2-4 the variables BMP, GIF, JPEG and TIFF were dropped and instead included the variable BITMAP (The construction of this variable is explained above in the variable descriptions).

The results show that compatibility with all the four file format standards increases the price by almost 26 percent. The other parameter estimates do not change very much but the standard errors are lower thereby improving efficiency of the estimates. The condition index falls but still remains beyond the recommended level of 20-30.

I therefore re-estimate the model by dropping the variables that are most insignificant (EPS and

Clipart). Column 3 of Table 2-4 shows the results when the least significant variables are dropped from the model. The precision of the estimates improves slightly and the condition index drops further (from 57 to 44) and is closer to the acceptable level of 30.

The construction of the firm dummies also presented some difficulties due to the fact that some of the firms were bought or merged with others. In the results presented here, the firm that marketed the software under its name took on a value of one in the years that it was marketing it. For example in 1993 Aldus Inc. sold an application by name Aldus PageMaker. Aldus PageMaker was later bought by Adobe and was marketed as Adobe PageMaker. So for the year 1993, the Aldus dummy takes on a value of one for observation corresponding to versions of PageMaker. In later years, the Aldus dummy takes on the value of zero and the Adobe dummy takes on the value of 1 for versions of PageMaker. The best way to get around this difficulty would have been to construct software specific dummies. But including software dummies in the regression resulted in perfect collinearity due to

the characteristics of the data. Some of the software products appeared only once in the sample and resulted in the regressors becoming linearly dependent when software dummies were included. Since I could not control for fixed effects for software, I tried alternative constructions of the firm dummies to check for robustness of results. When any application changed hands, instead of reflecting that change in the firm dummies, I continued to attribute the application to the original firm. So in the context of the previous example, all observations corresponding to PageMaker would have the Aldus dummy equal to 1 and all other firm dummies equal to zero in all the years. The results are robust to this alternative construction of firm dummies.

The robustness of the results and the consistent negative effect on prices of the JAVA variable therefore, indicates that Java was indeed effective in increasing the potential for competition in the market for graphics software applications. It indicates that application providers not only recognized the potential for future competition due to the

introduction of Java but also regarded it as a serious enough threat to strategically lower their prices. Thus, by increasing the potential for competition in the applications markets, Java successfully reduced the indirect network effects in the PC-OS market.

Also since Java did not eliminate the need for an underlying operating platform, it might be argued that it did not present an alternative platform from consumer's point of view and therefore, could have resulted in creating greater competition in the operating systems market as well.

2.5. Conclusion

In summary, this chapter contributes to the empirical literature on network effects and provides an indirect test of their impact on welfare. The results from the hedonic price regressions suggest that the potential availability of a platform-independent software application helped to decrease the prices of all graphics software. The results therefore suggest that despite the demand-side and supply-side economies of scale that are thought to exist in the software markets in general, these markets are quite competitive.

The significant increase in the number of software applications available for the OS/2 platform after the introduction of Java and the reduction in prices of all graphics software suggest that Java was indeed successful in reducing the indirect network effects in the PC operating systems market. Given the hypothesized importance of indirect network benefits in the PC-OS market, the results might be interpreted to suggest that this market might also be subject to competition.

Future research may be directed towards devising a direct test of whether or not positive feedback is indeed present in the PC operating systems market and whether such feedback always leads to inefficient market outcomes.

III. RACIAL DIFFERENCES IN SELF-EMPLOYMENT

3.1. Introduction

Historic studies of the relative economic status of the minority blacks in the United States have mainly focused on the gains made by blacks in real wage earnings. The factors leading to the changes in real black earnings and their relative importance for black economic progress have been studied at great length in Smith and Welch (1989), Donahue and Heckman (1991). Fairlie and Sundstrom (1999) document the long run trend in black unemployment rates and the persistence of the racial gap in unemployment rates, using census data from 1880-1990. The historical evolution of racial differences in self-employment rates and self-employment incomes however, has received much less attention in comparison.

Overall self-employment rates in the United States were falling for most of the twentieth century before starting to rise in the 1970's. Blau (1987) presents an analysis of the reasons for the rise in self-employment rates since the early 70's. While black

self-employment rates also followed the same trend as white self-employment rates, the black-white gap in self-employment rates remained almost constant throughout this period as documented in Fairlie and Meyer (1999) and Fairlie (2000). Since self-employment was thought to provide an opportunity to earn higher rewards than in wage/salary employment, twentieth century black leaders encouraged self-employment as an important means of closing the racial gap in wealth (J.E.K. Walker). In spite of regarding self-employment as a means of achieving faster economic progress, blacks continued to have the lowest self-employment rates among all racial and ethnic groups even almost a century later, in 1990.

Fairlie (1996) discusses the various theories offered to explain the ethnic and racial patterns of self-employment and offers some explanations for the low self-employment rates observed among African Americans. While a number of theories have been proposed in the sociology literature, the most common explanations tested empirically in the economics literature have focused mainly on the effects of

discrimination from various sources. It has been argued that the combined effect of labor market disadvantages and discrimination by employers pushes many minorities into self-employment while discrimination from consumers lowers the individual returns to self-employment. Borjas and Bronars (1989) develop a model of consumer discrimination and its effect on minority self-employment rates and incomes. Using U.S. census data from 1980 and a two-step procedure, they estimate self-employment probabilities and selectivity corrected self-employment earnings. They conclude that as a result of consumer discrimination, more able blacks are more likely to self-select into wage/salary employment while more able whites are more likely to choose self-employment.

This chapter reexamines the issue by replicating the empirical analysis in Borjas and Bronars (1989) with data from the 1980 census and also by expanding the analysis to include data from the 1970 census for comparison. A discrete choice probit model similar to Borjas and Bronars (1989) is employed to estimate the effects of employer and consumer discrimination on

self-employment rates and on the returns to self-employment. The results for 1980 when compared to the results in Borjas and Bronars (1989) indicate that the empirical analysis in Borjas and Bronars (1989) may be inaccurate due to lack of information on sample construction. In my attempt to replicate the sample construction as was described in Borjas and Bronars (1989), I found that adhering to the exact rules for sample selection results in a much larger sample size than was reported in Borjas and Bronars (1989). I used a ten percent random sample of self-employed whites, instead of the 33 percent reported in Borjas and Bronars (1989), in order to get the same sample size as reported in Borjas and Bronars (1989). A probable reason for the discrepancy in results and sample construction is that the sample used in Borjas and Bronars (1989) was constructed from a census extract that was originally stratified by ethnic group and immigrant status and may not have been truly random.

3.2. Background Literature

In a world without discrimination and with a perfectly random process of selection, the representation of any group in entrepreneurial activity should be proportional to the size and distribution of its labor force. This was one of the hypotheses tested in an early study by Higgs (1976). Focusing on participation of blacks in retail merchandizing during the period 1890-1910, Higgs examined the extent of the difference in participation rates of blacks, immigrants and native whites and explored the role of urbanization in promoting such participation. He found that when urban and regional distributions of ethnic groups are controlled for, participation rates were lowest for blacks and highest for immigrants with native whites falling in between. This indicated that even though urbanization and southern locations played a role in shaping participation in the merchant class, they do not explain the major differences in participation rates of different ethnic groups. Non-availability of individual level data prevented further examination of the role of group differentials in skill, age-sex

structure and occupational discrimination in explaining the differences in participation rates.

Theoretical models predict that labor market discrimination affects the decisions of those discriminated against, to enter self-employment. Discrimination in the labor market lowers the expected wage and therefore the opportunity cost of self-employment. Moore (1983) argues that if employers were the only source of racial discrimination then we would expect self employment rates to be higher for minority groups and that the black/white (female/male) earnings ratio would be higher among self-employed workers than their wage and salary counterparts.

Discrimination by employers and/or fellow employees could create a greater incentive for minorities to enter self-employment. The entry of more Blacks into self-employment as a result of discrimination by employers lowers the average level of entrepreneurial ability of self-employed Blacks compared to self-employed Whites. This in turn may be expected to lower the average returns to self-employment for the

minority group. Discrimination by employers however, should not lower the individual returns to ability in the self-employed sector. But if a black entrepreneur also faces discrimination by consumers and/or workers then the individual returns to self-employment are reduced, thus lowering the incentives for blacks to enter self-employment.

Borjas and Bronars (1989) show that the differentials by race in self-employment rates and incomes can arise in markets with consumer discrimination and incomplete information. Their theoretical model predicts that the income distribution of self-employed Blacks has a lower mean than the income distribution of self-employed Whites. The reason is that Black entrepreneurs have to compensate their consumers for their taste for discrimination by charging a lower price. Their theoretical model also predicts that more able blacks are less likely to self-select into self-employment than able whites i.e., blacks tend to negatively select into self-employment while whites positively select into self-employment. This is because in their model, high ability blacks are

assumed to be more likely to want to integrate (i.e. sell to white consumers) and therefore face a higher opportunity cost of choosing self-employment if they face discrimination by consumers. Using 1980 census data they find that observed racial differences in self-employment rates and incomes are consistent with their theoretical predictions.

In this chapter, I attempt to replicate the results of the consumer discrimination study by Borjas and Bronars with data from the U.S. census for the years 1970 and 1980. While the differences across race in selectivity, is emphasized as a main result of consumer discrimination in Borjas and Bronars (1989), a contradictory finding of no differences in selectivity across race does not necessarily imply the absence of consumer discrimination. In this chapter, I use an alternative method of quantifying the impact of consumer discrimination on self-employment earnings. The effect of discrimination on the decision of blacks to enter self-employment and on their earnings is estimated using the standard Oaxaca decomposition method (Oaxaca, 1973).

3.3. Econometric Model Specification

The empirical model in this section is analogous to the problem of self-selection in migration analyzed by Nakosteen and Zimmer (1980) and is analytically identical to the model presented in Borjas and Bronars (1989). It consists of an equation describing an individual's decision to enter self-employment or wage/salary employment and two income equations (one for self-employment and the other for wage employment). If self employment and wage employment are competing alternatives, a rational risk-neutral individual seeking to maximize his or her utility will, *ceteris paribus*, maximize the present value of net gains from choosing self employment or wage employment.

Let the self-employment and wage income equations be respectively:

$$\ln (Y_{si}) = Z_i\beta_s + \varepsilon_{si} \quad (1)$$

$$\ln (Y_{wi}) = Z_i \beta_w + \varepsilon_{wi} \quad (2)$$

where Y_{si} is self-employment income and Y_{wi} is wage/salary income and Z_i is a vector of demographic

characteristics. The error terms ε_{si} and ε_{wi} are assumed to be normally distributed with variances σ_s^2 and σ_w^2 respectively.

Let L_i represent the costs of entering self-employment as a proportion of income and assume that L_i may be represented as a linear function of some demographic characteristics and a random error term.

$$L_i = X_i \alpha_2 + \varepsilon_I \qquad (3)$$

Let I^* be the latent variable that represents the net returns to self-employment. Then,

$$I^* = \ln (Y_{si}/ Y_{wi}) - L_i$$

$$= Z_i(\beta_s - \beta_w) - X_i\alpha_2 + \varepsilon_{si} - \varepsilon_{wi} - \varepsilon_i$$

$$= H_i \gamma + u_i \qquad (4)$$

where H_i is a vector of all of the model's variables and $u_i = \varepsilon_{si} - \varepsilon_{wi} - \varepsilon_i$.

Equations (1) - (4) describe the structural form of the model. The endogenous variables are I^*, Y_s and Y_w. We do not observe I^*, but only

$I_i = 1$ if $I_i^* > 0$

$I_i = 0$ if $I_i^* \leq 0$

In addition we observe Y_s only for individuals who choose self-employment, *i.e.*, when $I_i^* > 0$ and we observe Y_w only for individuals who choose wage employment, *i.e.*, when $I_i^* \leq 0$.

If we assume that the error term in (4) is distributed normally with unit variance then we can estimate this reduced form decision equation by maximum likelihood probit methods. The income equations (1) and (2) cannot be estimated by OLS due to the selectivity bias resulting from the endogenous nature of the decision to choose self-employment. The conditional means of the income disturbance terms are non-zero and not constant for all observations:

$$E(\varepsilon_{si} \mid I_i = 1) = Cov(\varepsilon_{si}, u_i)\ [\phi(\Psi_i) / \Phi(\Psi_i)] \qquad (5)$$

$$E(\varepsilon_{wi} \mid I_i = 0) = Cov(\varepsilon_{wi}, u_i)\ [-\phi(\Psi_i) / 1-\Phi(\Psi_i)] \qquad (6)$$

where $\Psi_i = H_i\gamma/\sigma_u$

Due to self-selection of individuals into wage and self-employment, OLS estimates of the income equations are inconsistent and biased. In order to account for the self-selection, the income equations are modified

by incorporating the selectivity variables and adding error terms with zero means.

The corrected income equations are as follows:

$$\ln(Y_{si}) = Z_i\beta_s + \theta_s \lambda_{si} + \upsilon_{si} \quad \{ i/ I_i = 1\} \qquad (7)$$

$$\ln(Y_{wi}) = Z_i \beta_w + \theta_w \lambda_{wi} + \upsilon_{wi} \quad \{ i/ I_i = 0\} \qquad (8)$$

where $E(\upsilon_{si} \mid I_i = 1) = 0$, $E(\upsilon_{wi} \mid I_i = 0) = 0$, and

$\lambda_{si} = \phi(\Psi_i) / \Phi(\Psi_i)$ for $\{ i/ I_i = 1\}$

$\lambda_{wi} = -\phi(\Psi_i) / 1-\Phi(\Psi_i)$ for $\{ i/ I_i = 0\}$, and

$\theta_s = Cov((\varepsilon_{si}, u_i)\ \sigma_u$ and $\theta_w = Cov((\varepsilon_{wi}, u_i)\ \sigma_u$

OLS estimates of (7) and (8) above are known to be consistent. The error terms are heteroscedastic and must be corrected to get efficient estimates.

The main result of Borjas and Bronars (1989) was that the coefficients on the selectivity variables λ, be positive for Whites and negative for blacks in the self-employed sector. As shall be described in the next few sections, this result is not obtained in our study.

The first stage of estimation involves the maximum likelihood probit estimation procedure of the reduced form decision equation (4). Fitted values from stage one are used to calculate the inverse mills ratios (λ) which are then used in the second stage OLS estimation of the earnings equations (7) and (8).

3.4. Data and Estimation

The data used in this study are drawn from the 1970 'form I metro' and 1980 1% 'metro B' samples from the Integrated Public Use Micro Data Series (IPUMS) of the U.S. Census. The samples are restricted to employed males so as to be able to focus only on the choice between self-employment and wage employment. The model is estimated separately for blacks and whites in order to allow for different rates of returns to ability. For both 1970 and 1980, the samples were restricted to black and white males between the ages 25-64, residing in metropolitan areas, and employed in non-agricultural industries. The 1970 and 1980 samples consist of all self-employed black workers and 33 percent of black wage/salary workers; 3.3 percent of

white wage earners and 10 percent of white self-employed workers. Due to the choice based sampling, the probit regressions are weighted[19] to reflect the sample composition. In order to match the individual specific data with the local labor market characteristics, the sample is restricted to the 75 largest standard metropolitan statistical areas (SMSAs).

The source of the SMSA-specific characteristics is the County and city data book (supplement to the statistical abstract of the U.S) in 1970 and the published Census summaries (chapters A-D) for each of the fifty states in 1980. Tables 3-1 and 3-2 contain the means of the regression variables in 1970 and 1980 respectively.

---

[19] The weights are calculated as follows. For self-employed (black/white) individuals, the weights are calculated as the ratio of the (black/white) self-employment probability in the population to the (black/white) self-employment probability in the sample. Similarly for wage/salary-employed individuals, the weights are calculated as the ratio of population proportion of wage-employed (blacks/whites) to the sample proportion of wage-employed (blacks/whites).

Table 3-1. Variable Means 1970

Variable Name	1970Self - Employed Black	1970Wage Employed Black	1970Self - Employed White	1970Wage Employed White
EDUC < 12	0.564	0.587	0.330	0.350
EDUC = 12	0.235	0.273	0.291	0.322
EDUC 16+	0.119	0.055	0.245	0.187
AGE	44.195	41.275	45.839	42.299
Immigrant	0.030	0.029	0.104	0.069
Years in USA (If Immigrant)	0.403	0.372	1.793	1.065
Married, Spouse Absent	0.168	0.161	0.022	0.028
Married, Spouse Present	0.728	0.733	0.953	0.922
Health Limits Work	0.128	0.073	0.091	0.073
NorthEast	0.225	0.274	0.297	0.279
NorthCentral	0.245	0.291	0.252	0.293
WEST	0.128	0.104	0.217	0.205
FRACTION BLACK	0.171	0.169	0.126	0.126
FRACTION HISPANIC	0.055	0.046	0.056	0.053
FRACTION WHITE	0.819	0.820	0.861	0.862

Notes: EDUC < 12 is a dummy variable taking the value of one if the individual had less than 12 years of schooling. EDUC =12 is dummy variable indicating 12 years of schooling and EDUC 16 + indicates more than 16 years of schooling. Immigrant is a dummy variable indicating immigrant status. Northeast, NorthCentral and West are dummy variables indicating region. FRACTION BLACK, FRACTION WHITE AND FRACTION HISPANIC are defined as the percentage of population in the metropolitan area that is Black, White and Hispanic respectively.

Table 3-2. Variable Means 1980

Variable Name	1980 Self-Employed Black	1980 Wage Employed Black	1980 Self-Employed White	1980 Wage Employed White
EDUC < 12	0.346	0.342	0.176	0.201
EDUC = 12	0.297	0.366	0.272	0.338
EDUC 16+	0.198	0.123	0.376	0.273
AGE	44.368	40.367	43.731	41.256
Immigrant	0.075	0.059	0.094	0.094
YEARS IN USA	1.070	0.783	1.599	1.471
Married, Spouse Absent	0.195	0.222	0.083	0.114
Married, Spouse Present	0.693	0.617	0.830	0.773
Health Limits Work	0.076	0.052	0.053	0.048
Veteran	0.418	0.381	0.486	0.478
NorthEast	0.238	0.243	0.264	0.269
NorthCentral	0.200	0.242	0.210	0.246
WEST	0.152	0.117	0.255	0.221
FRACTION BLACK	0.182	0.184	0.131	0.134
FRACTION HISPANIC	0.090	0.079	0.084	0.079
FRACTION WHITE	0.702	0.713	0.757	0.761

Notes: The variable descriptions are the same as in Table 3-1.

3.5. Results

The regression results for each of the census years from 1970 and 1980 are reported in Tables 3-3 through 3-6. The probit results show that the probability of being self-employed increases with age but at a decreasing rate. The coefficient on age (proxy for experience) is positive in the earnings equations and is consistent with the results in the labor supply literature. Married men are less likely to be self-employed compared to single men, whereas the probability of being self-employed increases with the wife's education level for married men.

College graduates were more likely to be self-employed than college dropouts, high school dropouts and high school graduates of both racial groups. White men with a college degree were about 11 percent more likely to be self-employed than college dropouts in both 1970 and 1980. Black men with a college degree were about 33 percent more likely to be self-employed in 1970 and only about 20 percent more likely to be self-employed than college dropouts in 1980. Neither immigrant

status nor work disability had a statistically significant effect on the probability of being self-employed for white men. On the other hand, black men with a work related disability were more likely to be self-employed in both census years. The variables that were included to measure "enclave" effects as described in Borjas & Bronars (1989) - fraction Black, fraction Hispanic and fraction White, are not statistically significant. This result is consistent with the results in Borjas & Bronars (1989). No explanations are offered in Borjas & Bronars (1989) for the inclusion of these variables and for their apparent insignificance. I argue that these variables might be one of the several measures of the 'network effect' in these decisions. The apparent statistical insignificance of these variables might be due to a specification error, as will be argued in the next chapter. The following chapter develops an extended model of self-employment choice that includes these variables along with other measures of 'network effects' and rejects the econometric model of self-employment probability as estimated in this chapter.

The results from the selectivity corrected earnings regressions show that the returns to age (proxy for experience) are higher for wage/salary earners than for self-employed individuals of either race. Married men earned more on average than single men in both self-employment and wage-employment. The coefficients on the enclave variables, fraction Black, Hispanic and White are not statistically significant in the earnings regressions either. The coefficient on the selectivity variable lambda is negative for both salaried and self-employed individuals of both races. This result is contrary to the results obtained by Borjas and Bronars (1989). The results for the black sample are qualitatively similar to Borjas & Bronars (1989) but the results for the white sample are contradictory.

The reasons for the discrepancies in the results are two-fold. First, the predictions of the theoretical model in Borjas and Bronars are based on two critical assumptions. One assumption is that black consumers do not discriminate against a seller of their own race.

As we shall see in the next chapter, this was not necessarily always true. Black entrepreneurs often faced discrimination from their own race for several reasons pointed out in the next chapter. Two, the predictions depend on the untested assumption that more able blacks are more likely to want to integrate and try to attract white customers. The high ability blacks would therefore, face greater losses due to discrimination from consumers and therefore have a greater disincentive to choose self-employment. This assumption is crucial to the main argument of negative selection of blacks into self-employment in Borjas and Bronars (1989).

A second source of the discrepancy in results could be due to the fact that the white sample was constructed differently in this paper. (A 10 percent random sample of self-employed whites was used as opposed to Borjas & Bronars (1989) 33% random sample in order to get the same sample size as reported in their paper!).

Table 3-3. Probit estimates 1970 [20]

The dependent variable takes on a value of 1 if the individual is self-employed and zero otherwise. (t-ratios in parentheses)

Variable Name	Blacks 1970	Whites 1970
Educ < 12	-.075 (-.767)	.021 (.281)
Educ =12	-.023 (-.230)	.019 (.261)
Educ 16+	.329 (2.689)	.111 (1.433)
Age	.061 (2.948)	.067 (3.529)
Age Squared	-.0005(-2.262)	-.0005 (-2.496)
Immigrant	.642 (.388)	-.152 (-.149)
Years in USA (if Immigrant)	-.084 (-.341)	.020 (.141)
Years in USA, Squared	.0027 (.327)	.0003 (.067)
Married, Spouse Absent	-.104 (1.041)	.028 (.146)
Married, Spouse Present	-.358 (-2.265)	-.466 (-2.596)
Wife's Education	.022 (1.872)	.048 (4.641)
Health Limits Work	.252 (3.072)	.099 (1.253)
Fraction Black	5.216 (1.142)	-.066 (-.070)
Fraction Hispanic	.757 (1.451)	-.306 (-.694)
Fraction White	4.999 (1.092)	-.529 (-.618)
Region: NorthEast	.058 (.388)	.178 (1.578)
Region: NorthCentral	.026 (.201)	.110 (1.120)
Region: West	.181 (.851)	.149 (1.123)
# of children	.0157 (1.046)	.039 (2.503)
Income of other family members	-.58e-5 (-.846)	-.77e-5 (-1.357)
Constant	-7.728 (-1.657)	-2.082 (-2.063)

[20] The regressions also included the crime rate, median education level, median income level, per-capita government expenditure and unemployment rate in the SMSA. The 1980 regression also included a dummy for the individual's veteran status.

Table 3-4. Probit estimates 1980

The dependent variable takes on a value of 1 if the individual is self-employed and zero otherwise. (t-ratios in parentheses)

Variable Name	Blacks 1980	Whites 1980
Educ < 12	-.061 (-.726)	.005 (-.075)
Educ =12	-.069 (-.891)	-.059 (-.995)
Educ 16+	.194 (2.156)	.118 (1.958)
Age	.065 (2.981)	.049 (2.859)
Age Squared	-.0005 (-2.157)	-.0004 (-2.014)
Immigrant	.301 (.300)	.354 (.426)
Years in USA (if Immigrant)	-.045 (-.301)	-.076 (-.649)
Years in USA, Squared	.002 (.381)	.003 (.802)
Married, Spouse Absent	-.055 (-.595)	-.178 (-1.894)
Married, Spouse Present	-.435 (-2.358)	-.667 (-4.250)
Wife's Education	.039 (2.998)	.055 (5.372)
Health Limits Work	.169 (1.668)	.060 (.663)
Fraction Black	3.663 (1.512)	-.585 (-676)
Fraction Hispanic	3.377 (1.401)	-.144 (-.174)
Fraction White	3.067 (1.302)	-.144 (-.191)
Region: NorthEast	.118 (.843)	-.026 (-.252)
Region: NorthCentral	.046 (.347)	-.082 (-.852)
Region: West	.284 (1.653)	-.043 (-.438)
# of children	.017 (.859)	.018 (.987)
Income of other family members	-.37e-5 (-1.192)	-.79e-5 (-3.147)
Constant	-6.944 (-2.781)	-2.713 (-3.155)

Table 3-5. Earnings OLS Regression Estimates 1970

The dependent variable is natural log of weekly earnings. (t-ratios in parentheses)

Variable Name	Self-Employed Blacks	Self-Employed Whites	Wage/salary Employed Blacks	Wage/salary Employed Whites
Educ < 12	-.212 (-2.15)	-.305 (-5.39)	-.256 (-9.69)	-.236 (-9.79)
Educ =12	-.119 (-1.15)	-.206 (-3.72)	-.114 (-4.41)	-.076 (-3.26)
Educ 16+	.486 (2.98)	.269 (4.24)	.278 (5.33)	.214 (7.16)
Age	.041 (1.46)	-.006 (-.31)	.033 (4.65)	.032 (4.39)
Age Squared	-.0006 (-1.95)	-.00001 (-.07)	-.0003 (-4.57)	-.0004 (-4.52)
Immigrant	-.502 (-.39)	.086 (.16)	-.081 (-.17)	.074 (.17)
Years in USA if Immigrant	.014 (.07)	-.036 (-.47)	-.002 (-.03)	-.051 (-.85)
Years in USA, Squared	.001 (.14)	.001 (.35)	.0002 (.09)	.002 (.94)
Married, Spouse Absent	.242 (2.03)	.0235 (.16)	.079 (2.31)	.129 (2.23)
Married, Spouse Present	.484 (4.86)	.244 (2.03)	.224 (7.48)	.256 (7.50)
Health Limits Work	-.314 (-2.48)	-.226 (-3.56)	-.050 (-1.26)	-.137 (-3.75)
Fraction Black	-12.173 (-2.45)	-.608 (-1.18)	-.239 (-.80)	1.299 (1.08)
Fraction Hispanic	.557 (.91)	.344 (1.29)	-.098 (-.51)	-.118 (-.88)
Fraction White	-12.05 (-2.46)	-1.319 (-3.11)	-.445 (-1.67)	.670 (.58)
Region:NorthEast	.327 (3.50)	.109 (2.05)	.182 (6.70)	.114 (4.25)
Region:NorthCentral	.423 (4.82)	.239 (4.25)	.247 (9.88)	.212 (8.65)
Region: West	-.082 (-.44)	.061 (4.25)	.235 (5.86)	.214 (4.69)
Constant	16.297 (2.94)	8.124 (10.903)	4.389 (14.130)	3.301 (2.820)
LAMBDA	-.326 (-.89)	-.843 (-5.30)	.366 (.90)	-1.289 (-6.52)

Table 3-6. Earnings Regression Estimates 1980

The dependent variable is natural log of weekly earnings. (t-ratios in parentheses)

Variable Name	Self-Employed Blacks	Self-Employed Whites	Wage/Salary Employed Blacks	Wage/Salary Employed Whites
Educ < 12	-.343 (-3.771)	-.190 (-3.407)	-.212 (-8.341)	-.279 (-10.220)
Educ =12	-.204 (-2.399)	-.067 (-1.353)	-.091 (-3.905)	-.627 (-2.881)
Educ 16+	.159 (1.268)	.202 (3.928)	.169 (5.470)	.143 (5.703)
Age	.022 (.672)	.031 (2.005)	.039 (5.238)	.035 (5.050)
Age Squared	-.0004 (-1.050)	-.0003 (-2.082)	-.0004 (-5.440)	-.0003 (-4.566)
Immigrant	.751 (.660)	-.002 (-.003)	.159 (.525)	-.699 (-1.717)
Years in USA if Immigrant	-.099 (-.585)	.009 (.087)	-.052 (-1.154)	.078 (1.424)
Years in USA, Squared	.002 (.316)	-.001 (-.186)	.002 (1.497)	-.002 (-1.293)
Married, Spouse Absent	.138 (1.087)	.373 (4.398)	.113 (3.933)	.159 (4.532)
Married, Spouse Present	.181 (1.675)	.407 (5.808)	.211 (8.300)	.221 (8.383)
Health Limits Work	-.316 (-2.295)	-.253 (-3.696)	-.198 (-4.603)	-.142 (-3.844)
Fraction Black	-1.407 (-.482)	.233 (.310)	-.236 (-.795)	-.101 (-.321)
Fraction Hispanic	-2.119 (-.735)	.121 (.173)	-.865 (-2.985)	-.667 (-2.213)
Fraction White	-1.931 (-.713)	-.233 (-.359)	-.439 (-1.736)	-.452 (-1.649)
Region: NorthEast	.136 (1.600)	.142 (3.183)	.135 (5.821)	.074 (3.274)
Region:NorthCentral	.219 (2.423)	.257 (5.177)	.254 (11.136)	.186 (8.539)
Region: West	.017 (.085)	.063 (.934)	.169 (4.465)	.061 (1.837)
Constant	8.239 (2.437)	6.616 (8.236)	4.835 (15.681)	4.972 (16.757)
LAMBDA	-.582 (-1.641)	-.939 (-6.471)	-1.326 (-3.618)	-1.561 (-7.881)

Table 3-7 below shows the results of the decomposition of the earnings in the self-employed and wage-employed sectors. Since the earnings equations are corrected for selectivity, the decomposition of wage differentials contains a component that measures the differential due to self-selection[21] of individuals into these categories.

Table 3-7. Decomposition of Earnings Differential

	Self-Employed Sector 1980	Wage-Employed Sector 1980	Self-Employed Sector 1970	Wage-Employed Sector 1970
Actual ln(Earnings) Differential	.492	.349	.644	.429
ln(Earnings) Differential due to differences in endowments	.076	.083	.109	.096
Unexplained Differential	.826	.095	1.284	.026
Differential due to Selection	-.410	.172	-.749	.307
Unexplained differential as a % of Actual wage gap	197.89%	27.22%	199.37%	6.06%

[21] The gap due to selectivity in the self-employed sector is calculated as $(\theta_{sw}\ \lambda_{sw} - \theta_{sb}\ \lambda_{sb})$ where the subscript 'w' and 'b' refer to the race of the individual. The corresponding selectivity gap in the wage-employed sector is defined analogously.

The estimates show that the discrimination faced by wage/salary earners (employer/employee discrimination) increased over the decade of 1970-80, as shown by the percent of the gap that remains unexplained in the wage/salary sector (roughly 6 percent in 1970 and 27 percent in 1980). The discrimination faced by self-employed individuals (from consumers and/or workers) seems to have decreased over the decade (roughly 200 percent of the gap in self-employment earnings is unexplained by observed characteristics in 1970 versus only about 167 percent remains unexplained in 1980). The unexplained differential traditionally attributable to discrimination is greater than the actual difference in earnings due to the negative effect of selection.

The interpretation of the selectivity component is complicated and raises several interpretational issues as discussed in Neuman and Oaxaca (1999). The selectivity component may be further decomposed into three terms - one representing the effect of the racial differences in the parameters of the probit

selectivity equation on the earnings gap. A second component representing the effect of the racial differences in the (magnitude of) variables that determine self-employment status and a third component that reflects the effect of racial differences in the correlation between the selectivity equation error term and the earnings equation error term as well as the differences in earnings variability. Allocating these different components to discrimination and differences in endowments is tricky as described in Oaxaca and Neuman (2001). Five different methods of allocating these components to discrimination and endowments are discussed. Each of the methods allocates the earnings gap due to differences in the probit parameters to discrimination. This is problematic because the differences in the probit parameters reflect not only the differences in both the supply side and demand side valuations of personal characteristics they also reflect the differences in the variance of the error term in the probit equation. Since it is not clear whether such differences should be classified as discriminatory, an interpretation of the selectivity component is not attempted here.

Finally, to be able to quantify the extent of the effect of discrimination on self-employment probabilities, we use the white coefficients from the probit regression to predict the black self-employment and compare with the actual observed rates. Following Borjas and Bronars (1989) the predicted probability for blacks is calculated as follows:

$$P = \{\sum_i \Phi(Z_i \pi_w)\} \; / \; N$$

where Z_i is the vector of variables(levels for blacks) included in the probit and π_w is the vector of estimated probit coefficients in the white sample and N is the (black) sample size.

The difference between the predicted probability of self-employment for Blacks using white coefficients and the predicted probability using black coefficients is the gap that cannot be attributed to differences in the average productivity of the two groups.

The predicted probabilities are shown in table 3-8.

Table 3-8. Decomposition of Probability of Self-Employment

	1970	1980
Predicted Black Probability using White coefficients	9.62%	9.29%
Predicted Black probability using Black coefficients	4.24%	4.37%
Predicted White probability using White coefficients	11.25%	11.95%

As the above table shows, the difference in the observed probabilities of self-employment between blacks and whites seems to be largely due to the difference in the process of selection rather than the differences in the endowments of the two groups. If blacks faced the same selection process as whites, their predicted self-employment probability is much closer to that of whites. While 77 percent (calculated as (11.1-9.62)/(11.1- 4.75)) of the gap in self-employment rates was unexplained in 1970, about 73.5

percent of the difference in 1980 remained unexplained. The gap in the self-employment rates that remains after controlling for differences in endowments can be regarded as a combined measure of consumer discrimination and worker discrimination against minority entrepreneurs. The estimates show that this discrimination may have declined marginally over the decade of 1970-80.

3.6. Conclusion

The results do not support the main conclusions of Borjas and Bronars (1989) about differences across race in the type of selection. The main result in Borjas and Bronars (1989) is that when Black entrepreneurs face discrimination from consumers, the more able Blacks tend to choose wage/salary employment instead of self-employment. This result did not hold up when the study was replicated. My results indicate that Blacks experience qualitatively similar selection as Whites in both the self-employed and the wage/salary employed sectors. No difference in selectivity is found. The main source of this discrepancy appears to be due to the difference in the

construction of the samples. While Borjas and Bronars (1989) claim that the samples used in their analysis were random samples constructed by them, in an e-mail to this author they admit that the samples used were already stratified by immigrant status by another researcher and the rules of construction described in the paper may not reflect the actual construction of the samples.

While the differences across race in selectivity, is emphasized as a main result of consumer discrimination in Borjas and Bronars (1989), the contradictory finding of no differences in selectivity across race does not necessarily imply the absence of consumer discrimination. In this chapter, I use an alternative method of quantifying the impact of consumer discrimination on self-employment earnings. Using a standard Oaxaca decomposition method, the racial gap in self-employment earnings is decomposed into components reflecting differences in endowments and differences in the returns to individual characteristics attributable to discrimination by

consumers and/or workers. The results suggest that blacks face considerable discrimination from consumers and/or workers, which affects their decision to enter self-employment. About 83 percent (in 1970 and about 85 percent in 1980) of the actual earnings differential in the self-employment sector cannot be attributed to differences in endowments of the two groups.

Thus, a comparison of the results in this chapter with the results in Borjas and Bronars (1989) not only helps to identify some limitations and inaccuracies in the empirical analysis of Borjas and Bronars (1989) but also suggests that the model of consumer and employer discrimination might be incomplete. The next chapter extends the model to include several different measures of the individual's social network and attempts to identify the role of social networks in promoting self-employment while simultaneously accounting for several different explanations of the observed ethnic/racial differences in self-employment.

IV. RACIAL DIFFERENCES IN SELF-EMPLOYMENT RATES: THE ROLE OF SOCIAL NETWORKS

4.1. Introduction

The overall self-employment rate in the United States fell dramatically for most of the 20th century before it started to rise in the 1970's. Self-Employment rates differ substantially across ethnic/racial groups with blacks having the lowest self-employment rates. Fairlie and Meyer (1999) and Fairlie (2000) document the historic and recent trends in the overall self-employment rate as well as the trend in the racial gap in self-employment rates. Using data from the U.S. Census, they find that the recent upturn in the self-employment rates among African-Americans did not reduce the racial gap in self-employment rates. Instead, they find that the gap remained almost constant throughout the century. Using a simple inter-generational model, they predict that in the absence of continuing forces holding down black self-employment, black-white rates would converge quickly. Since self-employment has long been considered as an important means of closing the racial gap in wealth,

it is important to study the reasons for the persistence of this gap in self-employment rates. Several studies in the economics and sociology literature have tried to identify the possible reasons for the differential rates of self-employment across ethnic and racial groups. However, no single theory has been found satisfactory in explaining all the observed variation in self-employment rates across ethnic/racial groups.

Recent years have seen a growing interest among economists in the impact of social interactions on labor market outcomes. Network effects that occur either through social interactions or due to information problems have been shown to lead to "herd" behavior and "excess inertia" under certain circumstances (See Banerjee 1992, Bernheim 1994). It has been argued that network effects might explain phenomena such as the persistence of high levels of unemployment among some ethnic minorities and the persistence of ghettos. Social networks/interactions have also been shown to affect an individual's ability to obtain employment (Holzer 1987, Reingold 1999).

This chapter examines the role of social networks in explaining the decisions of individuals to enter self-employment. Social networks are defined in the sociology literature as an individual's network of friends, family and neighbors. There are several ways in which an individual's social network might affect his/her decision to become self-employed. First, self-employment requires access to start-up capital. Some of the possible sources of capital for starting a small business include family assets, loans from friends and/or family and loans from banks. Access to these sources depends heavily on the width and breadth of one's social network. Access to bank loans ideally should not depend on one's social network but in practice, such access might indirectly be affected if social interactions are a major means of sharing information about them.

Second, if self-employment is considered as an alternative to unemployment, then social networks might indirectly affect an individual's decision to become self-employed as it affects their access to

wage/salary employment through job referrals. Third, social networks might influence decisions to become self-employed as they provide individuals an opportunity to share and learn from the experiences of others. Several authors such as Du Bois (1899)[22], Cayton and Drake (1946), Frazier (1957) have emphasized that past inexperience in business was an important reason for the low rates of business ownership among blacks. Social networks might therefore be able to explain the persistence of the racial gap in self-employment. Allen (2000) examines the role of social networks in self-employment decisions and argues that it might explain gender differences in self-employment rates. However, there are no studies that empirically examine the role of

---

[22] "They [small minority businesses] are all alike hindered by three great drawbacks. First, the Negro never was trained for business and can get no training now; it is very seldom that a Negro boy or girl can on any terms get a position in a store or other business establishment where he can learn the technique of the work or general business methods. Second, Negro merchants are so rare that it is natural for customers, both white and colored, to take it for granted that their business is poorly conducted without giving it a trial. Third, the Negroes are unused to co-operation with their own people and the process of learning it is long and tedious." - W.E.B.

networks in explaining racial differences in self-employment decisions.

This chapter presents an empirical analysis to test whether or not social networks indeed play any role in an individual's decision to become self-employed and whether they might explain the persistent racial gap in self-employment rates.

4.2. Background Theory and Literature

Self-Employment, entrepreneurship or small business formation as it is variously described, has been the subject of several lines of recent research. Lucas (1978) and Kihlstrom and Laffont (1979) describe entrepreneurial choice as consequences of differing managerial abilities and risk attitudes respectively. Evans and Jovanovic (1989) and Holtz-Eakin, Joulfaian and Rosen (1994) examine the importance of liquidity constraints on entry into self-employment and the eventual survival of the enterprise. Various hypotheses have been proposed to explain the wide

---

Du Bois in 'The Philadelphia Negro' Chapter IX, pp122-123.

variation in self-employment rates across ethnic communities. The sociology literature provides some empirical and anecdotal evidence to support a host of theories that attempt to explain the ethnic and racial variation in self-employment rates (See Aldrich and Waldinger 1990). While each of the theories suggested are found to be consistent with the patterns of self-employment for some ethnic groups, they do not explain all the observed patterns for all ethnic/racial groups (See Fairlie 1996 for a discussion of the theories and their success/failure in explaining the self-employment patterns across various ethnic groups).

The following is a brief outline of the factors that have been hypothesized as important in encouraging entry into self-employment. These may be classified into two broad types - "push" factors and "pull" factors. The push factors include disadvantages in the labor market either due to a language problem (lack of fluency in English), health problems leading to work disabilities or due to discrimination by employers against minorities. All these factors lead to decreased opportunities in wage/salary employed

employment thereby pushing these individuals into self-employment. Bates (1998) argues that self-employment is often a form of under-employment for many immigrant Asians who are disadvantaged in the labor market, and provides some empirical evidence for the disadvantage theory among immigrant Asians.

The theories that emphasize "pull" factors include the following: a) The special demands theory states that ethnic minorities are better able to serve the unique and special tastes of their own ethnic groups. Therefore individuals belonging to any particular group will be more likely to be self-employed if there are a large number of co-ethnics residing in their neighborhood.

b) The special opportunity theory states that some ethnic groups are more likely to be self-employed than others as they might have a special advantage, such as having access to cheap labor from family members or fellow ethnics and/or access to loans from rotating credit associations for start-up capital from fellow ethnics. Once again this theory also predicts that individuals residing in areas with high concentrations

of fellow ethnics will be more likely to be self-employed

c) In the sojourner theory, immigrants are more likely to be self-employment because they aim to make/save more money and eventually go back to their native country.

d) Immigrants from countries with high rates of self-employment tend to choose self-employment. Borjas (1986), Light and Sanchez (1987), Fairlie and Meyer (1996) find that immigrants tend to have higher self-employment rates than native whites.

e) Experience sharing individuals learn of opportunities and share experiences through social interactions within their group; therefore, *ceteris paribus,* individuals belonging to groups with high self-employment rates are more likely to be self-employed than individuals belonging to groups with low self-employment rates.

Several factors may also be identified that discourage entry of ethnic and racial minorities into self-employment. Discrimination by consumers against minority businesses (Borjas and Bronars 1986),

discrimination by workers against minority employers, and discrimination by creditors against minority entrepreneurs (Blanchflower, Levine and Zimmerman 1998) are some factors that are identified as discouraging entry into self-employment. The economics literature has mainly focused on the differences between returns to self-employment and wage employment as the major explanatory factor while taking into account the effects of possible discrimination from various sources. Borjas and Bronars (1989), using 1980 U.S Census data, find that more able whites enter self-employment whereas more able blacks choose wage employment instead of self-employment. They argue that this is a consequence of discrimination by consumers against minority business owners. However, their results do not hold up to a careful replication as shown in chapter 3 of this dissertation. Blanchflower, Levine and Zimmerman (1998) present evidence on discrimination in credit markets. Dewey (1952) and Fishback (1992) provide some anecdotal and empirical evidence that workers exhibited distaste for working under a 'black boss' using data from the early part of the twentieth century.

Several studies have examined the role of social interactions, and social learning in various settings. Benabou (1993) shows that when social interactions that occur within a neighborhood affect human capital investment decisions, it can lead to large neighborhood discrepancies in the level of human capital formation. Glaeser, Sacerdote and Scheinkman (1996) show that social interactions might be an important factor for explaining the incidence of crime in neighborhoods. Allen (2000) examines the role of social networks in self-employment decisions and argues that it might explain gender differences in self-employment rates.

4.3. Model

I employ a discrete choice model that incorporates the impact of social interactions as in Brock and Durlauf (2000), to identify the role of social networks in self-employment decisions while simultaneously accounting for the various hypotheses described in the last section.

Consider the choice of individuals to be either self-employed or employed for a wage/salary. An individual's utility from choosing any given alternative is assumed to depend on the individual's own characteristics, the characteristics of the neighborhood that the individual resides in and on the average choice of other individuals in the neighborhood.

$$U_i\ (SE) = \alpha_{se} X_i + \beta_{se} Y_{n(i)} + S(P_{se}) + \varepsilon_{i,se}$$

$$U_i\ (WE) = \alpha_{we} X_i + \beta_{we} Y_{n(i)} + S(P_{we}) + \varepsilon_{i,we}$$

$U_i(SE)$ denotes the individual's utility from choosing self-employment and $U_i(WE)$ denotes the utility from choosing wage employment. X_i is a vector of individual characteristics and $Y_{n(i)}$ is a vector of characteristics describing the individual's neighborhood (which means the set of other individuals who influence *i* through interactions). S(.) denotes the social component of the individual's utility and is a linearly increasing function of the proportion 'P' of all individuals in the neighborhood that choose the same alternative. 'ε' is a random utility component that is assumed to be independently and identically distributed according to the double exponential distribution. Given these assumptions, and that an individual will choose self employment only when $U_i(SE) \geq U_i(WE)$, the probability that (s)he is self-employed is given by the following expression.

$$\text{Prob}_{se}=\frac{\exp((\alpha_{se}X_i+\beta_{se}Y_{n(i)}+S(P_{se}))}{\exp((\alpha_{se}X_i+\beta_{se}Y_{n(i)}+S(P_{se}))+\exp((\alpha_{we}X_i+\beta_{we}Y_{n(i)}+S(P_{we}))}$$

and the probability of choosing wage employment is given by

$$\text{Prob}_{we}=\frac{\exp((\alpha_{we}X_i+\beta_{we}Y_{n(i)}+S(P_{we}))}{\exp((\alpha_{se}X_i+\beta_{se}Y_{n(i)}+S(P_{se}))+\exp((\alpha_{we}X_i+\beta_{we}Y_{n(i)}+S(P_{we}))}$$

P_{se} is the proportion of all individuals in the neighborhood that have chosen self-employment and P_{we} is the proportion that has chosen wage/salary employment. As is standard with logistic models, the parameters for both the choices cannot all be identified independently. We therefore, normalize the parameters of the choice of wage/salary employment to zero, so that the probabilities of choosing self-employment and wage/salary employment are now given by:

$$\text{Prob}_{se} = \frac{\exp((\alpha_{se} X_i + \beta_{se} Y_{n(i)} + S(P_{se}))}{1 + \exp((\alpha_{se} X_i + \beta_{se} Y_{n(i)} + S(P_{se}))} \quad \text{and}$$

$$\text{Prob}_{we} = \frac{1}{1 + \exp((\alpha_{se} X_i + \beta_{se} Y_{n(i)} + S(P_{se}))}$$

The self-employment probability may be estimated by the standard logistic regression procedure where the dependent is a dummy variable indicating whether the individual is self-employed or not. The independent variables in X_i include individual characteristics such as age, experience, education, work related disabilities, immigrant status etc., that are thought

to influence the choice between self-employment and wage/salary employment. The vector $Y_{n(i)}$ contains neighborhood level characteristics such as the median income level, median education level, per-capita government expenditure, unemployment level in the neighborhood etc. The neighborhood characteristics for all individuals within the same neighborhood remain the same while the individual characteristics are different for each individual.

Finally, I assume a linear functional form for the social component of the utility $S(P_{se})$. The self-employment level in the neighborhood serves as a measure of the size of the social network. This measure, while is the same for all individuals within the same neighborhood, it varies across neighborhoods.

In general, this model is identified as long as the vector of neighborhood characteristics and the vector of individual characteristics do not contain regressors that are linearly dependent and the mean self-employment level varies across neighborhoods.

Also I assume that the choice of neighborhood is exogenous and that the group of individuals that influence an individual's choice through social interactions is known precisely.

4.4. Data and Econometric implementation

The data used are extracted from the one-percent samples of the Integrated Public Use Micro data Series (IPUMS) of the U.S. Census for the years 1940-1980[23]. I restrict our sample to employed males between the ages of 16-64 and residing in the 75 largest metropolitan areas and working in non-agricultural industries. The samples constructed in each census year are stratified by self-employment status and by race - the samples of 1940, 1970 and 1980 consist of ten percent of all whites who were self-employed, 3.3 percent of all whites who were employed for wage/salary, 33 percent of all blacks who were employed for wage/salary and all blacks who were self-employed. Due to the choice-based sampling, the logit regression is weighted

[23] 1940 1% sample, 1950 1% sample, 1970 form I metro sample, 1980 1% metro B sample.

appropriately[24] to ensure that the predicted probabilities reflect the population self-employment rates. The sample in 1950 is constructed slightly differently. The 1950 white sample consists of 33.33 percent of all self-employed white males and 10 percent of wage/salary employed white males whereas the black sample is unweighted and includes all self-employed and wage-employed black males that were asked to report their level of education.[25]

In the econometric implementation of the model specified above, the individual characteristics controlled for include age, education, immigrant

---

[24] The weights are calculated as follows. For self-employed (black/white) individuals, the weights are calculated as the ratio of the (black/white) self-employment probability in the population to the (black/white) self-employment probability in the sample. Similarly for wage/salary-employed individuals, the weights are calculated as the ratio of population proportion of wage-employed (blacks/whites) to the sample proportion of wage-employed (blacks/whites).

[25] In 1950, the question on level of education was only asked of only one individual from each household selected. The sample of individuals for whom this variable is included was therefore roughly a flat sample of 1-in-330. See IPUMS documentation for more detail. Special sample -line weights were used to account for this in 1950.

status, work disability status[26], marital status, number of children, wife's current employment status, wife's current self-employment status and wife's education[27]. The other characteristics help to test various theories of self-employment determination. For example, immigrant status if positive and statistically significant would offer support for the 'sojourner theory' and the disadvantage theory mentioned in the previous sections.

The neighborhood variables correspond to the metropolitan area[28] of the individual's residence and include the median education[29], median annual family

---

[26] Information on whether or not the individual suffered from any work-related disability was available only in 1970 and 1980 samples.

[27] Since only one person in the household was asked to report the level of education in 1950(see earlier footnote 3), and the selected unit of observation in our sample is employed males, wife's education level was unavailable for our sample in 1950.

[28] The metropolitan area is the smallest geographical area that can be identified in the census samples used. The data corresponding to the 1960 census year were not included in the analysis as the public use samples available for that year identified only the state of residence. For purposes of comparison and to avoid too much aggregation, it was decided not to include this year in the analysis.

[29] The source of the metropolitan area characteristics is the county and city data book (supplement to the

income, the overall crime rate[30], overall unemployment rate in metropolitan area, unemployment rate of the individual's own racial group, per-capita government expenditure in the metropolitan area[31], the proportion of the population in the metropolitan area that is Black, White, Chinese, Japanese and other Asians[32]. These neighborhood characteristics allow us to verify which of the various roles of social networks as

---

statistical abstract of the U.S) of 1940, 1950 and 1970 and the published census summaries (chapters A-D) for each of the fifty states in 1980.

[30] The overall crime rate was available for census years 1970 onwards and therefore could not be controlled for in the earlier years.

[31] The per-capita government expenditure data for 1940 is derived from the statistical abstract of the United States information on city government finances. They pertain to the major city in metropolitan area and therefore underestimate the government expenditures in the metropolitan area. The information provided in the statistical abstract pertains to the year 1936 and the total governmental cost payments, and operation, maintenance and interest expenses. Per-capita expenditure was calculated as (Total governmental cost payments - operation, maintenance and interest costs)/population in metropolitan area. Population in metropolitan area was taken from the published national summary of the decennial census of population.

[32] It was not possible to control for Hispanic origin, in all the years as this information was not consistently available in the earlier census samples of 1940 and 1950.

outlined in the previous section, are supported by the data. For example a finding that the probability of an individual choosing self-employment increases with the proportion of individuals in the area belonging to the same racial/ethnic group would support the special demands hypothesis, special opportunity hypothesis and the consumer discrimination hypothesis.

Finally, I control for the Black, White and Asian self-employment rates in the metropolitan area. If it is found that the self-employment rate of the individual's own group is statistically significant and positive, it might indicate support for the experience sharing hypothesis. The regressions also included industry dummies in order to control for any industry specific factors that might promote self-employment. When industry is not controlled for, most of the coefficients on individual characteristics and the network size have consistent signs even though they are smaller in magnitude. But the neighborhood variables, such as the per-capita government expenditure, median income etc., are insignificant and the crime rate has a counter-intuitive positive and

significant coefficient. When industry is controlled for however, crime rate is not statistically significant and some of the other neighborhood variables have the predicted signs and are significant. A likelihood ratio test conducted, rejects the restricted model without the industry dummies.

Identification Issues: As with other studies that attempt to analyze the role of social networks, this study suffers from several issues of identification. The first issue of identification arises because we do not know which group is relevant for defining the social interaction structure. When the relevant group is unknown, its structure has to be inferred from the data along with the strength of the interactions within the group. This presents an identification problem that is difficult to resolve. As outlined in Manski (2000) and elaborated by Jencks and Mayer (1990), when an individual's choice is found to be correlated with the mean choice of the group, it cannot be identified whether such correlation is due to social learning effects or simply due to some other

unobservable characteristics that are common to all individuals in that group. Therefore, when the relevant group structure is unknown, whether any inferences can be drawn is not known.

Various studies have used and/or suggested some ways of circumventing this difficulty. Each of the methods involves using extraneous information to identify the relevant group of interactions so that the group effect may be identified. Brock and Durlauf (2000) suggest an attempt to record and identify respondent's perspectives on which groups matter as a way of getting around this issue. Woittiez and Kapteyn (1998) use information elicited from a survey about the respondent's reference group and compare to reference groups that were constructed ex-post by the authors based on similarity of individual characteristics. They find that even though the two groups were highly correlated, the groups constructed based on survey responses performed better in predicting the labor supply behavior of females. Brock and Durlauf (2000) also note that under circumstances when the relevant group for interactions is unknown but we are

interested in the interaction effects of particular groups, this analysis might be applied but subject to accounting for the effects of using groups which only approximate actual interaction environments. In this case since we do not know whether the metropolitan area is the relevant neighborhood, the model suffers from the identification problem. However, since the data do not allow identification of smaller neighborhoods, we use the relevant racial group in the metropolitan area as a first approximation to the relevant reference group. The identification of smaller neighborhoods may only be crucial for small retail or grocery stores. The choice of the metropolitan area as the relevant neighborhood therefore, may be defensible as a good enough approximation for most businesses.

A second but related issue of identification arises when there is endogenous self-selection of individuals into neighborhoods. Evans, Oates and Schwab (1992) show that when endogeneity of the choice of peer group is accounted for, the estimated peer group effects disappear. Once again several papers have used various methods to overcome the difficulty of identification.

In studying the role of social interactions on unemployment, O'Regan and Quigley (1996) focus only on teenagers since their choice of location is most likely to be dictated by the choices of their parents and therefore may be assumed to be exogenous. In this chapter, since the relevant group is defined at aggregated at the level of the metropolitan area, the problem of endogenous choice of location may be reasonable assumed to be not as serious as when the choices are aggregated at the neighborhood level.

A final comment about identification in the specified model- the principal measure of the social learning effect in this model is the self-employment rate of the individual's own group in their neighborhood (metropolitan area). This is, however, also a measure of potential competition for the individual considering self-employment. It is not possible to separately identify the effects of competition and of social networks in this framework. However, since the two effects are expected to be opposite in direction, it is possible to identify the dominant effect. If the coefficient on this measure were positive, it would

indicate that the network effect dominated the competition effect and if it were negative, it would indicate that the competition effect is the dominant effect.

4.5. Results

Tables 4-1 below shows the descriptive statistics in each of the census years considered.

Table 4-1 Variable Means for employed Black males

VARIABLE/CENSUS YEAR	1940	1950	1970	1980
Years of Schooling	6.83	7.62	10.19	11.96
Age	40.63	41.07	40.26	39.58
Immigrant	0.03	0.03	0.03	0.06
Married, Spouse Absent	0.09	0.09	0.11	0.16
Married, Spouse Present	0.86	0.88	0.82	0.71
Wife's Education (in years)*	6.59	N.A	8.96	8.63
Wife is Employed*	0.22	0.29	0.43	0.46
Wife is Self-Employed*	0.01	0.01	0.01	0.01
No. of children	1.34	1.21	1.70	1.30
Health limits work	N.A	N.A	0.08	0.05
Northeast	0.27	0.26	0.26	0.23
Midwest	0.28	0.31	0.28	0.24
West	0.03	0.07	0.11	0.13

Variable Means for employed White males

VARIABLE/CENSUS YEAR	1940	1950	1970	1980
Years of Schooling	9.28	10.36	12.05	13.23
Age	41.58	41.36	41.24	40.25
Immigrant	0.23	0.15	0.08	0.09
Married, spouse absent	0.04	0.04	0.05	0.08
Married, spouse present	0.93	0.94	0.89	0.81
Wife's education (in years)*	8.63	N.A	10.39	10.13
Wife is employed*	0.13	0.20	0.36	0.43
Wife is self-employed*	0.01	0.01	0.02	0.03
No. of children	1.42	1.35	1.53	1.21
Health limits work	N.A	N.A	0.07	0.05
Northeast	0.45	0.41	0.26	0.26
Midwest	0.31	0.29	0.30	0.25
West	0.11	0.15	0.21	0.23

Notes: * Wife's education, and wife's employment status variables were averaged over the entire sample that included single men as well.

Consistent with evidence documented by several authors earlier, the racial gap in average schooling declined during the latter part of the twentieth century. Self-employed males of either race were more likely to be older and better educated than their wage-employed counterparts. Figures 4-1 and 4-2 show the descriptive trend in years of schooling for both races by self-employment status.

Figure 4-1. Average Years of Schooling - all employed Black males

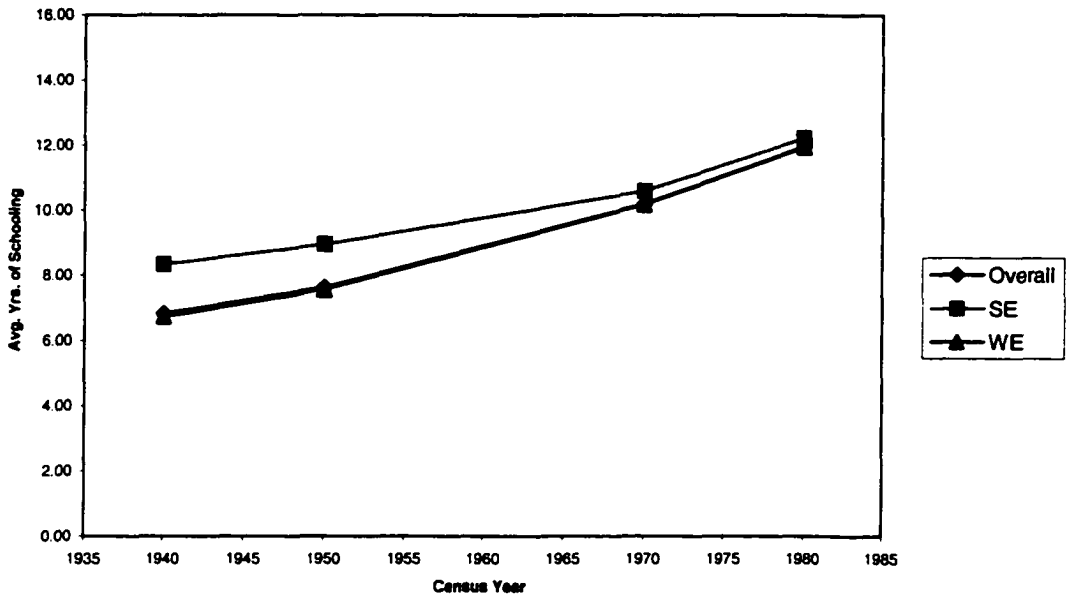
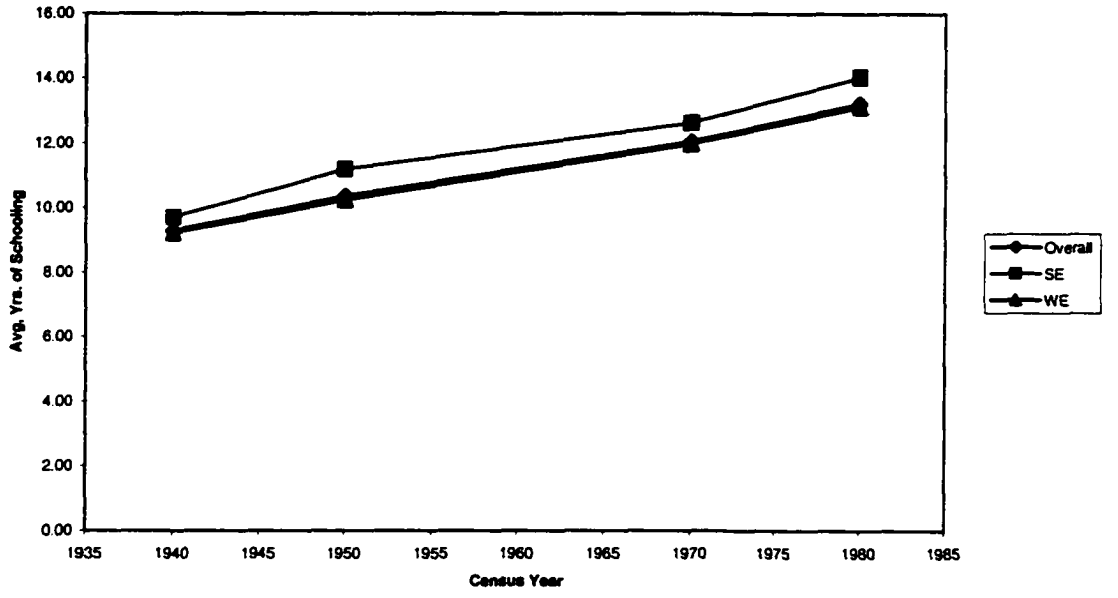


Figure 4-2. Average Years of Schooling - all employed White males



On average self-employed males of either race were likely to have had more years of schooling. However, the trend shows that the gap in average schooling between self-employed and wage/salary employed males converged among the blacks while it remained more or less constant among the white males.

Immigrants formed a greater percent of the self-employed workforce than the wage-employed workforce of both racial groups. However, racial differences are observed in the trend of the composition of the self-

employed population that was immigrant. In 1940, more than 35 percent of all self-employed white males were immigrants whereas less than 25 percent of wage/salary employed white males were immigrants. This ratio steadily declined over the years and by 1980 the proportion of self-employed whites that were immigrants was almost identical to the proportion of wage/salary-employed whites that were immigrants. See figures 4-3 and 4-4 below.

Figure 4-3. Percent Immigrant (employed White males)

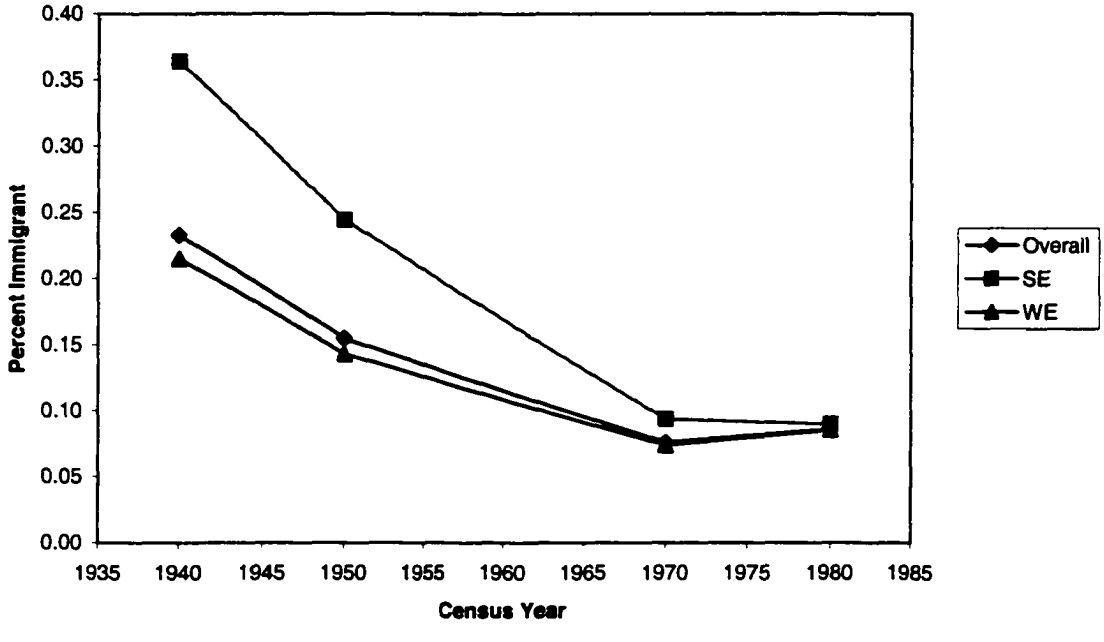
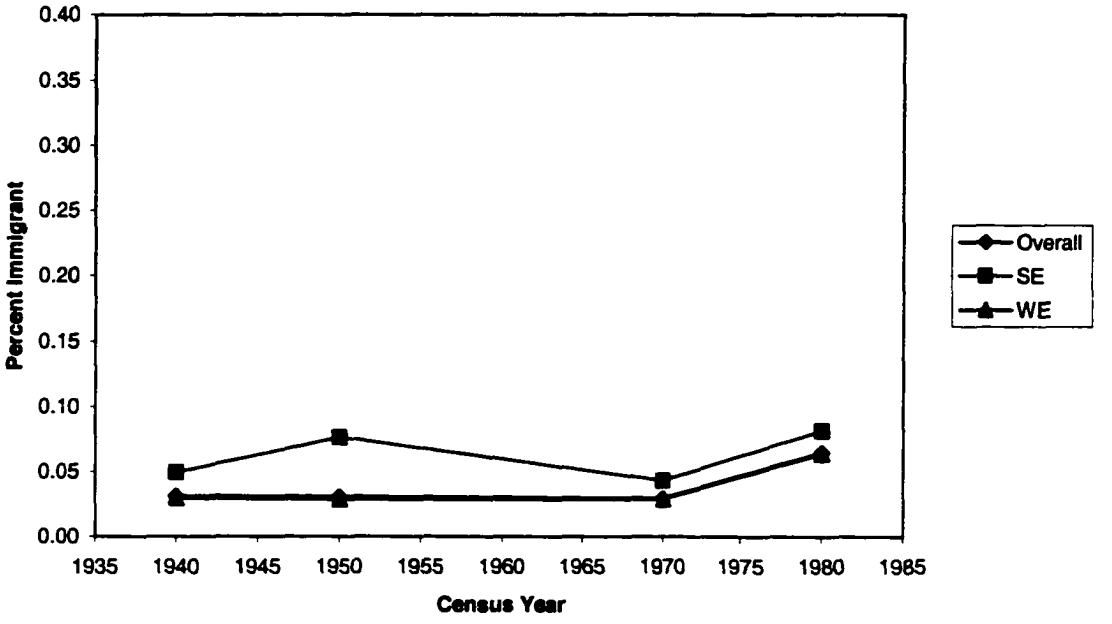


Figure 4-4. Percent immigrant (employed Black males)



The corresponding trend for the black males is different with the immigrant proportions of the self-employed black workforce remaining higher than the immigrant proportion of the overall black workforce thus suggesting racial differences in the role of immigrants in self-employment.

Table 4-2 shows the composition of self-employed and wage/salary employed males by race and industry. It is interesting to note that there are no significant racial differences in the trend of industrial

composition for both employment categories. The proportion of self-employed blacks and whites in the construction industry increased steadily from 1950 onwards while the corresponding proportion of wage-employed declined.

Table 4-2. Percent of all employed Black males by Industry and Census Year.

INDUSTRY/CENSUS YEAR	1940	1950	1970	1980
Mining	2.46%	1.44%	0.24%	0.35%
Construction	20.40%	9.44%	8.42%	6.78%
Manufacturing-durable	13.97%	21.97%	22.41%	18.57%
Manufacturing - non-durable	8.88%	12.60%	9.36%	9.02%
Transportation	11.12%	10.53%	9.65%	8.68%
Telecommunication	0.26%	0.10%	0.61%	1.15%
Utility	1.49%	1.47%	1.73%	2.15%
Wholesale	1.97%	3.42%	4.35%	4.31%
Retail	11.50%	10.72%	9.96%	8.97%
Finance	4.03%	3.48%	3.43%	4.23%
Professional	4.55%	5.04%	10.08%	13.04%
Public administration	4.73%	7.42%	11.38%	13.77%
Services	13.18%	11.19%	8.00%	8.44%

Percent of all employed WHITE males by Industry and Census Year

INDUSTRY/CENSUS YEAR	1940	1950	1970	1980
Mining	1.01%	1.19%	0.63%	0.71%
Construction	10.10%	8.61%	8.23%	8.67%
Manufacturing - durable	20.82%	20.62%	22.27%	21.10%
Manufacturing -non-durable	14.84%	13.91%	9.85%	7.92%
Transportation	7.61%	8.08%	6.48%	6.03%
Telecommunication	0.79%	1.35%	1.10%	1.77%
Utility	1.76%	1.11%	1.28%	1.62%
Wholesale	3.97%	6.42%	6.13%	5.90%
Retail	14.54%	14.07%	10.82%	10.74%

Table 4-2 continued.

Finance	4.69%	4.51%	5.76%	6.07%
Professional	4.83%	5.01%	9.75%	11.50%
Public administration	6.54%	8.09%	9.90%	9.12%
Services	7.15%	5.57%	7.11%	8.19%

The proportion employed in the retail sector declined in both categories of employment and for both racial groups. Figures 4-5 through 4-8 show the trends in industrial composition of self-employed and wage/salary employed males respectively. In 1940 the proportion of the black labor force that was self-employed in the construction, transportation and services sector was higher than the corresponding proportions for whites and the trend continued until 1980. The proportion of the black labor force that was self-employed in the retail and manufacturing sectors was lower in 1940 but by 1980 the black proportion had caught up in the retail and non-durable manufacturing. In contrast however, the black self-employed were more equally represented in the finance and professional services in 1940 but they lagged behind after 1970.

The industry trends in the wage/salary-employed sector are somewhat different. Black males who were under-represented in the public administration and manufacturing sectors, caught up with whites in the manufacturing sector and end up over-represented in the public administration sector. As in the self-employed sector, wage/salary employed blacks were over-represented in the construction, transportation and services industries in 1940 but by 1980 they were more equally represented in each of these industries.

Fig. 4-5. Industry shares _ Self-Employed White Males

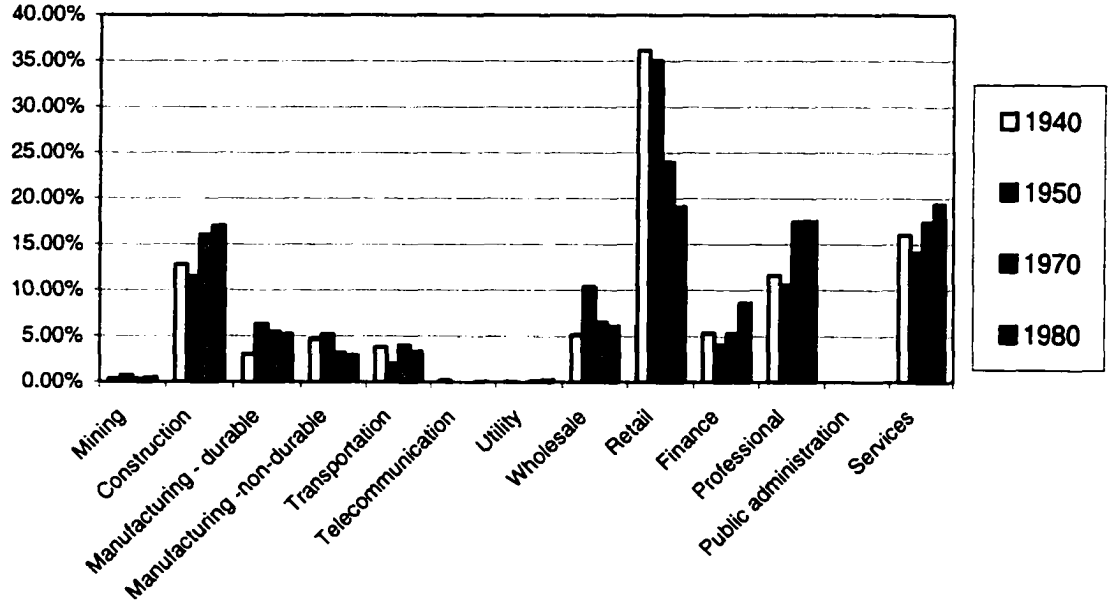


Fig. 4-6. Industry shares _ Self-Employed Black Males

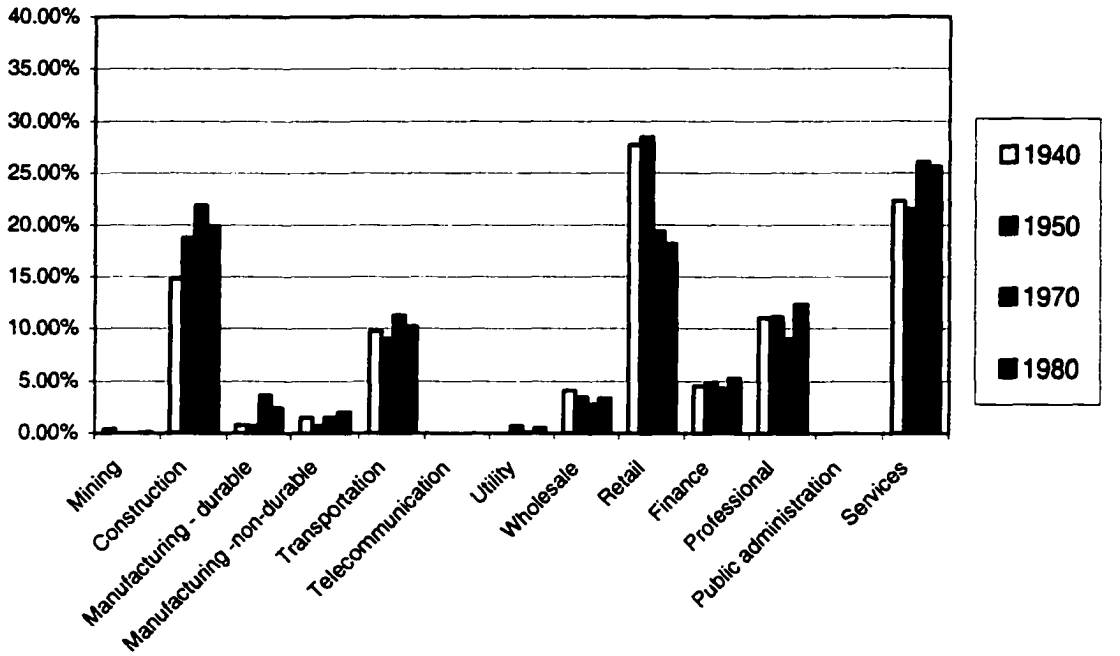


Figure .4-7. Industry shares - Wage/Salaried Black Males

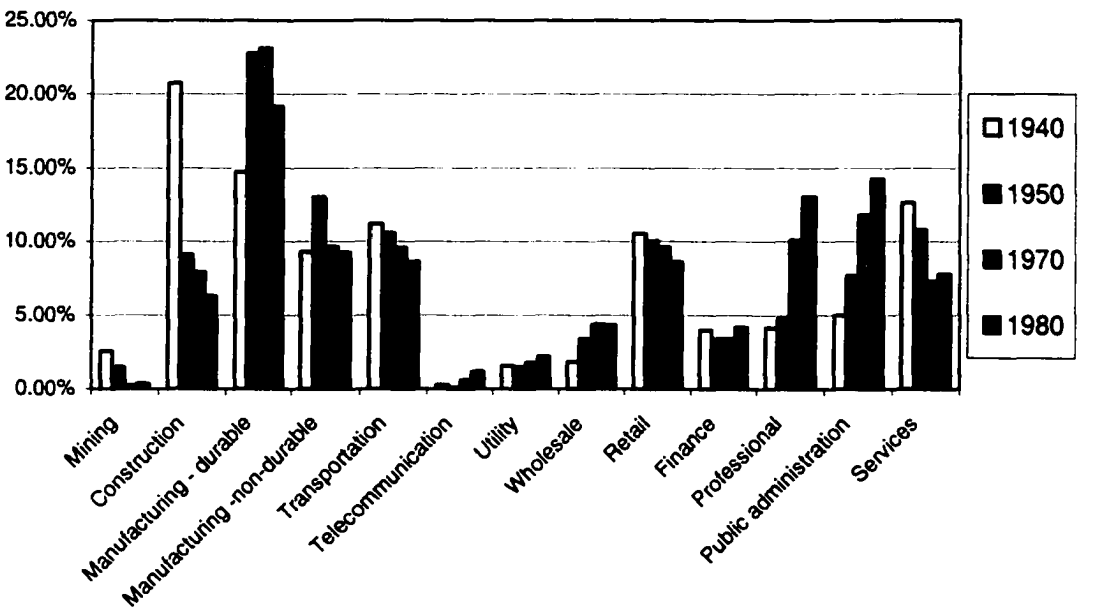
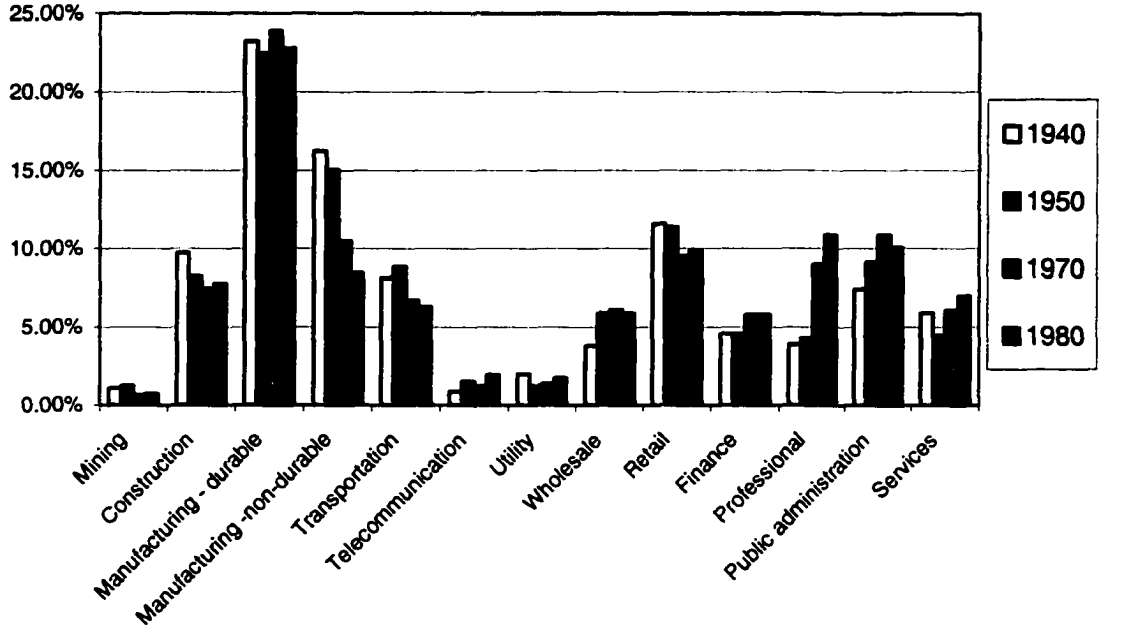


Figure 4-8. Industry sharee _Wage/Salaried White Males



Regression Results

I ran separate regressions for blacks and whites in order to allow for differences, if any, in the process determining self-employment status. The results, shown in tables 4-3 and 4-4, suggest that the probability of being self-employed for both racial groups increases with age but at a decreasing rate. It also increases with the number of years of education. Married men, in general, are less likely to be self-employed than single men. The coefficient on immigrant status is

positive and statistically significant in most years indicating that in general, immigrants are more likely to be self-employed than natives. This finding offers some support for the 'sojourner' hypothesis and the labor market disadvantage hypothesis and is consistent with earlier literature (Borjas 1986 and Bates 1998). Immigrant whites were seven percent more likely to be self-employed than native whites in 1940. By 1970 immigrant whites were only about one percent more likely to be self-employed than native whites. A decade later in 1980, immigrant whites were slightly less likely (0.2 percent) than natives to choose self-employment. The reasons for this downward trend remain to be explored. We could speculate that one reason why immigrants tend to choose self-employment is due to the disadvantages faced by them in the labor market either due to problems with fluency of language or due to discrimination. There may also be fewer barriers to self-employment for immigrants. As either of these reasons becomes less important over the years, they could contribute towards the downward trend in the participation of immigrants in the self-employed sector.

Table 4-3. Marginal Effects - Black Males

VARIABLE↓ YEAR→	1940	1950	1970	1980
INDIVIDUAL CHARACTERISTICS				
Intercept	-2.558**	-3.183**	-0.534**	-0.275**
Age	0.007**	0.007**	0.006**	0.005**
Age squared	-5.3E-5**	-5.5E-5**	-5.8E-5**	-4.5E-5**
Years of Schooling	0.005**	0.006**	0.003**	0.003**
Immigrant	0.029**	0.051**	0.018**	0.005**
Married, spouse absent	-0.030**	0.006**	-0.004**	-0.003**
Married, spouse present	-0.074**	0.021**	-0.038**	-0.042**
Wife's schooling (in years)	0.003**	---	0.003**	0.003**
Wife is employed	0.002	-0.031**	-0.005**	-0.001
Wife is self-employed	0.022**	0.109**	0.059**	0.080**
No. of children	-0.004**	-0.003**	0.001**	0.001**
Health Limits Work	----	----	0.009**	0.014**
REGION DUMMIES				
Northeast	-0.058**	0.040**	-4.5E-5	0.004*
Midwest	-0.043**	0.045**	0.007**	0.008**
West	0.078**	0.133**	0.016**	0.013**
NEIGHBORHOOD CHARACTERISTICS				
Crime rate	---	---	-3.7E-7	2.5E-7
Change in population(%)	---	---	-2.0E-4**	-0.009
Median Education (years)	0.001	-0.005**	-0.006**	0.015**
Median Income (000's of $)	0.097**	0.040**	-0.0003	0.001*
Per capita government expenditure (000's of $)	0.0003**	8.3E-5**	1.2E-5	-2.6E-7
Unemployed (%)	-0.244**	-0.001*	-0.004**	-0.002**
Black unemployed (%)	0.178**	-0.001**	0.001**	0.0004**
Percent Black	0.021*	0.029**	0.002*	-0.003**
Percent White	0.021*	0.025**	0.002*	-0.003**
Percent Chinese	0.007	0.034**	0.002	-0.006**
Percent Japanese	-0.041**	-0.006	-0.007**	-0.002*
Percent other Asians	-0.001	-0.063**	0.0001	-0.004**
White self-employed (%)	-0.001**	0.003**	-0.001**	-0.001**
Black self-employed (%)	-0.003**	0.013**	0.009**	0.005**
Asian self-employed (%)	0.0002**	1.11E-5	0.0002**	-3.7E-5
Pseudo-R-Square	0.1834	0.3526	0.2119	0.2163
No. of Observations	2245	1079	6826	7632

Table4-4. Marginal Effects - White Males

VARIABLE↓ YEAR→	1940	1950	1970	1980
INDIVIDUAL CHARACTERISTICS				
Intercept	2.854**	4.673**	-0.452**	-0.934**
Age	0.012**	0.011**	0.016**	0.015**
Age squared	-9.1E-5**	-1.0E-4**	-1.4E-4**	-1.4E-4**
Years of Schooling	0.010**	0.007**	0.005**	0.008**
Immigrant	0.072**	0.029**	0.009**	-0.002*
Married, spouse absent	0.045**	0.018**	0.012**	0.016**
Married, spouse present	0.044**	0.026**	-0.027**	-0.034**
Wife's schooling (years)	-0.001**	---	0.005**	0.004**
Wife is employed	0.024**	-0.017**	-0.027**	-0.019**
Wife is self-employed	0.103**	0.138**	0.161**	0.136**
No. of children	0.004**	-0.001**	0.005**	0.003**
Health Limits Work	---	---	0.010**	0.004**
REGION DUMMIES				
Northeast	0.008*	0.024**	0.002	-0.019**
Midwest	-0.027**	-0.018**	-0.005*	-0.022**
West	-0.009*	-0.024**	0.007**	0.002
NEIGHBORHOOD CHARACTERISTICS				
Crime rate	---	---	-5.9E-6**	-9.1E-7*
Change in population (%)	---	---	2.5E-4**	-0.087**
Median Education (years)	0.008**	-0.003**	-0.023**	0.010**
Median Income (000's of $)	-0.027**	-0.009**	0.002**	-0.0019**
Per capita government expenditure (000's of $)	3.9E-4**	-1.7E-4**	6.8E-5**	4.2E-7
Percent Unemployed	1.583**	-0.006**	-0.013**	-0.003**
White unemployed (%)	-1.981**	-0.001	0.015**	0.003**
Percent Black	-0.037**	-0.052**	-0.026	-0.053
Percent White	-0.037**	-0.052**	-0.001	6.0E-5
Percent Chinese	-0.025**	-0.064**	-0.006**	-0.003*
Percent Japanese	-0.017*	-0.058**	-0.007**	0.004**
Percent other Asians	-0.070**	0.048**	0.008**	-0.003**
White self-employed (%)	-2.4E-4	-0.001**	0.003**	0.006**
Black self-employed (%)	-0.001**	-1.4E-4*	0.004**	0.003**
Asian self-employed (%)	-5.1E-5*	-8.4E-5**	3.4E-4**	-6.3E-4**
Pseudo R-square	0.2763	0.2908	0.2496	0.2297
No. of Observations	4238	1657	6901	8158

Correspondingly for blacks, immigrant black males were roughly three percent more likely to be self-employed than native blacks in 1940 and 1.7 percent more likely in 1970. In 1980, immigrant blacks were also more likely to choose self-employment but only by 0.5 percent. Individuals with a work-related disability were more likely to be self-employed (roughly by one percent), consistent with the labor market disadvantage hypothesis.

The Impact of Neighborhoods

The neighborhood variables are statistically significant and the coefficient signs are consistent with the special demands hypothesis, the special opportunities and the consumer discrimination hypothesis. Each of these hypotheses predicts that the higher the proportion of the individual's own ethnic/racial group in the neighborhood, higher the probability that (s)he chooses self-employment. A positive coefficient on the variable measuring the percent of the population in the metropolitan area that belongs to the individual's own group therefore, is consistent with each of these hypotheses.

Per-capita government expenditure has a very small but statistically significant and positive effect on the probability of self-employment for both racial groups in the earlier years up to 1970. Blacks were more likely to be self-employed if they lived in a neighborhood with a large proportion of blacks. They were less likely to be self-employed if the proportion of Japanese increased in the neighborhood. The proportion of Whites and other Asians in the neighborhood generally had a positive effect on the self-employment probability of Blacks except in 1980. An increase in the overall unemployment rate in general depressed the probability of self-employment for both blacks and whites. However, an increase in the black unemployment rate increased the probability of self-employment for blacks, suggesting that blacks may choose self-employment as an alternative to unemployment. In general, the neighborhood coefficients reveal support for the labor market disadvantage hypothesis and the special demands hypothesis.

The Impact of Networks

The presence of a higher proportion of self-employed Asians in the neighborhood in general depressed the probability of self-employment for both Blacks and Whites suggesting that Asian businesses competed with both Black and White enterprises rather than promote them. The presence of a higher proportion of self-employed Blacks however, increased the probability of self-employment for Black males. The effect of the level of Black self-employment in the neighborhood is positive and statistically significant for Blacks in all the years except 1940. The positive coefficient suggests that the social network effect was stronger than the competition effect in the years after 1940.

In 1940, the significantly negative coefficient (on Black self-employment rate) suggests that the competitive effect dominated the social learning effect. This could be due to a combination of factors. First, the entry and success of black businesses in merchandizing and manufacturing businesses was limited by their past inexperience and lack of capital. As a

result, they were mostly confined to the provision of services that were traditionally black occupations such as of barbers, beauticians, undertakers, morticians etc., in which they competed vigorously amongst themselves. Second, black businesses complained bitterly about lack of patronage from their own community and their inability to co-operate with their own race.

> "Negro businessmen insist that they face five main competitive difficulties: (1) difficulty in procuring capital and credit, (2) difficulty in getting adequate training, (3) inability to secure choice locations on the main business streets, (4) lack of sufficient patronage to allow them to amass capital and to make improvements, (5) inability to organize for co-operative effort.
>
> These circumstances have resulted in a situation in which Negroes have found it extremely difficult to compete with white businessmen in the same field as to prices and service. This, in turn, tends to reinforce the stereotype that Negroes are not good businessmen. In order to meet the competition, Negro businessmen and community leaders stress dogmas of racial solidarity in an effort to amass capital and patronage…" -- Cayton and Drake (1946)

> "Over and over, Negroes in Bronzeville reveal a conflict between the economic imperative of "making ends meet" and the social demands of "race pride." … Poor people need credit. Negro merchants on the whole are unable to grant it. This forces the Negro housewife to avoid the colored grocer…Colored housewives often … charge that the quality and variety of stock in Negro stores are poor… Colored storekeepers are also accused of general

> inefficiency and ineptitude.." - Cayton and Drake (1946)

Another reason for the observed dominant effect of competition over co-operation for both blacks and whites in 1940 is the fact that the years of Depression had forced a lot of people that had lost a means of livelihood into self-employment.

> "The impact of the Depression combined with the fierce competition for good locations, for credit and capital, also resulted in an accentuation of racial antagonisms... Many of the more successful Negro businessmen criticize their less successful competitors on the ground that the latter do not take their business seriously, and thus give all colored enterprises a bad name. [They] were particularly antagonistic to "depression businesses" - the numerous small stores that between 1929 and 1938 sprang up in houses, basements, and old buildings." - Cayton and Drake (1946)

The rest of the census years saw a positive effect of the neighborhood black self-employment rate on black self-employment probability. A one percent increase in the black self-employment rate in the metropolitan area increased the self-employment probability of an average black male in the neighborhood by roughly 1.3 percent in 1950, by 0.9 percent in 1970 and by 0.5 percent in 1980. Thus, the social network effect was

stronger in the years after World War II and promoted self-employment among black males.

The decade of the 1940's had seen a significant increase in black businesses and co-operatives. This was mainly due to the fact that "the war [World War II] provided more than a million black men [that served in the military during the war] for the first time with savings that could be used as venture capital…" Since the beginning of the twentieth century, both the black political leadership and the religious leaders emphasized the importance of co-operative ventures as a means of pooling capital for promoting black business and the black economy. Despite their efforts however, blacks found it particularly difficult to organize effective and long-lasting cooperative ventures. The cooperative movement however made great strides in the late 30's and 40's. In the period between the world wars, there was greater economic cohesion among blacks on the periphery of the nation's economy. Black colleges and churches also established cooperatives during the 1940's (J.E.K. Walker 1998). Thus the combined effects

of the war and the temporary success of efforts to organize cooperatives and credit unions, might explain the strong network effect found in 1950 onwards.

The co-operative movement however, started losing steam after World War II and may also account for the weakening of the social network effect in the years after 1950.

> "Nationally, however, there was a post-World War II decline in co-op ventures and membership. While cooperatives provided economic advantages to their members, blacks, as loyal Americans, were determined to be a part of the capitalist free-enterprise system, pursuing individual avenues to wealth, even when it was to their disadvantage. And so cooperative enterprises, which provided benefits to the poor, were viewed by some blacks as socialist." - J.E.K. Walker (1998)

The individual probability of self-employment for blacks was negatively correlated with the White and Asian self-employment rates in 1970 and 1980 suggesting greater competition among the racial groups in these later years.

The corresponding results for white males indicate that competitive concerns dominated the self-employment decisions of white males in the earlier

years of 1940 and 1950 while the network effect steadily gained greater importance by 1970 and 1980. The self-employment probability of the average white male decreased with an increase in the self-employment rates of all racial groups in 1940 and 1950 whereas it increased by roughly 0.3 percent in 1970 and 0.6 percent in 1980. This indicates that social networks played a differential role not only across racial categories but also within the same racial group over time.

Finally, table 4-5 shows results from a decomposition of the racial gap in self-employment rates. The decomposition method used is similar to the one used in Borjas and Bronars (1986) and in chapter 3 of this dissertation. The decomposition reveals some interesting results. More than 100 percent of the racial gap in self-employment prior to 1970 is explained by the difference in individual and neighborhood endowments. The results indicate that blacks would have experienced much lower self-employment rates if they had faced the same self-employment structure as whites prior to 1970. This is

consistent with the disadvantage theory that discrimination in the labor market or other disadvantages such as language problems pushed the minorities into self-employment thus, lowering the average levels of ability in self-employment.

Table 4-5. Decomposition of Racial gap in Self-Employment Probabilities

	1940	1950	1970	1980
I. Predicted Black self-employment probability	9.10%	6.41%	4.24%	4.37%
II. Predicted White self-employment probability	15.78%	15.26%	11.25%	11.95%
III. Predicted Black self-employment probability if face the same coefficients as Whites	5.03%	5.20%	12.93%	10.23%
IV. Racial gap in self-employment rates (II-I)	6.69%	8.85%	7.00%	7.71%
V. Gap due to difference in endowments (II - III)	10.72%	10.05%	-1.68%	1.72%
VI. Gap due to difference in coefficients (slope parameters) (III - I)	-1.41%	-1.21%	8.69%	5.86%

From 1970 onwards, the results indicate that a large part of the gap is not due to differences in endowments but due to differences in self-employment determination. This might indicate the relative

success of the civil rights legislation of the 1960's in reducing the disadvantages in the labor market and making wage employment more likely and attractive for blacks and thus discouraging self-employment where the bias from consumers and workers could have continued.

4.6. Conclusion

In this chapter I attempted to test whether social networks play any role in the self-employment decisions of individuals and whether they might be able to explain the persistent racial gap in self-employment rates. Given that self-employment has long been considered as an important means of closing the racial gap in wealth, if social networks were found to be important in encouraging self-employment, it would have important implications for policy formulation. The results indicate that social learning is an important determinant of individual decisions to become self-employed. Further, social networks might have played a differential role not only across racial categories but also within the same racial group over time. Therefore social interactions and differences

therein, might potentially explain the persistence of the racial gap in self-employment rates.

Further work is required however, before any policy recommendations can be made. Future studies would have to overcome the limitations of this study. The cross-sectional nature of the data necessitated a static formulation. It was not possible to control for characteristics such as the length of time in business, previous business experience and asset levels. As with other empirical studies of self-employment in economics, I have no direct measures of managerial/entrepreneurial ability, risk tolerance, ease of access to credit etc. Further, the census data used do not allow for identification of smaller neighborhoods that might better approximate the real structure of social interactions. Identification of smaller neighborhoods might allow for better inferences to be drawn from an analysis of the role of social networks.

V. CONCLUSION

While network effects in general have been the subject of many studies, there have been only a few empirical studies that examine their importance in specific markets. The theoretical models, in general predict that the presence of network effects leads to conformity of behavior and very often, to inefficient market outcomes. However, the predictions depend heavily on the assumptions made regarding the underlying preferences and expectations of individual participants in the markets. The importance of empirical evidence cannot be overstated especially since any policies aimed at correcting market outcomes cannot be based purely on untested hypotheses. This dissertation contributes to the scant empirical literature on the subject of network effects.

This dissertation specifically examines the impact of network effects in two separate markets - computer software markets and the self-employment decisions of individuals in labor markets.

While there have been some empirical studies of the extent of network effects in computer software markets, there have been no direct or indirect tests of their implications for social welfare. A common prediction of most theoretical models of the impact of network effects on technological adoption is that the presence of network effects often leads to de-facto standardization and might result in lock-in of an inferior technology. Even though there is no non-controversial empirical evidence of such inefficient technological lock-in to date, the idea seems to be widely accepted mainly on the basis of anecdotal evidence. The use of this argument in a major antitrust case against computer software maker Microsoft Corporation, further emphasizes the need for an empirical test of the hypothesis. Hard empirical evidence is difficult to obtain however, as the data required for direct tests of this hypothesis are largely unavailable to researchers.

In the second chapter of this dissertation I devised an indirect test of this hypothesis by taking advantage of the natural experiment afforded by the

introduction of the programming language Java. Java made it possible for software programmers to write a single program that would run on any operating system enabled with a Java Virtual Machine. It therefore, had the potential to increase the level of competition in the software applications market and thereby eliminate the disadvantage faced by competing operating systems in obtaining the "critical mass" required for successful adoption by consumers.

Anecdotal evidence suggests that Java was not fully successful in fulfilling its potential- possibly due to several reasons including but not limited to the following. The possibility that consumers were locked into Microsoft's technology, the possibility that Java was actually an inferior technology as programs written in Java tended to be much slower in execution and the possibility of anti-competitive impact of some of Microsoft's actions directed at defending its dominant position.

The empirical model devised in chapter 2 of this thesis, tests whether Java was successful in creating

more competition in the applications market by examining its impact on the prices of graphics application software. The results of the test indicate that consumers and firms expected potential competition from Java-based applications and adjusted their demand and supply behaviors accordingly. Therefore any subsequent failure of Java cannot be attributed to "excess inertia" in software markets. This result not only has important implications for the ongoing antitrust trial of Microsoft but also for the broader question of whether or not a superior technology would always be successful in the presence of network effects. If firms expected potential competition from Java-based applications and reacted accordingly, as the evidence from chapter 2 suggests, then it could be argued that consumers and firms did not consider themselves locked into Microsoft's technology.

While some empirical studies of the strength of network effects have been conducted in computer software markets, the extent of network effects in the labor market decisions of individuals remains largely

unexplored. Chapter 3 of this dissertation replicates a key paper in the literature on the impact of consumer discrimination on minority self-employment decisions. A comparison of the results points to the limitations and inaccuracies in the previous studies of racial differences in self-employment. A closer examination of the model indicates that it may be incomplete and possibly mis-specified due to the exclusion of social network effects. Chapter 4 extends the empirical model to include several different measures of social networking and estimates the relative strength of these network effects in the self-employment decisions of individuals. Given that self-employment has long been considered an important means of closing the racial gap in wealth, if social networks were found to be important in encouraging self-employment, it would have important implications for policy formulation. Inclusion of several different network measures allowed for a reasonable identification and examination of the role of networks in self-employment decisions of individuals. The results indicate that social learning is an important determinant of individual decisions to become self-

employed. Further, social networks might have played a differential role not only across racial categories but also within the same racial group over time. Given the historic disadvantages faced by Blacks in business, social interactions and differences therein might potentially explain the persistence of the racial gap in self-employment rates.

Future studies on network effects would have to overcome the limitations of this study. A direct test of the hypothesis of technological lock-in would need micro data on individual purchases of technology over time that would allow estimation of structural models of consumer choice. Future studies of social networks would have to incorporate better information regarding the neighborhood relevant for social interactions.

In summary, this dissertation explores the role of network effects in two markets. In the computer software markets, where network effects have received a good deal of attention, the results of the study indicate that network effects may not be as persistent

or as damaging as previously thought. The evidence suggests that the presence of network effects may not offer as high barriers to entry as is generally believed and therefore has important policy implications. In labor markets and in self-employment decisions in particular, where the impact of network effects has received much less attention, this dissertation provides the first such test and quantification of the impact of network effects on racial differences in self-employment decisions. The results indicate that social networks might have a larger impact on self-employment decisions than previously acknowledged and may help explain the persistence of the racial gap in self-employment rates. As the empirical results indicate the importance of the social network effect in promoting self-employment seems to be changing over time and across races. Therefore any efforts by policy makers to promote minority self-employment would have to consider the effects of both networking and competition on self-employment decisions.

APPENDIX A

Correlation Analysis

Subset of data with the Street Price

Name of Software	Month	Year	List Price	Estimated Street Price
Adobe Illustrator 4.02	OCT	1994	695	370
Macromedia Freehand 5.0	SEP	1995	595	389
Adobe Photoshop 3.0	JAN	1995	895	549
CorelDraw 6.0	AUG	1995	695	450
Fractal Design Painter 4.0	APR	1996	549	375
Micrografx Designer 6.0	APR	1996	350	299
Macromedia Freehand 7	APR	1997	595	400

Correlation Coefficient (rho) of List Price with Street Price = 0.9620, t-ratio =9.326. Reject null hypothesis that rho=0

Correlation (rho) of Discount(street price/list price) with quality and compatibility variable

Variable	Discount	t-ratio	decision
Inst_base	0.441	1.299	fail to reject rho=0
java	-0.022	-0.06	fail to reject rho=0
Draw	-0.117	-0.31	fail to reject rho=0
edit	0.027	0.072	fail to reject rho=0
pres	0.075	0.199	fail to reject rho=0
EPS	0.488	1.479	fail to reject rho=0
GIF	0.004	0.012	fail to reject rho=0
JPEG	0.004	0.012	fail to reject rho=0
bitmap	0.004	0.012	fail to reject rho=0
multimedia	0.075	0.199	fail to reject rho=0
TEXT	-0.080	-0.21	fail to reject rho=0
data	0.075	0.199	fail to reject rho=0
RAM	0.284	0.784	fail to reject rho=0

APPENDIX B

Construction of Installed base measure

Market Share[a] data extracted from the publicly available Exhibit 1 of the Department of Justice in the Microsoft Antitrust Trial.
http://www.usdoj.gov/atr/cases/ms_exhibits.htm

OS[b,c]/YEAR	1993	1994	1995	1996	1997[e]
Microsoft[d]	93	93	90	94	95
IBM OS/2	3	4	7	3	3

Notes:

a. Worldwide shares
b. Operating Systems used in a single user client and PC operating Environment
c. Shares do not total 100% due to exclusion from this table of other Intel and Intel-based UNIX operating systems.
d. Includes Microsoft 16-bit and 32-bit Windows and MS-DOS.
e. The market shares for 1997 are forecasts.

Worldwide Shipments in millions of units

OS/ YEAR	1993	1994	1995	1996	1997
Microsoft[1](millions)	23.5	33.4	42	51.9	n.a
IBM OS/2[ii]	1.011	2.598	1.340	1.639	

Notes:

i. Microsoft's operating systems from DOJ trial exhibit 439. www.usdoj.gov/atr/cases/ms_exhibits.htm

ii. OS/2 shipments calculated by combining the market share data from the previous table and Microsoft's shipment data from the table above. Example: In 1993, Microsoft's shipments of 23.5 million units represent a 93 percent market share. Therefore OS/2's 3 percent share translates into (23.5 /0.93) *0.03 = 1.011 million units of shipments.

Installed base in 1993[1] (millions)

Microsoft(includes all MS-OS)	6
OS/2	111

Notes: 1. Source: Computer Reseller News; International Data Corporation; OS Overview; August 22, 1994; Pg. 223.

Installed base for each of the other years from 1994-97 was calculated as the sum of the previous year's installed base and the previous year's shipments. So for instance installed base for all Microsoft's Operating systems in 1994 would be 111 + 23.5 = 134.5 million.

APPENDIX C

Sample construction and data sources for the empirical analysis in Chapters 3 and 4.

Sample Construction:

1940

Source and restrictions: One percent Census sample restricted to employed males aged 25-64 residing in the 75 largest metropolitan areas and working in non-agricultural industries.

Stratification: ten percent of all whites who were self-employed, 3.3 percent of all whites who were employed for wage/salary, 33 percent of all blacks who were employed for wage/salary and 100 percent of all blacks who were self-employed.

1950

Source and restrictions: One percent Census sample restricted to sample line individuals that were asked to report their education level. Other restrictions include employed males aged 25-64 and residing in the 75 largest metropolitan areas and working in non-agricultural industries.

Appendix C continued.

Stratification: 33.33 percent of all self-employed white males and 10 percent of wage/salary employed white males. all self-employed and wage-employed black males that were asked to report their level of education.[33]

1970

Source and restrictions: One percent Census sample restricted to employed males aged 25-64 residing in the 75 largest metropolitan areas and working in non-agricultural industries.

Stratification: ten percent of all whites who were self-employed, 3.3 percent of all whites who were employed for wage/salary, 33 percent of all blacks who were employed for wage/salary and 100 percent of all blacks who were self-employed.

---

[33] In 1950, the question on level of education was asked of only one individual from each household selected. The sample of individuals for whom this variable is included was therefore roughly a flat sample of 1-in-330. See IPUMS documentation for more detail. Special sample-line weights were used to account for this in 1950.

Appendix C continued.

1980

Source and restrictions: One percent Census sample restricted to employed males aged 25-64 residing in the 75 largest metropolitan areas and working in non-agricultural industries.

Stratification: ten percent of all whites who were self-employed, 3.3 percent of all whites who were employed for wage/salary, 33 percent of all blacks who were employed for wage/salary and 100 percent of all blacks who were self-employed.

Sources of other metropolitan area characteristics:

The source of the metropolitan area characteristics is the county and city data book (supplement to the statistical abstract of the U.S) of 1940, 1950 and 1970 and the published census summaries (chapters A-D) for each of the fifty states in 1980.

The per-capita government expenditure data for 1940 is derived from the statistical abstract of the United States information on city government finances. They

pertain to the major city in metropolitan area and therefore underestimate the government expenditures in the metropolitan area. The information provided in the statistical abstract pertains to the year 1936 and the total governmental cost payments, and operation, maintenance and interest expenses. Per-capita expenditure was calculated as the difference between total governmental cost payments and operation, maintenance and interest costs, divided by population in metropolitan area. Population in metropolitan area was taken from the published national summary of the decennial census of population.

The overall crime rate by metropolitan area was obtained from the Sourcebook of Criminal Justice Statistics published by the U.S. Department of Justice, Bureau of Justice Statistics. The first issue of this periodical was in 1973.

REFERENCES

Aldrich, Howard E. and Roger Waldinger, (1990) "Ethnicity and Entrepreneurship," *Annual Review of Sociology*, Vol.16, pp 111 - 135.

Allen, David W., (2000) "Social Networks and Self-Employment," *The Journal of Socio-Economics*, Vol. 29, pp 487 - 501.

Arthur, W.B., (1989) "Competing Technologies and Lock-In by Historical Small Events," *Economic Journal*, Vol.99, pp 116-131.

Banerjee, Abhijit V., (1992) "A Simple Model of Herd Behavior," *Quarterly Journal of Economics*, Vol.107, Issue 3, pp 797 - 817.

Bates, Timothy (1998) "Exiting Self-Employment: An Analysis of Asian-Immigrant Owned Small Businesses," *CES Discussion Paper*, U.S. Bureau of the Census, Washington D.C.

Benabou, R. (1993), "Workings of a City: Location, Education, and Production," *Quarterly Journal of Economics*, Vol. 108, Issue????, pp 619-652.

Benabou, Ronald (1996), "Equity and efficiency in Human Capital Investment", *Review of Economic studies*, Vol. 63, Issue 2, pp 237-264.

Bernheim, Douglas B., (Oct. 1994) "A Theory of Conformity," *The Journal of Political Economy*, Vol. 102, Number 5, pp 841-877.

Bikhchandani, Herschleifer and Welch (Oct 1992) "A Theory of Fads, Fashion, Custom and Cultural Change as Informational Cascades," *The Journal of Political Economy*, Vol. 100, Issue 5, pp 992 - 1026.

Blanchflower, David G., Phillip B. Levine, and David J. Zimmerman (1998), "Discrimination in the Small Business Credit Market," *National Bureau of Economic Research Working Paper #6840*

Blau, David M., (1987), "A Time-Series Analysis of Self-Employment in the United States", *The Journal of Political Economy*, Vol. 95, Number 3, pp 445-467.

Blume, M., and Steven N. Durlauf (2000), "Interactions Based Models", Working Paper, University of Wisconsin-Madison.

Borjas, George (1986), "The Self-Employment Experience of Immigrants," *Journal of Human Resources*, Vol. 21, pp 487 - 506.

Borjas, George and Stephen Bronars (1989), "Consumer Discrimination and Self-Employment," *The Journal of Political Economy*, Vol. 97, Number 3, pp 581 - 605.

Bresnahan, T.F., and S. Greenstein (1999) "Technological Competition and the Structure of the Computer Industry" *The Journal of Industrial Economics*, Vol. 47, Issue 1, pp 1-40.

Brock, W., and Steven N. Durlauf (2000), "Discrete Choice with Social Interactions", Working Paper, University of Wisconsin-Madison

Cayton, Horace R. and St. Clair Drake (1946) Black Metropolis, London: Jonathan Cape.

Church, J., and N. M. Gandal (1992) "Network Effects, Software Provision and Standardization" *The Journal of Industrial Economics*, Vol. 40, Issue 1, pp 85-103.

Church, J., and N. M. Gandal (1993) "Complementary Network Externalities and Technological Adoption" *International Journal of Industrial Organization*, Vol. 11, Issue 2, pp 239-260.

Cowan, R. (1990) "Nuclear Power reactors: a study in technological lock-in." *The Journal of Economic History*, Vol. 5, Issue 3, pp 541-567.

Cusumano, M., Y. Mylonadis and R. Rosenbloom (1992) "Strategic Maneuvering and mass-market dynamics: the triumph of VHS over Beta" *Business History Review*, Vol. 66, pp 51-94.

David, P.A. (1985) "Clio and the Economics of QWERTY." *American Economic Review*, Vol.75, Issue 2, pp 332-337.

Dewey, Donald (1952), "Negro Employment in Southern Industry," *The Journal of Political Economy*, Vol. 60, Issue 4, pp 279-293.

Donohue, J. and Heckman, J.,(1991,) "Continuous Versus Episodic Change: The Impact of Civil Rights Policy on the Economic Status of Blacks," *Journal of Economic Literature*, Vol. 29, Issue ????, pp 1604-1643.

Du Bois, W.E.B. (1899) The Philadelphia Negro, Philadelphia: University of Pennsylvania.

Economides, Nicholas and Charles Himmelberg (1995), "Critical Mass and Network Size with Application to the U.S. Fax Market", August 1995, mimeo Stern School of Business, NYU.

Ellison, Glenn and Edward L. Glaeser (Oct. 1997) "Geographic concentration in U.S. Manufacturing Industries: A Dartboard Approach," *The Journal of Political Economy*, Vol. 105, Number 5, pp 889-927.

Ellison, Glenn and Drew Fudenberg, (1993) "Rules of thumb for Social Learning," *The Journal of Political Economy*, Vol.101, Issue 4, pp 612 - 643.

Ellison, Glenn and Drew Fudenberg, (1995) "Word-of-mouth Communication and Social Learning," *Quarterly Journal of Economics*, Vol.110, Issue 1, pp 93 - 125.

Evans, David S. and Boyan Jovanovic (1989) " An Estimated Model of Entrepreneurial Choice under Liquidity Constraints," *The Journal of Political Economy*, Vol. 97, Issue 4, pp 808 -827.

Evans, William N., Wallace E. Oates, Robert M. Schwab, (1992) "Measuring Peer Group effects: A study of Teenage Behavior," *The Journal of Political Economy*, Vol.100, Issue 5, pp 966 - 991.

Farrell, J. and G. Saloner (1986), "Installed Base and Compatibility: Innovation, Product Preannouncements, and Predation." *American Economic Review*, Vol. 76, Issue 5, pp 940-955.

Farrell, J. and G. Saloner (1992), "Converters, Compatibility and the Control of Interfaces," *Journal of Industrial Economics*, Vol. 40, Issue 1, pp 9-35.

Fairlie, Robert W., (1996), Ethnic and racial entrepreneurship - A study of historical and contemporary differences, New York & London : Garland Publishing Inc.

Fairlie, Robert W.,(1999), "The Absence of the African-American Owned Business: An Analysis of the Dynamics of Self-Employment," *Journal of Labor Economics*, Vol.17, Issue 1, pp 80-108.

Fairlie, Robert W., (2000), "Recent trends in Ethnic and Racial Self-Employment", *University of California, Santa Cruz, Working paper.*

Fairlie, R., and B. Meyer (1999)., "Trends in Self-Employment Among White and Black Men: 1910-1990", *Working Paper.*

Fairlie, Robert. W., and William A. Sundstrom (1999), "The emergence, persistence, and recent widening of the racial unemployment gap", Industrial and Labor Relations Review, Vol.52, Issue 2, pp 252-270.

Fishback, Price V., (1992) Soft coal, hard choices: the economic welfare of bituminous coal miners, 1890-1930, New York: Oxford University Press Inc.

Foray, D. (1997) " The Dynamic implications of increasing returns: Technological change and path dependent inefficiency." *International Journal of Industrial Organization*, Vol.15, Issue 6, pp 733-752.

Frazier, E. Franklin (1957): The Negro in the United States, 2nd Edition, New York: MacMillan.

Gandal, N.M. (1994), "Hedonic Price Indexes for Spreadsheets and an Empirical Test for Network Externalities." *RAND Journal of Economics* Vol. 25, Issue 1, pp 160-170.

Gandal, N.M. (1995) "Competing Compatibility Standards and Network Externalities in the PC Software Market."

Review of Economics and Statistics, Vol. 77, Issue 4, pp 599-608.

Gandal, N.M., S. Greenstein and D. Salant (1999) "Adoptions and Orphans in the Early Microcomputer Market" *The Journal of Industrial Economics*, Vol. 47, Issue 1, pp 87-105.

Gandal, N.M., M. Kende, and R. Rob (2000) " The Dynamics of Technological Adoption in Hardware/Software Systems: The case of Compact Disc Players." *RAND Journal of Economics*, Vol. 31, Issue 1, pp 43-61.

Gilbert, R.J. (1995), "Networks, Standards and the Use of Market Dominance: Microsoft (1995)," in J. Kwoka and L. White eds. The Antitrust Revolution: The Role of Economics, 3rd edition, Oxford: Oxford University Press.

Gilbert, R. J., and M. L. Katz (2001), "An Economist's Guide to U.S. v. Microsoft", *Journal of Economic Perspectives*, Vol. 15, Issue 2, pp 25-44.

Glaeser, E., B. Sacerdote and J. Scheinkman (1996), "Crime and Social Interactions," *Quarterly Journal of Economics*, Vol. 111, Issue 2, pp 507-548.

Hall, R.E. (1999), "Toward a Quantification of the Effects of Microsoft's Conduct." Hoover Institution and Department of Economics, Stanford University Working Paper, December 1999.

Higgs, R., (1976) "Participation of Blacks and Immigrants in the American Merchant Class, 1890-1910: Some demographic Relations," *Explorations in Economic History*, Vol. 13, Issue 2, pp 153-164.

Holtz-Eakin, D., David Joulfaian, Harvey S. Rosen., (1994) "Sticking it Out: Entrepreneurial Survival and Liquidity Constraints," *The Journal of Political Economy*, Vol. 102, Issue 1, pp 53 - 75.

Holzer (1987), "Informal Job Search and Black Youth Unemployment", *The American Economic Review*, Volume 77, Issue 3, pp 446-452.

Jencks, Christopher and Susan E. Mayer (1990), " The Social Consequences of Growing Up in a Poor Neighborhood", in Inner City Poverty in the United States, Lawrence E. Lynn, Michael G.H. McGeary, Editors, Washington D.C., National Academy Press 1990.

Katz, M.L. and C. Shapiro (1985) "Network Externalities, Competition, and compatibility." *American Economic Review*, Vol.75, Issue 3, pp 424-440.

Katz, M.L. and C. Shapiro (1986) "Technology Adoption in the Presence of Network Externalities." *The Journal of Political Economy*, Vol.94, Issue 4, 822-841.

Kihlstrom, Richard and Jean-Jacques Laffont (1979), "A General Equilibrium Theory of Firm Formation Based on Risk-Aversion," *The Journal of Political Economy*, Vol.87, Issue 4, pp 719 - 748.

Klein, B.,(2001) "What can a Dominant Firm Do to Defend Its Market Position?" *Journal of Economic Perspectives*, Vol. 15, Issue 2, pp 45-62.

Kretschmer, T. (2001) " Upgrading and Niche Usage of PC Operating Systems" Mimeo

Liebowitz, S.J. and S.E. Margolis (1990) "The Fable of the Keys." *Journal of Law and Economics*, Vol.33, Issue 1, pp 1-25.

Liebowitz, S.J and S.E. Margolis (1994) "Network Externality: An Uncommon Tragedy." *Journal of Economic Perspectives*, Vol.8, Issue 2, pp 133 - 150.

Liebowitz, S.J and S.E. Margolis (1995a). "Path Dependency, Lock-in and History." *Law, Economics and Organization*, Vol.11, pp 205-225.

Liebowitz, S.J and S.E. Margolis (1995b) "Are Network Externalities a New Source of Market Failure?" *Research in Law and Economics*, Vol.17, pp 1-22.

Lopatka, J. and W. H. Page, (1995) "Microsoft, Monopolization, and Network Externalities: Some Uses and Abuses of Economic Theory in Antitrust Decision making." *Antitrust Bulletin*. Vol.40, pp 317-70.

Lucas, Robert E. (1978), "On the Size Distribution of Business Firms," *Bell Journal of Economics*, Vol. 9, pp 508 -523.

Manski, C., (2000) "Economic Analysis of Social Interactions", *Journal of Economic Perspectives*, Volume 14, Issue 3. pp 115-136.

Matutes, C. and P. Regibeau (1992), "Compatibility and Bundling of Complementary goods in a Duopoly," *Journal of Industrial Economics*, Vol. 40, Issue 1, pp 37-54.

Moore,R.L.,(1983) "Employer Discrimination: Evidence from Self-employed Workers," *Review of Economics and Statistics*, Vol.65, pp 496-501.

Nakosteen,R.A and Zimmer,M.,(1980) "Migration and Income: The Question of Self-Selection," *Southern Economic Journal*, Vol.46, pp 840-851.

Neuman, S., and R.L. Oaxaca (2001) "Estimating Labor Market Discrimination with Selectivity Corrected Wage Equations: Methodological Considerations and An Illustration from Israel," *Working Paper*.

O'Regan, K.M., J.M. Quigley, "Teenage Employment and the Spatial Isolation of Minority and Poverty Households," *The Journal of Human Resources*, Vol. 31, Issue 3, pp 692-702.

Oaxaca, R.L., (1973) "Male-Female Wage Differentials in Urban Labor Markets," *International Economic Review*, Vol.14, pp 693-709.

Reingold, David A.(1999) "Social Networks and the Employment problem of the Urban Poor," *Urban Studies* Vol.36, Number 11 pp 1907-1932.

Ruggles.S and Sobek M et. al. Integrated Public Use Microdata Series: Version 2.0 Minneapolis: Historical Census Projects, University of Minnesota, 1997. http://www.ipums.umn.edu.

Saloner, G., and A. Shepard (1995) "Adoption of Technologies with Network effects: An Empirical Examination of the Adoption of automated Teller Machines." *RAND Journal of Economics*, Vol. 26, Issue 3, pp 479-503.

Smith, J.P. and Welch,F.R., (1989) "Black Economic Progress After Myrdal," *Journal of Economic Literature*, Vol. 27, pp 519-564.

Walker, Juliet E.K. (1998), The History of Black Business in America: Capitalism, Race, Entrepreneurship, Macmillan Library Reference USA, New York 1998.

Whinston, M.D., "Exclusivity and Tying in U.S. v. Microsoft: What We Know and Don't Know", *Journal of Economic Perspectives*, Spring 2001, pp 63-80.

Witt, U. (1997), " 'Lock-in' vs. 'Critical Masses' - Industrial Change Under Network Externalities." *International Journal of Industrial Organization*, Vol. 15, pp 753-773.

Woittiez, Isolde and Arie Kapteyn.,(1998) "Social Interactions and Habit Formation in a Model of Female Labour Supply," *Journal of Public Economics*, Vol.70, pp 185 - 205.