

10-3-2011

# Three Essays on Entrepreneurial Motivation, Entry, Exit and Monetary Rewards

Dmitriy Krichevskiy

*Florida International University*, [Krichevskyd@gmail.com](mailto:Krichevskyd@gmail.com)

Follow this and additional works at: <http://digitalcommons.fiu.edu/etd>

---

## Recommended Citation

Krichevskiy, Dmitriy, "Three Essays on Entrepreneurial Motivation, Entry, Exit and Monetary Rewards" (2011). *FIU Electronic Theses and Dissertations*. 519.

<http://digitalcommons.fiu.edu/etd/519>

This work is brought to you for free and open access by the University Graduate School at FIU Digital Commons. It has been accepted for inclusion in FIU Electronic Theses and Dissertations by an authorized administrator of FIU Digital Commons. For more information, please contact [dcc@fiu.edu](mailto:dcc@fiu.edu).

FLORIDA INTERNATIONAL UNIVERSITY

Miami, Florida

THREE ESSAYS ON ENTREPRENEURIAL MOTIVATION, ENTRY, EXIT AND  
MONETARY REWARDS

A dissertation submitted in partial fulfillment of the

requirements for the degree of

DOCTOR OF PHILOSOPHY

in

ECONOMICS

by

Dmitriy Krichevskiy

2011

To: Dean Kenneth G. Furton  
College of Arts and Sciences

This dissertation, written by Dmitriy Krichevskiy, and entitled Three Essays on Entrepreneurial Motivation, Entry, Exit and Monetary rewards, having been approved in respect to style and intellectual content, is referred to you for judgment.

We have read this dissertation and recommend that it be approved.

---

Jie Mi

---

Cem Karayalcin

---

Jesse Bull

---

Peter Thompson, Major Professor

Date of Defense: October 3, 2011

The dissertation of Dmitriy Krichevskiy is approved.

---

Dean Kenneth G. Furton  
College of Arts and Sciences

---

Dean Lakshmi N. Reddi  
University Graduate School

Florida International University, 2011

## DEDICATION

This dissertation is dedicated to my wife, my parents and my brother. Without their tremendous support no dissertation would have been possible. Thank you for believing in me.

## ACKNOWLEDGMENTS

First, I want to thank my mentor and a committee chair Dr. Peter Thompson, whose patience, understanding and support helped me enormously throughout this journey. I wish to thank him sincerely for teaching me rigor, doubt and independent thinking.

I wish to thank the members of my committee, Dr. Jie Mi, Dr. Cem Karayalcin and Dr. Jesse Bull, for their support and guidance. I learned a lot from Dr. Karayalcin and Dr. Bull both inside and outside the classroom. I am grateful for many stimulating conversations with both of them. I also want to thank D. Jie Mi for always being available, responsive and ready to help.

I would like to thank all the professors and staff of the Department of Economics for making my time there a very enjoyable experience.

I would like to acknowledge financial support from Kauffman and Lumina Foundations, Department of Economics, Department of Education and FIU University Graduate School in the form of Dissertation Year Fellowship.

Finally, special thanks to my wife, my parents and my brother making my education a top priority long before I valued it myself.

ABSTRACT OF THE DISSERTATION  
THREE ESSAYS ON ENTREPRENEURIAL MOTIVATIONS, ENTRY, EXIT AND  
MONETARY REWARDS

by

Dmitriy Krichevskiy

Florida International University, 2011

Miami, Florida

Professor Peter Thompson, Major Professor

This dissertation analyzes rewards and motivations of self-employment. In light of recent research contributions of Barton Hamilton (2000), which find entrepreneurship not as financially rewarding as wage work, my dissertation attempts to both verify and explain this claim. The first essay proposes a theoretical model of evolution of erroneous earnings expectations on part of a nascent entrepreneur. Inability to observe, survey, and take into account all of the returns to entrepreneurship prior to business entry creates a biased set of beliefs on part of the potential entrants. Using Bayesian learning, a nascent entrepreneur starting out with correct perception of profit distribution arrives at erroneous beliefs by incorporating limited information collected from existing businesses. An observed distribution of surviving businesses would exhibit higher earnings because of previous, unobserved, business failure entrepreneur get an overly positive view of her profit potential. Hence, the chapter offers a unique method of modeling overconfidence.

The second essay undertakes dynamic empirical comparison of earnings received by business owners and their wage counterparts. Using Survey of Income and Program Participation (SIPP) I examine both short and long run returns to entrepreneurship comparing these rewards to wage earners returns. I pay particular attention to transitions into and out of business ownership. I estimate entire earnings distribution. To characterize dynamic aspect of changes to individuals'

earnings I split the income distribution into five income quintiles and follow survey participants over the period of seven years. I find that period-to-period transitions to be Markovian. I find business tenure to be short, business ownership is costly in the short and rewarding in the long run.

The third essay considered different reporting schemes applied to the self-employed. It is another empirical investigation of entrepreneurial earning uses Panel Study of Income Dynamics (PSID). I find entrepreneurs while reporting lower than wage workers earnings enjoy significant consumption premiums. I observe evidence of income underreporting by entrepreneurs. This finding suggests a need for better earning comparison metrics and proposes to use consumption rather than income metrics for future comparisons.

## TABLE OF CONTENTS

CHAPTER	PAGE
1. Introduction.....	1
2. Limited Information and Entry Decision: A source of Entrepreneurial Overconfidence .....	7
2.1 Introduction.....	7
2.2 Literature Review.....	9
2.3 Model.....	13
2.3.1 Prior Uncertainty.....	17
2.3.2 Number of Signals .....	18
2.3.3 Noisy Signals .....	18
2.3.4 Variations in truncation point .....	19
2.4 Dynamics and empirical Suggestions .....	22
2.4.1 Number of Signals .....	25
2.4.2 Variances .....	27
2.5 Conclusion .....	28
3. How risky is Business Ownership?.....	31
3.1 Introduction.....	31
3.2 The Data.....	33
3.3 Earning Transitions in the Short and Long Run .....	36
3.3.1 Earning Distributions.....	36
3.3.2 Income Transitions and Trends at High Frequency .....	38
3.3.3 Income Transitions and Trends at Four-Month Frequency.....	44
3.3.4 Long-Run Consequences of Business Creation .....	48
3.4 Conclusion .....	54
4. A Consumption-Based Measure of the Monetary Rewards to Entrepreneurship .....	57
4.1 Introduction.....	57
4.2 Literature Review.....	60
4.3 Data.....	68
4.4 Group Differences, Underreporting and the Long Term Effects of Self-Employment .....	70
4.4.1 Descriptive Statistics and Group Differences .....	70
4.4.2 Personal Income and Expenditure Comparisons .....	73
4.4.3 Underreporting.....	75
4.4.4 Long Run Effects .....	78
4.5 Conclusion .....	81
List of References .....	85
Appendices .....	90
VITA.....	107



## LIST OF FIGURES

FIGURE	PAGE
1. Individual Income Comparison Wage Workers to Business Owners .....	36
2. Earning Densities for Business Owners and Wage Workers .....	37
3. Wage Earners Convergence Speed Example .....	51
4. Successful Business Convergence Speed Example .....	52
5. All Business Convergence Speed Example .....	52
6. Income and Consumption Densities for Wage Workers and the Self-employed.....	73
7. Long-Run Effects of Self- Employment .....	80

## LIST OF TABLES

TABLE	PAGE
1. Unconditional Transition Rates .....	42
2. Conditional Transition Rates .....	42
3. 4M Unconditional Transition Rates .....	45
4. 4M Conditional Transition Rates.....	46
5. Descriptive Statistics of Demographic Variables .....	71
6. Assets and Debt Comparison .....	72
7. Expenditure and Self-Employment.....	74
8. Income and Self-Employment .....	75
9. Evidence of Income underreporting.....	76
10. Additional evidence of Income underreporting. (10th percentile).....	77
11. Long Run Effects of Self-Employment (Median Regression).....	80

LIST OF ABBREVIATIONS

TABLE	PAGE
4M.....	Four Months
CBO.....	Congressional Budget Office
CPA.....	Certified Public Accountant
CPS.....	Current Population Survey
CSF.....	Consumer Survey of Finance
CWHS.....	Continuous Work History Sample
DHEW.....	Department of Health, Education and Welfare
ID.....	Identifier
IPS.....	Im-Pesaran-Shin
IRS.....	Internal Revenue Service
KFS.....	Kauffman Firm Survey
NFIB.....	National Federation of Independent Business
OLS.....	Ordinary Least Squares
PSID.....	Panel Study of Income Dynamics
SESTAT.....	the Scientists and Engineers Statistical Data System
SIPP.....	Survey of Income and Program Participation
USGAO.....	United States General Office of Accounting

## **I. Introduction**

Entrepreneurship is considered to be a vital element of a market society. Although entrepreneurial undertakings predate economics as a science, our understanding of motivations, rewards and entrepreneurial decision-making remains very limited. Some people end up being better entrepreneurs than others. Economists have created the term “entrepreneurial ability” term to reflect the differences. However, without individuals attempting entrepreneurship this entrepreneurial ability, strongly correlated with success of business venture, remains unobserved and thus non-measurable. At the same time society places a great emphasis on promoting and supporting business ownership. Entrepreneurial vigor, creativity, inventiveness, and dedication are credited with being the main driving force behind innovation, technological progress, economic growth, and a major building block of any progressive society.

Even the motivations for entrepreneurship remain unclear and contested. Adam Smith, the father of economic science, gave self-interest a central role in our society. He stipulated that it is self-interest that motivates people to act. In Smith’s view, business owners pursue profits. Opportunity to gain is the sole motivator for organization or increase in production. Increases in scientific rigor together with emergence of sophisticated statistical techniques, theoretical models and modern computing have enabled researches to document rewards received by both wage earners and the self-employed. Recent findings by Hamilton (2000) suggest that self-employment is not as lucrative as once believed. In fact, Hamilton found self-employment to be less rewarding than wage work. This finding stood at odds with another common belief, popularized by Frank Knight (1923), stating that entrepreneurs take on the risk of business ownership in hopes of earning more than their wage counterparts.

The voluntary nature of career choices, including self-employment, puts a big question mark on Smith’s and Knight’s assertions. How can a profit-seeking individual choose a risky venture knowing there are no rewards? Three possibilities arise from this risk-earnings paradox.

First, Adam Smith could have been wrong about entrepreneurial motivations. Something other than profit could motivate people to pursue self-employment. Second, entrepreneurs may have been unaware about their true profit prospects. Finally, modern findings (Hamilton, 2000; Moskowitz and Vissing-Jørgensen, 2002) suggesting that entrepreneurship does not pay might have missed something. Entrepreneurial motivation is hard to measure as it is both subjective and highly dependent on the success of a given venture. My research makes no attempt to measure or test entrepreneurial motivations. I believe Mr. Smith got it right nearly 250 years ago. My research takes another step in the direction of understanding profit-seeking motivations for entry, exit and rewards of entrepreneurship, suggesting and testing the possibility of misinformed entry and missing rewards.

In the first chapter I begin by putting forward a model in which entrepreneurs are profit seekers, estimating their future earnings potential and entering if those estimates exceed their wage opportunity costs. The process by which those expectations are formed is modeled explicitly with particular attention paid to the quality of information nascent entrepreneurs receive. The nature of the information available at the time of entry skews expectations of nascent entrepreneur and creates an overly optimistic estimate of profit potential. Inability to observe, survey, and take into account all of the returns to entrepreneurship prior to business entry creates a biased set of beliefs on the part of potential entrants. Using Bayesian learning, our nascent entrepreneur, starting out with the correct perception of the profit distribution, arrives at erroneous beliefs by incorporating limited information collected from existing businesses. An observed distribution of surviving businesses would exhibit higher earnings because of previous, unobserved, business failures; entrepreneurs get an overly positive view of their profits potential. The magnitude of the bias depends on the strength of prior beliefs, the number and variance of observations collected, the length of period dedicated to information collection prior to entry, and on the age of a particular industry. My model predicts that waiting and collecting earnings

information prior to entry reduces bias. Additionally, entirely new industries with no prior failures provide more balanced and less biased profit estimates for potential entrepreneurs. Having a strong prior belief would reduce this bias if the belief is consistent with the true distribution of earnings. Gathering more information in the same time period rather than having a longer observation span increases the bias. Industries where survival is toughest would create the most bias.

The second essay undertakes a dynamic empirical comparison of earnings received by business owners and their wage counterparts. I hypothesize that, if there is a short term cost associated with undertaking business ownership, it must be offset by the long term benefits if wage workers are to switch to entrepreneurship. Using the Survey of Income and Program Participation (SIPP), I examine both short- and long-run returns to entrepreneurship, comparing these rewards to wage earners' returns. I pay particular attention to transitions into and out of business ownership. I estimate the entire earnings distribution for every month in the seven years covered by the data. To characterize the dynamic aspect of changes to individuals' earnings I split the entire income distribution into five income quintiles and follow survey participants over the period of seven years using both month-to-month and quarter-to-quarter transitions. I find period to period income transitions to be Markovian regardless of observation frequency. For short term transitions there are four groups of people at any point of time: there are people who remain in either wage or business sector from one period to the next. There are two additional groups representing people transitioning from the wage sector to the business sector and vice versa. I calculate explicitly short term income mobility odds for each of the groups in the sample. I find business tenure to be short and business ownership to be costly in the short run, reflected by less favorable income mobility odds faced by agents transitioning back into wage sector paired with a high likelihood of business failure. I test my hypothesis of long-run rewards to business ownership by making use of the Markov property of period-to-period transitions. The nature of

the limiting distributions restricts my long run analysis to three groups: successful business owners, wage workers, and all entrepreneurs regardless of business success. I derive the corresponding limiting distributions for three groups of agents. I find business ownership to be rewarding in the long run, reflected by better income improvement odds faced by all entrepreneurs, successful or otherwise, when compared to wage workers.

In my third essay I question the accuracy of most wage to self-employment earnings comparisons. I hypothesize that, because of different tax code metrics and earning reporting schemes, a significant portion of earnings received by business owners fall through the cracks. Essentially there are core differences in both what and how earnings are reported as income. In the case of wage earners, their corresponding employees report wage rates, benefits and taxes on behalf of the wage workers, whereas for entrepreneurs a more complicated self-reporting scheme is implemented. The entrepreneur has an option of investing a freely chosen percentage of earnings back into his or her business. This portion of entrepreneurial earnings would not qualify as earnings kept. Additionally, business owners can deduct many expenses as monies spent on business. The line between business expenses and pure consumption is often blurry. For example, a business lunch paid by the business owners and deducted as business expense is common practice for many business relationships including sales, negotiations, and employee relationships. At the same time this business lunch leads to a clear set of savings on part of the business owner who would spend less on his personal food expenditure. A company car is another example of a business expense that the entrepreneur can use however she sees fit without it being reflected on her earnings. Hence, to examine the true benefits of being self-employed, consumption has to be considered. I claim that there is a substantial consumption premium enjoyed by the entrepreneur even if one were to exclude all the pure business expenses from the comparison. I find consumption metrics to be more universal with one notable downside: generally, consumption data are only available for the households rather than individuals.

Therefore, the definition of self-employment has to be a changed to definition of a self-employed household rather than individual. I deem any household containing at least one person who is an entrepreneur a self-employed household.

To test the income underreporting hypothesis I examine the 2003 version of the Panel Study of Income Dynamics (PSID), with which I compare wage earners to the self-employed using both income and household consumption data. I find a significantly large consumption premium enjoyed by households who have at least one entrepreneur as a member. Using established facts about entrepreneurial saving behavior together with PSID data, I find significant evidence of income underreporting on the part of the entrepreneur. In short, knowing that entrepreneurs save more than their wage counterparts, one would expect an interaction of income and self-employment when being regressed on consumption to produce a negative sign. If the sign of this term is positive, then it must be that the income data are faulty, essentially implying income underreporting. I detect income underreporting using this method and applying it to various parts of the income distribution via quintile regression. As a robustness check, I repeat the underreporting search using an alternative assumption and also detecting income underreporting. Assuming that all agents surveyed by PSID have their true marginal propensity to consume unrelated to their employment type and using ten years of income data from PSID, I detect changes in the empirical marginal propensities to consume occurring together with transitions into self-employment. I attribute these changes in the empirical marginal propensity to consume to changes in the reporting metric as it is applied upon an individual's entry into self-employment. These changes in reporting schemes distort the corresponding earnings data, creating biased income comparisons. Hence, I suggest using consumption rather than income metrics whenever there is a wage-worker entrepreneur earning comparison.

In this dissertation I make the following three contributions to the existing literature: First, I create theoretical groundwork for an overconfident entrepreneur whose initial beliefs are



correct. Previous models of overconfident entrepreneurs introduced overconfidence as an exogenous factor. My model suggests that the proper support of entrepreneurial undertakings involves relating a complete picture, including failure, to nascent entrepreneurs. Both the business school case study approach and the Small Business Administration's support structures heavily favor successful cases. This is dangerous and can lead to excessive failure and inefficiencies. Second, I observe an overwhelming number of business tenures that are very short. I document short-run costs of returning to the wage sector reflected by increased odds of moving down in the income distribution. At the same time I find that business undertaking has some benefits in the long run reflected by improved odds of moving up in the limiting income distribution. To my knowledge, this is the first time anyone has estimated long-run costs and benefits via limiting distributions. Hence, business ownership is costly in the short run and rewarding in the long run. Finally, I detect underreporting of income by entrepreneurs. This underreporting of earnings raises questions about the reliability of previous earnings comparisons and suggests a need for better earnings comparison metrics. I propose a different, consumption-based metric for comparing benefits accruing to entrepreneurs and wage workers. Using the proposed consumption metric, I document the rewards to entrepreneurship and compare them to wage work returns. I find that households who have at least one member who is self-employed to be better off in terms of consumption.

## **II. Limited Information and Entry Decision: A source of Entrepreneurial Overconfidence**

### **1. Introduction**

One of the major stumbling blocks for researchers studying entrepreneurship has been the fact that average returns to business ownership are too low. More specifically, entrepreneurs (on average) earn less than their wage earning counterparts (Hamilton 2000). This finding is true both for initial earnings and for earnings growth differentials. This paper seeks to provide an explanation for these differentials and for their persistence. The realization of this stark earnings difference is quite recent (older studies either did not find a difference or could not estimate self-employed earnings reliably). It is mostly because of the fact that true profits of entrepreneurs are an elusive variable. One of the common beliefs about business ownership is that founders often underreport their true profits by overstating business expenses. Some early studies (Brock & Evans, 1986; Rees & Shah, 1986; Borjas & Bornars, 1989; Evans & Leighton, 1989) found no significant differences in earnings between the self-employed and wage earners. In contrast, a more recent and more detailed examination has shown clear differences (Hamilton, 2000).

There are several problems from which previous studies suffered. First, the superstar theory proposed by Rosen (1981) may play a big role here; this theory suggests that wage to profit comparisons of averages may not be an appropriate measure since profits are highly skewed towards few superstars at the right tail of profit distribution. An equally important problem is that data used in all of these papers lack important information such as the length of business ownership (Hamilton 2000). Finally, net profits (as reported to the IRS) are used as a typical measure of self-employment income. Reporting less income to the IRS leaves a business' owners with more earnings to keep. Hamilton (2000) took several steps towards addressing these concerns. For example, he tracked business ownership over ten years, which allowed him to construct a better earnings approximation for the self-employed. This has enabled him to get a

better picture of business' profits since choices to sell the venture reflected some of the unreported true business values. Hamilton used the 1984 Survey of Income and Program Participation (SIPP), which tracks 8,771 male school leavers aged 18-65 over a ten year period (leavers implies all males exiting high school with a diploma, or otherwise). In addition to better approximating business' earnings Hamilton examined the business population more closely (by means of quantile regression) rather than a simple average comparison to address any superstar theory concerns. As a result, he found that self-employed earnings were dominated by incomes of wage earners with similar characteristics (such as experience, age, education, etc.). Two years later another study confirmed this result (Moskowitz and Vising-Jorgensen 2002). These findings beg the question: Why would anyone choose to become an entrepreneur?

This paper offers an explanation of this seemingly suboptimal choice. It points to the importance of information in the decision-making process of potential entrepreneurs. Information plays a vital role when evaluating the expected profits of a venture. Any agent who enters an industry must initially hold beliefs that their expected future profits will outweigh any entry costs. At the same time, these expected profits should compare favorably to the highest valued alternative. After entry, an agent compares her realized profit to her previous expectations, and updates her beliefs. In this paper I model this process (of entry, and consequential beliefs update) in order to gain insight into the entry and exit dynamics of entrepreneurship. In my model, whenever there are some informational limitations the analysis of future profit potential produces an overestimated value of expected profits. As a result there will be a number of agents entering erroneously, only to exit as soon as actual profits are realized and beliefs are updated. This need not be the case for all agents. As long as there are some who rely on erroneous information, overentry and inefficiency will prevail. As a result, there is a continuous overentry driven by new agents with incorrect expectations, which essentially explains why some entrepreneurs are overconfident. The paper proceeds as follows: Section 2 reviews the literature and examines

where this theory fits in; Section 3 develops the model; Section 4 discusses potential implications and tractable avenues for empirical examination; and, Section 5 concludes and summarizes the contributions of this paper.

## **2. Literature Review**

This paper offers a mechanism that may be responsible for the overconfident behavior of the entrepreneur. In this model, an entrepreneur behaves in what seems to be an irrational manner because there is a mistake in the information used for making an entry decision. This model neither excludes nor proves any of the previously posed explanations wrong. Other explanations for this entrepreneurial behavior and observed wage differentials include taste-for-variety, risk aversion, and overconfidence. All of the proposed explanations have some legitimacy and are not mutually exclusive. One of the most obvious ways to explain this seemingly irrational behavior is to assume that there is some additional utility that entrepreneurs get from business ownership. Hamilton himself concluded: “selfemployment earning differential reflects entrepreneurs’ willingness to sacrifice substantial earnings in exchange for nonpecuniary benefits of owning a business.” (Hamilton, 2000)

Another, the oldest explanation, dates back to Adam Smith, Schumpeter and Knight defines the role of an entrepreneur as the risk bearer. This was first formally modeled by Kihlstrom and Laffont (1979). In Kihlstrom and Laffont’s model, agents differ in their risk preferences; consequently, more risk tolerant individuals choose to be entrepreneurs. This model separates individuals into two groups (workers and entrepreneurs) based on risk. Some inefficiencies may arise from the same assumptions and mechanisms which are the driving force behind their conclusions. Since all agents (including entrepreneurs) have different risk preferences, in equilibrium entrepreneurs choose to produce different amounts of goods. This may lead to inefficiency. Another potential source of inefficiency is that wage earners do not bear

any risk at all; thus, if there are risk averse (but able) entrepreneurs, they do not get to put themselves to the best use.

Most recent and the closest to my proposed model explanation is that of the overconfident entrepreneur. In this view, the entrepreneur is not able to estimate her chances for success correctly because of the fact that she holds erroneous beliefs about her ability. Analogous to the Winner's Curse (Thaler 1992), the most overconfident person undertakes the path of business ownership. The agents not pursuing business ownership are those that are able to estimate their chances for success correctly; hence, realizing that it is not worth it, they choose not to pursue business ownership. Overconfidence theory has an entrepreneur evaluating her chances for success, which in this case means assessing her own ability.

Overconfidence has been studied for some time by psychologists, economists and business scholars. Psychologists have long known that people overestimate low probabilities (chances of winning a lottery) and underestimate high probabilities (chances of getting lung cancer for smokers). Studies of overconfidence in psychology goes back to Kahneman and Tversky (1973), with more recent experiments finding that people update their beliefs only mildly whenever presented with new and relevant information, which leads to overconfidence (Brenner et al. 1996). Business literature has also found that status quo bias is an impediment to incorporating relevant information into the decision making-process and therefore leads to biases (Samuelson and Zeckhauser 1988). Rabin (1996) wrote a bridge paper to expose economists to many of these psychological findings, including that of overconfidence. Lowe and Ziedonis (2006) found that entrepreneurs continue unsuccessful development efforts much longer when compared to corporations. But their study did not include length of time in the industry as a variable. This might simply imply that corporations are farther along their learning curve. It also seems that there is a simple selection problem associated with Lowe and Ziedonis (2006).

Namely, better start-ups are now actually corporations. In general, the results of studies linking entrepreneurs and overconfidence are mixed.

Brennan and Torous (1999) study the behavior of investors and find that investors tend to be overconfident (all investors); this is mainly driven by less than optimal portfolio diversification. Barber and Odean (2001) find men to be more confident than women. Results on whether overconfidence is socially efficient are mixed as well. Bernardo and Welch (2000) put forward a model in which the overconfidence of some individuals is essential for group survival. Puri and Robinson (2005) find that overconfident individuals choose to work more, which may be socially optimal. Eric Van Den Steen (2004) formalized choice-driven overoptimism in his paper. He models the evolution and persistence of the overconfidence with most overconfident agents pursuing entrepreneurship. The source of overconfidence in all of the above models is not explicitly modeled--some agents are assumed to be overconfident without explaining how they got there. For example, in Van Den Steen's model, agents have different priors (*ex ante*) which lead them to have different posterior beliefs (some end up forming overly optimistic expectations).

In my model, agents end up making suboptimal choices derived from the same set of prior beliefs, which is a correct set of beliefs (correct in the sense that it holds information about the true distribution of an industry's profits). Essentially, in my model, all agents know something about the true profit distribution. They all know the variance, but what is left to estimate is the expected earnings on which an entry decision is to be made. In my model, just like in Steen's, some agents end up forming an overly optimistic expectation of their profit potential. Biased posterior beliefs are a product of the limited information sample (I assume limited information is a product of competition) collected and analyzed by agents. An agent is evaluating her chances of success; her own ability is kept outside of the picture. Instead, the primary factor responsible for the estimation of future profits is an estimate derived from a sample of profit observations

collected from the industry. In reality, there are estimates of one's own ability and industry profit potentials for any agent contemplating entry. However, my model concentrates on the assessment of the state of industry's profits. More specifically, I look at the potential entrant evaluating the overall profitability of a given industry.

As in previous theories, this paper seeks to provide an explanation for the earnings differentials phenomena. In my model, overentry occurs whenever there is some profit information hidden from an entrepreneur who is contemplating entry. This information is hidden as a result of competition in the industry, which weeds out less successful firms. Over time there are some firms that draw low profits from the profit distribution, whenever these profits are too low to remain in the industry these firms exit. Hence, the remaining population of firms will have a higher average profit. Considering the population of existing firms in the estimation of future profits leads to a substantial bias since all the firms that ever had existed must be included in a true population.

Focusing attention on the problem of limited information has most recently been undertaken by Jerker Denrell (2003). He shows how undersampling may lead to systematic bias. The main argument of his paper is that firms and managers learn from each other, but in practice managers tend to pay more attention to the best managerial practices and to the best managers. This practice implies that the resulting sample is not representative of the whole manager/firm population. My model draws from the same idea. My goal is to explain the behavior of entry; specifically the entrepreneur's decision process. I model the Bayesian learning processes explicitly. I first show, how even based on the correct prior information, bias arises and persists. I then explore the properties of the bias as well as possible remedies.

### **3. Model**

To evaluate the role of information in entrepreneurial entry it is useful to first ask whether potential entrants try to gauge their chances of success using some sort of information gathering process. There is already evidence suggesting that an entrepreneur conducts a search prior to opening a business (Cooper, Folta & Woo, 1995). My paper examines the type and length of search among 1176 new entrepreneurs in a survey conducted by the National Federation of Independent Business (NFIB). Cooper et al. (1995) found that potential entrepreneurs with no relevant experience conducted an extensive search prior to entry. These researchers also found that a significant part of entrepreneurs' searches involved surveying other business owners. The second finding of Cooper et al. is that entrepreneurs with extensive experience conducted a less thorough search. A similar search process, along with a resulting entry decision, is what my model has at its core. In order to gauge agents' profit potential, entrepreneurs need to collect profit information from existing businesses to update their prior beliefs about the profitability of a venture. My model uses a Bayesian framework, where variances serve as weights in an updating process; a larger weight on prior information makes new information less valuable.

What about the finding that asserts experienced entrepreneurs search less? In essence, a more experienced entrepreneur has already collected a significant amount of information, which now serves as her prior distribution in the updating process. Thus, an experienced potential entrepreneur will spend fewer resources in going after additional information. Important evidence found by Cooper et al. is that the all entrepreneurs undertake this assessment (search) prior to entry. My model posits that this assessment is a primary criterion in the decision-making process. Cooper et al. essentially examined the evidence for the process of assessment prior to entry, which is in line with the mechanism that I propose. Cooper et al. states that the only theory (Lord & Maher 1990) which sets out to explain information search is not consistent with the evidence. Cooper et al. (1995) goes on to point out that Lord and Maher's (1990) theory suggested the



complete opposite to the evidence collected by Cooper et al. While it is not a primary goal of this model to produce a new search theory, the results concerning search are well in line with this evidence provided by Cooper et al. (1995).

To model this mechanism by which prior beliefs are updated using new information, several assumptions need to be made. First, I assume that all agents are equal in their ability. Hence, the only factors on which the entry decisions are to be made are prior beliefs about the profitability of a venture and any new information (collected by an agent) which can improve upon those beliefs. For simplicity, I assume that there are two distributions involved in the mechanism: distribution of profits (unknown to the observer) and a distribution of signals. Both of them are normal. Agents are able to estimate future profits based on signals. However, the individual profit draw, which an agent gets upon entry, remains unknown. Let the profit distribution take the following form:

$$p \sim N(\mu, \sigma_\mu^2) \tag{1}$$

Agents who choose to enter draw profits from this distribution. Assume that, while the whole distribution cannot be seen by people contemplating entry, the variance of the distribution ( $\sigma_\mu^2$ ) is common knowledge. This information is known to all agents and is essentially an unbiased (correct) prior belief. An entrepreneur who draws profit from this distribution will earn this profit from the first period after entry onwards forever. Incumbent firms communicate their true profits to potential entrants via signals  $s_i$  (whether they want to or not); these signals are stochastically related to agents' true profits. I assume that firms cannot think strategically when it comes to their signals: they have to report their profits honestly. This way, the incumbents cannot misinform the observer and deter entry. Assume a signal is characterized by:

$$s_i = p + \varepsilon_i \tag{2}$$

with  $\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$ . Here, I also assume that the precision of the white noise is known to the potential observer. A collection of these signals could be viewed as the new information

available to an entrant. These signals are used in a likelihood function to update prior beliefs and to form posterior beliefs. Posterior profit expectations could be estimated, which will enable agents to make an informed entry choice. The driving assumption behind the conclusions is the following: I assume that only operating firms can send these signals to an observer. Firms that have exited (in previous periods) no longer post this information. In reality, this information is probably obtainable but it is costly. I make these costs prohibitive in order to simplify the analysis and to have cleaner results. Immediately, it is clear that agents could differ in two respects: 1) agents know that this inaccessible information (profits of firms that have failed) exists; and, 2) agents are not aware of the fact that some information is missing. For simplicity, call them “Informed” and “Uninformed” agents, respectively. I further assume that entrepreneurs cannot make buying offers to one another, making it impossible for informed agents to take advantage of the uninformed. Since the distribution of profits is truncated from the left it essentially hides a portion of signals from an observer. Another important assumption is that those firms which perform poorly fail and exit. This supposition enables me to simplify exit decisions to a one dimensional analysis in which only profits and alternative wages matter to entrepreneurs. Entrants end up using a simple decision rule: getting high enough profit means staying while drawing a low profit forces them to immediately exit. Let  $\beta$  represent the truncation point separating failed firms from incumbents. An observer collects  $n$  signals from a Normal distribution with an unobserved mean and  $\sigma_\varepsilon^2$  variance.

Applying the standard Bayesian formula (e.g., de Groot, ch. 9), the posterior belief of an uninformed agent is:

$$E(p|s_1, s_2, \dots, s_n) \equiv \mu_n = \frac{\mu\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + n\sigma_\mu^2} + \frac{\sigma_\mu^2 \sum_{i=1}^n s_i}{\sigma_\varepsilon^2 + n\sigma_\mu^2} \quad (3)$$

Because the uninformed agent believes she is well informed she just assumes that the observed  $n$  signals are drawn from a standard normal distribution. At the same time, an informed agent is well aware of the fact that the distribution is truncated such that  $s_i \geq \beta$ . She is, however,

interested in a true value of  $\sum_{i=1}^n s_i$ , which would have been observable if the distribution was not truncated. The expected value of the truncated signals is:

$$E[s] = \hat{p} + \sigma_\varepsilon \frac{e^{-1/2 \frac{(\beta - \hat{p})^2}{\sigma_\varepsilon^2}}}{\int_{(\beta - \hat{p})/\sigma_\varepsilon}^{\infty} e^{-1/2 t^2} dt} \quad (4)$$

where we denote agent's maximum likelihood estimate of profit as  $\hat{p}$ . This estimate is based solely on the information available to an observer. Then,  $\hat{p}$  is given by the solution to (4), (see, for example Selvin (1976, eqn [2])), so the expected value of observed signals is defined implicitly by:

$$\hat{p} = \frac{1}{n} \sum_{i=1}^n s_i - \sigma_\varepsilon \frac{e^{-1/2 \frac{(\beta - \hat{p})^2}{\sigma_\varepsilon^2}}}{\int_{(\beta - \hat{p})/\sigma_\varepsilon}^{\infty} e^{-1/2 t^2} dt} \quad (5)$$

Let  $\bar{s} = \frac{1}{n} \sum_{i=1}^n s_i$ , and let  $g(\bar{s})$  denote the above solution. The decision maker knows (or is able to derive) (5). She will then use this estimate in a similar fashion (as in (3)) to obtain:

$$E(p | s_1, s_2, \dots, s_n : s > \beta) = \frac{\mu \sigma_\varepsilon^2}{\sigma_\varepsilon^2 + n \sigma_\mu^2} + \frac{\sigma_\mu^2 n g(\bar{s})}{\sigma_\varepsilon^2 + n \sigma_\mu^2} \quad (6)$$

An informed agent will use all of the available information (including relevant information about this truncation), and she then incorporates all of this information in her estimate. Equation (6) provides the correct Bayesian posterior belief about profit potential. This correct belief is accurate in the sense that it is unbiased. Bias is then determined by subtracting (6) from (3):

$$B_n = \frac{n \sigma_\varepsilon \sigma_\mu^2 (\bar{s} - g(\bar{s}))}{(\sigma_\varepsilon^2 + n \sigma_\mu^2) \int_{(\beta - \hat{p})/\sigma_\varepsilon}^{\infty} e^{-1/2 t^2} dt} \quad (7)$$

*Proposition 1: If the likelihood function is truncated from the left, then the observer who uses Bayesian updating will overestimate his/her expected profits. ( $B_n > 0$ )*

Note immediately that  $B_n = \frac{n\sigma_\varepsilon (\bar{s} - g(\bar{s}))}{\int_{(\beta-\bar{p})/\sigma_\varepsilon}^{\infty} e^{-1/2t^2} dt} * \frac{1}{(\sigma_\varepsilon^2 + n\sigma_\mu^2)}$  with  $\frac{n}{(\sigma_\varepsilon^2 + n\sigma_\mu^2)} > 0$ , and

$$\text{and } \bar{s} - g(\bar{s}) = \frac{\sigma_\varepsilon e^{-1/2 \frac{(\beta-\bar{p})^2}{\sigma_\varepsilon^2}}}{\int_{(\beta-\bar{p})/\sigma_\varepsilon}^{\infty} e^{-1/2t^2} dt} > 0 \text{ for all parameters implying}^1 B_n > 0 \blacksquare$$

The above proposition states that whenever a potential entrepreneur is observing an industry and does not take previous exiting firms (and their profits) into consideration, she is bound to end up with an overly optimistic estimate for her profit potential.

i) *Prior Uncertainty*

Bias has several parameters that affect its magnitude. I first consider prior beliefs and how the quality of these beliefs affects the resulting estimate. Prior information is captured by  $\sigma_\mu^2$ , which reflects the quality of prior beliefs. Differentiating bias with respect to  $\sigma_\mu$  is straight forward and yields:

$$\frac{\partial B_n}{\partial \sigma_\mu} = \frac{2ne^{-\frac{(\bar{p}-\beta)^2}{2\sigma_\varepsilon^2}} \frac{\sqrt{2}}{\sqrt{\pi}} \sigma_\mu \sigma_\varepsilon^3}{(n\sigma_\mu^2 + \sigma_\varepsilon^2)^2 (1 + \text{Erf}\left[\frac{\bar{p}-\beta}{\sqrt{2}\sigma_\varepsilon}\right])} \quad (8)$$

This effect is strictly positive for all values of parameters (i.e.,  $\partial B_n / \partial \sigma_\mu > 0$ ). This means that the greater one's prior uncertainty, the worse the bias. This arises from the simple fact that a reduction in the precision of correct information ( $\sigma_\mu^2$  is the variance of the true profit distribution) forces an observer to put more weight on the biased information captured by the likelihood function. One can think of this as a situation in which this agent has an initial idea about profit potential. The profit potential is denoted by an interval. This agent then goes on and

---

<sup>1</sup> For a more formal discussion of posterior moments arising from truncated likelihood, see Balakrishnan and Ma (2001).

collects more information. But, the wider the interval the more uncertain is the agent about what to make of this information. Thus, she trusts this newly collected (biased) information more as a result.

ii) *The number of Signals*

Does collecting more signals help? Intuition suggests that a deeper search may return better (less biased) results. As it turns out, collecting more signals does not help ( $\partial B_n / \partial n > 0$ ). Collecting more signals means that an observer is actually more misinformed. This is due to the fact that collecting signals from truncated distributions forces an observer to place more weight on the information carried by these signals. Therefore, increasing the number of signals does not alleviate the bias because all other signals are drawn from this incomplete distribution, which is misleading. Thus, getting larger samples of these signals is not only unhelpful in reducing the degree of misinformation, but it even makes the problem worse.

iii) *Noisy Signals*

If the collected signals are very noisy, then the correlation between an agent's profit and the profits of incumbents is weak. Knowing that the noise level is high, an observer will not know what to make of this information and will put less weight on the information extracted from signals; this should reduce resulting bias. However, if truncation is severe (truncation point is to the right of the mean), then it may actually help to have less noise. The overall direction of this parameter change is unclear. To see which restrictions are needed to provide a definitive, direction consider the entire expression:

$$\begin{aligned} \frac{\partial B_n}{\partial \sigma_\varepsilon} = & \frac{2e^{-\frac{(\beta-\hat{p})^2}{2\sigma_\varepsilon^2}} \frac{(\hat{p}-\beta)^2}{2\sigma_\varepsilon^2} (\beta-\hat{p})\sigma_\mu^2}{\pi\sigma_\varepsilon^2(n\sigma_\mu^2+\sigma_\varepsilon^2)(1+\text{Erf}\left[\frac{\hat{p}-\beta}{\sqrt{2}\sigma_\varepsilon}\right])^2} - \frac{2e^{-\frac{(\hat{p}-\beta)^2}{2\sigma_\varepsilon^2}} \sqrt{\frac{2}{\pi}}\sigma_\mu^2\sigma_\varepsilon^2}{(n\sigma_\mu^2+\sigma_\varepsilon^2)^2(1+\text{Erf}\left[\frac{\hat{p}-\beta}{\sqrt{2}\sigma_\varepsilon}\right])} + \frac{e^{-\frac{(\hat{p}-\beta)^2}{2\sigma_\varepsilon^2}} \sqrt{\frac{2}{\pi}}\sigma_\mu^2}{(n\sigma_\mu^2+\sigma_\varepsilon^2)(1+\text{Erf}\left[\frac{\hat{p}-\beta}{\sqrt{2}\sigma_\varepsilon}\right])} + \\ & + \frac{e^{-\frac{(\hat{p}-\beta)^2}{2\sigma_\varepsilon^2}} \sqrt{\frac{2}{\pi}}(\hat{p}-\beta)^2\sigma_\mu^2}{\sigma_\varepsilon^2(n\sigma_\mu^2+\sigma_\varepsilon^2)(1+\text{Erf}\left[\frac{\hat{p}-\beta}{\sqrt{2}\sigma_\varepsilon}\right])} \end{aligned} \quad (9)$$

Severe truncation, while mathematically feasible, is not going to be considered in this model. I assume that  $\beta < \hat{p}$ . A situation in which  $\beta > \hat{p}$  would imply an entrepreneur is observing an average firm and that this average firm is failing. This should mean that no one at all would enter. A closer look (see Appendix 1 for Proofs) reveals that  $\beta < \hat{p}$  is not a sufficient condition for the overall effect to be positive. There is a necessary and sufficient condition (see Appendix 1) which ensures that the derivative is positive.

$$\frac{\partial B_n}{\partial \sigma_\varepsilon} > 0 \quad \text{if and only if} \quad \frac{e^{-\frac{(\beta-\hat{p})^2}{2\sigma_\varepsilon^2}} \frac{(\hat{p}-\beta)}{\sqrt{2\pi}(1-\text{Erf}\left[\frac{\beta-\hat{p}}{\sqrt{2}\sigma_\varepsilon}\right])}} + \frac{2\sigma_\varepsilon^4}{n\sigma_\mu^2+\sigma_\varepsilon^2} > \sigma_\varepsilon^2 + (\hat{p} - \beta)^2 \quad (10)$$

If (10) is satisfied, then as signals get noisier the overall bias will get worse. This comes from the fact that the difference between  $\bar{s}$  and  $g(\bar{s})$  gets larger as  $\sigma_\varepsilon^2$  gets larger. At the same time, the above condition (10) identifies a range of parameters for which this derivative is negative (if both inequalities are reversed). In this case an observer will place less weight on the signals (incomplete information, which is misleading) thus improving her posterior beliefs.

iv) *Variations in truncation point.*

If a truncation point is moved to the left, leading to better information, then it should be obvious that there should be less bias as a result. While it is true for the severe truncation case, it is not always the case for mild truncation (see Appendix 1). An additional condition is needed to insure that increases in truncation increase the bias; namely,

$$\frac{\partial B_n}{\partial \beta} < 0 \quad \text{if and only if} \quad \sqrt{2\pi} (\hat{p} - \beta) \left( 1 + \text{Erf}\left[\frac{\hat{p}-\beta}{\sqrt{2}\sigma_\varepsilon}\right]\right) > 2\sigma_\varepsilon e^{-\frac{(\beta-\hat{p})^2}{2\sigma_\varepsilon^2}} \quad (11)$$

While these findings examine the bias and its properties, the factors that determine this truncation threshold for a particular industry remain outside of this framework. Nonetheless, these are very important. What determines the profits threshold necessary for a particular business? Why do these thresholds differ, and what does that mean for entrants? These questions are not addressed directly by the Bayesian mechanism of this model, but there are obvious determinants of the success thresholds for any particular business. These thresholds are higher for businesses where high entry fees need to be recouped. Increasing the sunk costs of entry decreases the number of entrants. Thus the number of those entering erroneously decreases as well. They (agents who entered misguidedly) then exit upon the first profit realization. I take higher entry costs to mean that it takes larger profit estimates to fool agents into entry (because agents would expect to recoup these costs over time). A number of agents will enter, only to realize that their expectations were too high. But many will remain in business since entry costs become sunk costs after an entry decision is made. Not exiting and not recouping costs implies migrating to the left side of the profit distribution. We should observe lower rates of failure for high entry costs (such as buying capacity settings – cars, planes manufacturing) as well as for expensive red-tape businesses (nuclear energy, natural resource extraction). More importantly, this becomes a self-sustaining mechanism; industries with low failure rates will continue to have low failure rates, and industries with high failure rates will sustain high failure rates. In case of the low entry costs industries, agents who drew low profits exit (unlike those in high sunk costs cases). Consequently, to the observer it seems that those remaining agents earn a good profit and this observation, paired with low costs of entry, should motivate a lot of entry and as a result more failure as well. These high failure rates will have a larger portion of industry hidden from some observers. This creates a large number of overly optimistic entrants; they enter the industry and fail, thus creating high failure rates.

Even though my model predicts this self-feeding mechanism for any given industry, there are changes in these thresholds for the number of industries (Dunne et al. 1989). Clearly, there are other mechanisms responsible for these changes. The case could be made for both inter and intra-industry differences and changes in thresholds. Truncation points could change over time due to some exogenous factors such as foreign competition. If we were to examine an industry which has gone through a severe drop in profits (due to foreign competition) it should have higher standards for firms' survival. Meaning, only the best would remain which would essentially create a new, higher truncation point. In such cases we should see a change in hazard rates as well. Immediately after a shake-out an observer will see fewer firms earning profits (keep in mind that an observer did not monitor the industry prior to shake-out). Fewer incumbent firms imply smaller samples from which estimation is done. Entering entrepreneurs are less misinformed compared to the same industry populated by more firms. Some examples of these types of industries include shoes, steel, knives, wood mills, fisheries, etc. Another factor that may change profit thresholds is some regulatory change which would make entry tougher or easier. Any of these exogenous shocks which change industry dynamics, profit levels, and fixed costs. These shocks will lead to new hazard rates. All of these should change failure rates by varying  $\beta$  in distribution of signals and thus affecting the degree of bias.

These are all static comparisons, and saying what happens over time does not fit in well with the current set up of the model. In order to switch to a dynamic setting, we need an agent who can observe a change in the industry and/or who observes several periods (essentially updating her beliefs more than once). As a result of numerous instances of updating, she should be closer to the correct estimate. This is because she observes as a failure as well as success.



#### 4. Dynamics and Empirical Suggestions

This current setup is limited to a one-shot “update-and-decide model”. Even though agents have prior beliefs, and they update those beliefs upon collecting new information, there is nothing to suggest continuous updating. One natural extension is to see what the value of waiting is. I now proceed to extend the model in this direction. Let us consider the very same updating done over “ $t$ ” periods. Simply, updated expectations for profit potential in period one becomes a prior belief in period two. This dynamic can be captured by the difference equations as follows:

$$\mu_{t+1} = \frac{\mu_t \sigma_\varepsilon^2}{\sigma_{\mu(t)}^2 n + \sigma_\varepsilon^2} + \frac{\sigma_{\mu(t)}^2 n \bar{s}}{\sigma_{\mu(t)}^2 n + \sigma_\varepsilon^2} \quad (12)$$

and

$$\sigma_{t+1} = \frac{\sigma_\varepsilon^2 \sigma_{\mu(t)}^2}{\sigma_{\mu(t)}^2 + n \sigma_\varepsilon^2} \quad (13)$$

Note that in the above equations “ $t+1$ ” subscripts denote posterior values and “ $t$ ” subscripts denote prior values. The likelihood function, which includes the latest sample mean and variance, is free of time subscripts to avoid confusion. Thus,  $\bar{s}$  captures an average value of signals from the newly collected sample in this current period. It is important to note that this sample contains the profits from all firms, including those that drew lower than expected profits and are exiting in the same period. One would expect that  $\bar{s}$  in the early periods of observation would contain incumbent firms as well as new entrants and new exits. As observation continues, prior beliefs become less biased and each successive posterior estimation improves in quality. Potential entrepreneurs can now see exiting firms by virtue of observing the industry dynamics given by (12) for numerous periods, without them having to enter. Observing numerous periods should improve an observer’s understanding of the true profit potential. It should be immediately clear that (13) contains only correct information and therefore will not be a source of bias. Recall that the variance of the prior is known to all agents and serves as a correct belief; henceforth, I

drop time subscripts on variances. Without loss of generality let us assume that the initial condition for (12) yields number one (it makes no difference, as long as any finite number is chosen). Thus, solving (12) yields a solution for some arbitrary period  $\tau$ :

$$\mu_\tau = -\frac{\bar{s}\sigma_\varepsilon^2\left(\frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + \sigma_\mu^2 n}\right)^{\tau-1}\left(1 - \left(\frac{\sigma_\varepsilon^2 + \sigma_\mu^2 n}{\sigma_\varepsilon^2}\right)^\tau\right)}{\sigma_\varepsilon^2 + \sigma_\mu^2 n} + \left(\frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + \sigma_\mu^2 n}\right)^{\tau-1} \quad (14)$$

One would expect bias to disappear in this case if observation continues infinitely. As our potential entrepreneur begins her observation in any period, she is initially misinformed. Her first estimate will be severely influenced by the fact that the industry has been around for some time in conjunction with the assumption that previous failures are not taken into consideration. But as she continues to monitor the industry, her estimates improve (since now complete dynamics are visible). Eventually, the updating process outweighs initial misinformation. There is new and complete information. Thus, if an agent begins her observation at any given period “s” and continues to update her beliefs as in (12), she is then updating her  $\mu_s$  which is initially biased by observing  $\bar{s}$ . Note, this average  $\bar{s}$  is less biased after numerous periods and represents an average drawn from the complete signal distribution because all of the signals from the beginning of observation are now included. But will this bias be eliminated completely? Taking the limit of equation (14) yields:

$$\lim_{\tau \rightarrow \infty} \mu_\tau = \bar{s}.$$

This is an intuitive result. It states that longer observation of an industry reduces bias. An agent observing the complete lifespan of an industry is not misinformed at all. This is an important point in and of itself. Consider the beginning of any industry: agents that entered have not been misled by years of failure. Agents that have started at time zero have no information to rely on. And, in this case this information is not misleading to early entrants. One must be careful about the term “early entrant”. The term “early entrant” could mean two different things. First,

an early entrant is an agent who has entered an industry in the early stages of the industry's lifecycle. Second, an early entrant is an agent who decided to wait less before to making an entry decision. This model states that the former would be less misinformed by definition while the latter is really misinformed. She can use some observing time to remedy the information driven bias (reduce the bias by observing the industry's dynamics).

The number of periods for which information is missing matters as well. An agent entering some industry in its infant stage does not have many firms that had previously exited. Take an extreme case where there are few firms in the industry and not one firm has failed in the previous few periods. An observer contemplating entry into this industry is not misinformed at all; in fact, she observes the entire profit distribution. Whereas an industry that has been able to weed out a number of firms over its lifetime potentially misleads an observer, suggesting that incumbent firms' profits are representative of the whole population. In this case an agent can improve the quality of her expectations by observing several periods completely (failures included).

i) *Number of Signals*

I now turn my attention to other possible changes that may take place over time. I first consider changes in the number of incumbent firms. The number of signals could either change over time due to some exogenous factors or as a result of internal entry and failure dynamics. Exogenously dictated differences are, for example, those industries which have different sizes and thus different number of signals *ex ante*. These industries can be compared to one another. Alternatively, the number of signals could change with time due to some outside factors. Resulting differences in the number of incumbent firms will change hazard rates from one industry to the next based on different  $n$ . In the case where the number of signals increases due to the internal dynamics of an industry, it has to be driven by larger number of firms drawing profits

above the exit threshold, put it more simply, there is a decrease in failure rates. At the minimum, this should imply that the profit distribution's left tail is being pulled in (same result could be due to the whole distribution stretching up, and pulling both tails in), in which case less distributional mass falls below the exit threshold. We should see less profit variance in this type industry and more firms similar to one another. The whole industry should begin to look more homogeneous, with fewer small firms and fewer large firms in term of profits. This intuition could only be used if the overall profit distribution is unchanged otherwise (i.e., profits for the industry and entry rates will remain fixed). If, however, there is something else going on which effects the shape or the size of a profit distribution (i.e. changes in entry and exit rate or movement within distribution) the above conclusions cannot be reached. For example, in a growing industry an increasing number of signals could be the outcome of expansion. In this case, not much could be predicted about changes in distribution. In the case of a declining industry, changes in the distribution of profits are also an outcome of incumbent firms capturing the market share of exiting firms. So it is difficult to analyze these dynamics since for most industries the life cycle tends to be increasing at first and declining thereafter (Horvath et al 2001, Klepper and Simons 1993). It is then likely that some exogenous factors are responsible for this single-peak evolution process.

A more interesting case to consider is that of industries which are different in size due to market demand. Larger industries (in terms of number of firms) should have more misinformed entrants when compared to smaller industries. This scenario holds true for industries with low or no entry costs and should be especially true for industries populated by small firms. Smaller firms tend to be self-funded. Therefore, entry choices are made by an entrepreneur based on search rather than by capital providers like venture capitalists or banks based on their assessment. Differences can arise from the fact that a person contemplating entry collects information from existing firms and updates her beliefs in Bayesian fashion. Thus, the number of signals matters.

Consequently, in the case of banks and venture capitalists they (capital providers) have their own information and their own criteria for funding. A search conducted by an entrepreneur matters less. Banks and venture capitalists have their own assessment formulas and techniques based on a much wider sample. Banks and venture capitalists have been observing an industry for much longer than any potential entrepreneur. Hence, having this larger quantity of signals drawn from truncated distribution will create a more biased posterior belief about potential profitability of a future venture. At first glance, the current evidence (correlating the number of firms to hazard rates) does not support this intuition (comparison was made based on Baldwin et al. 2000 study). Perhaps this suggests that a search is more local and would not be reflected in nationwide statistics about the size of an industry. Maybe more controls would need to be implemented (concentration ratios, industry life cycles, and macroeconomic variables). Obviously, clustering effects and geographical distribution will vary greatly depending on an industry. A look into a local population (city and metro areas) could be a more informational examination of this theory. Study examining search before entry do indeed point to a more local search (Cooper et al. 1995).

*ii) Variances*

Changes in variances also have interesting effects on posterior beliefs. Altering either one or both the variance of the prior beliefs or the variance of new signals being collected will result in changes to posterior expectations. Over time increases in the prior's variance means that new information from one period to the next is diminishing in quality (since a posterior in period one becomes a prior in period two). The prior variance is the "correct" information in the sense that it is made up of unbiased beliefs about the true population and of all the information gathered from observing a certain number of time periods completely (an observer must have seen a full population of signals since the initiation of industry observation). These complete observations include profits of all successful and failed firms from the beginning of observation. Increases in

this variance means that this correct information has decreased in quality and, as a result, it will be trusted less. An important conclusion is that earlier entrants (in the sense of entering in the early days of an industry) are misinformed less. Thus, there should be more successes among them. This model does not produce mechanisms which would be responsible for this change in variances; these mechanisms are responsible for a lot of industry dynamics and need to be investigated. Nonetheless, even if one was to take these changes as completely exogenous and apply it to this model, hazard rates will change. This can be empirically tested by a close examination of the particular profit distributions.

Literature on the life cycles of an industry has been pointing to the fact that many industries evolve in a single peak fashion, with the number of firms and profits initially going up and then steadily declining over time (Horvath et al 2001, Klepper & Simons 1993). Klepper and Simons (1993) also show early entrants to be more successful when compared to the later ones. And finally, the literature points to industry concentration decreasing during the period of growth, and increasing in decline period (Horvath et al 2001, Klepper and Simons 1993). This is consistent with my model's predictions about failure in conjunction with changes in prior variance. As prior information decreases in quality (precision drops), more misinformed agents enter and immediately fail. Incumbents then get the market share of the exiting firms. Clustering and geography may once again prove to be of importance here. As an industry grows (both geographically and in product variety), it is harder to keep track of all of the relevant information out there (since search is found to be local by Cooper et al 1995). This implies the quality of prior beliefs decreases.

Increasing the new signal's variance is a more helpful method of reducing the bias. Agents collecting and analyzing signals will trust (put less weight) on this "incorrect" (incorrect in the sense that the bias arises from the fact that it is not complete) information consequently

reducing the bias. In this case earlier (in terms of industry's age) entrants will have lower failure rates compared to the late ones because of more unobserved failures over time.

Another interesting story is the interaction of the two variances. For example, if both are increasing but the variance from new signals increases faster (the variance of prior information need not be increasing at all), then it could lead to a situation similar to the observed industry's life cycle. More specifically a shakeout will occur as the weights of variances switch rankings. Incorrect information enters into a decision making process as a more important component; later, the "correct" information takes over as a more important factor leading to a decrease in entry. The results from one industry to the next will vary because of different fixed entry costs, but as long as there are some entry costs a number of agents will mistakenly enter (hoping to regain fixed costs over time); these agents stay even though the costs cannot be recouped, but they move to the lower tail of the profit distribution. Recall, in the case of no entry costs, agents just stay or exit as soon as profits are revealed. So compared to the no entry costs case, a different fraction of entering agents will immediately exit. For that reason, fixed costs will also have a significant role in determining hazard rates in any particular industry.

## **5. Conclusion**

My model shows how seemingly irrational behavior of new entrants can be explained by a lack of complete information. The overestimation of future earnings potential and subsequent over entry is driven by the assumption that income information containing profit levels of firms who have gone under is not accessible. Whenever an observer contemplating entry is not aware of the limited nature of information, she would be bound to overinflate her estimate of earnings potential. This result arises even though the entrant has a correct prior belief. This correct information gets distorted during Bayesian updating, involving an analysis of incomplete information. I also find that increasing the number of signals during sampling is likely to make

the resulting bias worse. On the other hand, collecting information about the entire profit population (observing for several periods prior to entry) helps to reduce bias. For any agent having more precise information about prior beliefs helps, while having more precise information about current profit distributions is likely to make the bias larger. Dynamic changes in system parameters such as the variances of prior and/or signal distributions, and the number of incumbents produce results which may explain life cycles of an industry as well as the different hazard rates observed. Empirical verification is needed where the hazard rate would serve as a dependent variable. The lifespan of an industry, number of incumbents, and quality of information should serve as independent variables. This paper serves as a pilot investigation into the role of information for the study of entrepreneurship.

Information plays a vital role in any kind of decision-making process and it cannot be neglected. If we assume that entrepreneurship is a resource which needs to be supported, then policymakers need to aid business owners' to maximize their success. This paper suggests that potential entrepreneurs need to be better informed about their true chances of success. Business school provides an excellent example: most case studies examine successful firms, successful entrepreneurs, and choices that proved to be successful; it may therefore create a false picture of business potential. Some of these "educated" students go on to open their own companies. Beliefs about profit potentials are heavily distorted by these few superstar entrepreneurs (Bill Gates, Warren Buffet and the like); this implies that even estimates of average chances for success may not be the best prior belief. The same logic applies to high school sports, where many students count on becoming professional athletes. Lending institutions chose to fund an entrepreneur or not. A person who does not get a loan attributes her failure to the lender's lack of faith in her success. Better informed people contemplating startups at and before the lending stage (regardless of loan approval), will produce more optimal talent allocation and fewer bias driven failures. This does not imply that everyone wishing to open their own companies is capable of doing so. This



model can and should be extended in both a theoretical direction and verified empirically. An investigation incorporating Bayesian updating, limited information and entrepreneurial entry decision with the general equilibrium model is a natural extension of this research. Adding strategic interactions, where agent can misinform each other to deter entry, as well as a set up where repeated entrepreneurs are possible and different agents will sell businesses to one another because of different expectations are also useful avenues for further investigation.

### **III. How risky is business ownership?**

#### **1. Introduction.**

The failure of countless many suggests that business ownership is a risky undertaking and the extent of the risks faced by entrepreneurs is the subject of this paper. While company workers can face termination risk incurring lost wages (and, perhaps, partially vested 401k company contributions), the closing of a business potentially induces much bigger losses for the entrepreneur. Entrepreneurship has always been associated with risk,<sup>2</sup> but do entrepreneurs enjoy risk or do they tolerate it in exchange for larger expected payoffs?

Kihlstrom and Laffont's (1979) seminal model, in which risk preference plays an important role in agent's choices between entrepreneurship and wage employment, motivated a number of empirical studies evaluating the risk-return tradeoff for entrepreneurs. These studies have not arrived at a consensus of an opinion. Some studies (e.g., Rosen and Willen, 2002; Quadrini, 2005; Gort and Lee, 2007) provide evidence that expected entrepreneurial earnings exceed those of wage earners with similar characteristics, which is consistent with a standard story of risk-averse entrepreneurs obtaining higher expected returns in exchange for greater risk. Others (e.g., Hamilton, 2000; Moskowitz and Vissing-Jørgensen, 2002) find lower mean or median earnings for entrepreneurs, leading us either to invoke a preference for risk or the existence of nonpecuniary benefits obtained from business creation.

A number of reasons – including income definitions, sample definitions, and duration— are collectively responsible for these diverse findings. For example, Gort and Lee's finding of a premium to entrepreneurship is based on the SESTAT data, a notably high-tech and high-income sample of employees and entrepreneurs; in contrast, Hamilton's results are based on the SIPP, a

---

<sup>2</sup> Interestingly, the early meaning of 'entrepreneurship' had nothing to do with business or self-employment, and referred instead to risky behavior in general. For example, Lemaire de Belges (1510) described the Trojans as entrepreneurs. Only later did the definition narrow to refer only to risk takers associated with business creation and ownership (Cantillon, 1755; and famously Knight, 1921).

relatively low-income sample. Rosen and Willen (2002) also establish their standard mean-variance tradeoff in earnings data from the Panel Study of Income Dynamics (PSID), but they restrict their entrepreneurial sample to individuals that have continued in self-employment for at least three years (i.e., those that are more successful than average). In stark contrast, Moskowitz and Vissing-Jørgensen (2002)'s conclusions derived from an analysis of the returns to investment portfolios. Even within samples, the reported return to entrepreneurship appears to vary. Hamilton's conclusions also depend on how self-employment income is measured: for example, only two of his three measures yield a mean less than the average wage. Like Gort and Lee, Braguinsky and Ohyama (2007), study SESTAT data; but in contrast to Gort and Lee, they find that, while the mean return to self-employment is strongly positive in skill-concentrated occupations, it is strongly negative in occupations that require only general skills.

This paper contributes to our understanding of the risk and return to entrepreneurship by studying individual income transitions along with transitions into and out of self-employment. My paper is of course not the first to estimate income transitions in this context: Quadrini (1999) and Bradford (2003) calculated transition matrices for two- and three-period income transitions, from which they found that persistence in entrepreneurship is correlated with higher earnings. Unfortunately, their dynamics provide only limited insight into the effect of entrepreneurship on lifetime earnings, since they are based on only one- or two-step long-term transitions. This paper extends their approach with a closer examination of transitions over 84 high-frequency time periods.

My paper offers a new perspective on risk. By analyzing a large national dataset I am able to follow thousands of individuals over a period of seven years and compare business owners to wage earners. I not only estimate and track over time unconditional means and income distributions, I track *individual* earnings mobility through five distinct quintile earnings classes. I also characterize the process governing transitions from one period to the next, which allows me

to estimate transition probabilities reliably. I find that average entrepreneurial earnings dominate those of wage earners. However, the higher average earnings of entrepreneurs are skewed by a few very successful individuals at the top of the distribution: the *median* entrepreneur is strictly worse off monetarily compared to wage earners.

Since returns to entrepreneurship likely increase with persistence, I undertake a dynamic analysis of income to see if the entrepreneurship income penalty evident in the short and medium terms can be recovered in the long run. The transitions from one period to the next are stationary and have the Markov property. The estimated Markov transition matrices are therefore invertible, enabling me to obtain empirical estimates of the ergodic distribution of earnings. Contingent on keeping one's business, entrepreneurs are strictly better off than those staying in the wage sector. However, few entrepreneurs manage to survive. Entrepreneurs who return to the wage sector turn out to face very uncertain outcomes, and may do markedly worse or markedly better than those who remained as wage workers all the time. On average, however, a period of entrepreneurship followed by a return to wage work reduces earnings. Business ownership is relatively risky, and business failure is costly.

The remainder of the paper is structured as follows: In Section 2, I describe the data and note some limitations; in Section 3, I describe my findings; and in Section 4 I offer some conclusions.

## **2. The Data**

I use the Survey of Income and Program Participation (SIPP), a large national panel survey that was created in the late 1970's by the Department of Health, Education and Welfare (DHEW). Early panels contained overlapping waves of 20,000 households with a new panel beginning each year. The survey was redesigned in 1996, widening the survey to include non-overlapping panels of 37,000 households initially, with the first panel running for 4 years and

every subsequent panel running for 3 years. Some other major changes introduced in 1996 were: an expansion of the scope of information collected, improved imputation procedures, and computer-assisted interviewing. These changes enhanced longitudinal consistency and widened the scope and quality of the information collected.

The most attractive feature of SIPP is that it is a large panel dataset focused primarily on earnings. There are several other large national surveys with earnings data, such as the Current Population Survey (CPS), the Panel Study of Income Dynamics (PSID), and the Kauffman Firm Survey (KFS), but for my purposes the SIPP has several advantages over them. First, the KFS only surveys businesses, making it inappropriate for this paper. The CPS is a repeated cross-sectional survey of households, so changes in income for individuals and assets cannot be observed over time. The PSID is a large household survey (approximately 9,000 households) that conducts annual interviews, whereas the SIPP conducts interviews every four months to estimate monthly data, and has a more detailed inquiry into assets and business variables of the households. Indeed, one of the biggest strengths of the SIPP is that it collects very detailed high-frequency information pertaining to income, assets and wealth of individuals, families and households. The importance of different assets and especially different income categories cannot be understated. Income can be examined in detail to observe where the earnings come from (this is especially valuable when examining business income and how it translates into business owners' incomes). The SIPP includes detailed information on many income categories such as wage income, business income, real estate, and financial assets income. This ensures that business earnings, oftentimes not reported in traditional earnings categories, are not understated.

I combine the 1996 and 2001 panels resulting in seven years of monthly data. The 1996 panel contains twelve waves, while the 2001 panel has nine. I also create an individual level dataset by combining household IDs (sample identifiers) with the IDs of individuals in each of the households, allowing me to go inside the household to make individual and subfamily

earnings comparisons. I replicated the analysis for the households, families, sub-families, and individuals. Since the results were almost identical, this paper reports earnings dynamics at the level of the individual.

I restrict my sample to working-age men (18-65) in non-farm sectors.<sup>3</sup> Women were excluded from the sample because their labor force participation rates and incomes are often family dependent. I further restrict the sample to people who remained active in the survey throughout the periods covered by both the 1996 and the 2001 panels. Finally, I eliminate the non-response survey subjects, as well as subjects whose income information was missing or incomplete for more than two periods. The final sample consists of 14,381 men interviewed over 84 months. Around nine percent of the sample were business owners at some point during the observation period.

In order to estimate individual income transitions, I group income into 5 quintiles. Additionally, I created a binary variable for business ownership, where a value of 1 was assigned to a person who owns at least one business and 0 to everyone who does not own a business. I also created variables indicating which month of the survey a business was opened, and variables that capture time, length, and the number of times any person occupied any given quintile. Using these variables I am able to track movements of individuals among quintiles while relating these movements to changes in business ownership status.

---

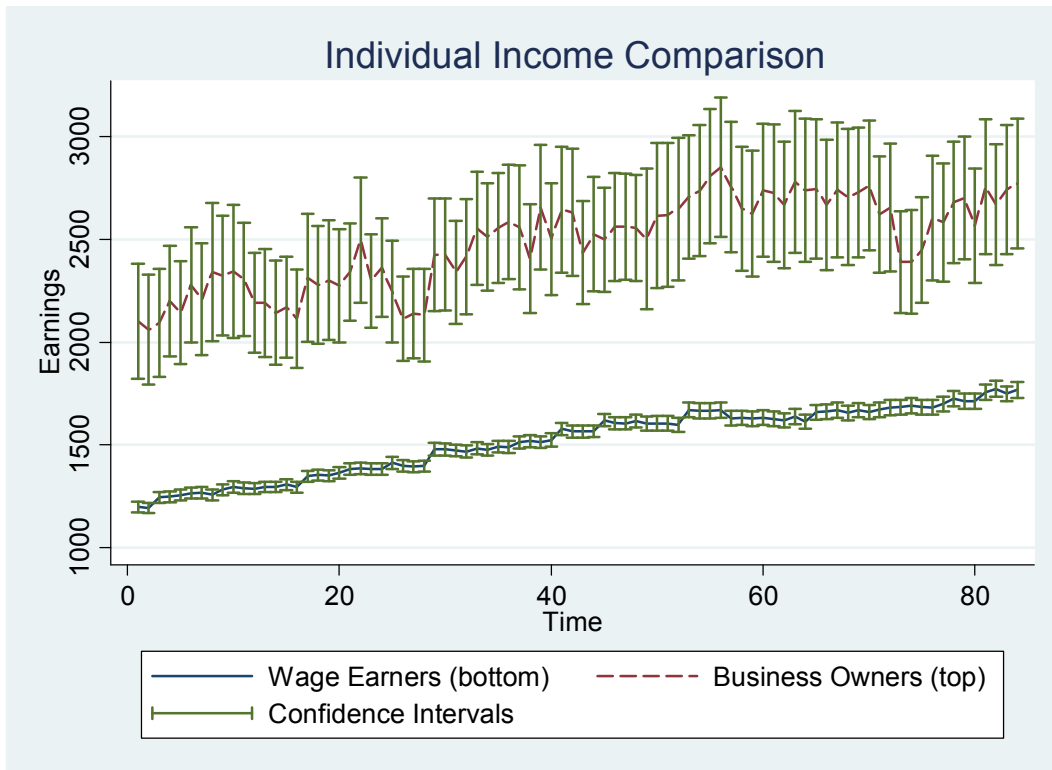
<sup>3</sup> I exclude farmers for two reasons. First, farm ownership is likely driven by quite different dynamics than non-farm business ownership. Second, the earnings of farmers are often heavily dependent on government transfers, which is not a desirable feature in earnings comparisons.

### 3. Earning Transitions in the Short and Long Run

#### 3.1 Earning Distributions

Figure 1 plots the mean earnings in each sample period for business owners and wage earners. Consistent with the findings of Quadrini (1999), average earnings of business owners consistently exceed those of wage earners.<sup>4</sup> In results not reported here, similar differences were found also at the household and family levels, while at the subfamily level the difference was not statistically significant.

**Figure 1. Individual Income Comparison Wage Workers to Business Owners**

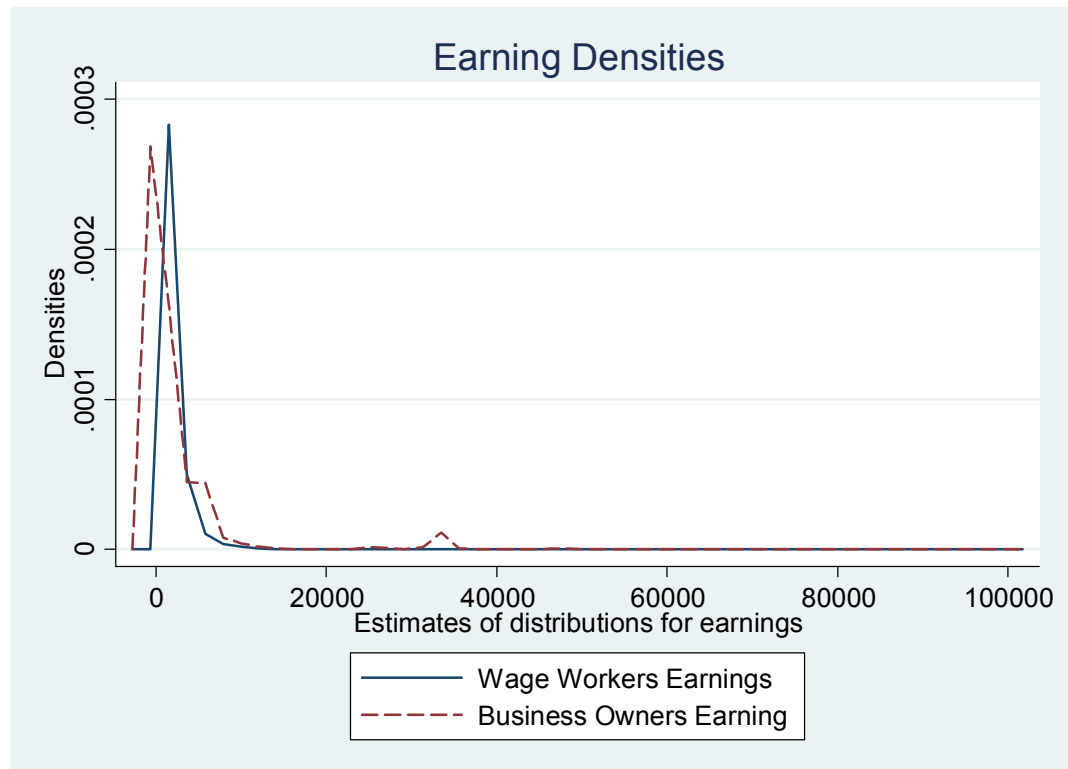


Such raw comparisons, of course, tell us little about the expected return to entrepreneurship for an individual. Indeed, in stark contrast to Quadrini, Hamilton has found that after conditioning on observable attributes, the earnings of the self-employed are generally lower

<sup>4</sup> A Kolmogorov-Smirnov test confirms that earnings of business owners are significantly different from those of wage earners.

than the earnings of wage workers. Moreover, as Figure 2 indicates, the higher mean earnings for business earnings are largely driven by the high earnings of a small minority of presumably unusually successful entrepreneurs.

**Figure 2. Earning Densities for Business Owners and Wage Workers**



Following Hamilton, I am interested in earnings expectations at the individual level. While Hamilton compared the earnings of business owners with their observationally-equivalent wage earning counterparts, I take the rather different approach of estimating individual earning transitions.



### 3.2 Income Transitions and Trends at High Frequency

To capture the dynamics of income transitions I discretize income into quintiles obtained from the income distribution of all observations. I create these categories for each of the 84 months in the sample. Consequently there are 83 transitions for each individual in the sample, although not all of the individuals have complete monthly income data. There are five income classes and two employment types. I will therefore estimate a single  $10 \times 10$  transition matrix, which I will call the unconditional matrix; this matrix provides the rates of transitions between each earnings quintile  $\times$  employment type. I estimate separately four  $5 \times 5$  transition matrices, one for each of: agents who are business owners across both periods, agents who are wage workers in both periods, agents who switch from wage work to business ownership, and agents who switch from business ownership to wage work. I will refer to these four matrices as the set of conditional matrices.

Transition probabilities need to have the Markov property in order for us to obtain meaningful estimates from panel data. However, missing values limit my ability to test for unit-roots, and only two panel data procedures, Im-Pesaran-Shin (IPS) (Im, Pesaran et al. 2003) and Fisher-type (Choi 2001) tests, are available. Both tests rely on asymptotic assumptions and are justified as  $T \rightarrow \infty$ , but this is reasonable given our long panel of 84 periods. The details of both tests are given, along with the results, in Appendix 2.A. Both tests yield results consistent stationarity in the transition probabilities.

Table 1 provides estimates of the mean unconditional monthly transition rates, while the four conditional transition matrices are given in Table 2. To highlight the conditionality of Table 2, the four matrices have been separated by gray borders.<sup>5</sup> I estimate Markov probabilities corresponding to a transition process from one period to the next. I combine Markov transition

---

<sup>5</sup> These are not, of course, separate estimates. It is in fact easy to derive the four sub-matrices in Table 2 from the numbers in Table 1.

probabilities into one 10x10 matrix (Table 1). The estimated probabilities represent estimates for any person in the sample transitioning from one period to the next. I also estimated the confidence intervals for these conditional transitions; those are given in Appendix 2.B.

Consider first the unconditional transitions. To interpret Table 1, consider the upper row, which reports the transitions of individuals that in the current period are business owners in the lowest income quintile. Some 75.1 percent of these individuals remain as business earners in the same earnings category a month later, while 0.58 percent remain in the lowest earnings quintile same. Summing the two numbers, I find that 75.7 percent of these agents remain in the same earnings quintile. Summing the numbers in columns 6 through 10, I find that 2.45 percent of business owners in the lowest earnings quintile had returned to wage work the next month. The numbers are similar for each of the five initial earnings categories

The rate of exit from business ownership is much greater than the rate of entry into business ownership. For example, only 0.23 percent of low-income wage workers switched to business ownership the following month. This is as expected in a stationary environment, because wage employment is much more common than business ownership. To see this, let  $p$  be the proportion of low-income agents that are wage workers. Then, if flows into and out of business ownership are equal, it must be the case that  $0.23p = 2.45(1-p)$ . This requires that  $p=0.91$ , or that the incidence of entrepreneurship among this income group is about nine percent; this number is about right.

Although the majority of agents remain in the same income group, Table 1 also reveals a considerable amount of volatility. At first, this month-to-month earnings volatility may seem high (especially when looking at the wage earners); however, several explanations and similar findings point to these estimates as both reasonable and in line with related prior research. Although earnings volatility studies using the methods used in this study have not been done before, there are several, more general, earnings volatility studies completed recently. Using the same data

over a longer time period and employing slightly different sample selection criteria the Urban Institute has recently published a study aiming explicitly at measuring income volatility and changed to income volatility over time (Acs, Loprest, and Nichols 2009). Researchers find similar magnitudes of earnings volatility for American households. In another recent investigation the U.S. Congress has commissioned a similar study into changes of earning volatility. The work, completed in 2007 by the Congressional Budget Office, also found similar earnings volatility using the Continuous Work History Sample (CWHHS) dataset. One notable disagreement between these two works is that the former finds a gradual increase over time in earnings volatility while the latter does not. More importantly, both reports find volatility similar to that described in this paper. Different long run volatility trends found by the two papers arise from different samples, methodology and time length of the surveys.

There are at least three reasons for this observed high earnings volatility: First, bi-weekly pay systems create irregularities in monthly income recording. Simply put, dividing 26 time periods between 12 months is usually done without splitting any of the pay periods to equalize monthly payments. This, of course, creates two months with three paychecks rather than two – thus contributing to some of the observed volatility. A second reason for the volatility is the recall error of the respondents resulting from them having to recall information from the previous four months to report monthly observations. This is also known as the “seam bias”, where respondents are more likely to assign an event to the first time period rather than any other (see, for example Hill 1987). Lastly, there are entire industries for which volatility is driven by seasonal demand fluctuations. For people occupied in tourism, services, retail, real estate etc. earnings vary greatly between months within the high demand season and other times. Often, there are small mini-seasons such as winter holidays and the spring break. For people employed in sectors sensitive to severe swings in demand, earnings will vary greatly for both the self-employed and wage earners.

Bi-weekly pay is the most likely of the three phenomena to affect self-employed and wage workers differently. At the same time it is hard to make any conjectures without asking respondents explicitly what kinds of wage scheme workers employ vs. what kind of earnings disbursement scheme is implemented by business owners. Income misreporting is also to blame for some inconsistencies. Additionally, there are life events such as college graduation, disability, bonuses, unpaid family leave, alimony etc. which change possible earnings. These events influencing earnings' volatility are not measured well in the SIPP; however, there is no reason to suspect them of causing systematic biases: wage workers and entrepreneurs are all subject to those events. Hence, it is reasonable to proceed with wage worker – business owner earnings comparison.

Table 1 suppresses an especially important aspect of earnings volatility that is made clear in Table 2. The likelihood of staying in the same income category from one period to the next is much higher for people who either stayed in the same job or continued with business ownership. Changing occupational type induces large probabilities of moving up and down one or more income categories. Uncertainty associated with the career changes is clearly reflected in the top right and bottom left panels of the tables. Consider a simple metric, by which an event with over fifty percent probability of being realized is considered likely and an event with less than fifty percent probability is termed unlikely. Using this metric, it is clear that for people continuing in their respective employment type (i.e., wage or business ownership) remaining in the same income category is likely for all income quintiles, whereas people switching between wage work and business ownership (in either direction) are more likely to move to other quintiles of the income distribution. However, the conditional matrices in Table 2 reveal that even continuing in wage or self-employment is associated with considerable earnings volatility. The findings in Tables 1 and 2 are in line with those from similar studies (Acs, Loprest, and Nichols 2009; Congressional Budget Office 2007).

**Table 1. Unconditional Transition Rates**

		Business Owners (t+1)					Wage Earners (t+1)				
		1	2	3	4	5	1	2	3	4	5
Business Owners (t)	1	75.09	10.17	4.31	3.34	4.66	0.58	0.43	0.35	0.39	0.70
	2	8.93	70.84	10.48	3.67	3.38	0.49	0.65	0.49	0.37	0.69
	3	4.20	9.60	68.43	10.84	4.45	0.54	0.54	0.50	0.58	0.33
	4	3.23	2.70	8.51	70.47	12.39	0.49	0.42	0.68	0.68	0.42
	5	2.43	1.56	2.97	6.39	83.95	0.44	0.39	0.56	0.68	0.63
Wage Earners (t)	1	0.04	0.02	0.04	0.05	0.08	61.06	12.12	8.99	8.68	8.93
	2	0.01	0.02	0.03	0.03	0.05	10.62	57.23	14.46	8.97	8.57
	3	0.03	0.02	0.02	0.03	0.08	7.88	14.11	52.55	16.56	8.72
	4	0.03	0.04	0.02	0.04	0.06	7.58	8.34	15.57	53.01	15.33
	5	0.03	0.03	0.03	0.05	0.07	7.54	7.82	8.10	14.88	61.47

**Table 2. Conditional Transition Rates**

		Business Owners (t+1)					Wage Earners (t+1)				
		1	2	3	4	5	1	2	3	4	5
Business Owners (t)	1	76.97	10.42	4.42	3.42	4.77	23.81	17.46	14.29	15.87	28.57
	2	9.18	72.80	10.77	3.77	3.48	18.18	24.24	18.18	13.64	25.76
	3	4.30	9.84	70.17	11.12	4.56	21.67	21.67	23.33	23.33	13.33
	4	3.32	2.77	8.75	72.42	12.73	18.31	15.49	25.35	25.35	15.49
	5	2.50	1.60	3.05	6.57	86.28	16.22	14.41	20.72	25.23	23.42
Wage Earners (t)	1	16.67	7.14	19.05	22.62	34.52	61.19	12.14	9.01	8.70	8.95
	2	9.84	13.11	18.03	22.95	36.07	10.63	57.32	14.48	8.99	8.59
	3	15.58	11.69	12.99	18.18	41.56	7.90	14.13	52.64	16.59	8.74
	4	16.67	20.51	8.97	21.79	32.05	7.60	8.36	15.59	53.10	15.35
	5	14.61	13.48	14.61	23.60	33.71	7.55	7.83	20.02	14.91	61.60

Agents who did not experience an occupational change have the most income stability and, among them, agents in the top and bottom quintiles are the most stable. For agents transitioning into business, the likelihood of moving into the top quintile is fairly large (over 30 percent for all quintiles). The same odds do not apply for agents going back into the wage sector. Agents finding out that business ownership is not for everyone are more likely to experience a less favorable outcome when going back into the wage sector. This can be deemed a short run cost of business failure, and may arise because of, *inter alia*, gaps in the resume created by non-verifiable self-employment and the stigma of failure caused by closure of one's business. For business owners, the odds of going into the top quintile are best for agents from the lower two and top quintiles of the income distribution. It is no surprise that people from the top quintile in one category would transition into a top quintile of another group. -- these are most likely high-end entrepreneurs. But observing people from the lowest quintiles moving up in income is surprising. This finding suggests that, perhaps, some valuable experience is gained by agents in the two lower income groups while they are self-employed allowing them to perform better upon their return to the wage sector. Alternatively, these could have been people with well-paid jobs who thought they can do better on their own, only to find out otherwise and who are now going back to their jobs. Despite good outcomes for some people, going back into the wage sector is on average costly when compared to the odds of higher earnings faced by wage workers.

Monthly transitions provide a useful insight into earnings fluctuations, especially when occupational transitions are involved, as they produce estimates of immediate welfare changes for the individuals. Clearly, choosing to start a business is associated with better immediate odds of improving one's earnings than is transition in opposite direction. This is expected, because as at least a portion of the transitions back into wage employment is an outcome of business failure. However, as these estimates are subject to distortions resulting from seasonal fluctuations, "seam bias" and bi-weekly payment schemes, they should be interpreted cautiously. Nonetheless, as

these three distortions apply to business owners and wage workers alike, comparisons between the two groups are warranted.

Nonetheless, I attempt to reduce the magnitude of these biases and check robustness of these findings by aggregating observation periods into a longer time periods. Aggregating into an annual observation is difficult as the partial non-response for many agents in a given year reduces the number of observations significantly. Annual income is not reported in this survey and has to be aggregated from the monthly interview data. Aggregating income into annual data would only be meaningful for agents present for all the interviews in a given year. All others would have to be dropped, severely reducing the size of the sample. Non-response is nonrandom and, as such, it would increase the attrition bias. Individuals experiencing income drops are more likely to leave the panel (Acs, Loprest, and Nichols 2009). To avoid issues of increased attrition associated with the annual data, I choose to examine four-month intervals. Four-month intervals are the most attractive because the interviews take place every four months. Thus, every respondent interviewed is asked about four months of data and if there is no response it is the case for all four months. Using the four month income data will reduce the “seem bias” the most. I report transitions using wave-to-wave (four months) data.

### **3.3 Income Transitions and Trends at Four-Month Frequency**

I repeated the analysis after replacing the monthly transition rates with four-monthly (hereafter 4M) transition rates and monthly income with 4M income. The data are constructed by taking eight sequential monthly observations. The monthly income data reported each quarter are summed for the each quarter. Summed values give rise to new 4M income distributions. I split these new distributions into five quintiles for each 4M period, analogs to the monthly analysis. The resulting data are the twenty-one 4M income distributions split into five income quintiles. Agents for whom 4M data is incomplete are dropped from the analysis since including them

would give rise to erroneous income data. Transitions into and out of business are recorded as agents report their employment status for each 4M period. Because interviews take place every four months, the vast majority of agents report the same status for the entire 4M period. For agents reporting changes in business ownership status, I use last month of the 4M period as an indicator of their business status. I report resulting 4M transition estimates in Tables 3 and 4, below. As another robustness check, I run the same unit root tests for 4M transitions.

Unsurprisingly, I do not detect an autoregressive component in transitions.

**Table 3. 4M Unconditional Transition Rates**

		Business Owners (t+1)					Wage Earners (t+1)				
		1	2	3	4	5	1	2	3	4	5
Business Owners (t)	1	8.64	0.65	0.78	1.56	1.10	18.18	18.44	17.01	17.08	16.56
	2	0.46	5.10	0.59	1.27	1.27	13.03	20.50	20.16	18.94	18.68
	3	0.38	0.87	5.87	1.19	1.61	13.97	17.64	19.91	18.79	19.77
	4	0.51	0.72	1.02	6.39	1.83	13.03	19.51	19.72	19.00	18.25
	5	0.55	0.74	1.06	1.19	6.90	13.28	18.00	20.97	18.47	18.85
Wage Earners (t)	1	0.41	0.55	0.67	0.73	1.09	50.33	11.71	11.92	11.38	11.21
	2	0.55	0.88	1.01	1.16	1.66	13.86	19.66	20.64	20.63	19.94
	3	0.46	0.87	0.98	1.14	1.74	12.90	19.64	21.17	21.03	20.06
	4	0.43	0.87	1.07	1.22	1.72	12.78	19.62	21.19	20.88	20.20
	5	0.51	0.90	1.09	1.30	1.75	12.66	19.24	20.60	21.08	20.88



**Table 4. 4M Conditional Transition Rates**

		Business Owners (t+1)					Wage Earners (t+1)				
		1	2	3	4	5	1	2	3	4	5
Business Owners (t)	1	67.86	5.10	6.12	12.24	8.67	6.45	22.58	22.58	25.81	22.58
	2	5.34	58.74	6.80	14.56	14.56	22.22	22.22	19.44	16.67	19.44
	3	3.87	8.80	59.15	11.97	16.20	18.75	25.00	34.38	9.38	12.50
	4	4.87	6.88	9.74	61.03	17.48	33.33	15.15	18.18	15.15	18.18
	5	5.27	7.10	10.14	11.36	66.13	22.22	0	33.33	25.93	18.52
Wage Earners (t)	1	20.00	20.00	4.00	28.00	28.00	52.13	12.13	12.34	11.78	11.61
	2	41.67	8.33	12.50	16.67	20.83	14.63	20.75	21.79	21.78	21.05
	3	25.00	13.89	11.11	16.67	33.33	13.61	20.72	22.33	22.18	21.16
	4	31.58	15.79	26.32	10.53	15.79	13.50	20.73	22.38	22.06	21.33
	5	21.05	23.68	21.05	10.53	23.68	13.41	20.37	21.81	22.32	22.10

There is an especially striking contrast between Tables 1 and 3. The probabilities of staying in business after four months drop significantly. This is a very telling result as it points to some other factor – not captured here. I investigate the average time span and individuals spends in a particular quintile (not reported here) and find wage earners spend more than double the time in each income category when compared to business owners. This is to be expected, as many more business owners survive two months as compared to those with over four months of business tenure, leading them to move about the income distribution. Reasons for the short business tenure could be numerous, among which expectations mismatch is probably the most important factor. These erroneous expectations could be a mix of earnings and costs misjudgments leading to an early exit. As a result of this phenomenon business owners are likely to transition quickly back into wage employment. For business owners in the lowest tail of the income distribution, the transition back into wage work is likely to be into the two lowest income

quintiles. Business owners transitioning from the second and third quintiles are more likely to see a small improvement, reflected in improved odds of upward income mobility. Business owners from the top two quintiles are more or less equally likely to end up in any of the top four quintiles. Wage workers in the lowest quintile of income distributions are the worst off, as their chances of staying in the lowest quintile are highest. For wage workers in top four quintiles, there is more mobility and the chances of moving to any other one of the top four quintiles are about equal. It is also clear that business owners from the lowest quintile fare better than their wage counterparts from the same. In other words, if entry costs are not too high, undertaking business ownership is not a bad idea, especially for wage workers in the two lowest income categories.

Table 4 is similar to Table 2, although there are some differences of note. In terms of similarities, it is clear that surviving businesses are relatively stable in both tables. The biggest difference is how unstable income is for wage workers. Unlike the month-to-month transitions, wave-to-wave transitions show a lot more volatility for wage workers. The only stable income category for the wage worker is the lowest quintile of the distribution where agents are more likely to stay rather than transition to another quintile. For the remaining top four quintiles wage workers face about the same odds of moving to any one of the top four quintiles. Previously mentioned biases are applicable here as well, with perhaps the exception of the “seam bias.” At the same time, eight-month periods are more likely to be affected by job changes, bonuses, unpaid leaves and other events of a similar nature when compared to the two-month time spell. This is not the case for business owners as they cannot leave their businesses and can enjoy bonuses as part of earnings on continuous bases.

Earnings, conditioning on the survival of the business, are more stable in the business sector. This is not surprising because of going back to wage work is generally associated with poor performance of the business. Hence, business owners who are doing well, reflected by their

choice to remain in business, face better odds of doing well in the future. At the same time, the true risk of entering into business is displayed in the unconditional estimates (Table 3) rather than in conditional estimates (Table 4), because the former incorporate the likelihood of business' survival and business survival is not very likely. A clear conclusion emerges: business ownership is risky. At the same time, it seems that income from wage work is not as stable as is commonly believed. Even with the short tenure in business the attempt of entrepreneurial undertaking may have positive influence in future earnings. On one hand, there is immediate improvement in earnings odds for people entering into business ownership. On the other, there is a penalty for agents going back into the wage sector. One question remains open: Who is better off in the long run? In the next section I estimate terminal outcome of these monthly income transitions.

### **3.4 Long-Run Consequences of Business Creation**

One of the most attractive features about estimating transition matrices is that doing so allows us to make conjectures about the long-term consequences of entrepreneurship at horizons well beyond those covered by the sample data. In this section, I make use of the transition matrices in Table 1 and 2 to explore the following questions.

- How do the earnings of agents that create a business in period 0 compare at some future period  $t$  with the earnings of agents who never create a business?
- How do the long-term earnings of agents that create a business in period 0 and that who survive in business compare at some future period  $t$  with the earnings of agents who never create a business?

These questions can be answered by straightforward manipulation of the estimated transition matrices. To obtain the distribution of earnings of agents that never create a business, I use the lower right matrix from Table 2, beginning with the steady-state wage distribution. To

consider the evolution of the earnings distribution of business creators who never fail, I make use of (i) the steady-state wage distribution and the relative distribution of transitions from the bottom half of the first column of Table 2 to calculate the initial business earnings distribution, and then I use the matrix in the upper left of Table 2.

Then, given an initial distribution vector  $\mathbf{d}$  and a transition matrix  $P$ , the distribution of earnings by time  $t$  is given by the repeated ( $t$  times) transition through the probability matrix<sup>6</sup>

$$P = \begin{bmatrix} p_{11} & \dots & p_{51} \\ \vdots & \ddots & \vdots \\ p_{15} & \dots & p_{55} \end{bmatrix},$$

where  $p_{ij}$  denotes an actual probability of transitioning into a  $i^{\text{th}}$  income quintile from  $j^{\text{th}}$  income quintile at any point of time. These probabilities are calculated from actual movements of all agents through the distribution vector  $\mathbf{d}$  over the entire life of the panel. After  $t$  transitions the probability matrix is given by

$$P^t = \begin{bmatrix} p_{11} & \dots & p_{51} \\ \vdots & \ddots & \vdots \\ p_{15} & \dots & p_{55} \end{bmatrix}^t$$

and as  $t \rightarrow \infty$  limiting distribution can be derived by solving for  $p_i$   $i = 1, \dots, 5$  the following system of six equations:

$$\begin{cases} p_1 = p_1 \cdot p_{11} + p_2 \cdot p_{21} + p_3 \cdot p_{31} + p_4 \cdot p_{41} + p_5 \cdot p_{51} \\ p_2 = p_1 \cdot p_{12} + p_2 \cdot p_{22} + p_3 \cdot p_{32} + p_4 \cdot p_{42} + p_5 \cdot p_{52} \\ p_3 = p_1 \cdot p_{13} + p_2 \cdot p_{23} + p_3 \cdot p_{33} + p_4 \cdot p_{43} + p_5 \cdot p_{53} \\ p_4 = p_1 \cdot p_{14} + p_2 \cdot p_{24} + p_3 \cdot p_{34} + p_4 \cdot p_{44} + p_5 \cdot p_{54} \\ p_5 = p_1 \cdot p_{15} + p_2 \cdot p_{25} + p_3 \cdot p_{35} + p_4 \cdot p_{45} + p_5 \cdot p_{55} \\ p_1 + p_2 + p_3 + p_4 + p_5 = 100 \end{cases}$$

I then to derive the limiting distributions for wage workers, all agents attempting business (this includes both failing and successful businesses), and the successful business owners. The

---

<sup>6</sup> See for Example Clark, B. and Disney, R. (1985) "Probability and Random Processes: A First Course with Applications", 2<sup>nd</sup> Edition. John, Wiley and Sons: 218-245

corresponding solution vectors are given by **W** for wage workers, **B** for all agents attempting business and **S** for successful business owners:

$$\mathbf{W} = \begin{pmatrix} 17.77102 \\ 19.73451 \\ 20.00963 \\ 20.98742 \\ 21.49742 \end{pmatrix}; \quad \mathbf{B} = \begin{pmatrix} 18.00877 \\ 18.50449 \\ 18.58422 \\ 20.42565 \\ 24.47687 \end{pmatrix}; \quad \mathbf{S} = \begin{pmatrix} 15.77287 \\ 15.99597 \\ 16.93581 \\ 18.73297 \\ 32.56238 \end{pmatrix};$$

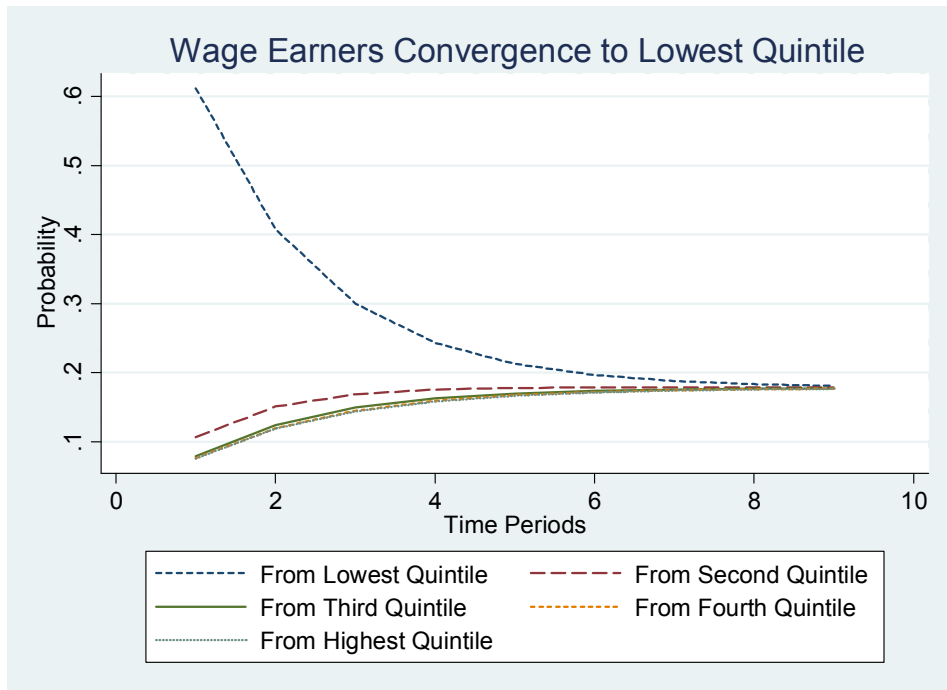
Notice that one desirable estimate of a limiting distribution is missing. One may want to figure out the limiting distribution for failed business owners explicitly. However, this is not possible as a result of the nature of these transitions. To get to a limiting distribution agents have to transition through the same probability matrix repeatedly until it converges on the limiting distribution vector. Clearly, failing business owners transition back into the wage sector and would go through the corresponding wage workers' matrix in the following period. Hence, repeated transition through "failing business" matrix is not possible by the definition of business failure.

Successful business owners have a much higher probability (almost 33 percent), compared to two other groups chances, of ending up in the top quintile of income distribution. Their odds of ending up in the top *two* quintiles, at 57 percent, are also much higher than is the case for the other two groups. Unsurprisingly, perhaps, success in business is financially rewarding. This is especially true since the decision of exit is endogenous and any agent performing below her expectations can simply quit and go back to wage work. A more important comparison is that of chances which wage workers face to those trying their luck at business ownership. This is especially true since people undertaking a risk of business ownership mostly are not aware of their entrepreneurial ability, which they can only find out by giving business ownership a try. Comparing all business to all workers vectors quintile by quintile shows that for most part agents attempting business ownership are slightly better off in the long run in terms of

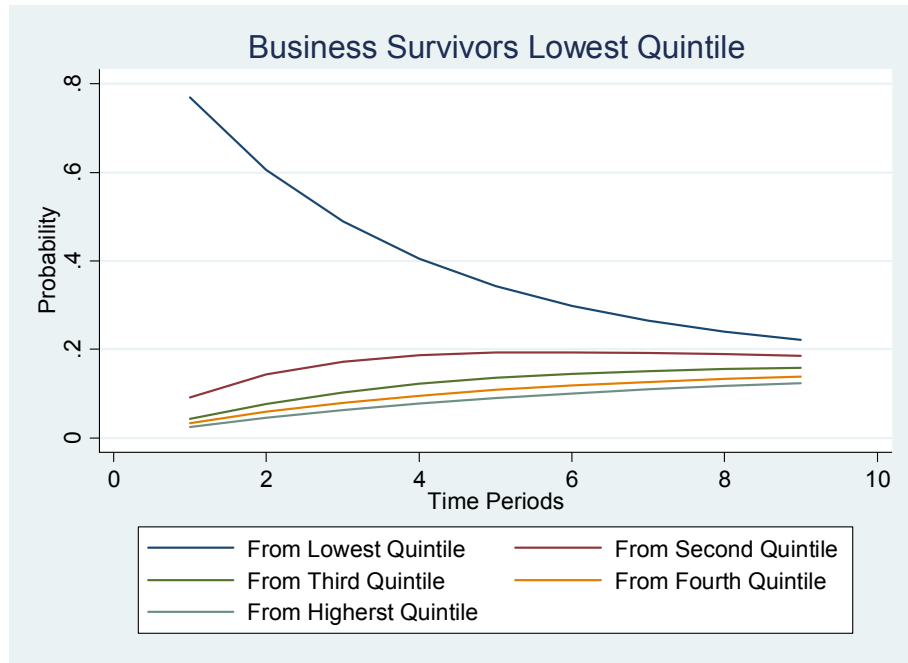
their odds of getting into the highest income category. People attempting business ownership have 24 percent chance to getting into the top quintile compared to 21 percent for the wage workers. For wage workers the odds of ending up in two top income quintiles are about 42.5 percent, whereas for business owners those are 44.9 percent. At the same time business undertakers have a slightly higher, less than half a percentage point, chance of finishing in the lowest quintile of the income distribution when compared to the chances faced by wage workers.

Naturally, one would like to know how fast the probabilities approach these limiting distribution estimates. I calculate the transition matrices for each of the groups for 9 periods. By the 9<sup>th</sup> transition through the matrices of Tables 1 and 2, the income distributions have converged very close to the calculated values of the limiting distribution. There are, however, some differences in convergence speed; most notably, surviving business owners exhibit slower convergence rates. Figures 3 ,4 and 5 depict convergence to selected quintiles.

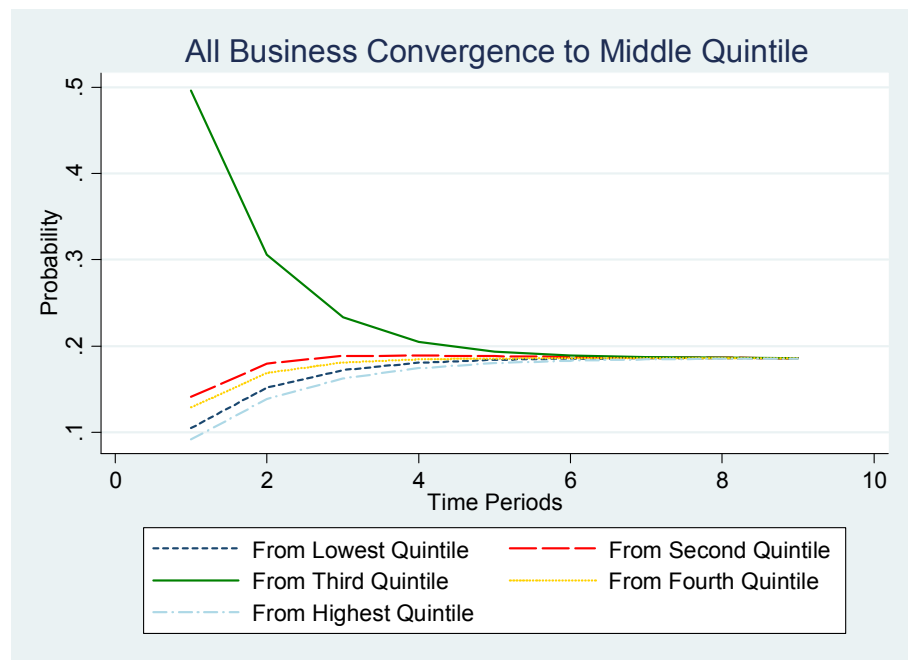
**Figure 3. Wage Earners Convergence Speed Example**



**Figure 4. Successful Business Convergence Speed Example**



**Figure 5. All Business Convergence Speed Example**



From the convergence rates shown in the figures, it is clear that the limiting distribution is reached fairly quickly. The odds for surviving businesses remain more uncertain for a longer period of time. In other words, the income category of a successful business owner remains dependent on his or her initial earnings quintile for a longer period of time than is the case for the other groups. The odds for the remaining two groups settle on their corresponding limiting values fairly quickly. This dichotomy adds to another dimension of stability of wage work (since the **B** vector is also driven by people returning into the wage sector). Business success while likely to be rewarding, eventually, takes a bit longer for all entrepreneurs to be facing settled odds depicted by the limiting distribution vector. Within nine months of entering monthly transition matrices, wage workers and all business entrants face specific odds of their final income quintile destination. Successful entrepreneurs take longer than 9 months to arrive at their limiting distribution values. At the same time it is clear that it does not take many years for all participants to get to their respective limiting distribution vectors. These clearly established odds are the most attractive feature of limiting distributions as it does not matter where one starts from, within short period of time (9 months for wage workers and all people trying entrepreneurship, and a bit longer for successful entrepreneurs), agents in the same group face the same income transition odds. This phenomenon makes limiting distribution vectors a very attractive tool, which can be used for group comparisons.

What is most important is that there is no monetary penalty for giving business ownership a shot. In other words, rational, risk neutral individual, should take her chances at business ownership as she would expect to be at least as well off as continuing employment in the wage sector. This is especially true if we are to believe that all agents are not aware of their entrepreneurial ability *ex ante*, an assumption supported by the existence of very short business tenures. Moreover, such short business tenure is not problematic if there is no income penalty for trying business ownership.



#### 4. Conclusion

My paper contributes to the wage-worker entrepreneur earnings comparisons by undertaking a careful examination of the dynamical aspect of earnings. Individuals' movement within the income distribution is examined for both wage workers and entrepreneurs with particular attention paid to switches from one type of employment to the other. Dividing the income distribution into quintiles helps in capturing the movement between those income classes, as it can be characterized explicitly. Transitions from period to period are found to exhibit the Markov property. These immediate transitions can be considered as short-run income changes. In the short run, I find it is beneficial to enter self-employment but costly to transition back into the wage sector. This is the cost of entrepreneurial failure.

Additionally, the long-term likelihood of income transitions movements is estimated via the derivation of limiting distributions. These limiting distributions represent steady state type of probabilities of transitioning from one income category into the other faced by groups of agents: wage workers, all people trying business ownership out, and successful business people. Convergence onto the steady state of probabilities depicting transitions from one income category into another is fairly fast – meaning the long run realization is likely to occur anywhere within six month to two years. One notable exception is that successful business owners take longer to arrive at their respective limiting distribution. Hence, the odds of long term earnings changes take longer to settle for the established business owners.

The main finding of this study is that business ownership is risky. The risk, however, is mainly reflected in the very short tenure of business ownership and drastic changes to one's earnings associated with changes in employment type. This can be viewed as a short term penalty paid by failing business owners. Entrepreneurship *per se* is not found to exhibit more income volatility than wage work. Earnings volatility is the highest when transitioning from wage to

business sector and back again. Another important finding is that taking on business ownership carries no long term future earnings penalty; moreover, it is rewarding (in terms of future income) to give entrepreneurship a try. Agents entering into entrepreneurship, regardless of their future success, face slightly better chances to end up in the upper tail of the income distribution in the long run. This combination of short term costs and long term rewards is one of the main contributions of this paper. The other is the actual estimate of when the long run is, finding that they wait for a steady state probabilities of changes in one's income is fairly short.

The volatility found in this paper is on a par with other recent investigations (Acs, Loprest, and Nichols 2009; Congressional Budget Office 2007). Nonetheless, the volatility itself should not be taken entirely at face value because at least two systematic biases influence the estimates. First, seam bias is a result of the SIPP survey design where monthly data is collected in batches every four months. This introduces a four-month period for which surveyed people need to account by recollecting relevant information. Recollection is imperfect and suffers from seam bias, where surveyed individuals are more likely to attribute any changes to the first month rather than any other. Another issue is the actual reimbursement schemes commonly employed. The fifty-two week pay system creates systematic imperfections with an unequal number of pay periods for different months. Other factors such as interest, dividends payments, tax returns and tax payments also skew the pay schedule. I undertook a four-month analysis as a robustness check and as means to reduce seam bias. Volatility in earnings persists nonetheless. I attribute this persistence to short business tenure and drastic earnings changes associated with that along with other biases mentioned above. So long as these biases do not effect wage workers and entrepreneurs differently a meaningful comparison can be made. This is because it is not the volatility itself but rather the comparison of wage workers to entrepreneurs that is the subject of this paper.

Extensions to this work could proceed in two directions: First, longer time periods, larger panels and other datasets could be examined to compare probability estimates. This is a very useful exercise because of the importance of what the limiting distributions tell us. Limiting distributions enable researchers to look at outcomes beyond the life of any panel. If earnings are the primary motivation for business ownership, limiting distributions are the best estimates of what eventually awaits the entrepreneur. This information would be welcomed by entrepreneurs, policymakers, financial institutions, and academicians. Improving these estimates is of tremendous value. A second direction for future work is to better understand the reasons behind short business tenures. Cross-sectional views of the entrepreneurial population miss many dynamic components of the entrepreneurial population. An important feature, which has previously received little attention, is uncovered in this paper. Most people claiming to be business owners are in and out of the entrepreneurial status very quickly. Improving our understanding of this dynamic could help reduce erroneous entry, improve success rates, improve financing and guarantee schemes, and improve overall social efficiency.

## **IV. A Consumption-Based Measure of the Monetary Rewards to Entrepreneurship**

### **1. Introduction**

My third paper examines the monetary returns to entrepreneurship. I compare rewards of self-employment to those of wage employment. This is, of course, an exercise that has been carried out many times before. However, previous studies have typically overlooked a potentially important source of bias that arises from the propensity of entrepreneurs to underreport their income. As a result, evidence that entrepreneurship is not a monetarily rewarding undertaking, and that other nonmonetary rewards must be important (Hamilton 2000; Moskowitz and Vissing-Jorgensen 2002), may be misleading. In this paper I produce indirect evidence of underreporting of income in survey data, and I provide alternative measures of monetary rewards to entrepreneurship based on consumption data. My paper also estimates the effects of previous business ownership on current consumption. Contrary to some recent investigations, I find entrepreneurship to be a financially rewarding activity.

Numerous obstacles present themselves when measuring the earnings of entrepreneurs. First, the distribution of income appears to have “fat tails”, suggesting that OLS estimates are likely unreliable.<sup>7</sup> In this paper, I employ a quintile regression method, which is a better metric for contrasting earnings distributions produced by two groups, such as wage workers and the self-employed (Hamilton 2000). Second, because the returns to self-employment consist of salary and accumulated business equity, the income of entrepreneurs is hard to define and track. Business equity is hard to measure unless the researcher examines the entire lifespan of a given firm, and even then, dissolution of a business only reveals its equity value at that particular point in time. In other words, an entrepreneur could have received more (or less) funds if she was to sell the

---

<sup>7</sup> The heavy right tail of this income distribution produces support for the superstar theory (Rosen 1981)

business earlier. The issue is further complicated by entrepreneurs who stay in business for periods beyond the boundaries of any data collection survey. However, by tracking consumption instead of income, it may be possible to capture the wealth effect by observing increased spending of the entrepreneur. Additionally, lifestyle comparisons obtained from consumption measures are likely to be more reliable because they do not rely on income, which, even if not hidden, is computed differently from wage income because of different legal metrics and reporting methods by entrepreneurs.

At the same time some benefits of self-employment are very hard to measure, regardless of the metric at hand. Much more serious complications arise if a researcher attempts to measure less obvious business benefits, some of which include use of business assets for personal use, or claiming personal expenditures as business expenses. Nonmonetary benefits of wage employment, such as fringe benefits, are much easier to measure because they appear as business costs on their employers' books. This drawback associated with difficulty of measuring nonmonetary benefits of self-employment is also better addressed through consumption comparisons: the researcher can detect differences in particular consumption categories in which there is no overlap with business expenditures – housing expenditures, vacations, money spent on clothing, etc.

The consumption comparison approach carries significant benefits in assessing and addressing both expense overstatements and income underreporting. By examining expenditures in a number of discrete categories and subcategories, I can allocate particular expenditures into business, individual consumption, or both categories, thus facilitating better estimates of earnings spent on personal consumption. Overstatements of business expenditures, together with hiding of income by the self-employed, comprise the underreporting gap. The underreporting gap is quite significant given that U.S. General Office of Accounting states that entrepreneurs are responsible for 68% of the \$48 billion underreported gap in 1987 (USGAO 1990). It is a striking number

considering that the self-employment rate remained below 10% throughout the 1980s (Hipple 2004).

In this paper, I show that consumption among entrepreneurs is strictly greater than consumption of wage earners, even after controlling for race, education, experience, family size, and wealth.

This is in striking contrast to recent investigations, claiming that entrepreneurship does not pay (Hamilton 2000; Moskowitz and Vissing-Jorgensen 2002). These investigations have relied on

income data in order to reach their conclusions without any adjustments for income

underreporting, which is best detected by examining consumption data. To detect income

underreporting, one has to either assume (or calculate) saving rates for each of the groups (self-employed and wage earners), or assume that an individual's propensity to save is constant over

time. I rely heavily on previous estimates of saving rates in assessing income underreporting.

Noting that entrepreneurs save more than their wage counterparts (Caner 2003; Dynan, Skinner et

al. 2004; Siman 2008) by creating an interaction term of income and self-employment, I am able

to detect underreporting of earnings. Higher saving rates by the entrepreneurs should then lead to

a negative sign of the interaction term when regressing expenditure of the self-employed on

income, demographics, and the interaction term. Using the Panel Study of Income Dynamics

(PSID), I show repeatedly a positive, highly significant sign on the interaction term, suggesting

extensive underreporting of income. I also detect underreporting using a second assumption

(constant individual marginal propensities to save) by observing changes in sign of an interaction

term (income and self-employment status) with change of employment status.

I then proceed to compare wage rewards to entrepreneurial rewards without using income

by switching to consumption metrics. Not surprisingly, consumption comparisons shows greater

benefits for the entrepreneurs than do income comparisons. These results might of course be

contaminated by sample selection problems. Sample selection bias could arise because a sample

of current entrepreneurs consists of those who either just started or those who chose to remain

self-employed and are thus deemed to be successful entrepreneurs. However, to form an accurate interpretation of entrepreneurial earnings, it is important to examine the earnings of all subjects, not just the successful ones. I therefore also estimate effects of *previous* business ownership on current consumption in order to ameliorate my estimates. I find evidence that those remaining in business enjoy the largest consumption premium but even those who are no longer self-employed, on average, get a small increase in their consumption. Additionally, making sure the sign of the interaction term switches for individuals at the same time as changes in their individual self-employment status (*i.e.*, going back into wage sector, or becoming self-employed), ensures that underreporting accounts for sign of the interaction term and not some fixed effect, which could be associated with people who are more likely to undertake entrepreneurship.

The rest of the paper is structured as follows: in Section 2, I review the literature, in Section 3 I describe the data, in Section 4, I describe the analysis, and Section 5 concludes the paper.

## **2. Literature Review**

Measuring earnings is important to help researchers and policymakers understand the entrepreneur. In the most simple scenario, policymakers want to be able to identify able entrepreneurs *ex ante* in order to provide the future self-employed the necessary support. The entrepreneur is the Superman of economics, he is credited with fostering innovation, creating jobs, keeping production efficient, and producing exactly what consumers desire.

As Adam Smith noted,

[w]hen an independent workman, such as a weaver or shoemaker, has got more stock than what is sufficient to purchase the materials of his own work, and to maintain himself till he can dispose of it, he naturally employs one or more journeymen with the surplus, in order to make a profit by their work (Smith 1776).

If this career choice is motivated by profit, as in Smith's vision, then it is a very straight forward case where uncertainty of business profits is rewarded by a larger than wage profit (profit for Adam Smith is the combination of interest and risk rewards). Hence, the expectation is that entrepreneurial earnings are higher than what the entrepreneur can get otherwise by earning a less risky wage working for someone else. There could be other nonmonetary reasons as to why people choose these careers. However, it seems that alternative explanations are proposed only after self-employment is found not to be financially rewarding, because otherwise profit is the most natural explanation for why people choose entrepreneurship.

The inquiry into earnings comparison originates in the desire to empirically verify several models explaining why some people choose entrepreneurship. One of the first rigorous theoretical models to explain entrepreneurship was developed by Kihstrom and Laffont (1979) who stipulated that risk tolerances were responsible for selection into an entrepreneurial. Essentially, a random parameter in the production function imposes some uncertainty on the output, thus attracting more risk tolerant agents to business ownership. Unfortunately, risk is hard to measure empirically, although several attempts have been made (Hersch and Viscusi 2001; Cramer, Hartog et al. 2002; Sanarelli and Vivarelli 2007; Krichevskiy 2009). Nonetheless, Kihlstrom and Laffont opened up the floodgates of theoretical models and empirical investigations by formally modeling what Adam Smith in 1770s and Frank Knight in the 1920s stipulated about the entrepreneur (Smith 1776; Knight 1916-1921).

Three new classes of models have subsequently emerged: 1) investments and agency models (Lazear and Moore 1984) claiming that earnings profiles differ from wage workers to self-employed because entrepreneurs do not need to be provided incentivized to perform; 2) matching and learning models (Jovanovic 1982) where entering firms do not know what the true costs are and thus have to rely on some prior beliefs to make an entry choice. Upon entry, costs are revealed and firms update their beliefs about their costs (this could be viewed as firm quality)



and only the best survive, while those that made an erroneous entry choice exit; 3) alternatively, either the beliefs (Steen 2004) or the information used in updating (see Chapter I) could be incorrect, or the utility is not fully measured because there could also be nonpecuniary benefits to business ownership. If an entrepreneur is neither overoptimistic nor misled about the profit potential upon entry, but earns less profit compared to wage work, it is reasonable to assume that she derives her utility elsewhere (e.g. the benefit of being in control). However, these conclusions are usually drawn only after the realization that entrepreneurship is not as monetarily rewarding as wage work. In all three models, mentioned above, rewards matter and the next step is to empirically measure those rewards. Nonpecuniary benefits aside, which group has higher earnings?

There are papers claiming that entrepreneurship does not pay (Hamilton 2000; Moskowitz and Vissing-Jorgensen 2002), as well as other works claiming that entrepreneurs are better off (Rees and Shah 1986; Borjas and Bronars 1989; Brock and Evans 1989; Evans and Leighton 1989; Quadrini 1999). The varied conclusions are caused by numerous complications associated with defining an average entrepreneur, choice of household vs. individual comparison, measuring entrepreneur's profits (business equity in particular), choice of hourly wage vs. annual income comparisons, and by a choice of control variables. For example, Quadrini (1999) reports higher household annual incomes for families whose head of the household is self-employed and those who have any family member as the self-employed. The main goal of Quadrini's paper is to measure wealth accumulation by the entrepreneurs; he uses both Panel Study of Income Dynamics (PSID) and Consumer Survey of Finance (CSF) data. However, Quadrini does not control for either education or hours worked; he also does not measure business equity. In contrast, Hamilton (2000), finds that entrepreneurs get both lower initial earnings and lower earnings growth. Hamilton uses the Survey of Program and Income Participation (SIPP) and

estimates hourly earnings while also controlling for education and tracking changes in business equity.

My paper uses annual income and annual consumption measures. I do not think hourly earnings are the best metric for measuring monetary rewards to employment, because the nature of work is different for the self-employed. First, the self-employed determine their work hours on the basis of the demands of the business, and are oftentimes not able to replace their expertise with that of employees. There is also no supervisor to make sure they only take, say, thirty minutes for lunch, meaning reported work hours are not necessary actual work hours (since business owners come and go as they please). Because of the absence of oversight, entrepreneurs may feel that they spend an entire workday at their business while in-between they might have managed to pick up their kids from schools or accomplish other personal tasks. Most wage jobs in the private and public sectors do not allow such flexibility.

Second, most wage workers (especially in the lower half of the income distribution) would be happy to work overtime, thus accumulating more annual hours, but the choice is not theirs to make (business owners who have to pay higher wage rates for hours exceeding 40 hour per week are often discouraged from providing additional work hours to their workers). I draw a distinction between controls such as education, experience, race, and family size versus hours worked, with the former being a universal measure across population and the latter being a very different metric depending on the group in question. Unlike Hamilton, I therefore choose annual earnings measures instead of hourly earnings.

Hamilton also addresses distributional differences by running quintile regressions, since the distribution of earnings for the self-employed is noted to have fat tails, this idea first emerged as a superstar theory put forward by Rosen (1981). Noting these distributional differences I also employ quintile regressions in my comparisons. Unfortunately, neither Quadrini nor Hamilton addresses the income underreporting issue, which is likely to change both of their estimates.

Some researchers maintain that income underreporting is partially circumvented because most research considers surveys that ask income questions directly and that do not rely on income tax data (Hamilton 2000). The idea is that net profits are not reported correctly but so long as econometricians have reported salary draw, overstatements of expenses do not appear in the data. This may not be true if the salary draw is also not reported honestly. My dataset contains both income reported to the IRS and income reported to the surveyors. The two income measures are very similar and produce virtually identical regression estimates; thus, income reported to the IRS is not different from that which is reported to researchers. Hence, if a person has reasons to misrepresent her income for the IRS, she is just as likely to do the same when it comes to interviews with government-funded surveys.

The finding that self-employed underreport their income is not new. It has long been suspected that self-employment provides an opportunity for tax evasion and underreporting. This problem is not as rampant among wage earners whose income is reported by their employers. Thus, wage earners are not necessarily made of higher moral fiber than wage workers; they are simply deprived of the tax evasion opportunity. To address the widespread underreporting by the self-employed several measures have been proposed (Klepper and Nagin 1989; Feinstein 1991; Adreoni, Erard et al. 1998; Schuetze 2000). One way to overcome the problem and estimate someone's income is to use consumption (expenditure) estimates (Pissarides and Weber 1989) derived from parametric (Baker 1993) and nonparametric (Tedds 2005) measures. These expenditure estimates give a researcher an idea as to how much a person should be earning (on average) given her level of consumption. I explore this avenue by examining consumption differences between entrepreneurs and wage earners. I measure distributional differences when comparing several measures of consumption for wage earners and the self-employed. I also find evidence of underreporting of income by entrepreneurs. Income underreporting places a big question mark on the traditional income comparisons among wage earners and the self-employed.

I suggest building consumption measures instead of outright income comparisons. To detect underreporting by the self-employed, I have to rely on the well-supported assumption that entrepreneurs save more (Quadrini 1999; Caner 2003; Dynan, Skinner et al. 2004; Siman 2008).

In fact, wealthy individuals save more regardless of occupation (Reiter 2004). Wealth itself is correlated with entrepreneurship, or with so-called entrepreneurial ability (Hurst and Lusardi 2004); entrepreneurs obtain higher wealth precisely because they save more. Entrepreneurs invest heavily in their own business (Moskowitz and Vissing-Jorgensen 2002), which is oftentimes driven by either the information asymmetry or the expense and unavailability of external financing. Continued re-investment into an entrepreneur's own business could also hide some earnings from the observer.

It appears that while, earning a lot less than their wage counterparts, entrepreneurs save a lot more. While it has been found that entrepreneurs accumulate more wealth (Quadrini 1999), most of the studies of saving patterns used income as a variable (Caner 2003; Siman 2008), which may be problematic if income information is biased because to commonly overstated business expenses and understatement of income. But in a recent careful examination of savings Dynan et al., (2004) found that saving rates are almost identical for samples which included and excluded entrepreneurs. While zeroing in on the saving rates of the self-employed was not their explicit goal, by examining several large datasets (Panel Study of Income Dynamics, Survey of Consumer Finances, and Consumer Expenditure Survey) and looking at the long term (10 years) income averages they were able to side-step the issue of underreporting by entrepreneurs in establishing saving rates (over the periods of time extending beyond business tenure), even while using income. Thus, it is fair to assume when doing consumption comparisons that entrepreneurs do not save less than their wage counterparts.

Armed with the fact that entrepreneurs save more, it is worth re-examining the income-wealth relationship. Two components of the income-wealth relationship and findings about

savings still pose some problems: (1) it appears that entrepreneurs need to save a lot of their income to make up in wealth. This is paradoxical since it is hard to imagine people with a smaller income saving more (in absolute terms) than those whose incomes are significantly larger and (2) much more troubling, income estimates for entrepreneurs are not reliable if there is any underreporting on the part of the self-employed. Mindful of these underreporting issues, I re-examine the income of the self-employed and compare it to the income of wage-workers. To circumvent the issue of income underreporting, I design consumption measures to better estimate levels of disposable income. I find that controlling for demographic variables and assets that self-employed consume more, regardless of whether total expenditure, housing values and expenditures, car values, or utilities expenditures are used as a measure. As a result, this paper contributes to the existing self-employed vs. wage-worker income comparison debate as it points to significant monetary rewards associated with self-employment that are not captured by traditional income measures.

There is also a bigger issue for economists to tackle, related to the core question of who reaps more benefits from their respective jobs, the self-employed or the wage worker? To understand who benefits more, income has been used as a means of obtaining desired goods and services. However, this may not be fully appropriate when comparing self-employed individuals with wage workers, even if neither group attempts to hide its incomes from the IRS (or surveys). Because business owners have the power to decide how business assets are used, and can potentially use these assets for personal benefits (e.g. transportation using the company car), it might be worthwhile to include these benefits into calculations of received earnings. Granted, sometimes similar benefits are available to the wage workers, for example a cab driver who does not own a vehicle may still go and do his grocery shopping using the taxicab. However, these benefits are at the discretion of the business owner, and in many cases must be reported as income. In other words, if a business has a nice car, beach condo etc., business owners receive the

same utility from using these resources as private owners do, and thus they ultimately choose to use their cars, properties and other assets as they see fit. Unlike the intangible and unquantifiable psychological rewards of business ownership, resource use is tangible and should be measured. I argue that in light of the scope for underreporting, which is further complicated by definitional difficulties (in terms of what is a legitimate business expense and what is not) of income itself, consumption instead of income can be used to contrast monetary benefits of employment. Although consumption seems like a natural proxy for earnings, several challenges remain. First issue is the inability of researcher to separate true business expenditures from personal use consumption. For example, a self-employed person may purchase an expensive car to later write it off as a business expense. While this car could be helpful in signaling quality of the self-employed to potential clients, making it a very useful business tool, the owner also enjoys this high quality product. For example, when hiring a Certified Public Accountant (CPA) a customer may perceive an expensive car driven by the CPA as a signal of experience, quality, or sign of fewer instances of IRS audit, all of which are measures of success. In this example the expensive car aids the entrepreneur (CPA) in attracting business clientele making it a useful business asset. However, at the end of the day, when she goes to the stores, malls, and restaurants, she enjoys the quality of the car in the pure consumption sense.

Another issue arises with self-selection of people into self-employment. It may be necessary to have collateral in order to secure the funding crucial for establishment of a business. Hence, by observing that more of the self-employed own their own housing we cannot conclude that these differences are because of larger incomes as the wage workers with large incomes may be kept out of the self-employment because of lack of collateral. Accumulation of collateral can either come from bequest or from saving over time. So are the entrepreneurs (or prospective entrepreneurs) different when it comes to lifelong saving behavior?

To ensure that these entrepreneurial individuals do not have some fixed effect associated with their saving behavior and thus causing the income-entrepreneurship interaction term to serve as a proxy for this fixed effect: I look at the long-term history of business ownership. I create a number of terms interacting indicators for prior business ownership with current consumption. Observing the sign of the interaction terms, contingent on current business ownership only (essentially observing change in sign with the change of self-employment status), I find additional re-assurance that it is underreporting and not some fixed effect that is responsible for the sign of the significant interaction term. This method has not been previously used in the underreporting literature which has previously assumed the same marginal propensity to consume for everyone (Pissarides and Weber 1989). The only assumptions that I make are that 1) consumption is reported truthfully for everyone, 2) marginal propensities to consume remain constant for all agents for the duration of the study, and 3) self-employed individuals have higher saving rates, as has been repeatedly demonstrated (Quadrini 1999; Caner 2003; Dynan, Skinner et al. 2004; Siman 2008).

### **3. Data**

I use the Panel Study of Income Dynamics (PSID) as it offers a number of asset variables, income variables and most recently (starting with 2005 data) a number of detailed consumption variables. The PSID is designed to track individuals, but it also collects income and expenditure information on their spouses and, to a lesser degree, on other family members living with them. While my dataset does not contain a long history of consumption, it does have business ownership history, which I use to estimate the long-term effects of business ownership. By examining how current consumption levels differ among wage earners some of whom were previously self-employed but no longer are, I can see how long these differences persist after a person returns to wage employment.

My dataset consists of detailed observations for 2005, with demographic and income variables going as far back as the 1995 wave of PSID. I examine consumption to produce results comparable to income estimates of Hamilton (2000), I restrict my sample to males. Following convention, I exclude farmers and anyone outside the 18-65 age range. I also exclude people earning practice income, since they are often re-classified into self-employment as they become partners in their firms, but they are not self-employed in the conventional sense. The resulting sample consists of 4,797 people. Of these, 708 are what I will call self-employed (14.77%). This number is unusually high. Participants of the PSID survey are asked if they are working, and if "yes", then the following options are given: (i) for someone else, (ii) themselves or (iii) both? Only seven percent of those surveyed work for themselves. However, since I am examining consumption and household consumption records are not disaggregated, it is possible to determine whose income went towards what (in terms of purchasing goods and services by members of the household), I use another variable in PSID in deciding whether the person is a part of the self-employed group. If any person answers positively to the question of whether they, or anyone else in their household, are self-employed, I place them in the self-employed group. I create similar variables for other years of the survey going back to 1995.

The PSID underwent major revisions in 1997, starting bi-annual interviews instead of previously conducted annual interviews. Hence, my dataset includes 1995, 1996, 1997, 1999, 2001, 2003, and 2005 interview variables. Demographic variables are carried through unless there is a change. I use demographic variables for the current year (2005) as the controls in the analyses. I also construct variables capturing whether a business owner is present in the household since the previous survey. In other words if there was a self-employed person in 1995-97 but since then he or she either went back to the wage sector or left the household, I create a variable capturing this change in order to trace the effects of previous business tenure on current consumption.



It is not immediately obvious which consumption measure should be employed. I create several consumption measures with total annual consumption being the broadest – it includes total annual expenditure on housing, transportation (car leases, gasoline, parking, insurance, and public transportation), utilities (net of subsidies), food, clothing, child-care, adult-care, schooling, home repair, vacations, entertainment, and total donations. In the case of transportation, cars are often claimed as business expenses, but they are consumption goods as well. I keep expenditure on cars as consumption in my sample and exclude utility vehicles, which are used more for business than as consumption goods. I replicated my analysis excluding pickup trucks as well. This exclusion did not alter the results to any significant degree, so I re-introduced the pickup truck category into the data set, while excluding utility vehicles. In other measures of consumption, I used housing or utilities as the dependent variables only – these are much narrower definitions that oftentimes do not detect significant differences between the two groups.

Income in my analysis includes wages, interest, royalties, rent, dividends and both public and private transfers. This variable encompasses head of the household income and spouse's income as well. Additional income brought in by other family members both present at the household and transferring money from elsewhere is also accounted. I also repeated the entire analysis using taxable income, which is asked separately in the survey. The results were practically identical and are not reported here.

#### **4. Group Differences, Underreporting and the Long Term Effects of Self-Employment**

##### **4.1 Descriptive Statistics and Group Differences**

I first set out to explore population differences between the self-employed and wage workers. I then examine their respective consumption profiles controlling for these differences. The two populations differ along every measured dimension. First, the average self-employed

person is older, more likely to be married, with more work experience and education; he has greater self-reported income and he also consumes more. I conducted non-parametric Kolmogorov-Smirnov tests (Conover 1999) on expenditure, income, education, and experience distributions. In each case, the distribution of the self-employed dominates those of wage workers, with  $p$ -values near zero. In my sample 86 percent of business owners are married compared to only 78 percent of married wage workers. The self-employed also differ from wage earners along racial, educational, and experience lines as well (Table 5). As for distribution differences the findings are well in-line with the expectations given that self-employed are more experienced, and older.

Another dimension along which the self-employed differ markedly from wage workers is in the amount of household assets and debt. The self-employed own more of all types of assets, including housing assets, stocks, and certificates of deposit. At the same time, the self-employed carry more debt (Table 6). Assets are necessary to obtain business financing, which is reflected in the greater housing equity held by the self-employed. Additionally, both assets and business loans generate debt since all of these are likely to be financed.

**Table 5.** *Descriptive Statistics of Demographic Variables*

	Self-Employed	Wage Workers
Mean age	46.46	40.65
No children	54.06	51.19
2 or more kids	9.37	11.81
High school diploma only	26.85	34.59
2 year college	17.65	19.82
4 years +	15.54	7.49
Race white	64.57	82.13
10 yrs or less experience	40.91	53.11
20 yrs or more experience	21.52	16.5

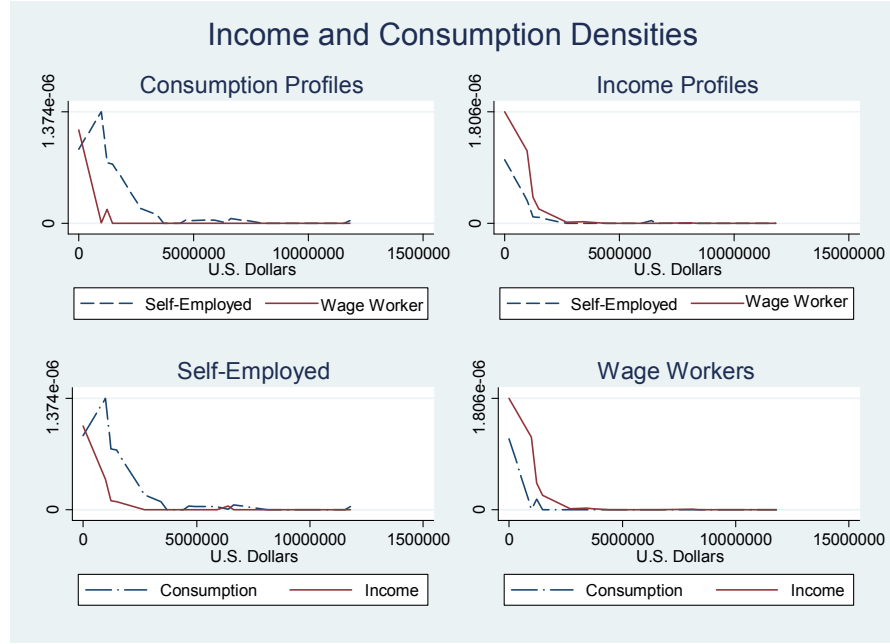
Values are percentage of total population

**Table 6** *Assets and Debt Comparison*

	Self-Employed	Wage Workers
Mean CD Savings	\$31,154	\$8,710
Median CD Savings	\$5,000	\$1,000
Percent of Population with CD $\geq$ 25,000	24.19	9.04
Mean Housing Equity	\$142,575	\$94,768
Median Housing Equity	\$74,500	\$53,000
Percent of Population with Housing equity > 500,000	8.2	2.09
Mean Family Debt	\$16,844	\$7,709
Median Family Debt $\geq$ 10,000	\$2,000	\$500
Percent of Population with Family Debt $\geq$ 25,000	13.12	8.39

Significant demographic group differences create two very different income and consumption profiles. At the same time, even the raw plots of the distributions suggest some disparity in the reliability of income reporting. Figure 6 depicts estimated consumption and income distributions for the two groups. It demonstrates a within group income-consumption contrast revealing interesting patterns. In the case of the self-employed, it seems that consumption outpaces income, suggesting some underreporting of income taking place. There is no reason for which an entire group of people would be consuming beyond their current income – unless the group in question is the unemployed. Yet it seems that self-employed households enjoy consumption well beyond the levels allowed by the earnings they receive. At the same time, merely observing these differences of distributions is not sufficient to deduce that there is any underreporting of income.

**Figure 6. Income and Consumption Densities for Wage Workers and the Self-employed**



## 4.2 Personal Income and Expenditure Comparisons

Unconditional distributions provide only a limited insight and are not entirely suitable for the comparison of rewards to employment – especially given what we know about differences in education, experience, and financial wellbeing among the two populations. Thus, I produce conditional estimates of both incomes and expenditures controlling for education, experience, and family size. I regress demographic variables and the self-employed dummy on both expenditure and the total consumption as in (1) and (2)

$$E_i = \beta_0 + \mathbf{X}'_i \beta_i + Z_i + \epsilon_i \quad (1)$$

$$I_i = \beta_0 + \mathbf{X}'_i \beta_i + Z_i + \xi_i \quad (2)$$

where  $E_i$  is expenditure,  $I_i$  is income,  $\mathbf{X}_i$  is a vector of demographic variables and  $Z_i$  is a self-employment dummy. I produce both OLS estimates and quantile regression estimates at the 25<sup>th</sup>,

50<sup>th</sup>, and 75<sup>th</sup> percentiles, mindful of the fat tail distributions documented by Hamilton (2000).

Tables 7 and 8 provide the results.

**Table 7** *Expenditure and Self-Employment*

Variable	OLS	Quantile Regression		
		.25	.50	.75
<i>Education</i>	8.77*** (2.55)	3.32*** (.20)	5.52*** (.30)	9.54*** (.93)
<i>Race White</i>	-18.17 (14.52)	6.51*** (1.28)	7.76*** (1.72)	10.28** (4.39)
<i>Experience</i>	-0.84 (.74)	.095 (0.067)	.50*** (.09)	1.20*** (.21)
<i>Family size</i>	55.24*** (11.69)	-7.50*** (.88)	-6.86*** (1.37)	-10.64** (4.17)
<i>Self-Employed</i>	35.51* (18.34)	11.74*** (1.61)	19.74*** (2.17)	39.99*** (5.54)
<i>Constant</i>	2.48 (35.65)	-19.55** (2.90)	-36.38*** (4.23)	-61.70*** (12.71)
<i>R-squared</i>	0.013	0.04	0.048	0.05

Dep. Var: Total Expenditure in thousands.  $N=3082$   
\*\*\* $\alpha=0.01$ . \*\* $\alpha=0.05$ . \* $\alpha=0.1$

It is immediately clear that the choice between mean or median comparisons, as well as the choice of consumption versus income, yields very different results. Quantile regressions produce a slightly better fit when comparing the  $R$ -squared of the OLS to the pseudo  $R$ -squared of any of the quantile regressions; Figure 1 also shows severe skewness and non-normality, which also suggests quantile regressions are more appropriate. Consequently, I shall pay more attention to them than to the results obtained via OLS. However, using either specification, it is clear that self-employed earn more. Moreover, regardless of the methodology, expenditure premiums of self-employment are larger than income premiums. This suggests underreporting on part of the

self-employed, especially since we would expect the opposite results given higher saving rates of the self-employed.

**Table 8** *Income and Self-Employment*

Variable	OLS	Quantile Regression		
		.25	.50	.75
<i>Education</i>	10.25*** (1.17)	4.61*** (.24)	6.80*** (.32)	8.54*** (.53)
<i>Race White</i>	24.75*** (6.81)	13.05*** (1.60)	16.11*** (1.90)	17.45*** (2.62)
<i>Experience</i>	.79** (.33)	.26*** (.088)	.46*** (.09)	1.06*** (0.12)
<i>Family size</i>	-9.18** (5.93)	-5.90*** (1.20)	-7.72*** (1.66)	-9.52*** (2.53)
<i>Self-Employed</i>	30.79*** (8.77)	-.57 (2.04)	5.32** (2.45)	16.47*** (3.40)
<i>Constant</i>	-83.59*** (16.39)	-35.96*** (3.52)	-45.62*** (4.59)	-46.29*** (7.33)
<i>R-squared</i>	0.038	0.065	0.084	0.095

Dep. Var: total income in thousands.  $N=3601$ .  
\*\*\* $\alpha=0.01$ . \*\* $\alpha=0.05$ . \* $\alpha=0.1$

### 4.3 Underreporting

A closer examination of the data suggests that some income claims made by the surveyed self-employed individuals are not likely to hold up to scrutiny. For example, 51.8 percent of males and their spouses who claim to be self-employed only (these are the households where everyone is self-employed) claim to have received exactly zero income from their self-employment in 2005. This clearly is not likely. Additionally, it would mean severe dissavings on the part of these households. To confirm my suspicion of underreporting I regress demographic

variables such as race, individual's education, experience, family size, total income and the interaction term between income and business ownership on the total expenditure. My model specification is

$$E_i = \beta_0 + \mathbf{X}'_i\beta_i + I_i + D_i + \varepsilon_i \quad (3)$$

where  $E_i$  is expenditure,  $X_i$  is the vector of demographic variables,  $I_i$  is individual's income, and  $D_i$  is the interaction term. I find strongly significant positive coefficients on the income x business ownership interaction term (see Table 9). This suggests that the self-employed are underreporting their income since their higher saving rates should lead to a negative coefficient if they were to report their income honestly.

**Table 9** Evidence of Income underreporting.

Variable	OLS	Quintile Regression		
		.25	.50	.75
<i>Income</i>	.000065445 (.0000458526)	.0001690939*** (.0000458526)	.0003322333*** (.0000042246)	.0005700435*** (.0000052913)
<i>Education</i>	7.596428*** (2.801553)	2.178698*** (.1462655)	2.922443*** (.2578926)	4.339161*** (.5867257)
<i>Race</i>	-18.17901 (16.32578)	3.811877*** (.9476383)	3.198972** (1.50196)	5.118825* (2.890641)
<i>Experience</i>	-.986501 (.8130209)	0.03359129 (.04890513)	.2768124*** (.07486109)	.6390559*** (.1382626)
<i>Family size</i>	72.16536*** (13.83903)	-5.674421*** (.7051715)	-5.218422*** (1.267869)	-6.752845* (2.818167)
<i>Interaction Term</i>	.0002034619*** (.0000692232)	.0000526269*** (.0000045018)	.0000328126*** (.0000063774)	.0002601576*** (.0000083956)
<i>Constant</i>	15.85689 (39.23889)	-10.37368*** (2.093101)	-16.34522*** (3.610893)	-25.89748*** (8.075975)

Dep. variable expenditure in thousands ( $N=2705$ ; =0.023; =0.084; =0.099; =0.11)  
 \*\*\* $\alpha=0.01$ . \*\* $\alpha=0.05$ . \* $\alpha=0.1$

In Table 10, I show repeatedly (using specification (3) with variables representing current income interacted with indicators for self-employment status in previous years) a positive sign on the interaction term for current business owners and mixed signs for people who are no longer self-employed (see Appendix 3 for complete regression results). Finding mixed sign for the wage workers' interaction terms is not surprising since I do not have any particular expectation about the saving rates for the wage workers. In fact, the coefficient turns out to be negative for most years, whereas for current business owners the positive sign suggests continued income underreporting.

**Table 10** *Additional evidence of Income underreporting. (10th percentile)*

Interaction Term	Self-Employed 2005	Wage Worker 2005	Number of Cases Self Employed (SE) and Wage Workers (WW)
1995xCurrent Income	.0000807*** (.0000277)	.000177*** (.0000139)	SE 158 WW 817
1996xCurrent Income	.0000980*** (.0000276)	-.000169*** (.0000125)	SE 164 WW 839
1997xCurrent Income	.0000708** (.0000325)	-.000171*** (.0000096)	SE 171 WW 957
1999xCurrent Income	.0000381 (.0000280)	-.000197*** (.0000112)	SE 182 WW 1067
2001xCurrent Income	.0000697** (.0000318)	-.000197*** (.0000112)	SE 192 WW 1185
2003xCurrent Income	.0000113*** (.0000284)	-.000313** (.0000142)	SE 212 WW 1351
Dep. variable expenditure in thousands, *** $\alpha=0.01$ . ** $\alpha=0.05$ . * $\alpha=0.1$			
Other regressors: Total Income, Education, Experience, Family Size, Race			

The regressions in Table 10 show the sign of the interaction term to be contingent on business ownership. In my regressions the interaction term for the self-employed always has both positive sign. Table 10 presents results for the 10<sup>th</sup> quintile of the expenditure distribution using quantile regression. These results are robust to changes in specification with similar signs on the interaction coefficient for other quintiles. The regression signs for the lower 10<sup>th</sup> quintile are the same as those from 25<sup>th</sup>, 75<sup>th</sup>, 90<sup>th</sup>, or 95<sup>th</sup> quintiles (not reported here). But for the 10th percentile I find statistical significance for interaction term much more often (compared to 25<sup>th</sup>, 50<sup>th</sup>, or 75<sup>th</sup>



quintile). My suspicion is that 10<sup>th</sup> percentile shows a significant sign more often because of lower variances in the left tail of income distributions for both groups. The results in Table 10 lead to the conclusion that the self-employed do not have any particular tendencies associated with income underreporting as a fixed individual trait. To the contrary, they only underreport their income while they are in still in business.

#### 4.4 Long Run Effects

I now examine how previous business tenure affects current consumption. I regress the same demographic variables (excluding income) on current consumption, adding previous business ownership dummies. These dummies record previous years' business ownership while all other variables continue to report current consumption and current demographics. My model specification is

$$E_i = \beta_0 + \mathbf{X}'_i \beta_i + Z_i + \epsilon_i \quad (4)$$

where  $Z_i$  is a specific year of business ownership.

I run this specification for both types of individuals: those currently in the wage sector and those currently self-employed. I produce estimates for both mean (via OLS regression) and median (via quantile regressions). The results are very similar in nature with those agents who are still self-employed enjoying the largest consumption premiums. For both median and mean estimates the same cannot be concluded about the people who went back into the wage sector, as we cannot observe the reason for this change of status. Returning to the wage sector could be because of failure of business venture (bankruptcy for example) or it could be a successful sale of a business. There is a large body of literature addressing firms' exit and how it is related to success and failure (Headd 2003). The benefits of previous business ownership are highly correlated with current consumption (survival and success go hand in hand). At the same time, the only conclusive results for current wage workers who were previously self-employed is that,

on average (in both median and mean regressions), there is a positive shock to current consumption due to previous business ownership. Median estimation results are summarized in Table 11, whereas mean estimates (with the 95 percent confidence intervals) are shown in Figure 7.

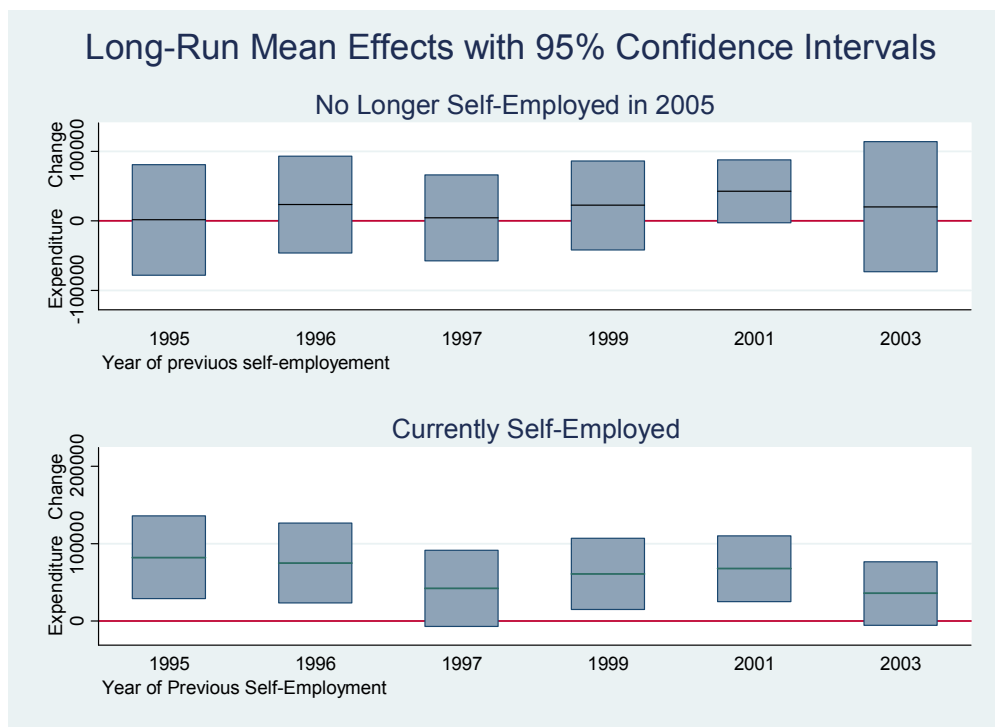
Table 11 reveals that even after controlling for demographic differences long term average effects are positive. On average, even after business tenure has completed, there are positive effects on consumption that extend for up to four years after the termination of a business. Even combined with the fact that for those that stayed in business effects are mostly larger, the finding of a lingering positive effect suggests that undertaking self-employment has rewards even if one was to go back into the wage sector. This suggests that on average, business termination is not failure when compared to wage work. This, "on average" statement needs to be taken at face value - while on average the effect is non-negative for many agents the change is negative nonetheless, because this average is not statistically significant and a large portion of the confidence interval is in the negative territory . Recent investigation of long term effects using income finds that an individual significant penalty (via income reduction) is paid by entrepreneurs returning to the wage sector (see Chapter II), yet in the long run it pays off to attempt business ownership - regardless of the success of the business venture. Using SIPP I previously found that people attempting business ownership have bigger chance of ending up in higher income classes. These effects are likely to be even more pronounced given income underreporting of the self-employed.

**Table 11. Long Run Effects of Self-Employment (Median Regression)**

Business Ownership Year	Self-Employed 2005	Wage Worker 2005	Number of Cases Self Employed (SE) and Wage Workers (WW)
1995 Business	28.57** (12.81)	18.18*** (5.65)	SE 185 WW 915
1996 Business	11.25** (4.89)	11.25** (4.89)	SE 192 WW 942
1997 Business	2.08 (9.89)	4.53 (4.77)	SE 202 WW 1073
1999 Business	1.96 (7.04)	1.22 (4.91)	SE 214 WW 1198
2001 Business	3.71 (3.83)	3.71 (3.83)	SE 226 WW 1330
2003 Business	.000011*** (.000028)	-.00031** (.000014)	SE 248 WW 1517

Dep. variable expenditure in thousands, \*\*\* $\alpha=0.01$ . \*\* $\alpha=0.05$ . \* $\alpha=0.1$   
 Other regressors<sup>8</sup>: Total Income, Education, Experience, Family Size, Race

**Figure 7. Long-Run Effects of Self-employment**



<sup>8</sup> Full regression results are reported in the Appendix 3

Excluding income from the regressions completely is likely to create an omitted variable bias, since income matters in the determination of an individual's consumption levels. But as it is underreported by the self-employed, including income will offset the findings associated with the self-employed and the regression dummy coefficient. Hence, removal of income and using a consumption comparison alone is more informative. I look at the coefficients of correlation for the relevant variables and try to determine the magnitude of the bias created by the omission. It could be assumed that the correlation coefficient for the wage workers is the unbiased estimate, which could be used to adjust the incomes of the self-employed. However, economists are not yet set on the systematic differences between the two groups and because of that we cannot assume that the relationship of income to consumption is the same. This is partially because of different spending patterns created by legitimate business expenses. Nonetheless, the coefficients of correlation themselves provide some evidence of the disparities, but even very big systematic differences among the two groups should not lead to the striking differences observed; I find a coefficient of 0.48 for the self-employed and 0.04 for the wage workers. Needless to say that both the difference over 10 times in magnitude and the 3% income consumption correlation for the wage workers seem unrealistic. Clearly, better estimates are needed. An ideal way to address this issue is via a long panel data examination with both income and consumption, which should create better coefficients for people who have switched from self-employment into the wage sector and back as their fixed effects could be controlled for. Nonetheless, this paper demonstrates the importance of underreporting issue, which needs to be carefully considered in any earnings comparison.

## **5. Conclusion**

The self-employed are financially better off compared to the wage workers even after controlling for their respective demographics. This finding is in stark contrast to some of the

recent literature on income comparisons (e.g., Hamilton 2000). The main reasons for this disparity are that this paper pays particular attention to the underreporting issue, largely ignored elsewhere, and the choice of annual income versus hourly wage comparisons. None of the recent income comparisons of self-employed and the wage worker controls for the income underreporting by the self-employed. I find significant evidence of income underreporting by the self-employed which casts doubts on conclusions stating that self-employment is not as financially rewarding as wage work.

Comparisons undertaken in this paper find that the self-employed are better off in terms of both income and consumption, even after controlling for education, experience, and demographic variables. Perhaps the entire income definitions need to be re-written whenever evaluating self-employment. Researchers need to be very careful in separating business expenses from those of pure consumption. Having said that, there is often no clear line about the proportions of good's used for many mixed purpose expenses (cars are a good example of this). In this paper I address the issue of mixed purpose expenses by reducing the vehicle category to exclude various types of cars, some of which appear to carry much smaller consumption utility and much larger business purpose. At the same time this paper does not address a full addition of the fringe benefits to the rest of the consumption due to data availability. Therefore, even the consumption measures created here are only limited representatives as both large categories, such as fringe benefits, as well as small categories of incidental spending and numerous services categories are not accounted for in the survey. There should be no situation in which self-employed people with similar characteristics earn less; this is because they can simply close shop and move into wage employment. The only scenario where the above situation can persist is when there are some benefits not captured by income. I argue that consumption captures some of the benefits not reflected by income (for example personal use of business assets). At the end of

the day, the self-employed consume more and are thus better off. These differences are both robust to changes in definitions of consumption and persistent across the two distributions. Annual earnings comparisons are a better metric because entrepreneurs can choose to work more hours whenever business conditions allow whereas wage earners mostly depend on the entrepreneurs (and managers) to make these choices for them. Because of current law structure workers are usually more motivated to work overtime (since they get paid more than their regular fare); in most cases, they are not able to work the hours that they want. Earnings, in the end of the day, are only the means necessary for the acquisition of goods and services. It is goods and services (current or future) that people are really after. Hence, consumption comparison is a more direct metric in comparison of benefits to employment across groups. For example, a much fuller picture of how taxing unemployment is could be drawn by examining individual's consumption before, during, and after the unemployment spell, as the actual hardships can be observed by looking at substitutions in consumption.

The long term effects of business ownership, while non-negative on average, do not seem to persist for people returning to the wage sector. This is because there could be various reasons responsible for the end of business tenure. At one extreme, a business may end because it was a complete failure, while at the other termination of a business could be the result of a successful sale by the entrepreneur.

Some limitations arise from consumption comparisons done on the household level. The self-employed, in this context, are households that have someone who is self-employed as a member. Any changes in household makeup can move the household into either wage or self-employment category, having nothing to do with either business origination or the termination of such venue. For the self-employed still in business, previous business ownership is correlated with current increase of consumption as one would expect. Unfortunately, since the PSID has only recently started to track expenditure variables, our ability to measure long-run effects is very

limited. This is the limitation of cross-section analysis, which can only be corrected with evaluation of the panel data.

Better income estimates need to be constructed in the future with income underreporting corrections made for the self-employed. The obvious difficulty lies in estimating the magnitude of underreporting. Perhaps a long panel data containing income, consumption, assets, and employment data needs to be examined, as the cross-section analysis is able to detect the underreporting issues but not correct for them.

## List of References

- Acs, G. Loprest P. et al. (2009) "Risk and Recovery: Understanding the Changing Risk to Family Incomes" The Urban Institute, Paper 14.
- Adreoni, J., B. Erard, et al. (1998). "Tax Compliance." *Journal of Economic Literature* 36: 818-860.
- Anderson, L. R. and C. A. Holt (1997). "Information Cascades in the Laboratory." *The American Economic Review* 87(5): 847-862.
- Astebro, T. (2003). "The Return to Independent Invention: Evidence of Unrealistic Optimism, Risk Seeking or Skewness Loving?" *Economic Journal* 113: 226-239.
- Astebro, T. and P. Thompson (2008). *Entrepreneurs: Jacks of all Trades or Hobos?* University of Toronto and Florida International University.
- Baker, P. (1993). *Taxpayer compliance of the self-employed: estimates from household spending data.* No. W93/14, Institute of Fiscal Studies.
- Barber, B. M. and T. Odean (2001). "Boys Will be Boys: Gender, Overconfidence, and Common Stock Investment." *The Quarterly Journal of Economics*, 116(1): 261-292.
- Belges, J. L. d. (1510-1514). *Illustrations de Gaule et singularitez de Troye.*
- Benz, M. and B. S. Frey (2004). "Being Independent Raises Happiness at Work." *Swedish Economic Policy Review* 11(2): 95-134.
- Bernardo, A. E. and I. Welch (2001). "On the Evolution of Overconfidence and Entrepreneurs." *Journal of Economics and Management Strategy* 10(3): 301-330.
- Borjas, G. and S. G. Bronars (1989). "Consumer Discrimination and Self-Employment." *Journal of Political Economy* 97: 581-605.
- Braakmann, v. N. (2007). *Differences in earnings distribution of self- and dependent employed German men – evidence from a quintile regression decomposition analysis.* University of Lüneburg Working Paper Series in Economics.
- Braguinsky, S. and A. Ohyama (2007): "Where does entrepreneurship pay? Disaggregating returns to self-employment in science and engineering workforce." Working paper, SUNY Buffalo.
- Brennan, M. J. and W. N. Torous (1999). "Individual Decision Making and Investor Welfare." *Economic Notes By Banca Monte dei Paschi Siena (Spain)* 28(2): 119-143.
- Brenner, L. A., D. J. Koehler, et al. (1996). "Overconfidence in Probability and Frequency Judgments: A Critical Examination."



- Brock, W. A. and D. S. Evans (1986). *The economics of small business: Their role and regulation in the U.S. economy*. New York, Holmes & Meier.
- Brock, W. A. and D. S. Evans (1989). "Small business economics" *Small Business Economics* 1(1): 7-20.
- Brockhaus, R. H. (1980). "Risk taking propensity of entrepreneurs" *Academy of Management Journal* 23(3): 509-520.
- Busenitz, L. W. and J. Barney (1997). "Differences between Entrepreneurs and Managers in Large Organizations: Heuristics in Strategic Decision-Making." *Journal of Business Biases and Venturing* 12(1): 9-30.
- Caner, A. (2003) *Savings of Entrepreneurs*. SSRN eLibrary Volume, DOI: 10.2139/ssrn.452020
- Cantillon, R. (1755). *Essai sur la nature du commerce en general*.
- Chamley, C. P. (2003). *Rational Herds: Economic Models of Social Learning*. Cambridge University Press
- Choi, I. (2001). "Unit root tests for panel data." *Journal of International Money and Finance* 20(2): 249-272.
- Congressional Budget Office. 2007. "Trends in Earning Volatility Over the Past 20 Years." Washington D.C. Congressional Budget Office. Doc 8007.
- Conover, W. J. (1999). *Practical Nonparametric Statistics*. New York, Wiley.
- Cooper, A. C., T. B. Folta, et al. (1995). "Entrepreneurial information search." *Journal of Business Venturing* 10(2): 107-120.
- Cramer, J. C., J. Hartog, et al. (2002). "Low risk aversion encourages the choice for entrepreneurship: an empirical test of a truism " *Journal of Economics Behavior & Organization* 48: 29-36.
- DeGroot, M. (2004). *Optimal Statistical Decisions*, Wiley.
- Dunne T. Roberts M. et al. (1989) "Plant turnover and gross employment flows in U.S. manufacturing sector" *Journal of Labor Economics* 7(1) 48-71.
- Dynan, K. E., J. S. Skinner, et al. (2004). "Do the Rich Save More?" *Journal of Political Economy* 112(2).
- Evans, D. S. and L. S. Leighton (1989). "Some empirical aspects of entrepreneurship." *American Economic Review* 79(June): 519-535.
- Feinstein, J. S. (1991 ). "An Econometric Analysis of Income Tax Evasion and its Detection." *Rand Journal of Economics* 25(1): 14-35.

- Fisher, R. A. (1930). The Moments of the distributions of normal samples of measures of departure from normality. Proceedings of the Royals Society of London.
- Ghiselli, E. E. (1974). "Some perspectives for industrial psychology." *The American psychologist* 29(2): 80-87.
- Gort, M. and S.-H. Lee (2007): "The rewards to entrepreneurship." Working paper, SUNY Buffalo.
- Greene, W. H. (2003). *Econometric Analysis* pp757-759. New Jersey, Prentice Hall.
- Hamilton, B. H. (2000). "Does Entrepreneurship Pay? An Empirical Analysis of the Returns to Self- Employment." *Journal Of Political Economy* 108(3).
- Headd, B. (2003). "Redefining Business Success: Distinguishing Between Closure and Failure." *Small Business Economics* 21(1): 51-61.
- Heaton, J. and D. Lucas (2000). "Portfolio Choice and Asset Prices: The Importance of Entrepreneurial Risk " *The Journal of Finance* 55(3): 1163-1198.
- Hersch, J. and W. K. Viscusi (1990). "Cigarette Smoking, Seatbelt Use, and Differences in Wage-Risk Tradeoffs " *The Journal of Human Resources* 25(2): 202-227.
- Hersch, J. and W. K. Viscusi (2001). "Cigarette Smokers as Job Risk Takers." *The Review of Economics and Statistics* 83(2): 269-280.
- Hill D. (1987) "Response Errors around the Seam: Analysis of Change in a Panel with Overlapping Reference Records." In 1986 Proceedings of the Section on Survey Research Methods, American Statistical Association (201-15). Alexandria VA: American Statistical Association.
- Hipple, S. (2004). "Self-Employment in the United States: an update." *Monthly Labor Review* July(13-23).
- Hurst, E. and A. Lusardi (2004). "Liquidity Constraints, Household Wealth, and Entrepreneurship." *Journal of Political Economy* 112(2): 319-347.
- Im, K. S., M. H. Pesaran, et al. (2003). "Testing for unit roots in heterogeneous panels." *Journal of Econometrics* 115(1): 53-74.
- Jerker Denrell,. (2003). "Vicarious Learning, Undersampling of Failure, and the Myths of Management." *Organizational Science* 14(3): 227-243.
- Johnson, N. L., S. Kotz, et al. (1994). *Continuous Univariate Distributions* pp156-164. New York, Wiley-Interscience.
- Jovanovic, B. (1982). "Selection and the Evolution of Industry." *Econometrica* 50(3): 649-670.

- Kahneman, D. and A. Tversky (1973). "On the psychology of prediction." *Psychological Review* 80: 237-251.
- Kihlstrom, R. E. and J.-J. Laffont (1979). "A General Equilibrium Entrepreneurial Theory of Firm Formation Based on Risk Aversion." *Journal of Political Economy* 87(4): 719-748.
- Klepper, S. and D. Nagin (1989). "The Anatomy of Tax Evasion." *Journal of Law, Economics and Organization* 5(1): 1-24.
- Knight, F. H. (1916-1921). *Risk, Uncertainty and Profit*. New York, Houghton Mifflin.
- Landier, A. and D. Thesmar (2003). *Financial Contracting with Optimistic Entrepreneurs: Theory and Evidence*.
- Lazear, E. P. and R. L. Moore (1984). "Incentives, Productivity, and Labor Contracts." *Quarterly Journal of Economics* 99(2): 275-296.
- Lord, R. and K. Maher (1990). "Alternative information-processing models and their implications for Theory, research, and practice." *Academy of Management Review* 15: 9-28.
- Lowe, R. and A. A. Ziedonis (2006). "Overoptimism and the Performance of Entrepreneurial Firms." *Management Science* 52(2): 173-186.
- Ma, Y. and N. Balakrishnan (2001). *On the Posterior Moments for Truncation Parameter Distributions and Identifiability by Posterior Mean for Exponential Distribution with Location Parameter*. Boca Raton, New York, London, Washington D.C., Chapman & Hall/CRC.
- Moskowitz, T. J. and A. Vising-Jorgensen (2002). "The Returns to Entrepreneurial Investment: A private Equity Premium Puzzle?" *The American Economic Review* 92(4): 745.
- Nakamura, T. (1980). "On the moment of positively truncated normal distribution." *Journal of Japan Statistical Society* 10: 139-144.
- Parker, S. C. (2004). *The economics of Self-Employment and Entrepreneurship*, Cambridge University Press.
- Pissarides, C. A. and G. Weber (1989). "An expenditure-based estimate of Britain's black economy." *Journal of Public Economics* 39(1): 17-33.
- Puri, M. and D. T. Robinson (2007). "Optimism and Economic Choice." *Journal of Financial Economics* 86(1): 71-99.
- Quadrini, V. (1999). "The importance of entrepreneurship for wealth concentration and mobility." *Review of Income and Wealth* 45(1): 1-59.
- Rabin, M. (1998). "Psychology and Economics." *Journal of Economic Literature* 36(1): 11-46.

- Rees, H. and A. Shah (1986). "An empirical analysis of self-employment in the U.K." *Journal of Applied Econometrics*(January): 95-108.
- Reiter, M. (2004). *Do the Rich Save Too Much? How to Explain the Top Tail of the Wealth Distribution*. Working Paper. Barcelona, Universitat Pompeu Fabra.
- Rosen, H. and P. Willen (2002). Risk, return and self-employment.
- Rosen, S. (1981). "The economics of superstars." *The American Economic Review* 71(5): 845-858.
- Samuelson, W. and R. Zeckhauser (1988). "Status Quo bias in decision making." *Journal of Risk and Uncertainty* 1(1): 7-59.
- Sanarelli, E. and M. Vivarelli (2007) *Entrepreneurship and the process of firms' entry survival and growth*. *Industrial and Corporate Change* Volume, DOI: 10.1093/icc/dtm010.
- Schuetze, H. J. (2000). "Taxes, economic conditions and recent trends in male self-employment: a Canada–US comparison." *Labor Economics* 7(5): 507-544.
- Selvin, S. (1976). "A Graphical Estimate of the Population Mean from Censored Normal Data." *Applied Statistics* 25(1): 8-11.
- Shumpeter, J. A. (1934). *A Theory of Economic Development; An inquiry into Profits, Capital, Credit, Interest, and the Business Cycle*. Cambridge Mass, Harvard University Press.
- Siman, E. (2008). *Wealth accumulation differences between wage-earning and entrepreneurial families: the role of active saving behavior*. Columbia, University of Missouri.
- Smith, A. (1776). *An Inquiry into the Nature and Causes of the Wealth of Nations*.
- Steen, E. V. d. (2004). "Rational Overoptimism (And Other Biases)." *The American Economic Review* 94(4): 1141-1151.
- Tedds, L. M. (2005). *Nonparametric expenditure-based estimation of income under-reporting and the underground economy*. Working Paper. McMaster University.
- Thaler, R. H. (1992). *The Winner's Curse*. New York, The Free Press.
- USGAO (1990). *Tax Administration: Profiles of Major Elements of the Tax Gap*. GGD-90-53BR. Washington DC, US General Accounting Office.

APPENDICIES

**Appendix 1**

i) Note that  $\partial B_n / \partial n$  is clearly positive since:

$$\frac{\partial B_n}{\partial n} = \frac{\sigma_\varepsilon \sigma_\mu^2 (\bar{s} - g(\bar{s}))}{(\sigma_\varepsilon^2 + n\sigma_\mu^2) \int_{(\beta-p)}^{\infty} e^{-1/2t^2} dt} \text{ with all parameters larger than zero and } E(\bar{s} -$$

$g(\bar{s}))$  free of  $n$  as well.

ii) In the case of  $\partial B_n / \partial \sigma_\varepsilon$ , there are no clear-cut signs of the derivative; parameters

produce both positive and negative results. Nonetheless, some bounds could be

established, for example in order for  $\partial B_n / \partial \sigma_\varepsilon > 0$  there is a necessary and

sufficient condition

*Proof:*

*Note that after taking expectations:*

$$\begin{aligned} \frac{\partial B_n}{\partial \sigma_\varepsilon} &= \frac{2e^{-\frac{(\beta-\mu)^2}{2\sigma_\varepsilon^2}} \frac{(\mu-\beta)^2}{2\sigma_\varepsilon^2} n(\beta-\mu)\sigma_\mu^2}{\pi\sigma_\varepsilon^2(n\sigma_\mu^2+\sigma_\varepsilon^2)(1+Erff\left[\frac{\mu-\beta}{\sqrt{2}\sigma_\varepsilon}\right])^2} - \frac{2ne^{-\frac{(\mu-\beta)^2}{2\sigma_\varepsilon^2}} \sqrt{\frac{2}{\pi}}\sigma_\mu^2\sigma_\varepsilon^2}{(n\sigma_\mu^2+\sigma_\varepsilon^2)^2(1+Erff\left[\frac{\mu-\beta}{\sqrt{2}\sigma_\varepsilon}\right])} + \frac{e^{-\frac{(\mu-\beta)^2}{2\sigma_\varepsilon^2}} \sqrt{\frac{2}{\pi}}n\sigma_\mu^2}{(n\sigma_\mu^2+\sigma_\varepsilon^2)(1+Erff\left[\frac{\mu-\beta}{\sqrt{2}\sigma_\varepsilon}\right])} + \\ &\frac{e^{-\frac{(\mu-\beta)^2}{2\sigma_\varepsilon^2}} \sqrt{\frac{2}{\pi}}(\mu-\beta)^2n\sigma_\mu^2}{\sigma_\varepsilon^2(n\sigma_\mu^2+\sigma_\varepsilon^2)(1+Erff\left[\frac{\mu-\beta}{\sqrt{2}\sigma_\varepsilon}\right])} (A) \end{aligned}$$

Now to see how can  $\partial B_n / \partial \sigma_\varepsilon > 0$  I re-arrange and simplify (A) equation to compare first two terms to the second two, the resulting necessary and sufficient condition is then given by:

$$\partial B_n / \partial \sigma_\varepsilon > 0 \quad \text{iff} \quad \frac{\frac{(\hat{p}-\beta)^2}{2e^{-\frac{(\hat{p}-\beta)^2}{2\sigma_\varepsilon^2}} (\hat{p}-\beta)}}{\sqrt{2\pi}(1-\text{Erf}\left(\frac{\hat{p}-\beta}{\sqrt{2}\sigma_\varepsilon}\right))} + \frac{2\sigma_\varepsilon^4}{n\sigma_\mu^2 + \sigma_\varepsilon^2} < \sigma_\varepsilon^2 + (\hat{p} - \beta)^2$$

From which it clearly follows that assuming  $\sigma_\varepsilon^2 > \frac{2\sigma_\varepsilon^4}{n\sigma_\mu^2 + \sigma_\varepsilon^2}$ ,  $(\hat{p} - \beta)^2 >$

$\frac{\frac{(\hat{p}-\beta)^2}{2e^{-\frac{(\hat{p}-\beta)^2}{2\sigma_\varepsilon^2}} (\hat{p}-\beta)}}{\sqrt{2\pi}(1-\text{Erf}\left(\frac{\hat{p}-\beta}{\sqrt{2}\sigma_\varepsilon}\right))}$  are a sufficient and necessary condition for above inequality to

hold ■

There is also a range of parameters for which  $\partial B_n / \partial \sigma_\varepsilon < 0$ . It is clear that reversing the inequality signs produces the two necessary and sufficient conditions for that range of parameters.

iii)  $\partial B_n / \partial \sigma_\mu > 0$  for all values of parameters, this follows from the fact that

differentiating Bias with respect to prior variance yields:

$$\partial B_n / \partial \sigma_\mu = \frac{2 e^{-\frac{(\mu-\beta)^2}{2\sigma_\varepsilon^2}} n^2 \sqrt{\frac{2}{\pi}} \sigma_\mu \sigma_\varepsilon}{(n\sigma_\mu^2 + \sigma_\varepsilon^2)(1 + \text{Erf}\left[\frac{\mu - \beta}{\sqrt{2}\sigma_\varepsilon}\right])} - \frac{2 e^{-\frac{(\mu-\beta)^2}{2\sigma_\varepsilon^2}} n \sqrt{\frac{2}{\pi}} \sigma_\mu^3 \sigma_\varepsilon}{(n\sigma_\mu^2 + \sigma_\varepsilon^2)^2 (1 + \text{Erf}\left[\frac{\mu - \beta}{\sqrt{2}\sigma_\varepsilon}\right])}$$

which can be easily simplified to:

$$\frac{\partial B_n}{\partial \sigma_\mu} = \frac{2 e^{-\frac{(\mu-\beta)^2}{2\sigma_\varepsilon^2}} \sqrt{\frac{2}{\pi}} \sigma_\varepsilon^3 n \sigma_\mu}{(n\sigma_\mu^2 + \sigma_\varepsilon^2)^2 (1 + \text{Erf}\left[\frac{\mu-\beta}{\sqrt{2}\sigma_\varepsilon}\right])}$$

It is immediately clear that  $\frac{\partial B_n}{\partial \sigma_\mu} > 0$ .

iv)  $\frac{\partial B_n}{\partial \beta}$  could also take on both positive and negative values depending on

parameters since differentiating with respect to the truncation point yields:

$$\begin{aligned} & \frac{\partial B_n}{\partial \beta} \\ &= \frac{2e^{-\frac{(\beta-\mu)^2}{2\sigma_\varepsilon^2}} \frac{(\mu-\beta)^2}{2\sigma_\varepsilon^2} n\sigma_\mu^2}{\pi(n\sigma_\mu^2 + \sigma_\varepsilon^2) \left(1 + \text{Erf}\left[\frac{\mu-\beta}{\sqrt{2}\sigma_\varepsilon}\right]\right)^2} - \frac{e^{-\frac{(\mu-\beta)^2}{2\sigma_\varepsilon^2}} \sqrt{\frac{2}{\pi}} (\mu-\beta) n\sigma_\mu^2}{\sigma_\varepsilon (n\sigma_\mu^2 + \sigma_\varepsilon^2) \left(1 + \text{Erf}\left[\frac{\mu-\beta}{\sqrt{2}\sigma_\varepsilon}\right]\right)} \quad (\text{B}) \end{aligned}$$

where in case of severe truncation ( $\beta > \mu$ ) it is clear that  $\frac{\partial B_n}{\partial \beta} > 0$ . In the case

of milder truncation ( $\beta < \mu$ ) this result is more difficult to obtain. More specifically

an additional condition needs to be imposed.

Rearranging (B), while assuming  $\beta < \mu$ , it becomes clear that

$$\frac{\partial B_n}{\partial \beta} > 0 \quad \text{iff} \quad 2\sigma_\varepsilon e^{-\frac{(\mu-\beta)^2}{2\sigma_\varepsilon^2}} > \sqrt{2\pi} (\hat{p} - \beta) \left(1 + \right.$$

$$\left. \text{Erf}\left[\frac{\mu-\beta}{\sqrt{2}\sigma_\varepsilon}\right]\right)$$

## Dynamics

*Note that while Equation 13 has no impact on bias we need to make sure solution exists and that is finite.*

$$\text{solution to } \sigma_{t+1} = \frac{\sigma_{\varepsilon}^2 \sigma_{\mu(t)}^2}{\sigma_{\mu(t)}^2 + n\sigma_{\varepsilon}^2} \quad \text{is} \quad \sigma_{\mu(t)} = \frac{(n-1)n\sigma_{\varepsilon}^2}{n^{t+1} - n^t\sigma_{\varepsilon}^2 + n^{t+1}\sigma_{\varepsilon}^2 - n}$$

*Assuming that initial condition constant is equal to 1. Clearly above solution converges to 0.*



## Appendix 2

### 2.A. Unit Roots Tests for Stationarity

The Im-Pesaran-Shin (IPS) test specifies a set of augmented Dickey-Fuller regressions of the form

$$\Delta y_{it} = \varphi_i y_{i,t-1} + \mathbf{z}'_{it} \gamma_i + \epsilon_{it}$$

where  $\varphi_i$  is a panel-specific constant with an underlining assumption that  $\epsilon_{it} \sim iid$  for all  $i$  and  $t$  (Im, Pesaran et al. 2003). Hypotheses specification is that the null hypotheses  $\varphi_i = 0$  for all  $i$ .

The alternative is that the fraction of panels that are stationary is larger than zero. Let that fraction be represented by number  $A$ . The IPS test allows for the heterogeneous panels with serially uncorrelated errors for  $T$  periods and  $N$  panels. An asymptotic assumption is that  $T \rightarrow \infty$ . Test statistics arises from the fact that as  $N \rightarrow \infty$ ,  $A/N$  is a non-zero value. Using the IPS as a base Fisher type tests combine the  $p$ -values from panel-specific test using inverse  $\chi^2$ , inverse normal, inverse logit, and inverse  $\chi^2$  transformed to suit  $N \rightarrow \infty$  (Choi 2001).

The test statistics for augmented Dickey-Fuller tests are as follows:

$$ADF_t = t_{\varphi=1} = \frac{\hat{\varphi} - 1}{SE(\varphi)}$$

$$ADF_n = \frac{T(\hat{\varphi} - 1)}{1 - \hat{z}_1 - \dots - \hat{z}_p}$$

Table A1. *Unit-root tests using the augmented Dickey-Fuller specification*

<i>Ho: All panels contain unit roots</i>		Number of panels	14381
<i>Ha: At least on panel is stationary</i>		Avg. number of periods	17.74
<i>AR parameter: Panel specific</i>		<i>Asymptotics:</i>	$T \rightarrow \infty$
<i>Panel means: Included</i>		<i>ADF regressions:</i>	1 lag
		<i>Statistic</i>	<i>P-value</i>
Inverse chi-squared (25424)	P	100076	1.000
Inverse normal	Z	44.4336	1.000
Inverse logit t(43894)	L	35.6421	1.000
Modified inv.chi-squared	Pm	-34.9105	1.000

I also employ Phillips-Perron (PP) type of specification as the other alternative specification of the unit-root structure. The regression for the PP test is:

$$\Delta y_{it} = \beta_i' \mathbf{D}_{it} + \pi_i y_{it} + u_{it}$$

where D is the vector of deterministic parameters (constant, trend etc.). The PP tests correct for any serial correlation and heteroskedasticity in errors  $u_{it}$  of the regression by modifying the test statistic  $t_{\pi=0}$  and  $t_{\hat{\pi}}$ . The statistics denoted  $Z_t$  and  $Z_\pi$  are given by

$$Z_t = \left( \frac{\hat{\sigma}^2}{\hat{\lambda}^2} \right)^{1/2} \cdot t_{\pi=0} - \frac{1}{2} \left( \frac{\hat{\lambda}^2 - \hat{\sigma}^2}{\hat{\lambda}^2} \right) \cdot \left( \frac{T \cdot SE(\hat{\pi})}{\hat{\sigma}^2} \right)$$

$$Z_\pi = T_{\hat{\pi}} - \frac{1}{2} \frac{T^2 \cdot SE(\hat{\pi})}{\hat{\sigma}^2} (\hat{\lambda}^2 - \hat{\sigma}^2)$$

with terms  $\hat{\lambda}^2$  and  $\hat{\sigma}^2$  as consistent estimates of the variance parameters

$$\sigma^2 = \lim_{T \rightarrow \infty} T^{-1} \sum_{t=1}^T E[u_t^2]$$

$$\lambda^2 = \lim_{T \rightarrow \infty} \sum_{t=1}^T E[T^{-1} S_T^2]$$

where  $S_T = \sum_{t=1}^T u_t$ .

Table A2. Unit-root tests using the Phillips-Perron specification

<i>Ho: All panels contain unit roots</i>		Number of panels	14381
<i>Ha: At least on panel is stationary</i>		Avg. number of periods	17.74
<i>AR parameter: Panel specific</i>		<i>Asymptotics:</i>	$T \rightarrow \infty$
<i>Panel means: Included</i>		<i>ADF regressions:</i>	1 lag
		<i>Statistic</i>	<i>P-value</i>
Inverse chi-squared (25424)	P	241076	1.000
Inverse normal	Z	26.5549	1.000
Inverse logit t(43894)	L	21.3670	1.000
Modified inv. chi-squared		-14.9941	1.000
Pm			

**2.B. Confidence Intervals and sample sizes for Table 2.**

Three numbers are reported in each cell. The top (bottom) number is the upper (lower) 95 percent confidence bound for the means given in Table 2. The middle numbers record the number of transitions observed.

**Table B1. Confidence Intervals and sample sizes for (one- month transition rates)**

<b>Panel A. Business owners in both periods</b>						
		Business Owners, $t+1$				
		$1_{(t+1)}$	$2_{(t+1)}$	$3_{(t+1)}$	$4_{(t+1)}$	$5_{(t+1)}$
Business Owners, $t$	$1_{(t)}$	75.3	9.2	3.6	2.7	3.9
		1,935	262	111	86	120
		78.6	11.6	5.2	4.1	5.6
	$2_{(t)}$	8	71	9.5	3	2.7
		219	1,737	257	90	83
		10.3	74.6	12	4.5	4.2
	$3_{(t)}$	3.5	8.6	68.3	9.8	3.7
		101	231	1,647	261	107
		5.1	11	72	12.4	5.4
	$4_{(t)}$	2.6	2.1	7.7	70.7	11.4
		85	71	224	1,854	326
		4	3.4	9.8	74.2	14
	$5_{(t)}$	2	1.2	2.5	5.8	85.2
		100	64	122	263	3,453
		3	2	3.6	7.3	87.3

<b>Panel B. Wage workers in both periods</b>						
		Wage Earners, $t+1$				
		$1_{(t+1)}$	$2_{(t+1)}$	$3_{(t+1)}$	$4_{(t+1)}$	$5_{(t+1)}$
Wage Earners, $t$	$1_{(t)}$	60.7	11.8	8.7	8.4	8.7
		22,888	4,542	3,370	3,254	3,348
		61.7	12.5	9.3	9	9.2
	$2_{(t)}$	10.3	56.8	14.1	8.7	8.3
		4,426	23,860	6,028	3,741	3,574
		10.9	57.8	14.8	9.3	8.9
	$3_{(t)}$	7.6	13.8	52.2	16.2	8.5
		3,329	5,959	22,196	6,993	3,685
		8.2	14.5	53.1	16.9	9
	$4_{(t)}$	7.3	8.1	15.3	52.6	15
		3,364	3,701	6,906	23,516	6,799
		7.8	8.6	15.9	53.6	15.7
	$5_{(t)}$	7.3	7.6	7.9	14.6	61.1
		3,40	3,529	3,656	6,718	27,758
		7.8	8.1	8.4	15.2	62

<b>Panel C. Wage workers establishing a business</b>						
		Business Owners, $t+1$				
		$1_{(t+1)}$	$2_{(t+1)}$	$3_{(t+1)}$	$4_{(t+1)}$	$5_{(t+1)}$
Wage Earners, $t$	$1_{(t)}$	8.5	1.5	10.5	13.5	24.1
		14	6	16	19	29
		24.8	12.8	27.6	21.8	44.9
	$2_{(t)}$	2.1	4.4	8.1	12.1	23.7
		6	8	11	14	22
		17.5	21.8	28	33.8	48.5
	$3_{(t)}$	7.3	4.3	5.3	9.4	30.3
		12	9	10	14	22
		23.9	19	20.7	27	52.8
	$4_{(t)}$	8.2	11.3	2.5	12.4	21.5
		13	16	7	17	25
		25.1	29.7	15.5	31.2	42.6
	$5_{(t)}$	7.1	6.2	7.1	14.6	23.7
		13	12	13	21	30
		22.1	20.7	22.1	32.6	43.7

<b>Panel D. Business owners transitioning in to wage work</b>						
		Business Owners, $t+1$				
		$1_{(t+1)}$	$2_{(t+1)}$	$3_{(t+1)}$	$4_{(t+1)}$	$5_{(t+1)}$
Business Owners, $t$	$1_{(t)}$	13	7.8	5.4	6.6	17.1
		15	11	9	10	18
		34.6	27.1	23.2	25.2	40
	$2_{(t)}$	8.6	13.6	8.6	5.1	14.9
		12	16	12	9	17
		27.7	34.9	27.7	22.1	36.6
	$3_{(t)}$	10.9	10.9	9.6	12.3	4.5
		13	13	12	14	8
		32.4	32.4	30.4	34.4	22.2
	$4_{(t)}$	9.1	6.9	15	15	6.9
		13	11	18	18	11
		27.5	24.1	35.7	35.7	24.1
	$5_{(t)}$	9.3	7.8	13.1	17	17
		18	16	23	28	26
		23.2	21.1	28.4	33.4	33.4

## 2.C. Confidence Intervals and sample sizes for 4M income transitions

Three numbers are reported in each cell. The top (bottom) number is the upper (lower) 95 percent confidence bound for the means given in Table 4. The middle numbers record the number of transitions observed.

**Table B1.** *Confidence Intervals and sample sizes for (4M transition rates, 4M income)*

<b>Panel A. Business owners in both periods</b>						
		Business Owners, $t+1$				
		$1_{(t+1)}$	$2_{(t+1)}$	$3_{(t+1)}$	$4_{(t+1)}$	$5_{(t+1)}$
Business Owners, $t$	$1_{(t)}$	61.3	2	2.7	7.6	4.7
		133	10	12	24	17
		74.5	8.2	9.5	16.9	12.6
	$2_{(t)}$	2.2	52	3.3	9.7	9.7
		11	121	14	30	30
		8.4	65.5	10.3	19.4	19.4
	$3_{(t)}$	1.6	5.5	53.4	8.2	11.9
		11	25	168	34	46
		6.1	12.1	64.9	15.8	20.5
	$4_{(t)}$	2.6	4.2	6.6	55.9	13.5
		17	24	34	213	61
		7.1	9.5	12.9	66.2	21.5
	$5_{(t)}$	3.3	4.8	7.5	8.5	61.9
		26	35	50	56	326
		7.3	9.4	12.8	14.2	70.3
<b>Panel B. Wage workers in both periods</b>						
		Wage Earners, $t+1$				
		$1_{(t+1)}$	$2_{(t+1)}$	$3_{(t+1)}$	$4_{(t+1)}$	$5_{(t+1)}$
Wage Earners, $t$	$1_{(t)}$	51.7	11.9	12.1	11.5	11.3
		29,160	6,785	6,903	6,590	6,496
		52.5	12.4	12.6	12	11.9
	$2_{(t)}$	14.3	20.4	21.4	21.4	20.7
		7,034	9,976	10,472	10,469	10,117
		14.9	21.1	22.2	22.1	21.4
	$3_{(t)}$	13.3	20.4	22	21.8	20.8
		6,895	10,499	11,315	11,241	10,720
		13.9	21.1	22.7	22.5	21.5
	$4_{(t)}$	13.2	20.4	22	21.7	21
		6,767	10,387	11,218	11,055	10,691
		13.8	21.1	22.7	22.4	21.7
	$5_{(t)}$	13.1	20	21.4	21.9	21.7
		6,524	9,910	10,613	10,859	10,754
		13.41	20.7	22.2	22.7	22.5

<b>Panel C. Wage workers establishing a business</b>						
		Business Owners, $t+1$				
		$1_{(t+1)}$	$2_{(t+1)}$	$3_{(t+1)}$	$4_{(t+1)}$	$5_{(t+1)}$
Wage Earners, $t$	$1_{(t)}$	3.1	3.1	0	9.1	9.1
		5	5	1	7	7
		36.9	36.9	12.3	46.9	46.9
	$2_{(t)}$	20.4	0	0	0.06	3.3
		10	2	3	4	5
		62.9	20.3	26.8	32.7	38.4
	$3_{(t)}$	10.1	2	0.3	3.9	17.2
		9	5	4	6	12
		39.9	25.8	21.9	29.5	49.5
	$4_{(t)}$	8.6	0	4.5	0	0
		6	3	5	2	3
		54.6	33.8	48.1	25.7	33.8
	$5_{(t)}$	7.5	9.5	7.5	0.3	9.5
		8	9	8	4	9
		34.6	37.8	34.6	20.7	37.8

<b>Panel D. Business owners transitioning in to wage work</b>						
		Business Owners, $t+1$				
		$1_{(t+1)}$	$2_{(t+1)}$	$3_{(t+1)}$	$4_{(t+1)}$	$5_{(t+1)}$
Business Owners, $t$	$1_{(t)}$	0	7	7	9.5	7
		2	7	7	8	7
		15.6	38.2	38.2	42.1	38.2
	$2_{(t)}$	8	8	5.9	3.9	5.9
		8	8	7	6	7
		36.5	36.5	33	29.5	33
	$3_{(t)}$	4.5	9.1	17	0	0.4
		6	8	11	3	4
		33	40.9	51.8	20.1	24.6
	$4_{(t)}$	16.4	2.2	4.3	2.2	4.3
		11	5	6	5	6
		50.3	28.1	32.1	28.1	32.1
	$5_{(t)}$	5.5	0	14.3	8.3	2.9
		6	0	9	7	5
		39	0	52.3	43.6	34.2

### Appendix 3

#### Regressions for Table 11

#### Additional Evidence of Underreporting (Median Estimates)

Dependent variable expenditure in thousands, \*\*\* $\alpha=0.01$ . \*\* $\alpha=0.05$ . \* $\alpha=0.1$

<u>A1. Current propensities to save using 1995 employment status.</u>		
	Current Wage Earners	Current Self-employed
<i>Income</i>	.000177722*** (.0000139353)	.0001760637*** (.0000280074)
<i>Education</i>	1.657036*** (.2229091)	.8809366 (.9877422)
<i>Race White</i>	1.955574* (1.185384)	3.24959 (2.936027)
<i>Experience</i>	-.0536193 (.06186582)	.1522846 (.2037608)
<i>Family size</i>	-4.55209*** (.9903618)	-2.519669* (1.390202)
<i>(Self-Employed 1995)x(2005 Income)</i>	.000177722*** (.0000139353)	.0000807228*** (.0000277231)
<i>Constant</i>	-12.2845*** (2.670906)	-6.404724 (15.07987)
	N=817, R <sup>2</sup> =0.089	N=158, R <sup>2</sup> =0.16

A2. Current propensities to save using 1996 employment status.

	Current Wage Earners	Current Self-employed
<i>Income</i>	.0001881638*** (.00001257319)	.0001595584*** (.0000281575)
<i>Education</i>	1.854621*** (.2482233)	.8614481 (.9150135)
<i>Race White</i>	2.927793** (1.410048)	2.456798 (2.872762)
<i>Experience</i>	.03436921 (.0685051)	.1516165 (.1855706)
<i>Family size</i>	-4.670029*** (.9261412)	-2.331599* (1.347432)
<i>(Self-Employed 1996)x(2005 Income)</i>	-.0001698983*** (.0000125199)	.0000980634*** (.0000276934)
<i>Constant</i>	-15.9463*** (3.210247)	-6.16006 (13.28564)
	N=839, R <sup>2</sup> =0.096	N=164, R <sup>2</sup> =0.16

A3. Current propensities to save using 1997 employment status.

	Current Wage Earners	Current Self-employed
<i>Income</i>	.0001901903*** (.0000123907)	.0001842529*** (.0000328447)
<i>Education</i>	1.077878*** (.1617026)	1.181003 (.9881792)
<i>Race White</i>	2.61863** (1.099963)	.9665809 (5.519584)
<i>Experience</i>	.07094096 (.05348281)	-.01802031 (.1775672)
<i>Family size</i>	-4.265715*** (.8894281)	-2.910628* (1.527428)
<i>(Self-Employed 1997)x(2005 Income)</i>	-.0001711012*** (.0000096425)	.000070879** (.0000325506)
<i>Constant</i>	-7.234556*** (2.093277)	-5.58392 (14.2219)
	N=957, R <sup>2</sup> =0.089	N=171, R <sup>2</sup> =0.16



A4. Current propensities to save using 1999 employment status.

	Current Wage Earners	Current Self-employed
<i>Income</i>	.0002188864*** (.0000143044)	.000258373*** (.0000279386)
<i>Education</i>	.5615408*** (.1646435)	.6577603 (1.196253)
<i>Race White</i>	2.718512** (1.099693)	-1.640338 (6.437674)
<i>Experience</i>	.02313125 (.05157134)	-.07083331 (.2362744)
<i>Family size</i>	-4.547023*** (.8895046)	-2.744733 (1.839639)
<i>(Self-Employed 1999)x(2005 Income)</i>	-.0001974567*** (.0000112482)	.0000381247 (.0000280574)
<i>Constant</i>	-.9092512 (2.119199)	1.817691 (16.97725)
	N=1067, R <sup>2</sup> =0.096	N=182, R <sup>2</sup> =0.16

A5. Current propensities to save using 2001 employment status.

	Current Wage Earners	Current Self-employed
<i>Income</i>	.0001911024*** (.0000164882)	.0002293289*** (.0000318041)
<i>Education</i>	.8929879*** (.1756677)	-.07335361 (1.176397)
<i>Race White</i>	2.397245** (1.055269)	-2.522549 (6.93356)
<i>Experience</i>	.0386792 (.05366305)	-.05446202 (.2110872)
<i>Family size</i>	-4.377966*** (.7391648)	-2.775846 (1.959667)
<i>(Self-Employed 2001)x(2005 Income)</i>	-.000011739 (.0000180819)	.0000697666** (.0000318041)
<i>Constant</i>	-4.705587** (2.161508)	13.48344 (16.28237)
	N=1185, R <sup>2</sup> =0.097	N=192, R <sup>2</sup> =0.16

A6. Current propensities to save using 2003 employment status.

	Current Wage Earners	Current Self-employed
<i>Income</i>	.0001943649*** (.0000142112)	.0001527923*** (.0000287006)
<i>Education</i>	.8216306*** (.1513525)	.7662472 (.9020232)
<i>Race White</i>	2.110387** (.8718046)	1.331547 (5.466565)
<i>Experience</i>	.05597175 (.04399514)	-.06633353 (.1992514)
<i>Family size</i>	-4.574083*** (.6220216)	-3.100908** (1.464956)
<i>(Self-Employed 2003)x(2005 Income)</i>	-.000313388** (.0000142112)	.0000113688*** (.0000284406)
<i>Constant</i>	-3.494536* (1.836009)	.6483869 (11.23985)
	N=1351, R <sup>2</sup> =0.077	N=212, R <sup>2</sup> =0.16

## Long Run Effects (Median Estimates)

Dependent variable expenditure in thousands, \*\*\* $\alpha=0.01$ . \*\* $\alpha=0.05$ . \* $\alpha=0.1$

A7. Effects of 1995 employment status on current expenditure.

	Current Wage Earners	Current Self-employed
<i>Education</i>	7.076*** (.6638936)	10.75564*** (2.684138)
<i>Race White</i>	8.312667** (3.350804)	5.244272 (20.56587)
<i>Experience</i>	.209333 (.1735715)	-.6293636 (.6929092)
<i>Family size</i>	-6.696444** (2.801582)	-8.889273 (11.63469)
<i>Self-Employed 1995</i>	18.18467*** (5.656639)	28.57373** (12.81543)
<i>Constant</i>	-54.56933*** (9.051145)	-84.50963 (42.65558)
	N=915, R <sup>2</sup> =0.050	N=185, R <sup>2</sup> =0.061

A8. Effects of 1996 employment status on current expenditure.

	Current Wage Earners	Current Self-employed
<i>Education</i>	7.508965*** (.6356214)	10.800*** (2.150845)
<i>Race White</i>	7.426621** (3.236627)	-1.895273 (15.52712)
<i>Experience</i>	.2998621* (.1652749)	-.1398182 (.5208083)
<i>Family size</i>	-7.151724** (2.691952)	-21.42982* (11.07528)
<i>Self-Employed 1996</i>	11.25234** (4.895187)	28.06273*** (9.957066)
<i>Constant</i>	-60.25434*** (8.640316)	-82.30836 (33.05751)
	N=942, R <sup>2</sup> =0.053	N=192, R <sup>2</sup> =0.063

A9. Effects of 1997 employment status on current expenditure.

	Current Wage Earners	Current Self-employed
<i>Education</i>	4.857238*** (.5156421)	9.74377*** (1.848756)
<i>Race White</i>	6.122857* (3.132357)	6.303385 (12.97612)
<i>Experience</i>	.2480952 (.1569289)	.5830769 (.4479618)
<i>Family size</i>	-8.289571*** (2.783942)	-8.897693 (10.04055)
<i>Self-Employed 1997</i>	4.538 (4.771968)	2.083077 (9.8912681)
<i>Constant</i>	-24.17914*** (7.28194)	-78.05108 (27.67717)
	N=1073, R <sup>2</sup> =0.040	N=202, R <sup>2</sup> =0.060

A10. Effects of 1999 employment status on current expenditure.

	Current Wage Earners	Current Self-employed
<i>Education</i>	4.887877*** (.4741486)	7.911467*** (1.458155)
<i>Race White</i>	6.631234** (2.822728)	11.67733 (10.34467)
<i>Experience</i>	.4017284*** (.1447127)	.3026667 (.3707955)
<i>Family size</i>	-9.086362*** (2.500521)	-14.45367** (6.099329)
<i>Self-Employed 1999</i>	1.227868 (4.913084)	1.9696 (7.047)
<i>Constant</i>	-26.80052*** (6.677371)	-58.2656*** (22.05197)
	N=1198, R <sup>2</sup> =0.040	N=214, R <sup>2</sup> =0.056

A11. Effects of 2001 employment status on current expenditure.

	Current Wage Earners	Current Self-employed
<i>Education</i>	4.68387*** (.3860218)	8.039644*** (2.290889)
<i>Race White</i>	6.402304*** (2.227324)	1.044633 (16.38391)
<i>Experience</i>	.469087*** (.1145432)	.01542573 (.5648374)
<i>Family size</i>	-6.654967*** (1.688007)	-16.84731* (10.10809)
<i>Self-Employed 2001</i>	3.715022 (3.836145)	14.87653 (11.09489)
<i>Constant</i>	-26.50613*** (5.33793)	-50.58998 (34.7174)
	N=1330, R <sup>2</sup> =0.050	N=226, R <sup>2</sup> =0.057

A12. Effects of 2003 employment status on current expenditure.

	Current Wage Earners	Current Self-employed
<i>Education</i>	4.426252*** (.3441727)	7.710286*** (1.309546)
<i>Race White</i>	7.48526*** (1.95912)	6.867858 (8.996112)
<i>Experience</i>	.4786866*** (.1008178)	.4078571 (.33343)
<i>Family size</i>	-6.519347*** (1.50875)	-7.574286 (7.192842)
<i>Self-Employed 2003</i>	.02457672 (3.837547)	7.814 (6.44438)
<i>Constant</i>	-25.2937*** (4.744954)	-58.376*** (19.3729)
	N=1517, R <sup>2</sup> =0.037	N=248, R <sup>2</sup> =0.049

VITA

DMITRIY KRICHEVSKIY

February 1, 1976	Born, Moscow Russia
2004	B.A. Economics Florida International University Miami, Florida
2006	M.A. Economics Florida International University Miami, Florida
2005-2007	Teaching Assistant Florida International University Miami, Florida
2007-2008	Kauffman Fellow Florida International University Miami, Florida
2009-2010	Research Assistant Lumina Foundation Florida International University Miami, Florida
2011	Dissertation Year Fellowship Florida International University Miami, Florida