

Of speculators, migrants and entrepreneurs

Essays on the economics of trying your fortune



STOCKHOLM SCHOOL
OF ECONOMICS
HANDELSHÖGSKOLAN I STOCKHOLM

EFI, The Economic Research Institute

EFI Mission

EFI, the Economic Research Institute at the Stockholm School of Economics, is a scientific institution that works independently of economic, political and sectional interests. It conducts theoretical and empirical research in the management and economic sciences, including selected related disciplines. The Institute encourages and assists in the publication and distribution of its research findings and is also involved in the doctoral education at the Stockholm School of Economics. At EFI, the researchers select their projects based on the need for theoretical or practical development of a research domain, on their methodological interests, and on the generality of a problem.

Research Organization

The research activities at the Institute are organized into 21 Research Centres. Centre Directors are professors at the Stockholm School of Economics.

EFI Research Centre:

Management and Organisation (A)
Centre for Entrepreneurship and Business Creation (E)
Public Management (F)
Information Management (I)
Centre for People and Organization (PMO)
Centre for Innovation and Operations Management (T)
Centre for Media and Economic Psychology (P)
Centre for Consumer Marketing (CCM)
Centre for Information and Communication Research (CIC)
Marketing, Distribution and Industrial Dynamics (D)
Centre for Strategy and Competitiveness (CSC)
Centre for Business and Economic History (BEH)
Accounting and Managerial Finance (B)
Centre for Financial Analysis and Managerial Economics in
Accounting (BFAC)
Finance (FI)
Centre for Health Economics (CHE)
International Economics and Geography (IEG)
Economics (S)
Economic Statistics (ES)
Law (RV)
Centre for Tax Law (SR)

Centre Director:

Sven-Erik Sjöstrand
Carin Holmquist
Nils Brunsson
Mats Lundeberg
Andreas Werr (acting)
Christer Karlsson
Guje Sevón
Magnus Söderlund
Per Andersson (acting)
Björn Axelsson
Örjan Sölvell
Håkan Lindgren
Johnny Lind
Kenth Skogsvik

Clas Bergström
Bengt Jönsson
Mats Lundahl
Paul Segerstrom
Anders Westlund
Johnny Herre
Bertil Wiman

Chair of the Board: Professor Carin Holmquist

Director: Associate Professor Filip Wijkström

Address

EFI, Box 6501, SE-113 83 Stockholm, Sweden • Website: www.hhs.se/efi/
Telephone: +46(0)8-736 90 00 • Fax: +46(0)8-31 62 70 • E-mail efi@hhs.se

Of speculators, migrants and entrepreneurs

Essays on the economics of trying your fortune

Milo Bianchi



STOCKHOLM SCHOOL
OF ECONOMICS
HANDELSHÖGSKOLAN I STOCKHOLM

Dissertation for the Degree of Doctor of Philosophy, Ph.D.
Stockholm School of Economics

KEYWORDS: Speculative bubbles, bounded rationality; immigrants' self-selection, political economy of immigration, immigration policy preferences; financial development; entrepreneurship, job satisfaction; neoclassical modeling.

©EFI and the author, 2007
ISBN NR 978-91-7258-724-3

PRINTED BY:
Elanders Gotab, Stockholm 2007

DISTRIBUTED BY:
EFI, The Economic Research Institute
Stockholm School of Economics
P O Box 6501, SE-113 83 Stockholm
www.hhs.se/efi

Contents

Acknowledgement	ix
Introduction	1
Chapter 1. Introduction	3
1. Speculators	4
2. Migrants	5
3. Entrepreneurs	5
Speculators	7
Chapter 2. Speculative Bubbles without Stupid Investors	9
1. Introduction	9
2. The model	16
3. Analysis	20
4. Bubbles and rationality	28
5. Conclusion	36
References	39
Migrants	43
Chapter 3. Immigration Policy and Self-Selecting Migrants	45
1. Introduction	45
2. The model	50
3. Analysis	52
4. Discussion and extensions	60
5. Conclusion	65
References	69
Entrepreneurs	75
Chapter 4. Financial Development, Entrepreneurship, and Job Satisfaction	77
1. Introduction	77
2. The Model	79
3. Analysis	83
4. Testing the model	86
5. Empirical evidence	87

6. Conclusion	93
7. Appendix	94
References	103
Chapter 5. Is Neoclassical Economics still Entrepreneurless?	105
1. Introduction	105
2. Entrepreneurship in the social sciences	107
3. Modeling entrepreneurship in neoclassical economics	108
4. Summary and assessment of the contributions	121
5. Useful models and useful policies	122
References	127

Acknowledgement

E' un libero cercare una parola leggera,
che dica tutto col peso di niente
e che ci sembri vera.
E benvenuto sia ogni abbaglio del cuore,
e benvenuto sia anche l'errore.
(Teresa de Sio)¹

Let me first thank a few people who have made this thesis possible, and my life in the past five years unique.

Thanks to many professors. I am grateful especially to Jörgen Weibull, who has guided my anarchic spirit with generosity and enthusiasm. I have truly appreciated his advice, typically offered with a metaphor or with a smile; his curiosity; his care for details. I am also indebted to Tore Ellingsen, who showed me that one can be broad and deep, and also how to punctuate formulas; to Philippe Jehiel, who has taught me there is no need for long sentences to communicate essential concepts; and to Abhijit Banerjee, who has been a profound source of inspiration, and accepted to be a kind discussant of this thesis. Many other professors have enriched my academic life, and I would like to mention in particular Mike Burkart, David Domeij, Martin Flodén, Magnus Henrekson, Giancarlo Spagnolo, Karl Wärneryd, and Muhamet Yildiz.

Thanks to the wonderful classmates. Ola, who showed me how cold Sweden can be, and he has been the *grande maestro* since the very first day; Henrik, with whom I feel like talking about myself as it seldom happens; Erik, who has taught me the essential words to survive in Sweden, and the distinctive smell of *rapsolja*; Per, who shares the taste for old-times stuffs, but he also provides a good excuse to wear new shoes; Björn, who gave me some unique experiences, like *snus* and *drikka*, and insisted on the importance of going out for lunch (which took me quite a while to understand).

Thanks to the Kista community. Mirco, honorary member, whom is nice to meet whenever we happen to be in the same city at the same time; Jaewon and Andrew (and Jun), who gave me a roof, and showed me how to cook fried noodles; Cristina, who always cries when something nice happens to me (or she is told so). And Ganesh,

¹ It is a free search for a weightless word, which tells everything as if it were nothing, and which seems true./ And may we welcome any heart blunder, and may we welcome even mistakes. (Own translation)

who has been a true companion of uncountable meals, drinks and walks. He still tries to put green chili into my pasta, and I am still not convinced that his *bhurji* is the way it should be, but anyway, these are typical old couple discussions...

Thanks to the job market group, and in particular to Cédric and Elena, who have shared the ups and downs of this intense period, and sometimes even made it fun (didn't we feel like rock stars?) Thanks to the Italian crew, Caterina, Dario, Emilio, Ettore, Margherita, Mema, Sara, Stefano, who have helped to preserve the taste of home; and to many who have made the days at school more amusing: Agatha, Birgit, Camilo, Chloé, Damian, Erik, Fredrik, Liam, Marcus, Marieke, Martin, Mauricio, Max, Mia, Robert (Bobby), Sam, and all those I forget to mention here.

Thanks to several places. It has been a great pleasure to be at SSE, also because of the friendly staff (let me thank in particular Anneli and Ritva). The departments of economics at UCL in London and at MIT have also been wonderful places for working and meeting people. Staying in the US has left me some understanding about economics, a few very good friends (Filippo, Giovanni, Nicolas), and a taste for real coffee, coke and California. I also acknowledge Ryanair, SAS, and Sterling, which provided kind hospitality while part of this thesis was written.

Thanks to my family, who has always tried to interpret my few words and understand what I do, why I do it far from home, and even answer my grandma's question: "But so, at the end, what is that Milo keeps searching?" They are where I come from, and it is easier to move around with these roots.

Thanks to my friends back home, in Sorbolo and nearby, who have instead given up trying to understand what I do. Despite pulling my leg as I am still studying, they always welcome me back at the bar as if I had never left.

Thanks (and a kiss) to Novita. Among other things, she has taught me to be a clown, and to collect moments.

Paris, September 2007.

Introduction

CHAPTER 1

Introduction

This thesis builds on a few questions that I have encountered in recent years. I started to consider that financial markets go up and down in the early 2000s, when a few friends lost their job, the school lost quite some money, and my little sister learned the word Nasdaq. Although I still skip the financial pages in the newspaper, I find these markets an ideal test bed for our theories on individual rationality and on its bounds.

Having decided to leave my hometown, and having continued to move from then on, I got to think about migration. In particular, I have asked myself who is more likely to migrate, and how the answer may depend on the way immigrants are treated in the receiving countries.

My interest in entrepreneurs might be due to my cautious and a bit lazy attitude. As such, I am attracted from people who invest their time and money in pursuing a business idea, and in getting other people involved into it. Even if I have always felt quite unfit for any entrepreneurial activity, I have discovered that sometimes being a researcher requires something similar.

This work collects my explorations into these issues. As such, it may reveal some idiosyncrasies of my approach to research. In spite of them, the title of the thesis draws a connection between the essays. They are about fortune; about people's opportunities to find it (or, if one prefers, to shape it). Speculators confide in their luck, when they bet on the market against other investors. Migrants try their fate, when they pack their stuff and move to a foreign country. Entrepreneurs trust their fortune, when they set up their business, and devote themselves to grow it. PhD students too taste their luck, when they move their first steps into research, playing with symbols to see what happens.

A second connection between these essays is methodological. The papers have striven to find the simplest way to convey an idea. Despite being almost entirely theoretical, their formal sophistication is not so high. This is not to cast doubts on the usefulness of rigorous modeling. On the contrary, it reflects my personal view that models are even more useful when they uncover their logic without obscure constructions or long calculus. And, to my taste, simple models tend to be beautiful, also

when they are as barebones as the ones presented below. Indeed, ancient Latins even believed that simplicity is the seal of the true ("*Simplex sigillum veri*"). While these papers are far from the truth, they still aim to be simple.

I leave the reader with the task of finding other red threads along the lines of this thesis, and I turn to a summary of the next Chapters.

1. Speculators

The second Chapter - "Speculative bubbles without stupid investors" - presents a model of speculative bubbles in which some investors have a *partial* understanding of the strategies of other investors. In particular, they understand only the aggregate selling and buying strategies of other traders along the bubble, without a precise understanding of the timing of these decisions. One reason may be that detecting the date of the crash is quite hard: crashes tend to occur even with no fundamental news event, in days that may appear similar to many others.

As a result, these investors form erroneous beliefs about when the crash will occur, and they are not able to sell just before it. This allows bubbles to arise even in a setting with finite horizon, no private information and in which all investors, throughout the game, are aware that the market is overvalued and bound to crash.

The resulting bubble equilibrium displays a number of features that have been documented in the empirical literature. First, the market becomes increasingly euphoric. High short-term returns induce some investors to trade in the speculative market, and to remain invested even longer than they had originally planned. Eventually, such investors overestimate the duration of the bubble, and fail to sell before the crash. On the other hand, fully rational investors are able to make great profits by riding the bubble for a while and exiting just before the endogenous crash.

Such equilibrium bubble is more likely to arise and to be longer the greater is the amount of people who can potentially enter the speculative market relatively to those who have already invested. That is, bubbles are characterized by a short supply of speculative stocks and by the involvement of many new and typically inexperienced investors.

Moreover, it turns out that the relation between bubbles and rationality is multifaceted, and in particular bubbles need not vanish if one increases the amount of rational traders in the market. In our model, such traders have the incentive to ride the bubble, and sometimes to ride it even longer than some not fully rational, but cautious, investor. Indeed, we show that in a setting with ambiguity averse agents, bubbles may last longer when the amount of fully rational traders increases.

2. Migrants

The third Chapter - "Immigration policy and self-selecting migrants" - develops a simple theory of self-selection into migration and immigration policy formation. In this Chapter, we first show that any immigration restriction affects the skill composition of the migration flows, as it influences benefits and costs of migration. In fact, such restrictions allow only the richest foreigners to migrate, and these tend to be high skilled; however, at the same time, they have a stronger deterrence on those with lower gains from migration, who may be low skilled.

We then show that such composition effect may drive the policy outcome in the receiving country. Hence, the common assumption that immigration restrictions only have a size effect, i.e. they just select from a given pool of applicants, may be misleading. In fact, size and composition effects have typically the opposite directions. The size effect is by definition random, hence it hits a group of foreigners proportionally to their propensity to migrate. In contrast, the composition effect tends to be stronger on the least represented group. Moreover, there are situations in which the composition effect prevails, i.e. the least represented group of migrants is, in absolute terms, the most sensitive to a policy change.

As it turns out, understanding the migration decision becomes crucial for analyzing the political economy of immigration. In fact, self-selection determines how different potential migrants respond to immigration policy changes, and so ultimately what are the effects of immigration restrictions in the receiving country.

These selection effects have interesting implications also with respect to natives' attitudes over immigration policy and to the government's optimal policy design. For example, they may lead the majority of natives to support a more restrictive policy even though current immigrants are not harmful for them. Moreover, they may explain significant immigration restrictions even in a world with no distributional concerns and no political economy distortions, as these restrictions may indirectly screen the most desirable type of migrants.

3. Entrepreneurs

The fourth Chapter - "Financial development, entrepreneurship and job satisfaction" - explores the mechanisms of financial development and occupational choice. More specifically, it studies the relation between financial development and labor market rents, i.e. systematic differences in job satisfaction that the self-employed report relative to employees. The exercise is motivated by the common assumption that such rents are due to financial barriers to entry into self-employment.

Indeed, the paper shows that it is quite the opposite. First, rents are not a universal phenomenon. The self-employed report higher job satisfaction than employees only in *more* developed countries. Second, rents *increase* with financial development, and this effect is stronger in less developed countries. Third, greater financial development increases rents *not* by making entrepreneurs richer, but by allowing them to enjoy an higher freedom in their work.

These findings can be interpreted with a simple occupational choice model in which self-employment may be either a profitable choice or the last resort to avoid unemployment. Financial development increases labor market rents as it favors both job creation and a better matching between talents and occupational choices. As a result, in more developed countries, entrepreneurs display higher satisfaction as they have chosen to be self-employed due to their particular motivation, rather than by lack of better opportunities.

Finally, the fifth Chapter - "Is neoclassical economics still entrepreneurless?" - shows how simple economic theory can be used to shed some light on complex phenomena, like entrepreneurship, which have long been considered impossible to represent in formal models. Specifically, it reviews a number of recent contributions in modeling entrepreneurs within a neoclassical framework, analyzing how, and to what extent, these models capture fundamental ingredients suggested in the social science literature. It shows how these approaches are important in stressing the main elements of a complex picture, despite that the entrepreneurial function, broadly perceived, eludes analytical tractability.

Speculators

CHAPTER 2

Speculative Bubbles without Stupid Investors

Milo Bianchi and Philippe Jehiel

ABSTRACT. We model speculative behaviors in a market with finite horizon and complete information, where all traders, throughout the game, are aware to be in a bubble. We introduce partially sophisticated investors, who understand only the aggregate selling and buying strategies of other traders along the bubble, without a precise understanding of the timing of these decisions. In the bubble equilibrium, such traders remain invested for too long, since a series of high prices lead them to overestimate the duration of the bubble. Fully rational investors ride the bubble and exit just before the crash. We show that a bubble is more likely to arise and to be longer the greater is the amount of people who can potentially enter the speculative market relatively to those who have already invested. We also show that increasing the share of fully rational investors need not make the bubble disappear. First, a minimal fraction of fully rational investors is needed to make other investors aware that the crash is about to occur and it is time to sell. Second, if agents are ambiguity averse, bubbles may last longer when the amount of fully rational traders increases, since rational investors face less ambiguity and so they may be willing to take longer positions in the speculative market.

Keywords: Speculative bubbles, bounded rationality.

JEL codes: D84, G12, C72.

1. Introduction

A speculative bubble is characterized by an abnormal increase in the price of an asset, followed by a crash. In such situations, the *only* reason behind prices above fundamentals is the expectation that the selling price will be even higher in the near future (Stiglitz, 1990). These phenomena have been widely documented (see Garber, 1990; Kindleberger, 2005), but their foundations remain largely unclear. Standard economic theory rules out purely speculative behaviors. Rational agents should not invest just before the crash; anticipating this, no one should invest on the day before, and so on... nobody should invest in the first place (Tirole, 1982).

⁰ Philippe Jehiel is at the Paris School of Economics and at University College London. We thank numerous seminar participants and in particular Cedric Argenton, Abhijit Banerjee, Mike Burkart, Gabrielle Demange, Daniel Dorn, Tore Ellingsen, Botond Koszegi, John Moore, Marco Ottaviani, Peter Raupach, Marcus Salomonsson, Johan Walden, Jörgen Weibull, Muhamet Yildiz and Robert Östling for useful discussions. Milo Bianchi gratefully acknowledges financial support from the Knut and Alice Wallenberg Foundation.

In this paper, we show that speculative bubbles may be explained by considering that some investors have a *partial* understanding of the strategies of other investors, and in particular, they fail to understand how these strategies depend on the history of trades. As a result, some investors form erroneous beliefs about the date of the crash, and they are not able to sell just before it. This allows bubbles to arise even in a setting with complete information and in which all investors, throughout the game, are aware that the market is overvalued and bound to crash.

Our approach builds on three related observations. First, private information does not seem a central ingredient in the explanation of bubbles. Second, speculation occurs even if agents are aware to be in a bubble. Third, in such situations, each investor tries to time the market, and this may be hard if the date of the crash does not appear fundamentally different from some other dates. Before previewing our main results, we now consider these arguments more in details.

We are interested in the emergence of bubbles in contexts with complete information both for theoretical and empirical reasons. No-trade theorems tell that private information alone should not lead to trade, as no investor should expect positive gains from trade (Milgrom and Stokey, 1982). Hence, even if agents have private information and hence heterogeneous beliefs, say regarding the date of the crash, speculative bubbles should not be sustained. If one adds subjective (and thus erroneous) views about how this information is distributed, then bubbles may occur. However, this line of reasoning raises the issue of where subjective priors come from, and why they survive in equilibrium.¹

Empirical observations too suggest that asymmetric information may not be the reason for the occurrence of bubbles. First, while it can be argued that in modern times the scope for private information is reduced, speculative bubbles do not seem to have vanished. Many factors, including technology and regulation, now allow virtually all information about the economic environment to be *a priori* available.² Still, the 1990s have been qualified as "the most tumultuous decade ever" in terms of speculative behaviors (Kindleberger, 2005). Second, bubbles emerge even in experimental asset markets, where the amount of information provided to traders is controlled by design. In a famous study, Smith, Suchanek and Williams (1998) document speculative behaviors in a market with finite horizon and in which the structure of the game, and

¹ As discussed in Dekel, Fudenberg and Levine (2004), Nash equilibria with subjective priors are not appealing from a learning perspective. If agents can learn the strategies of other agents as a function of their private information, then why these agents cannot learn how private information is distributed? In other words, why don't they learn the strategy of nature?

² Of course, this does not mean that agents process information in an optimal way, as this paper emphasizes.

the value of future dividends, is common knowledge. In their interpretation, bubbles arise from strategic uncertainty: agents cannot perfectly assess how other investors act upon the common information, i.e. what are their strategies.³

In our model, we follow the insight that some agents may not be able to understand fully their opponents' strategies. Still, we require that they have some understanding of the environment, emphasizing in particular that all investors, throughout the game, understand to be in a bubble. This approach requires that investors are not totally naive, and in particular it rules out explanations based on purely adaptive expectations. If agents were to base their investments simply extrapolating from past trends, then they would expect the price to increase with no bounds, and they would never understand to be in a bubble. We believe that extrapolative traders may exist, but we doubt that they are sufficiently many to produce considerable departures from fundamental values.

This view finds support not only in experimental asset market, but also in the field. Shiller (1989), for example, reports that just before the US stock market crash of October 1987, 84% of institutional investors thought that the market was overpriced; 78% of them thought that this belief was shared by the rest of investors and, still, 93% of them were net buyers.⁴ Thus, according to this story, most agents were aware to be in a bubble, and investments were solely driven by speculative motives.⁵

In such situations, the crucial issue is therefore to predict *when* the crash will occur, i.e. the timing of other investors' strategies. This is precisely the dimension over which boundedly rational investors make mistakes: they fail to distinguish different histories of trades when assessing other investors' strategies. Hence, they expect the same average strategy, in a sense made precise below, to be observed irrespective of previous trades. Their inability to distinguish different histories when assessing future demand and supply should capture the observation that the date of the crash tends to appear quite similar to many other days. Investors and analysts typically find it hard to understand in what sense that precise day was so special, and fundamentally different from the day preceding it. Even the systematic analysis by Cutler, Poterba and Summers (1989) concludes that "many of the largest market movements in recent years have occurred on days when there were no major news events."

³ These conclusions have proven robust to many variations of the original experiment (see the review in Porter and Smith, 2003).

⁴ The percentages for individual investors are approximately the same.

⁵ Speculative behaviors seem a well documented phenomenon, spanning from the South Sea bubble in 1720 described in Temin and Voth (2004) to the Internet bubble in 2000 considered by Brunnermeier and Nagel (2004).

We formalize these ideas by assuming that partially sophisticated agents understand only the aggregate buy and sell rates along the duration of the speculative market, and they lack a precise perception of the timing of such decisions. One reason for this may be that looking at historical data in similar speculative markets, aggregate information was more accessible or salient (to these agents) than other more detailed statistics about say the daily buy and sell rates.⁶

We further assume that partially sophisticated investors adopt the simplest theory of trade volumes and price dynamics that is compatible with their knowledge.⁷ They expect constant buy and sell rates throughout the duration of the speculative market, independently from the history of trades.

In equilibrium, these constant rates match the aggregate intensities averaged over time, as resulting from the actual sell and buy decisions.⁸ In this sense, the understanding of partially sophisticated investors is correct. Still, these rates are only a partial representation of their opponents' investment strategies. As a result, these investors may fail to identify accurately the date of the crash, and to sell just before it. This is why the speculative market need not unravel and bubbles may arise.

The bubble equilibrium emerging in our model shares many aspects of a typical speculative phenomenon, as described for example by Kindleberger (2005). The bubble displays first a series of high prices, due to excess demand for speculative stocks. Boundedly rational investors interpret such high prices as a good surprise, from which they infer that the bubble will last further.⁹ Thus, they are induced to trade in the speculative market, and to remain invested even longer than they had originally planned. In this way, boundedly rational investors become euphoric, they revise their expectations and so overestimate the duration of the bubble. Fully rational investors can exploit such expectations, so they feed the bubble for a while and exit just before the endogenous crash. The massive sale by rational investors reveals that good times are over, boundedly rational investors realize it is time to sell (actually, it may be too late), and this indeed leads to the crash.¹⁰

⁶ See Higgins (1996) for an exposition of the idea of accessibility in psychology, and Kahneman (2003) for an analysis of its implications for decision making.

⁷ In line with this view, Wittgenstein (1922) writes: "The process of induction is the process of assuming the simplest law that can be made to harmonize with our experience."

⁸ This is a bit reminiscent of the use of linear models in Sargent (1993).

⁹ Shiller (2000), documents this strong regularity in investors' attitudes. As the price increases, more people display "bubble expectations", i.e. the belief that, despite the market being overvalued, it will still increase for a while before the crash.

¹⁰ One may now invoke Friedman's critique: as boundedly rational investors on average lose money, in equilibrium they must be wiped away. A number of recent studies show that this argument is not so clear cut (see Blume and Easley, 2006; Kogan, Ross, Wang and Westerfield, 2006). In our model, we can think of a new population of such investors entering the market in every bubble episode.

We show that such a bubble is more likely to exist, and it can last longer, the greater is the amount of people who can potentially enter the speculative market relatively to those already invested in it. This allows a greater excess demand, higher volumes of trade and higher increases in the price. These observations fit well with the empirical literature (e.g. Cochrane, 2002; Kindleberger, 2005) documenting that successful bubbles are characterized by short supply of speculative stocks, initially concentrated in a few hands, and they involve a large number of new, typically inexperienced investors.

The model also provides some novel insights on the relation between bubbles and rationality, showing that in general bubbles need not vanish as one increases the share of rational investors in the market.

As one should expect from the no-trade theorem, bubbles cannot arise if all investors are fully rational. However, we show that some rational investors are required in order to make boundedly rational agents realize that the bubble is bursting earlier than expected. When exiting the market, rational investors create a negative shock, which reveals that the crash is about to occur and it is time to sell. In a sense, we stress that rational investors are required to break the bubble at some point, and thus sustain such equilibrium.

Second, we show that, under risk-neutrality, bubbles tend to disappear as the fraction of rational investors increases, while this need not be the case if we consider ambiguity averse agents. In the first case, a larger share of rational investors makes the expectations of partially sophisticated traders less optimistic, since in equilibrium they observe more people leaving the market. Moreover, rational investors need to leave earlier when they are many, as they need to find enough investors who buy their stocks. Both effects tend to reduce the maximal duration of the bubble. Under ambiguity aversion, however, these may be counterbalanced by a third one. Rational agents are now more prone to invest than boundedly rational one, since they are affected less by strategic uncertainty. As the share of rational investor increases, more people are entering the speculative market. Boundedly rational agents may then become more optimistic, and this may sustain longer bubbles.

The rest of the paper proceeds as follows. We now relate our approach to the existing literature. The basic model is presented in Section 2; the results on the existence and the characterization of bubble equilibria are derived in Section 3. In Section 4, we discuss bubbles and rationality of the market, and in Section 5, we conclude with some policy implications.

1.1. Related Literature. A number of other works have proposed alternative explanations of bubbles. The literature on rational bubbles describes departures from fundamentals in markets with fully rational investors, who have the correct model

in mind and make the best use of available information (see Blanchard and Watson, 1982). In such setting, however, the bubble component must be expected to grow over time, so each agent has to allow that with some (small) probability trades will occur at arbitrarily large prices. If one considers an economy with finite wealth, or with wealth that cannot grow as fast as the price of the asset, the maximal bubble has a finite upper bound, and the argument unravels. As shown in Tirole (1982), in a market with finite horizon, purely speculative trade relies on inconsistent plans and as such it is ruled out by rational expectations. The result is fundamental, as it shows that *any* theory of bubbles represents a departure from a fully rational world.¹¹

Several explanations of bubbles are based on investors' private information. However, private information alone is not enough. Allen, Morris and Postlewaite (1993) show that bubble may arise in a setting with a finite number of trading opportunities and rational agents only if one introduces ex-ante inefficiency, private information, short sale constraints, and lack of common knowledge of agents' trades. In such setting, bubble may be sustained even if everyone realizes that the stock is overpriced, since other investors' trades cannot be observed and so their beliefs are never common knowledge. A similar point is developed by Abreu and Brunnermeier (2003), where the limit to arbitrage stems from a coordination failure among rational arbitrageurs, who desire to stay invested as long as profitable, while facing different opinions on when the bubble will burst. Here the bubble is fed by some (unmodeled) "behavioral agents" and it can be sustained since its existence never becomes common knowledge.

In our model too, investors are driven by purely speculative motives. However, for the reasons explained in the Introduction, we want to abstract from private information and keep the existence of the bubble and other investors' trades as common knowledge. Moreover, since asymmetric information can induce speculative trade only if one adds some extra ingredients, there is need for more structured theories of where these extra ingredients come from (see Bianchi, 2007, for further elaboration on this point.)

Our paper follows the line of incorporating bounded rationality in a setting where no bubble could arise if all investors were fully rational. Our emphasis on cognitive heterogeneity builds on the insights from the wide literature on the limits to information processing,¹² which is now being applied also to the study of financial markets (see

¹¹ A popular explanation of bubbles is based on herd behavior. Indeed, a number of studies show that fully rational agents may decide to disregard their own information and follow the herd (see for example Banerjee, 1992). However, these studies typically consider investment decision made at a fixed price, which is not the case in financial market. Once the price is allowed to vary in response to agents' decisions, informational cascades become impossible, and the price correctly aggregates the common information (see Avery and Zemsky, 1998, and Cipriani and Guarino, 2005).

¹² These constraints have been widely documented by psychologists (see Rabin, 1998 and Kahneman, 2003 for economic applications) and they are part of the economics literature at least since

Hirshleifer, 2001).¹³ Along the lines of bounded rationality, De Long, Shleifer, Summers and Waldmann (1990*a*) introduce noise traders, who act like crazy but may make great profits in equilibrium, since rational arbitrageurs are limited by finite horizon and risk aversion. An essential difference with our approach is that expectations and behaviors of noise traders are exogenously given, so their trading decisions are mechanical. As already emphasized, we depart also from the literature on positive feedback traders (e.g. Cutler, Poterba and Summers, 1990; De Long, Shleifer, Summers and Waldmann, 1990*b*), where some agents form their expectations about future price dynamics only by extrapolating from previous trends. For the reasons exposed in the Introduction, we want to explain bubbles despite all investors understand, and it is actually common knowledge, that a crash will occur.

Lastly, bubbles can be explained if agents have heterogeneous beliefs. In such a word, investors know that there may be others with an higher valuation for the asset, and so they may be willing to speculate, i.e. to pay the asset more than what they would pay if they were forced to hold it forever (Harrison and Kreps, 1978). Scheinkman and Xiong (2003) consider a classical source of multiple priors: overconfidence. Overconfident traders overestimate the precision of their information, i.e. they believe to have a better understanding than the rest of investors. As such, these agents are willing to weight their private information more than the public information revealed in market prices, and so divergence in opinions about fundamentals may be persistent. This model is silent on how the difference in opinions among traders increases over time, so that one can predict increasingly high prices, and how overconfidence may disappear and so a crash may occur. More generally, we also model speculative trade driven by divergence in expectations, but the difference in opinions stems from the degree of sophistication, i.e. how precisely these agents understand others' strategies, rather than subjective priors. In this sense, our work is complementary to these approaches. Moreover, it avoids the issue, already discussed in the Introduction, of modeling equilibria with subjective priors.

Herbert Simon (1955). Some of these themes have been recently explored also in strategic settings (see Rubinstein, 1998 and the references therein and Jehiel, 1995; Jehiel, 2005; Jehiel and Samet, 2006).

¹³ Hirshleifer (2001) states: "Since time and cognitive resources are limited, we cannot analyze the data the environment provides us with optimally. Instead, natural selection has designed minds that implement rules of thumb ('algorithms', 'heuristics', or 'mental modules') selectively to a subset of cues [...]. Such heuristics are effective when applied to appropriate problems. But their inevitable biases can become flagrant when used outside their ideal domain of applicability. [...] People share similar heuristics, those that worked well in our evolutionary past. So on the whole we should be subject to similar biases."

2. The model

Our economy is populated by a unit mass of risk neutral individuals. Initially, a fraction K of individuals is endowed with cash, each of them owning $w > 0$, and a fraction $(1 - K)$ has stocks, each individual owning one stock. The value of cash is constant over time. Stocks pay no dividend, and their return is given only by changes in the price p_t . That is, we normalize their fundamental value to zero and concentrate on purely speculative assets. For simplicity, we assume that each agent can hold at most one stock and each stock is indivisible.¹⁴ Hence, player i 's profits are $\pi_i = 0$ if he always holds cash, and $\pi_i = (p_s - p_b)$ if he buys a stock at p_b and sells it at p_s .

At the beginning of each period $t = 1, 2, \dots$, individuals can simultaneously submit an order to buy or sell a stock. A price p_t is then announced and orders are cleared. Borrowing stocks or cash is not allowed, so the investment option for individual i in each period t is simply $\{buy, stay out\}$ if i holds cash at t , and $\{sell, stay in\}$ if i holds a stock at t .¹⁵ Let B_t denote the amount of people willing to buy at t , S_t those willing to sell, and $N_t = B_t - S_t$ the net demand. The trading price at each t is determined according to the rule:

$$(2.1) \quad p_t = f(N_t, p_{t-1}),$$

where $f : [K - 1, K] \times [0, w] \rightarrow [0, w]$ is increasing in N_t , with $f(0, p_{t-1}) = p_{t-1}$. In case that B_t differs from S_t , buyers/sellers in excess are chosen randomly, so the volume of trade in t is $V_t = \min\{B_t, S_t\}$.

The function $f(\cdot)$ together with the rationing scheme defined above are known to all agents, and they can be viewed as a reduced form model of price formation and trade allocation. Our specification assumes that the price in any period t exceeds the one at $t - 1$ if and only if there is excess demand at t , where, given borrowing constraints, the trading price can never exceed w . Equating demand and supply in period t is achieved by an anonymous rationing scheme. Whenever trade occurs in this market, we say that there is a bubble.¹⁶

¹⁴ The substance of our analysis would not change if stocks were perfectly divisible and everyone could spend his entire wealth in stocks.

¹⁵ We rule out more sophisticated investment strategies, where for example orders are made conditional on volumes or on other people's orders. Notice however that our analysis does not rely on the exclusion of short selling. As it will become clear, this would tighten the conditions defining our bubble equilibrium, without changing its structure.

¹⁶ The functioning of our market may remind the one in Kyle (1985), where agents submit a (market) order to buy or sell and the market maker increases the price over time according to the excess demand. For simplicity, we consider a market maker who does not take any position, but simply clears the market via rationing. Our analysis would be unaffected if the market maker could take any position. Rationing may capture the common observation that growth stocks are scarce during booms (as documented e.g. in Cochrane, 2002).

In each period, people in the speculative market may be hit by a shock and so need to sell immediately. Shocks are i.i.d. among agents; each individual shock occurs with probability $z > 0$, and it is permanent: an agent who exits the market stays out forever.¹⁷ Given that $(1 - K)$ is the amount of people in the speculative market,¹⁸ we can define $X \equiv z(1 - K)$ as the exogenous exit from speculation in each period, where we assume that $X < K$. We define the amount of exit from the speculative market in period t as E_t , which includes those who sell their stocks and those with cash who decide not to buy at t .¹⁹ Moreover, as it turns out to be the case in equilibrium, those agents who decide to exit the speculative market, never re-enter. Accordingly, we can define K_t as the amount of people who have cash and who have not yet exited the speculative market in period t : these are the only agents who can possibly contribute to the demand in this period.²⁰ Moreover, we write

$$(2.2) \quad B_t = \beta_t K_t, \text{ and } S_t = \sigma_t(1 - K),$$

where β_t and σ_t are the shares of potential buyers and sellers who want to buy or sell in period t , respectively. Accordingly, we have

$$K_{t+1} = K_t - E_t,$$

where

$$(2.3) \quad E_t = V_t + (1 - \beta_t)K_t,$$

and thus

$$(2.4) \quad K_{t+1} = \beta_t K_t - V_t.$$

Observe that if in period t the price drops due to excess supply, then $V_t = \min\{B_t, S_t\} = B_t$, and equation (2.4) gives $K_{t+1} = 0$. From then on there is no one who can enter the speculative market, which then closes. In this case, provided that $t > 1$, we say that a crash has occurred.

The same information is available to everybody. Once realized, trade orders B_t and S_t become common knowledge, hence in particular the mass of cash K_t available in

¹⁷ The fact that a small fraction of investors are hit by a permanent shock avoids that the speculative market increases indefinitely. Such shock can be thought as consequence of unforeseen liquidity needs, and it has not to be confused with noise trade.

¹⁸ This mass never changes over time due to the equalization of demand and supply.

¹⁹ The agents with cash who decide not to buy at t should not be confused with the agents who submit an order to buy but do not get any stock at t due to rationing. Only the former are considered to exit the speculative market.

²⁰ A crucial ingredient of our analysis is that K_t cannot increase over time. This is because the population of potential investors and their capital endowment is fixed and finite. As the price increases, less people can afford the stock, hence K_t decreases. Moreover, in equilibrium, no one is willing to invest when the price decreases. Hence, this structure gives our market a finite horizon.

period t is observed by all agents. Moreover, all agents rightly understand that traders who exit the speculative market never re-enter. What varies across individuals is their ability to process information, which determines their expectations about future prices. While some players are fully rational, as in standard economic models, we postulate that some other players have a partial rather than total understanding of the mechanics of other players' decisions. In particular, they fail to distinguish accurately one history of trades from another when forming their expectation.

For concreteness, we consider a setting with two extreme cognitive types: R and I . R -types are standard rational players, who understand perfectly well the dynamics of trade volumes and the associated prices. I -types instead are less sophisticated: they understand only the average sell and buy rates along the entire duration of the speculative market, as opposed to the exact patterns of these rates. As discussed in the Introduction, a rationale for I -types' limitation may be that the information about aggregate buy and sell rates, derived for example by historical data in similar speculative markets, is more accessible or salient than more detailed statistics, reporting say the daily buy and sell rates.²¹

We further assume that I -agents adopt the simplest theory of trade volume and price dynamics that is compatible with their knowledge. Specifically, we assume that I -agents expect the same buy and sell rates to occur throughout the duration of the speculative market. Irrespective of the history of trades, they expect a constant fraction of people with cash to buy and a constant fraction of people with stocks to sell.²² In equilibrium, these rates are required to correctly represent the average buy and sell rates over the entire duration of the speculative market.^{23,24}

We denote by r the proportion of R -players, while $(1 - r)$ is the proportion of I -players. At the beginning of $t = 1$, cash and stocks are randomly distributed across types. Hence, a fraction $r(1 - K)$ of R -types is given stocks and a fraction rK is

²¹ In this vein, one should view R -types either as having access to better and more intelligible statistics or as being more sophisticated in their information treatment.

²² Yet, as will become clear, the simplest theory need not correspond to the truth, and accordingly I -agents are viewed as bounded rational.

²³ The correctness of these statistics is consistent with our interpretation in terms of historical data, which suggests that to the extent that previous markets are similar, they should correspond approximately to the same average patterns.

²⁴ This idea follows very closely the spirit of Jehiel (2005) who assumes, in the context of extensive form games, that each player i is characterized by a partition of the set of nodes where other players move. This partition is a collection of subsets called analogy classes, and it determines how player i forms expectations about other players' strategies along the game. Player i assesses only the average behavior of his opponents within each analogy class, and expects this same average behavior to be played at each node of the analogy class. A similar idea is developed in static games of incomplete information by Eyster and Rabin (2005).

given cash. Similarly, I -types with stocks and cash are respectively $(1-r)(1-K)$ and $(1-r)K$.

We can now describe formally an equilibrium in our setup with agents of cognitive types R and I . We start by defining the expectations: for each type $\theta \in \{R, I\}$, we let $B_s^{\theta,t}$ be the expectation of an investor of type θ in period t about the demand in period s . Similarly, $S_s^{\theta,t}$ denotes the same expectation about the supply.

R -agents are fully rational and, as such, their equilibrium expectations are correct. In particular, if the true buy and sell decisions arising in equilibrium in period s are given by B_s and S_s , R -agents' expectations must satisfy:

$$(2.5) \quad B_s^{R,t} = B_s \text{ and } S_s^{R,t} = S_s \text{ for every } t \leq s.$$

We turn next to I -agents' expectations. Denoting with $T+1$ the last date in which the speculative market operates, which is determined endogenously in equilibrium, we can define the average buy rate $\bar{\beta}$ and the average sell rate $\bar{\sigma}$ for the sequence of buy and sell decisions arising in equilibrium from period $t = 1$ to $t = T+1$. That is, average rates are defined by

$$(2.6) \quad \bar{\beta}(T) = \frac{1}{T+1} \sum_{t=1}^{T+1} \beta_t, \text{ and } \bar{\sigma}(T) = \frac{1}{T+1} \sum_{t=1}^{T+1} \sigma_t.$$

Given these average rates, and definition (2.2), I -agents' expectations are defined by

$$(2.7) \quad B_s^{I,t} = \bar{\beta}(T) \cdot K_s^{I,t} \text{ and } S_s^{I,t} = \bar{\sigma}(T) \cdot (1-K) \equiv S^I(T) \text{ for every } t \leq s,$$

where $K_s^{I,t}$ denotes I -agents' expectations about the amount of cash available in period s , given the cash observed in t . That is, $K_t^{I,t} = K_t$ and, for $s > t$, we have

$$(2.8) \quad K_s^{I,t} = K_t - \sum_{t'=t}^{t'=s-1} E_{t'}^{I,t},$$

where, according to the definition (2.3), we write

$$E_{t'}^{I,t} = \min(B_{t'}^{I,t}, S_{t'}^{I,t}) + (1 - \bar{\beta})K_{t'}^{I,t}.$$

Given the formulation in equations (2.5) and (2.7), agents of all types form expectations about the price dynamics according to the function $f(\cdot)$ defined in equation (2.1), and the anonymous rationing scheme described above. With these expectations, each agent determines an optimal investment strategy, which specifies for every possible history of B_t and S_t whether this agent should sell his stock, given that he has one, or he should buy a stock, given that he has cash. An investment strategy profile specifies an investment strategy for every agent in the economy. We can now define an equilibrium in our setting.

DEFINITION 1. (**Equilibrium**): *An investment strategy profile is an equilibrium if, all along the equilibrium path, each agent's investment strategy is a best response to his expectations, as defined in equations (2.5) and (2.7).*

Remarks. Our definition of equilibrium is in the spirit of the rational expectation equilibrium in which, due to the dynamic nature of the interaction, beliefs and investment strategies must be optimally adjusted at every point in time. Yet, the beliefs of I -agents are not necessarily correct, as there is no reason in general that β_t and σ_t be constant over time (in fact, as we shall see, they are not in a bubble equilibrium). Note also that our definition only considers the incentives of agents on the equilibrium path, and not the adjustment of beliefs and strategies after a positive mass of agents have made non-equilibrium decisions.²⁵

3. Analysis

We focus on symmetric equilibria in pure strategies, where all investors of a given type and with a given endowment in period t follow the same pure strategy in such period. Observe that the existence of an equilibrium is not an issue as there is always the non-bubble equilibrium in which every agent exits the speculative market at the very first period.²⁶ Our interest lies in showing the possibility of bubble equilibria and characterizing the conditions for such equilibria to exist.

The problem faced by an individual of a given type is the same irrespective of whether he has cash or stock. That is, for any player $i \in \theta$ with cash and any player $j \in \theta$ with stock, i prefers to buy if and only if j prefers to stay in, and i prefers to stay out if and only if j prefers to sell. It follows that speculative trade can only occur between players of different types. Intuitively, trade occurs either among people with different needs, as described by the liquidity sellers X , or among those with different expectations, i.e. different types.

If player $i \in \theta$ submits an order to buy at t , there must be a period $s > t$ such that he expects a gain from selling at s . That is, we need

$$(3.1) \quad q_s^{\theta,t} \cdot p_s^{\theta,t} \geq p_t^{\theta,t},$$

²⁵ We could easily amend the solution concept to cover off the path optimizations and expectations. Yet, this would make the notation heavier (in particular, the state variable parameterizing the decisions should no longer be the calendar time t but the entire history of buy/sell decisions) without adding much economic insight. Moreover, since each individual agent has a negligible weight (there is a continuum of agents), our notion of equilibrium is in the spirit of the Nash equilibrium where no single agent can on his own move the system away from the equilibrium path.

²⁶ In this equilibrium, I -agents' expectations are correct, and their decisions to exit immediately is thus rational.

where $q_s^{\theta,t} = \min\{1; B_s^{\theta,t}/S_s^{\theta,t}\}$ is the expected probability of being able to sell in period s , as perceived in period t , and $p_s^{\theta,t}$ is the corresponding selling price. Otherwise, buying at t would not be profitable.²⁷

Our premise in describing the dynamics of the market was that if an agent decides to exit speculation, he never tries to re-enter. This premise is consistent since the set of potential buyers does not increase over time and so I -players never wish to re-enter. In fact, if they exit at some period t , than they must expect a crash to occur no later than at $t + 1$. At $t + 1$, even if they do not observe such crash, they expect that it will occur no later than at $t + 2$, as the amount of available cash has not increased, hence they stay out. Given that, R -players too prefer not to re-enter. In fact, if they sell at t , it must be that the price is lower at $t + 1$, i.e. $B_{t+1} < S_{t+1}$. Suppose they re-entered at $t + 1$, then $B_{t+1} < S_{t+1}$ means that I -investors are selling. However, given that they do not re-enter, no one will be willing to buy from $t + 2$ on, so it is not optimal to re-enter. We can express the result more precisely with the following Lemma.

LEMMA 1. *Given the expectations and price dynamics considered in Section 2, an agent who sells at t stays out of the speculative market from then on.*

PROOF. We prove the result for I -agents first. Suppose $p_{t+1}^{I,t} \geq p_t^{I,t}$. Then $B_{t+1}^{I,t} \geq S^I(T)$ and $q_{t+1}^{I,t} = 1$, hence $q_{t+1}^{I,t} \cdot p_{t+1}^{I,t} \geq p_t^{I,t}$ and they will not sell at t . If I -agents sell at t , then it must be that $p_{t+1}^{I,t} < p_t^{I,t}$, i.e. $B_{t+1}^{I,t} = \bar{\beta}K_{t+1}^{I,t} = \bar{\beta}(\bar{\beta}K_t - S^I(T)) < S^I(T)$. Since K_t cannot increase with t , then it must be that $B_s^{I,w} < S^I(T)$ for every $w, s > t$, hence I -agents will never buy again.

Now consider R -agents. As above, if they sell at t , then it must be that $p_{t+1} < p_t$ (we drop the expectations as these are fully rational), i.e. $B_{t+1} < S_{t+1}$. Hence, $K_{t+s} = 0$ and $B_{t+s} = 0$ for every $s > 1$. Since it will be impossible for them to sell after $t + 1$, they will not buy. \square

The Lemma implies that the price of the asset does not recover after having dropped. This allows a very simple characterization of the optimal investment strategies, as expressed in the following Proposition.

PROPOSITION 1. *A player $i \in \theta$ prefers to buy/stay in the market at t if and only if*

$$(3.2) \quad B_{t+1}^{\theta,t} \geq S_{t+1}^{\theta,t}.$$

PROOF. Given that each agent can hold at most one stock, he would like to buy when the price is at the minimum and sell when it is at the peak. Hence, in principle, he would wait if he thought the price would drop for a while before recovering, but this

²⁷ Without loss of generality, we assume that, when indifferent, a player decides to buy.

price dynamic is ruled out by Lemma 1. Hence, this Lemma implies that condition (3.1) is also sufficient for an agent i to buy. Moreover, notice that condition (3.1) is equivalent to

$$(3.3) \quad p_{t+1}^{\theta,t} \geq p_t^{\theta,t}.$$

In fact, given the price dynamics in equation (2.1), $p_{t+1}^{\theta,t} \geq p_t^{\theta,t}$ is equivalent to $q_{t+1}^{\theta,t} = 1$, and $p_{t+1}^{\theta,t} < p_t^{\theta,t}$ is equivalent to $q_s^{\theta,t} = 0$ for every $s \geq t + 1$. Writing equation (3.3) according to equation (2.1) gives the result. \square

Hence, the model endogenously generates short-termism: agents are concerned only with short-term speculative gains rather than with fundamentals. Everybody knows that the fundamental value of the asset is zero, any trade is driven by the resale option, and everybody understand that this resale option become worthless as soon as the market displays excess supply. Moreover, this proposition allows writing the optimal investment strategy for I -agents simply as follows.

COROLLARY 1. *A player $i \in I$ prefers to buy/stay in the market at t if and only if*

$$(3.4) \quad K_t \geq W(T),$$

where

$$(3.5) \quad W(T) \equiv \frac{(1 - K)\bar{\sigma}(T)(1 + \bar{\beta}(T))}{(\bar{\beta}(T))^2}.$$

PROOF. According to equation (2.7), condition (3.2) can be written as $\bar{\beta}K_{t+1}^{I,t} \geq \bar{\sigma}(1 - K)$, where by (2.8) we have $K_{t+1}^{I,t} = \bar{\beta}K_t - \bar{\sigma}(1 - K)$. Rearranging these equations gives the result. \square

Hence, the optimal investment strategy for I -agents is a function of the available cash in the economy K_t and of their expectations about the future buy and sell rates. Notice that these expectations, as defined in equation (3.5), depend on the equilibrium T , and not on the time t .

3.1. The bubble equilibrium. We can now show that, under conditions to be characterized in the next Subsection, there exists an equilibrium, which we call the bubble equilibrium, of the following form. In each period $t \leq T - 1$, everyone tries to enter the speculative market and no one wants to sell; in period T , I -investors buy and R -investors sell; at $T + 1$, everyone tries to sell but no one is willing to buy. The crash then occurs and the market closes.

Along this equilibrium, the buy and sell rates are

$$(3.6) \quad \beta_t = \begin{cases} 1 & \text{for } t \leq T-1, \\ 1-r & \text{for } t = T, \\ 0 & \text{for } t = T+1, \end{cases}$$

and

$$(3.7) \quad \sigma_t = \begin{cases} z & \text{for } t \leq T-1, \\ z+r(1-z) & \text{for } t = T, \\ 1 & \text{for } t = T+1. \end{cases}$$

According to equation (2.6), this equilibrium induces I -agents to expect the following buy and sell rates:

$$(3.8) \quad \bar{\beta}(T) = \frac{T-r}{T+1},$$

and

$$(3.9) \quad \bar{\sigma}(T) = \frac{Tz+r(1-z)+1}{T+1}.$$

This allows us to define $B_s^{\theta,t}$ and $S_s^{\theta,t}$ according to equations (2.5), (2.6), (2.7), (2.8). Besides, given the above specifications, the only variable remaining to endogenize is the duration T of the bubble.

As already noticed, Lemma 1 implies that the optimal investment strategy for each player i satisfies a threshold property. If i wants to sell/stay out at t , then he wants to sell/stay out for every $s > t$. Likewise, if i wants to buy/stay in the market at t , this reveals that he wanted to buy/stay in the market at each $s < t$. Hence, the equilibrium T is defined simply by three conditions. First, each player $i \in R$ has to prefer to buy at $T-1$, so we must have $p_T > p_{T-1}$, i.e. there cannot be excess supply at T . Second, each player $i \in I$ has to prefer to buy at T . Third, each player $i \in I$ has to prefer to sell at $T+1$.²⁸ Using equations (3.2) and (3.4), we can express the three conditions, respectively, in the next Lemma.

LEMMA 2. *The above investment strategy profile is an equilibrium if and only if there exists a T is such that*

$$(3.10) \quad B_T \geq S_T,$$

$$(3.11) \quad K_T \geq W(T),$$

and

$$(3.12) \quad K_{T+1} < W(T).$$

²⁸ Notice that the last two conditions imply that each player $i \in R$ prefers to sell at T .

The above conditions can be expressed in terms of our exogenous parameters using equations (2.4), (3.6) and (3.7). We have that

$$K_T = K - (T - 1)X,$$

and

$$K_{T+1} = (1 - r)(K - TX) - r(1 - K).$$

Condition (3.10) defines an upper bound on the duration of the bubble equilibrium. In fact, with simple algebra, it can be written as

$$(3.13) \quad T \leq \frac{K - r}{X(1 - r)} \equiv T_1.$$

Conditions (3.11) and (3.12) are instead implicit conditions on the bubble equilibrium. To see their structure, we first have to characterize the shape of $W(T)$, as expressed in the next Lemma.²⁹

LEMMA 3. *The function $W(T)$ is decreasing and convex in T . Moreover, $W(T) \rightarrow 2X$ as $T \rightarrow \infty$.*

Since K_T and K_{T+1} are linearly decreasing in T and they both tend to minus infinity as T goes to infinity, each of them can intersect $W(T)$ at most twice. Define T_2 as the largest root solving $K_T = W(T)$ and T_3 as the largest root solving $K_{T+1} = W(T)$.³⁰ Conditions (3.11) and (3.12) then require, respectively, that $T \leq T_2$ and $T > T_3$.

To summarize, let T_{\max} be

$$T_{\max} \equiv \min\{T_1, T_2\}.$$

Conditions (3.10) and (3.11) jointly require $T \leq T_{\max}$, which defines the maximum T before which R -players are able to profit from the bubble, selling at its peak. Selling after T_{\max} would imply selling together (or after) I -players, hence during the crash. However, not all $T \leq T_{\max}$ can be sustained as equilibrium. Condition (3.12) is satisfied for $T > T_3$: if R -players sell too early, I -players would not exit and the crash would not occur at $T + 1$, hence selling at T would not be optimal. Together with Lemma 2, this implies the following Proposition.

PROPOSITION 2. *Any T satisfying conditions (3.10), (3.11) and (3.12) can be sustained as a bubble equilibrium. In particular, conditions (3.10) and (3.11) define a maximal T and condition (3.12) a minimal T that can be sustained as a bubble equilibrium.*

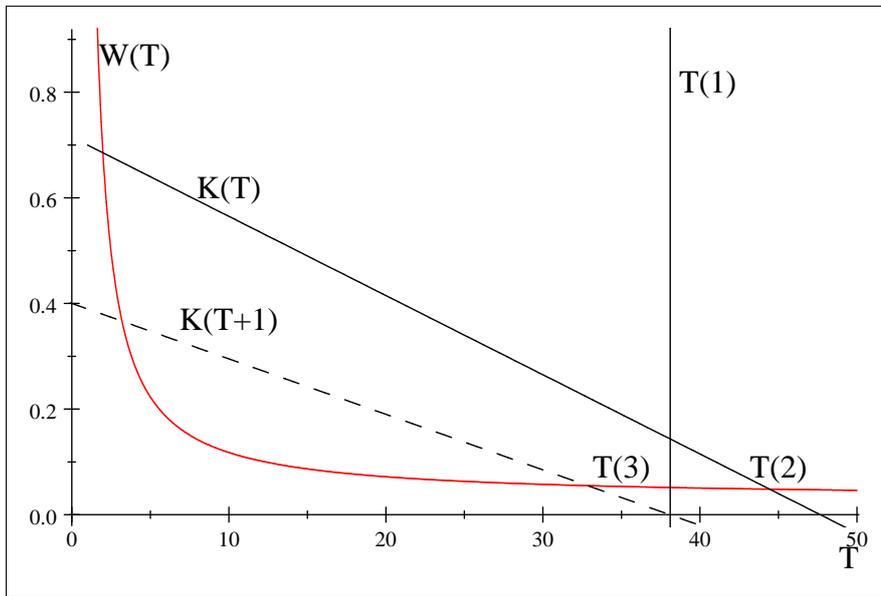
²⁹ Being only algebra, the proof is omitted.

³⁰ In the next Subsection, we give conditions for existence and uniqueness of such T_2 and T_3 . For now, we just assume that they exist.

The structure of this equilibrium is consistent with a number of empirical observations. First, our market becomes increasingly euphoric (as documented e.g. in Kindleberger, 2005). This is because I -investors initially observe more people than expected investing in the speculative market, and so bigger price increases. This leads them to revise upwards their expectations about the future demand and the date of the crash, and then eventually to remain invested for too long. Second, major investors do ride the bubble and are able to earn great profits by exiting at the right time. Moreover, their strategies are driven by a better understanding of the market rather than by the superior access to information. This is documented in a number of studies, like Temin and Voth (2004) and Brunnermeier and Nagel (2004).

Before characterizing the conditions for the existence of some T which can be sustained as a bubble equilibrium, it is useful to highlight the structure of conditions (3.10), (3.11) and (3.12) through a numerical example.

EXAMPLE 1. *Suppose that $z = 0.05$, $r = 0.3$ and $K = 0.7$. The functions $W(T)$ and K_T are plotted below in solid lines, while K_{T+1} is the dashed line. Hence, T_3 is given by the intersection of the dashed line with $W(T)$, while T_2 by the intersection of the solid line with $W(T)$. The vertical line is T_1 (which binds in this example). Substituting our values in equations (3.10), (3.11) and (3.12) we find that, up to integer approximations, $T_2 = 44$, $T_3 = 33$ and $T_1 = 38$.*



In fact, condition (3.11) requires $T \in [2, 44]$, condition (3.12) requires $T < 3$ or $T > 33$ and condition (3.10) $T \leq 38$. Hence, in this example, any $T \in [2, 3) \cup (33, 38]$ can be a bubble equilibrium.

3.2. Existence of a bubble equilibrium. In order to investigate the conditions for the existence of some T defining our bubble equilibrium, we now analyze equations (3.10), (3.11) and (3.12) in terms of our exogenous parameters K , z and r . We start with the following Lemma.

LEMMA 4. $T_3 < T_{\max}$ for every K , z and r .

PROOF. First, since $K_{T+1} < K_T$ for every T , we have that $T_3 < T_2$. Second, by definition, $K_{T_1+1} = 0$ and so condition (3.12) holds for sure at $T_1 + 1$. Hence, we also have that $T_3 < T_1$. \square

This observation, together with Lemma 3, implies that the bubble equilibrium exists if and only if $W(T)$ and K_T intersect at least once, i.e. if there exists a $T_2 \geq 1$ such that $K_{T_2} = W(T_2)$. In fact, when this is the case, T_{\max} can always be sustained as equilibrium. Hence, a *sufficient* condition for the existence of a bubble equilibrium is that $W(T)$ and K_T intersect once, and only once, as it is considered in the following Proposition.

PROPOSITION 3. *If $K \geq W(1)$ then a bubble equilibrium exists. The maximal duration of an equilibrium bubble is given by T_{\max} .*

It follows from this Proposition that a bubble equilibrium is more likely to exist when K is large relatively to $W(1)$. Given that $W(1)$ increases in z and r , the condition requires K to be large relatively to z and r , as we express in the following Corollary.³¹

COROLLARY 2. *There exist a K^* , r^* and z^* such that if $K \geq K^*$, or equivalently $r \leq r^*$ or $z \leq z^*$, then a bubble equilibrium exists.*

The conditions in the last Corollary, as well as the comparative statics emerging in the next Subsection, show that bubbles are supported by large K and small z and r . These relations are consistent with empirical evidence. A large K is in line with the observation that speculative stocks tend to be initially in short supply, and that bubbles are sustained by the large involvement of new investors. This allows greater excess demand, greater increases in the price and higher volumes of trade (see Cochrane, 2002; Kindleberger, 2005). A small probability of shock z implies that the fraction of potential investors decreases slowly, which is consistent with the fact that bubbles tend to display slow booms and sudden crashes (see Veldkamp, 2005). Finally, a small r emphasizes that these episodes tend to attract a large number of inexperienced investors (see Shleifer, 2000, Kindleberger, 2005). However, as we discuss in Section 4, the relation between bubbles and rationality is not so clear-cut: an increase in r need not make bubble vanish.

³¹ Again, the proof is just algebra and it is omitted.

3.3. The maximal equilibrium bubble. As noticed in Proposition 2, there need not be only one T satisfying conditions (3.10), (3.11) and (3.12). In general, rational investors need to solve a coordination problem in order to select a T at which they all sell. One natural candidate for such selection is T_{\max} , which is the largest T that can be sustained in equilibrium, and the one maximizing R -players' profits. It is then of interest to characterize such T_{\max} , and to investigate how it varies with our exogenous parameters.

First, irrespective of whether or not T_1 exceeds T_2 , the comparative statics are clear: both T_1 and T_2 increase in K and decrease in r and z . In fact, $W(T)$ decreases in K and increases in z and r , while K_T increases in K and decreases in z . Hence, when z and r are low and K is high, T can be large. This implies that the mistakes in I -players' expectations are sufficiently large, which in turn allows to sustain longer bubbles. We summarize with the following Proposition.³²

PROPOSITION 4. *The maximal equilibrium bubble T_{\max} increases in K and decreases with z and r .*

Despite the comparative statics are unambiguous, it may still be useful to characterize the conditions under which T_1 rather than T_2 defines T_{\max} . In particular, for the analysis in the next Section, we are interested in the effect of r . Intuitively, one should expect T_2 to be the binding constraint when r is low, and T_1 to bind when r is high. When rational investors are many, they need to exit the market at a point in which the cash in the economy is still very high, that is a point in which I -investors think that the date of the crash is far away. The following Proposition formalizes this intuition.

PROPOSITION 5. *$T_1 < T_2$ if and only if $r > \bar{r}$. The threshold \bar{r} is implicitly defined by $\bar{r} = G(\bar{r})$, where*

$$(3.14) \quad G(r) \equiv \frac{W(T_1) - X}{W(T_1) - X + 1 - K}.$$

PROOF. By definition of T_2 , $T_1 < T_2$ if and only if $W(T_1) < K_{T_1}$. By definition of T_1 , $K_{T_1}(1 - r) = S_T$ and $S_T = X(1 - r) + r(1 - K)$. Rearranging, we have equation (3.14). Now notice that $G(r)$ is increasing in $W(T_1)$, and $W(T_1)$ is increasing in r . Moreover, $G(0) > 0$ and $G(1) < 1$. Hence $r > G(r)$ holds for $r > \bar{r}$, where \bar{r} is uniquely defined by $\bar{r} = G(\bar{r})$. \square

It follows from the last Proposition that if r is smaller than the minimum of $G(r)$, then T_2 binds, so we have the next Corollary.

COROLLARY 3. *If $r \leq \frac{z}{z+1}$, then $T_{\max} = T_2$.*

³² These results can be formally obtained by implicitly differentiating the function $W(T) - K_T$.

PROOF. As shown in Lemma 3, $W(T) < 2X$ for every T . Since $G(r)$ increases in $W(T)$, we have that $G(r) > \frac{X}{X+(1-K)}$. Recalling the definition $X = z(1-K)$, we write $G(r) > \frac{z}{z+1}$. Hence if $r \leq \frac{z}{z+1}$, then $r < G(r)$, i.e. $T_{\max} = T_2$ \square

Finally, the Proposition allows to characterize T_{\max} when the probability of liquidity shocks is very small, which is an important special case we will reconsider in the next Section. It turns out that both T_1 and T_2 tend to infinity as z tends to zero, but T_2 exceeds T_1 , as expressed in the next Corollary.

COROLLARY 4. *If $z \rightarrow 0$, then $T_{\max} = T_1$.*

PROOF. If $z \rightarrow 0$, then $X \rightarrow 0$, $T_1 \rightarrow \infty$ and $W(T_1) \rightarrow 0$. Hence, $G(r) \rightarrow 0$, and condition (3.14) always hold. As a result, $T_{\max} = T_1$. \square

4. Bubbles and rationality

The results above emphasize that bubbles are more likely to be sustained when the fraction of rational investors is small. In fact, rational investors need to leave earlier when they are many, as they need to find enough investors to buy their stocks, which decreases T_1 . Moreover, a higher r means that more people leave the market along the equilibrium. This makes I -traders' expectations less optimistic, which in turn decreases T_2 . Hence, a larger share of rational investors tends to reduce the maximal duration of the bubble.³³ However, these investors are not always a stabilizing force in our model. As we now show, a more rational market need not display fewer or shorter bubbles.

4.1. Rational investors should not be too many. As in standard models, we cannot have bubbles if all investors are fully rational. In particular, in a bubble equilibrium, r has to be small enough, relative to K and z , so that all rational players are able to all sell at T . This condition defines T_1 , as expressed in equation (3.13). As we must have $T_1 \geq 1$, we need

$$r \leq \frac{K - X}{1 - X} \equiv r_{\max}.$$

Hence, we can state the following Proposition.

PROPOSITION 6. *If $r > r_{\max}$, then no bubble equilibrium exists.*

Notice that r_{\max} defines a necessary condition for the existence of a bubble equilibrium. In the previous Section, we characterized a sufficient condition, in terms of the upper bound r^* . As one expects, it can be shown that $r_{\max} > r^*$ for every K and z .

³³ None of these effects are considered in De Long, Shleifer, Summers and Waldmann (1990b), where feedback traders have exogenous expectations and no budget constraint. Indeed, they find the opposite result, i.e. more rational investors drive the prices further away from fundamentals.

4.2. Rational investors should not be too few. As expressed in Propositions 3 and 4, bubbles are more likely to arise and to last longer when the share of rational investors r is low. The extreme case of $r = 0$, however, deserves a closer look. First, notice that if all agents are of the same type, they share the same expectations. Hence, trade occurs only for exogenous reasons and $V_t = X$ for all $t \leq T$.

Second, and more interestingly, note that in such scenario I -players never realize that the crash occurs at $T + 1$, not even in period $T + 1$. By contrast, when there are sufficiently many rational investors, they give a negative shock to the market by exiting at T . At this point, I -investors realize they had overestimated the length of the bubble, i.e. the crash is going to occur earlier than they had predicted at T , when they bought for the last time. That is, after having observed R 's exit, I -traders are convinced that a crash is about to occur and they run to sell (actually, it is already too late). Such final panic phase is rather common in these episodes (see Kindleberger, 2005), and we now show that it requires r be not too small. Specifically, consider the following condition:

$$(4.1) \quad B_{T+1}^{I,T+1} < S_{T+1}^{I,T+1}.$$

Condition (4.1) ensures that, at the beginning of $T + 1$, just before the crash occurs, all investors expect the crash to occur next. Together with condition (3.11), this condition requires that I -investors' expectations about the date of the crash change between period T and period $T + 1$. That is, I -players understand that the plans they made at T were not accurate. In particular, they realize that the crash is occurring earlier than expected, as they have underestimated the amount of exits at T .

Adding condition (4.1) to our equilibrium conditions (3.10), (3.11) and (3.12), we can uncover an important role that rational players play in our model. From the previous discussion, it is clear that condition (4.1) requires some bad shock to occur in period T . Since the only source of such shock is that R -investors decide to exit, we need sufficiently many of them. It follows in particular that if $r = 0$, no equilibrium bubble can satisfy condition (4.1). I -investors would not receive any bad news along the equilibrium path, so they would never revise their plans and sell earlier than expected.³⁴ We can state this more precisely with the following Proposition.

PROPOSITION 7. *In a bubble equilibrium where condition (4.1) holds, we must have*

$$r > r_{\min},$$

where r_{\min} is implicitly defined by $r_{\min} = 1/T$.

³⁴ This result, which stresses that we need rational investors to break the bubble at some point, can be viewed as a complement to the literature noting that fully rational investors may artificially increase returns and initiate a bubble (De Long et al., 1990b).

PROOF. Condition (4.1) writes $\bar{\beta}K_{T+1} < S^I(T)$. Together with equation (3.11), this require $K_{T+1} = K_T - S_T - rK_T < K_{T+1}^{I,T} = \bar{\beta}K_T - S^I(T)$. That is, $(1-r)K_T + S^I(T) < \bar{\beta}K_T + S_T$. We now show that $rT > 1$ is equivalent to $(1-r) < \bar{\beta}$ and to $S^I(T) < S_T$. First, by the definition of $\bar{\beta}$, $(1-r) < \bar{\beta}$ writes $(T+1)(1-r) < (T-r)$ which gives $rT > 1$. Second, again by definition, $S^I(T) < S_T$ writes $(T-r)X + (1+r)(1-K) < (T+1)[r(1-K) + X(1-r)]$. Rearranging, this gives $Tr(1-K) - TrX > 1-K-X$, that is $rT > 1$. Hence requiring $(1-r)K_T + S^I(T) < \bar{\beta}K_T + S_T$ is equivalent to requiring $rT > 1$. Since for $r = 0$, $1/T > 0$, the last expression implicitly defines a unique r_{\min} such that (4.1) holds only if $r > r_{\min}$. \square

4.3. Ambiguity Aversion. Proposition 4 shows that, in a setting with risk neutral investors, bubbles tend to disappear as the fraction of rational investors increases. We now show that this need not be the case if we consider a setting with ambiguity averse agents.³⁵ In fact, in our model, uncertainty is purely strategic, i.e. it concerns solely the predictions of what other investors do. Hence, the uncertainty borne by each player depends on his ability to understand his opponents' equilibrium strategies. Perfectly rational players face no uncertainty, and so they may be willing to take longer positions than some I -investors. As a result, increasing the share of rational investors need not take the market closer to efficiency. Instead, by increasing the amount of investors in the speculative market, an higher r can make I -investors' expectations more optimistic, and allow to sustain longer bubbles.³⁶

In order to formally explore these intuitions, we enrich our setting by introducing ambiguity averse investors. In particular, consider two types of I -investors: L -investors are ambiguity neutral, as modeled in Section 2, and H -investors are ambiguity averse.³⁷ Specifically, we assume that H -agents admit that their model can be wrong. They believe that their predictions of β_t and σ_t can be mistaken by at most ε , so that the actual buy rate β_t will be in the interval $[\bar{\beta} - \varepsilon, \bar{\beta} + \varepsilon] \cap [0, 1]$ for every t , and similarly the actual sell rate σ_t will be in $[\bar{\sigma} - \varepsilon, \bar{\sigma} + \varepsilon] \cap [0, 1]$.³⁸ Furthermore, these players adopt the most cautious behavior: they assume the worst realizations of β_t

³⁵ Uncertainty (or equivalently ambiguity) describes situations where the perceived likelihood of some event cannot be represented by a probability measure over the possible states of the world. That is, we allow for the possibility that the information perceived by our investors is not accurate enough to provide them with a unique probability measure.

³⁶ In particular, in our model, differently from the noise traders literature (De Long et al., 1990a), increasing the risk bearing capacity of the economy drives the prices further away from fundamentals.

³⁷ For the reason just mentioned, attitudes toward ambiguity are not relevant for R -investors.

³⁸ The error term ε is here taken as given. One could for example endogenize this interval by letting it lie between the minimum and the maximum buy and sell rates observed along the equilibrium.

and σ_t , and choose the optimal strategy given this realization.³⁹ It follows that, in order to be part of speculation, they require a higher return to be compensated by the perceived uncertainty.⁴⁰ Investors of type L , who are assumed to be neutral towards ambiguity, can alternatively be thought as being unaware of the fact that they can make mistakes.⁴¹

We can now replicate our previous analysis, characterizing how in this setting the maximal sustainable bubble varies with the share of rational investors. First, as in equation (3.11), L -investors buy/stay in at t if and only if $K_t \geq W(T)$, while H -investors buy/stay in at t if and only if $K_t \geq W(T, \varepsilon)$, where

$$W(T, \varepsilon) \equiv \frac{(1 - K)(\bar{\sigma}(T) + \varepsilon)(1 + \bar{\beta}(T) - \varepsilon)}{(\bar{\beta}(T) - \varepsilon)^2}.$$

We can see that $W(T, \varepsilon)$ increases in ε , hence H -investors will always sell before L -investors.⁴² We want to define an equilibrium where H -investors sell at some \tilde{T} , fully rational investors sell at $T > \tilde{T}$, and L -investors sell at $T + 1$. As above, let r be the share of fully rational investors and $(1 - r)$ be that of boundedly rational ones. The latter can display either high or low ambiguity aversion, in proportion h and $(1 - h)$ respectively. We look for an equilibrium in which

$$\beta_t = \begin{cases} 1 & \text{for } t < \tilde{T} \\ 1 - (1 - r)h & \text{for } t = \tilde{T} \\ 1 & \text{for } t \in (\tilde{T}, T) \\ 1 - \frac{r}{1-h(1-r)} & \text{for } t = T \\ 0 & \text{for } t = T + 1, \end{cases}$$

and

$$\sigma_t = \begin{cases} z & \text{for } t < \tilde{T} \\ z + (1 - z)(1 - r)h & \text{for } t = \tilde{T} \\ z & \text{for } t \in (\tilde{T}, T) \\ z + (1 - z)\frac{r}{1-h(1-r)} & \text{for } t = T \\ 1 & \text{for } t = T + 1. \end{cases}$$

³⁹ Formally, we are assuming that these investors have a set of probability measures over the possible realizations of β_t and σ_t . Investors compute the minimal expected payoffs conditional on each possible prior, and decide the investment strategy corresponding to the maximum of such payoffs. This idea, which may be thought as an extreme form of uncertainty aversion, was formalized by Gilboa and Schmeidler (1989).

⁴⁰ Indeed, many authors have invoked ambiguity aversion as a possible resolution of the Equity Premium Puzzle (e.g. Chen and Epstein, 2002; Klibanoff, Marinacci and Mukerji, 2005).

⁴¹ One way to interpret these types is in terms of different degrees of confidence about cognitive ability, with H -investors being cautious and L -investors being overconfident. Under this perspective, we notice that our previous analysis is implicitly stressing the role of overconfidence, as I -investors do not consider the possibility that their forecast may be inaccurate (see Scheinkman and Xiong, 2003).

⁴² $W(T, \varepsilon)$ is simply obtained from (3.11) by replacing $\bar{\beta}(T)$ with $\bar{\beta}(T) - \varepsilon$, and $\bar{\sigma}(T)$ with $\bar{\sigma}(T) + \varepsilon$.

That is, for $t < \tilde{T}$ no investor wants to exit, so the volume of trade is given by liquidity shocks only, occurring with probability z . At \tilde{T} , high ambiguity averse agents, with mass $(1 - r)h$, leave the speculative market, selling to rational agents R and to low ambiguity averse agents L .⁴³ For $t \in (\tilde{T}, T)$ everyone in the market is willing to invest, as H -agents have left and they are not considered anymore. At T , rational investors exit speculation, selling to L -agents. At $T + 1$, L -agents realize that the crash is about to occur and they want to sell, while no one is willing to buy. The crash occurs and the market closes.

We can then replicate the analysis of Section 3, computing the average buy and sell rates induced by this equilibrium, and defining the maximal bubble T_{\max} . We are interested in how T_{\max} varies with r , for a given proportion of ambiguity averse agents h .⁴⁴ In order to simplify the exposition, from now on we assume that H -investors perceive enough uncertainty to be induced to sell immediately, i.e. that ε is large enough to have $\tilde{T} = 1$.⁴⁵ In addition to the effects already explored in the previous Section, which lead T_{\max} to decrease with r , we now have to consider that the proportion of ambiguity averse investors H decreases in r . This has a series of direct and indirect effects, making it possible that T_{\max} increases in r .

4.3.1. *The effects on T_1 .* Recall that T_1 is the latest period in which all rational investors can exit the market at the peak of the bubble, i.e. the last date with excess demand. That is, T_1 solves $B_T = S_T$. As in the previous Section, an higher r tends to decrease T_1 since it increases the supply S_T and decrease the demand B_T . On the other hand, an higher r also implies that the amount of ambiguity averse people H is lower, and this tends to increase T_1 . In fact, a lower H increases K_T , the amount of cash available at T , hence possibly allowing for a greater demand. Moreover, the amount of investors who enter the speculative market at \tilde{T} increases in the amount of investors selling at \tilde{T} , i.e. it increases in H . Hence, if H is small, S_T is small as a few

⁴³ Since the excess demand is allocated randomly between R and L , and shocks are also random, the proportion of R and L in the speculative market remains constant. That is, for $t \in (\tilde{T}, T)$, the proportion of R -investors in the market is

$$\frac{r}{r + (1 - h)(1 - r)},$$

and, similarly, the proportion of L -agents is

$$\frac{(1 - r)(1 - h)}{r + (1 - h)(1 - r)}.$$

⁴⁴ The effect of h is clear: T_{\max} decreases with h , that is the fraction of people leaving the market early.

⁴⁵ For example H -investors may think that β_t and σ_t are respectively drawn by distributions with mean $\bar{\beta}$ and $\bar{\sigma}$ and full support $[0, 1]$. As they are extremely ambiguity adverse, they assume $\beta_t = 0$ and $\sigma_t = 1$ for all t , so they exit as soon as possible. In other words, given that there is a one-to-one mapping between \tilde{T} and ε , we now consider \tilde{T} as an exogenous parameter of the model.

rational investors have stocks, and B_T is large as many low ambiguity averse still have cash, so T_1 can be larger. We have the following Lemma.

LEMMA 5. *Let*

$$\hat{r} \equiv 1 - \sqrt{\frac{1-K}{h(1-h)(1-X)}}.$$

Then,

$$\frac{\partial T_1}{\partial r} > 0,$$

for $r \leq \hat{r}$.

PROOF. In our equilibrium, S_T includes all R -investors with stocks at T and the exogenous sales X , while B_T includes all L -investors with cash at T . That is,

$$S_T = \sigma_t(1-K) = X + (1-z)(1-K)\frac{r}{1-h(1-r)},$$

and

$$B_T = \beta_t K_T = \frac{(1-r)(1-h)}{1-h(1-r)} \{K - (T-1)X - (1-r)h[(1-z)(1-K) + K - (\tilde{T}-1)X]\},$$

Hence, $B_T > S_T$ defines the following condition:

$$(4.2) \quad T < \frac{1}{X} \left[K + X - (1-X)(1-r)h - \frac{X - hX(1-r)}{(1-r)(1-h)} - \frac{(1-z)(1-K)r}{(1-r)(1-h)} \right] \equiv T_1.$$

After simple algebra, we can see that

$$(4.3) \quad \frac{\partial T_1}{\partial r} = \frac{1}{X} \left[h(1-X) - \frac{1-K}{(1-r)^2(1-h)} \right],$$

which is positive when

$$(4.4) \quad r \leq 1 - \sqrt{\frac{1-K}{h(1-h)(1-X)}} \equiv \hat{r}(h).$$

This proves the Lemma. \square

Hence, this condition tells that T_1 increases in r when r is sufficiently small. This is due to the fact that, as r increases, $\partial B_T / \partial r$ decreases more than proportionally than $\partial S_T / \partial r$. In fact, the positive effect on K_T , which is weighted by β_t , gets smaller as r increases. Moreover, as shown in Section 3.3, T_1 is the constraint defining T_{\max} when z is very small.⁴⁶ Hence, the following Corollary follows.

⁴⁶ Replicating the analysis of Section 3.3 in this setting gives that $T_1 < T_2$ if and only if r exceeds a threshold, which is implicitly defined by

$$r > \frac{(1-h)(W(T_1) - X)}{(1-h)(W(T_1) - X) + 1 - K}.$$

Hence, again, if $X \rightarrow 0$ the right hand side of the last equation tends to zero, and so $T_1 < T_2$.

COROLLARY 5. *If $z \rightarrow 0$, then T_{\max} increases in r for every $r \leq \hat{r}$.*

4.3.2. *The effects on T_2 .* Recall that T_2 is defined as the latest period in which I -investors believe it is profitable to enter the speculative market. Hence, T_2 is defined by the amount of available cash observed in the economy and by I -investors' expectations about future buy and sell rates. Formally, as seen, T_2 is the largest root solving $W(T) = K_T$. In the previous Section, with risk neutral investors, we had that T_2 unambiguously decreased in r as an higher r made expectations more pessimistic, i.e. it increased $W(T)$. Again, this need not hold now since, by changing r , we also affect the mass H of cautious investors who exit the market immediately.

The effects on T_2 are two. The first, already mentioned, is that the amount of available cash K_T increases in r . Hence, by this effect, an higher r pushes towards a larger T_2 . Second, an higher r influences L -players' expectations, as defined in $W(T)$. By decreasing H , it pushes towards more optimistic expectations, i.e. it decreases $W(T)$. In addition, decreasing H has an indirect effect. Given that exits are perceived in relation to the amount of people still in the market, a lower H increases the amount of people in the market at T , hence making the exit of rational players appear smaller. Hence, this may also induce more optimistic expectations and greater T_2 . We then have the following Lemma.

LEMMA 6. *Let*

$$\tilde{r} \equiv \frac{1}{h} \left[\sqrt{\frac{1-h}{h}} - (1-h) \right].$$

Then,

$$\frac{\partial T_2}{\partial r} > 0,$$

when either (i) $r \geq \tilde{r}$ or (ii) $z \rightarrow 0$ hold.

PROOF. Recall that $W(T)$ decreases in $\bar{\beta}(T)$ and increases in $\bar{\sigma}(T)$. The average buy and sell rates induced in this equilibrium are

$$(4.5) \quad \bar{\beta} = \frac{1}{T+1} \left[T - (1-r)h - \frac{r}{1-h(1-r)} \right],$$

and

$$(4.6) \quad \bar{\sigma} = \frac{1}{T+1} \left[1 + zT + (1-z) \left((1-r)h + \frac{r}{1-h(1-r)} \right) \right].$$

Observing equations (4.5) and (4.6), we can see that

$$\frac{\partial \bar{\beta}}{\partial r} > 0 \iff \frac{\partial}{\partial r} \left[(1-r)h + \frac{r}{1-h(1-r)} \right] < 0 \iff \frac{\partial \bar{\sigma}}{\partial r} < 0,$$

and hence

$$(4.7) \quad \frac{\partial W(T)}{\partial r} < 0 \iff \frac{\partial}{\partial r} \left[(1-r)h + \frac{r}{1-h(1-r)} \right] < 0.$$

Condition (4.7) requires

$$-h + \frac{1-h}{[1-h(1-r)]^2} < 0,$$

which writes

$$(4.8) \quad r > \frac{1}{h} \left(\sqrt{\frac{1-h}{h}} - (1-h) \right) \equiv \tilde{r}(h).$$

Hence, if $r \geq \tilde{r}$ then an increase in r unambiguously increases T_2 , since it increases K_T and decreases $W(T)$. This proves condition (i).

If $r < \tilde{r}$ instead, the effect is ambiguous, as both K_T and $W(T)$ increase. What matters is then the magnitude of the two effects. The marginal effect on K_T , which tends to increase T_2 , is $h[1-X]$. The marginal effect on $W(T)$ is small when T is large, i.e. when X is small. In fact, as $X \rightarrow 0$, $T_2 \rightarrow \infty$, $\bar{\beta} \rightarrow 1$, $\bar{\sigma} \rightarrow 0$. Moreover, differentiating $W(T)$ with respect to r , we see that

$$\frac{\partial W(T)}{\partial r} \rightarrow \frac{2(1-K)}{T+1} \left[-h + \frac{1-h}{[1-h(1-r)]^2} \right], \text{ as } T \rightarrow \infty,$$

so $\partial W(T)/\partial r \rightarrow 0$ as $X \rightarrow 0$. Hence, the effect on K_T always dominates for sufficiently small values of z . That is, T_2 increases in r when $z \rightarrow 0$, which is condition (ii). \square

4.3.3. The general case. The previous Lemmas may not be sufficient for determining how T_{\max} varies with r . In fact, it may be that T_1 is the binding constraint for $r > \tilde{r}$ and T_2 is binding for $r < \hat{r}$.⁴⁷ One way to show that T_{\max} may increase with r irrespective on whether T_1 or T_2 binds is to choose an r within the interval $[\tilde{r}, \hat{r}]$, provided this is not empty. With some algebra, one can see that \hat{r} always exceeds \tilde{r} when K is sufficiently close to 1.⁴⁸ Hence, the previous conditions can be jointly satisfied, and they define a set of sufficient conditions such that the maximal sustainable bubble is locally increasing in r . The next Proposition summarizes this result.

PROPOSITION 8. *If $r \in [\tilde{r}, \hat{r}]$, then*

$$\frac{\partial T_{\max}}{\partial r} > 0.$$

⁴⁷ Indeed, in Section 3.3, we showed that T_1 is binding when r is large.

⁴⁸ More precisely, we require $K \geq 1 - (1-X)(1-y)^2 y^2$, where $y \in (0,1)$ is defined as $y = \sqrt{(1-h)/h}$.

To sum up, the possibility that T_{\max} increases in r is in general assured when r is not too big nor too small. For small r , the constraint defining T_{\max} is T_2 . However, T_2 may not increase in r if r is too small. In this case, the amount of people in the market is small, exit are perceived as big and so expectations are more pessimistic. On the other hand, for large r , the constraint defining T_{\max} is T_1 . However, T_1 does not increase in r if r is too large. In fact, in this case, the amount of people buying at T would be small, hence the positive effect of having more cash available at T would be small. Hence, r has to take some intermediate values, within the interval defined in the last Proposition.

5. Conclusion

In this paper, we have modeled speculative behaviors in a setting where it is common knowledge that the stock market is overvalued and that a crash is bound to occur. In our model, traders face the same common information, but they differ in the ability to process it, and thus in the accuracy of their forecasts. We have shown that the existence of bubbles in such market does not require investors to be irrational, and we have provided a simple structure to partially sophisticated investors, who fail to understand the timing of other investors' strategies. The resulting bubble equilibrium displays a number of features that have been documented in the empirical literature: high short-term returns attracting many new and typically inexperienced investors, who form euphoric expectations about the duration of the bubble. These investors provide the capital necessary to fuel the bubble, hence allowing high volumes of trade and further price increases. At some point, fully rational investors exit and the crash occurs. Moreover, the analysis has emphasized in several ways that the relation between bubbles and rationality of the market is multifaceted, thereby qualifying the view that deviations from fundamentals tend to vanish if one increases the amount of rational traders in the market.

Our model is deliberately very simple, and surely open to many extensions and modifications. However, with due *caveats*, some policy implications can be drawn. First, the general idea behind our approach is that information availability *per se* need not be the solution for market efficiency. Instead, we emphasize information *accessibility*, which focuses on whether information is presented in a way to ease its interpretation. In our model, information is complete, but bubbles arise as some people face limitations in processing all the relevant aspects of such information. In this sense, a regulator concerned with market stability, or with people losing too much money, may consider issues of simplicity and interpretability of information, or even of information overload, rather than just increasing the amount of potentially available information.

Another way in which information is not necessarily promoting efficiency in our setting is that news may have a destabilizing effect. In fact, as seen, partially sophisticated investors get excited by observing unexpected increases in the price, which lead them to overestimate the duration of the bubble and stay invested for too long. According to our approach, media can play an important role in stimulating or undermining a speculative phenomenon.⁴⁹ If media make information more accessible, in the sense just described, they have a stabilizing effect as they allow inexperienced investors to be sufficiently informed and avoid exploitation. If instead media are just reporting short-term high returns, they can distort perceptions, and hence asset prices, by making people euphoric (as it happens in our model). Under this light, we can see that inexperienced investors would benefit from ignoring the news and committing to some investment strategy. In fact, having observed apparently good realizations in the early phase of the bubble, these investors revise their expectations in the wrong way. If such commitment were possible, the bubble would not arise.

Third, we have seen that having a larger fraction of rational traders need not lead closer to market efficiency. This is due to their willingness to ride the bubble, and sometimes to ride it longer than some not fully rational, but cautious, investors. In our model, increasing the risk-bearing capacity in the economy drives the prices further away from fundamentals, as people who perceive less uncertainty are willing to speculate more. Efficiency would instead be achieved by making investors aware that their predictions can be imprecise, and that apparently strange observations may not be the result of chance, but rather of a wrong model. Hence, in this sense, the really stabilizing force in our setting are not arbitrageurs, but people considering that they can make mistakes.

⁴⁹ The strong relation between media coverage and abnormal returns has been recently documented e.g. by Dyck and Zingales (2003) and Veldkamp (2006).

References

- Abreu, D. and Brunnermeier, M. K. (2003), ‘Bubbles and crashes’, *Econometrica* 71(1), 173–204.
- Allen, F., Morris, S. and Postlewaite, A. (1993), ‘Finite bubbles with short sale constraints and asymmetric information’, *Journal of Economic Theory* 61(2), 206–229.
- Avery, C. and Zemsky, P. (1998), ‘Multidimensional uncertainty and herd behavior in financial markets’, *American Economic Review* 88(4), 724–48.
- Banerjee, A. V. (1992), ‘A simple model of herd behavior’, *Quarterly Journal of Economics* 117, 797–817.
- Bianchi, M. (2007), ‘Speculative bubbles and rationality - a review of the theory’, Mimeo: Stockholm School of Economics.
- Blanchard, O. J. and Watson, M. W. (1982), ‘Bubbles, rational expectations and financial markets’, NBER Working Papers No. 0945.
- Blume, L. and Easley, D. (2006), ‘If you’re so smart, why aren’t you rich? belief selection in complete and incomplete markets’, *Econometrica* 74(4), 929–966.
- Brunnermeier, M. K. and Nagel, S. (2004), ‘Hedge funds and the technology bubble’, *Journal of Finance* 59(5), 2013–2040.
- Chen, Z. and Epstein, L. (2002), ‘Ambiguity, risk, and asset returns in continuous time’, *Econometrica* 70(2), 1403–1443.
- Cipriani, M. and Guarino, A. (2005), ‘Herd behavior and contagion in financial markets’, Mimeo: George Washington University and University College London.
- Cochrane, J. H. (2002), ‘Stocks as money: Convenience yield and the tech-stock bubble’, NBER Working Papers No. 8987.
- Cutler, D. M., Poterba, J. M. and Summers, L. H. (1990), ‘Speculative dynamics and the role of feedback traders’, *American Economic Review* 80(2), 63–68.
- Cutler, D., Poterba, J. and Summers, L. (1989), ‘What moves stock prices?’, *Journal of Portfolio Management* 15(3), 4–12.

- De Long, J. B., Shleifer, A., Summers, L. H. and Waldmann, R. J. (1990a), 'Noise trader risk in financial markets', *Journal of Political Economy* 98(4), 703–38.
- De Long, J. B., Shleifer, A., Summers, L. H. and Waldmann, R. J. (1990b), 'Positive feedback investment strategies and destabilizing rational speculation', *Journal of Finance* 45(2), 379–395.
- Dekel, E., Fudenberg, D. and Levine, D. K. (2004), 'Learning to play bayesian games', *Games and Economic Behavior* 46(2), 282–303.
- Dyck, A. and Zingales, L. (2003), 'The media and asset prices', Mimeo: Harvard Business School and University of Chicago.
- Eyster, E. and Rabin, M. (2005), 'Cursed equilibrium', *Econometrica* 73(5), 1623–1672.
- Garber, P. M. (1990), 'Famous first bubbles', *Journal of Economic Perspectives* 4(2), 35–54.
- Gilboa, I. and Schmeidler, D. (1989), 'Maxmin expected utility with nonunique prior', *Journal of Mathematical Economics* 18, 141–153.
- Harrison, J. M. and Kreps, D. M. (1978), 'Speculative investor behavior in a stock market with heterogeneous expectations', *Quarterly Journal of Economics* 92(2), 323–36.
- Higgins, E. T. (1996), 'Knowledge activation: Accessibility, applicability and salience', in E. T. Higgins and A. E. Kruglanski, eds, *Social psychology: Handbook of basic principles*, New York: Guilford., pp. 133–168.
- Hirshleifer, D. (2001), 'Investor psychology and asset pricing', *Journal of Finance* 56(4), 1533–1597.
- Jehiel, P. (1995), 'Limited horizon forecast in repeated alternate games', *Journal of Economic Theory* 67, 497–519.
- Jehiel, P. (2005), 'Analogy-based expectation equilibrium', *Journal of Economic Theory* 123(2), 81–104.
- Jehiel, P. and Samet, D. (2006), 'Valuation equilibria', Mimeo: UCL and PSE.
- Kahneman, D. (2003), 'Maps of bounded rationality: Psychology for behavioral economics', *American Economic Review* 93(5), 1449–1475.
- Kindleberger, C. (2005), *Manias, panics, and crashes: a history of financial crises*, Wiley and Sons.
- Klibanoff, P., Marinacci, M. and Mukerji, S. (2005), 'A smooth model of decision making under ambiguity', *Econometrica* 73(6), 1849–1892.
- Kogan, L., Ross, S., Wang, J. and Westerfield, M. (2006), 'The price impact and survival of irrational traders', *Journal of Finance* 61(1), 195–229.
- Kyle, A. S. (1985), 'Continuous auctions and insider trading', *Econometrica* 53(6), 1315–35.

- Milgrom, P. and Stokey, N. (1982), 'Information, trade and common knowledge', *Journal of Economic Theory* 26(1), 17–27.
- Porter, D. P. and Smith, V. L. (2003), 'Stock market bubbles in the laboratory', *Journal of Behavioral Finance* 4(1), 7–20.
- Rabin, M. (1998), 'Psychology and economics', *Journal of Economic Literature* 36(1), 11–46.
- Rubinstein, A. (1998), *Modeling Bounded Rationality*, MIT Press.
- Sargent, T. (1993), *Bounded Rationality in Macroeconomics*, Oxford University Press.
- Scheinkman, J. A. and Xiong, W. (2003), 'Overconfidence and speculative bubbles', *Journal of Political Economy* 111(6), 1183–1219.
- Shiller, R. J. (1989), *Market Volatility*, MIT Press.
- Shiller, R. J. (2000), 'Measuring bubble expectations and investor confidence', *Journal of Psychology and Financial Markets* 1(1), 49–60.
- Shleifer, A. (2000), *Inefficient Markets*, Oxford University Press.
- Simon, H. A. (1955), 'A behavioral model of rational choice', *Quarterly Journal of Economics* 69(1), 99–118.
- Smith, V. L., Suchanek, G. L. and Williams, A. W. (1988), 'Bubbles, crashes, and endogenous expectations in experimental spot asset markets', *Econometrica* 56(5), 1119–51.
- Stiglitz, J. E. (1990), 'Symposium on bubbles', *Journal of Economic Perspectives* 4(2), 13–18.
- Temin, P. and Voth, H.-J. (2004), 'Riding the south sea bubble', *American Economic Review* 94(5), 1654–1668.
- Tirole, J. (1982), 'On the possibility of speculation under rational expectations', *Econometrica* 50(5), 1163–1182.
- Veldkamp, L. (2005), 'Slow boom, sudden crash', *Journal of Economic Theory* 124(2), 230–257.
- Veldkamp, L. (2006), 'Media frenzies in markets for financial information', *American Economic Review* 96(3), 577–601.

Migrants

CHAPTER 3

Immigration Policy and Self-Selecting Migrants

Milo Bianchi

ABSTRACT. We build a simple theory of self-selection into migration and immigration policy formation. We show that any immigration policy affects the skill composition of the migration flows, and this effect may drive the policy outcome in the receiving country. For example, restricting immigration when it is low skilled may worsen immigrants' self-selection and thus the receiving country skill distribution. Hence, understanding the migration decision becomes crucial for analyzing the political economy of immigration. For example, these composition effects may lead the majority of natives to support further restrictions even though current immigrants are not harmful for them, and they may explain significant immigration restrictions even in a purely utilitarian world.

Keywords: Immigrants' self-selection; political economy of immigration; immigration policy preferences.

JEL codes: J61, F22, O24, D78.

1. Introduction

It is well known that those who decide to migrate are not a random sample of their home country population. Incentives to migrate, and resources to pay the migration costs, vary with skills. Moreover, immigration restrictions significantly influence benefits and costs of migration, and thereby immigrants' skill composition. Hence, understanding immigrants' self-selection becomes crucial for predicting the outcome of immigration restrictions in the receiving country, since self-selection determines how different potential migrants respond to immigration policy changes.

This paper builds a simple framework to analyze the interaction between self-selection into migration and the determination of immigration policy. Indeed, it is commonly understood that various effects of migration, both in sending and receiving

⁰ This is a substantially revised version of Bianchi (2006). I am indebted to Abhijit Banerjee, Guido Friebel, Sergei Izmailkov, Anna Maria Mayda, Elena Paltseva and especially Jörgen Weibull for helpful comments and precious encouragement. I have also benefited from seminar discussions at MIT, Stockholm School of Economics, Uppsala University; at the EEA Meeting (Vienna) and the XVII Villa Mondragone International Economic Conference (Rome). Financial support from the Jan Wallander and Tom Hedelius and the Knut and Alice Wallenberg Foundations is gratefully acknowledged.

countries, vary significantly with immigrants' characteristics.¹ However, the relation between immigration policy and immigrants' skill composition remains largely unexplored.

Our framework allows a unified treatment of a series of questions, which we show are deeply interrelated. First, what drives the decision to migrate, i.e. which conditions are likely to lead to high vs. low skilled migration. Second, how receiving countries' policies affect this decision, i.e. what is the relation between immigration restrictions and immigrants' skill composition. Last, what are the implications for receiving countries, i.e. how immigrants' self-selection needs to be considered in order to predict the effects of a policy, and then its political support.

More specifically, we model migration in a world with two countries. The sending country has a heterogeneous population of individuals, called foreigners, who decide whether to work at home or migrate to the receiving country. Foreigners differ in skills and wealth, so each of them faces different incentives to migrate, and different resources to pay for it. High skilled tend to have higher wealth and so higher ability to incur the migration costs. High skilled, however, need not be the ones with the highest gain from migration, since returns to skills may be higher in the sending country. Accordingly, immigrants' self-selection is driven both by incentives and by wealth constraints.

The receiving country has a heterogeneous population of workers, called natives, and a standard labor market where immigrants compete with similarly skilled natives and complement those with different skills. Natives support a policy that maximizes their equilibrium wages, so high skilled aim at increasing the supply of low skilled immigrants, while low skilled push for the opposite. According to these preferences, and to the weight attached to different groups in the population, the receiving country government sets the immigration restrictions.

In particular, immigration restrictions increase the costs migrants have to pay to enter and work in the receiving country. Given that incentives and resources to migrate differ across foreigners, the policy influences immigrants' skill composition. On the one hand, higher costs allow only the richest foreigners to migrate, and these tend to be high skilled. On the other, restrictions have a stronger deterrence on those with lower gains from migration, thereby increasing immigrants' skill composition if and only if returns to skills are higher in the destination country.

In this setting, we show that any immigration policy affects both size and skill composition of the migration flow. While the latter effect is typically overlooked, it may indeed *reverse* the immigration policy outcomes, as predicted by the size effect

¹ Restricting the attention to the economics literature, see Borjas (1994), Friedberg and Hunt (1995), Chiswick, Lee and Miller (2005) on the labor market effects; and Storesletten (2000), Lee and Miller (2000) on the effects on public finance.

only. Hence, it is generally not safe to assume that immigration restrictions just select from a given pool of applicants, and so they act independently from the migration decision. On the contrary, the forces behind the decision to migrate may be crucial also for understanding the effect of a given policy in the receiving country.

In fact, size and composition effects have typically opposite directions. The size effect, whereby restrictions reduce flows for a given skill composition, is by definition random, hence it hits a group of foreigners proportionally to their propensity to migrate. In contrast, the composition effect tends to be stronger on the least represented group.

Moreover, the composition effect is dominant when the least represented group of migrants is, in absolute terms, the most sensitive to a policy change. For this to be the case, migration costs need to be small, so that the population of migrants who respond to policy changes is sufficiently heterogeneous. If only one group of foreigners migrates, being they the richest or the most motivated, then by definition there is no composition effect.

Given that the strength of composition and size effects depends on the level of immigration restrictions, these restrictions may have non-monotone effects on the receiving country. Suppose for example that returns to skills are higher in the destination country and the two countries have similar skill compositions. Immigrants are then positively self-selected and more skilled than natives. Increasing restrictions improve immigrants' skill composition and so they tend to improve the receiving country skill distribution. At the same time, they reduce the size of high skilled migration and this tends to worsen the receiving country skill distribution. The resulting effect is an inverted-U: for low levels of cost, the composition effect tends to be stronger; after some point, the size component takes over. For high migration costs, few low skilled foreigners migrate, so further restrictions just lower the size of high skilled migration.

These effects turn out to have interesting implications also with respect to natives' attitudes over immigration policy and to the government's optimal policy design. First, the relation between natives' skills and support for a given policy may appear counterintuitive. When the composition component is stronger, we may observe a group of natives supporting a more restrictive policy even though current immigrants complement them, and vice versa. For example, low skilled natives may oppose immigration even if current immigrants are not competing with them, since further restrictions would increase immigrants' average skills, the receiving country skill ratio and hence low skilled wages. In this case, it is self-selection, rather than immigrants' skill composition *per se*, that drives the relation between natives' skills and immigration policy preferences.

Second, the government's program has now to account for immigrants' self-selection. As in standard models, we show that, fixing immigrants' skill composition, free immigration maximizes the total surplus in the receiving country. Once one considers the composition effect, however, some restrictions to immigration become optimal even in a world with no distributional concerns and no political economy distortions. In fact, restrictions may be an indirect way to screen the most desirable type of migration.

The rest of the paper proceeds as follows. Section 2 presents the baseline model and Section 3 derives our main results. We then elaborate further on migration costs, network effects and time consuming bureaucracies (Section 4.1); returns to skills and labor market discrimination (Section 4.2); immigration policy preferences, fiscal policy and political economy issues (Section 4.3). In Section 5, we conclude by drawing some policy implications. Before that, however, we briefly confront with the existing literature.

1.1. Related literature. The present paper lies between two streams of literature: the supply side of immigration, dealing with the migration decision and immigrants' self-selection; and the demand side, dealing with citizens' preferences over immigration and immigration policy formation. At a general level, the major novelty of the paper is the focus on the interaction between demand and supply, in order to show that, by considering each side in isolation, one may draw erroneous conclusions both on self-selection and on the effects of immigration policy in the receiving country.²

More specifically, we model the migration decision building on the basic theory of migration as human capital investment, which goes back to Sjaastad (1962). The most influential work in this literature is probably by Borjas (1987), who emphasizes the role of different returns to skills across countries in shaping self-selection into migration. This model typically predicts negative self-selection, since wage inequality tend to be higher in developing countries. Some recent empirical literature, however, documents that immigrants may be positively self-selected even when returns to skills are higher in the sending country (see e.g. Chiquiar and Hanson, 2005; Akee, 2005; Brucker and Defoort, 2006). We reconcile these findings by enriching the standard setting in two important and highly unexplored aspects. First, we introduce wealth constraints. These are more likely to bind for low skilled and so they may drive self-selection.³

² The only paper considering this interaction is, to my knowledge, Belletini and Berti Ceroni (2005). Assuming that immigrants are positively self-selected, they argue in favor of a high immigration quota. By reducing wages in the receiving country, this would increase immigrant quality and maximize national income.

³ Hatton and Williamson (2004) state: "When dealing with selection, the immigration literature tends to stress income incentive [...But] changes in selection can be best explained by changes in the costs of the move and the capital constraints on it". Exceptions are the theory of illegal migration in Friebel and Guriev (2004) and the work by Lopez and Schiff (1998), who focus on the effect of trade

Second, and more importantly, we stress the role of immigration policies, which can affect immigration costs and so immigrants' skill composition.⁴

The demand side has traditionally received less attention.⁵ In line with a number of recent studies on the determinants of individual preferences over immigration policy, we relate individual skills to such preferences using a standard factor-proportion analysis, where immigrants are assumed to compete in the labor market with similarly skilled natives and to complement natives with different skills (see Scheve and Slaughter, 2001; Mayda, 2004; O'Rourke and Sinnott, 2004; Hanson, Scheve and Slaughter, 2005). Differently from this literature, in which immigrants' skill composition is fixed to past levels, we model individual preferences whereby natives realize that such composition is affected by immigration restrictions.

Our paper may contribute also to the relatively small literature on the political economy of immigration.⁶ These works consider only on the determination of quotas, but there may be reasons to explore the role of migration costs as well. First, any restriction to immigration entails, at least indirectly, monetary costs. Second, the exercise is useful even if one considers literally such policy as a tax on immigrants. In fact, historically, the first intervention to limit and select immigration flows in the US and Canada acted on costs rather than on quotas. Monetary incentives were removed, and they were substituted by lengthy procedures and head taxes for admission.⁷ Perhaps more importantly, immigration taxes have recently received attention in policy debates (see for example Legrain, 2007), but to my knowledge their effects have not been explored in a formal model.

liberalization in the sending country in a modified Heckscher-Ohlin model with heterogeneous labor force, migration costs and financing constraints.

⁴ The fact that migration costs can partly be a policy variable is recognized also in Clark, Hatton and Williamson (2002), who assume that lower quotas indirectly imply higher costs for migrants. However, their analysis, similarly to Mayda (2005), is focused on the volume of immigration flows and does not address the relation between policy and skill composition of immigrants.

⁵ For example Borjas (1994) states that "the literature does not yet provide a systematic analysis of the factors that generate the host country's demand function".

⁶ See for example Benhabib (1996), who explores how the median voter determines minimal capital requirements for admission, and Epstein and Nitzan (2005) and Facchini and Willman (2005), who use a lobbying model to explain the formation of immigration quotas.

⁷ Timmer and Williamson (1998) report that the United States, for example, introduced a head tax of 50 cents per migrant in 1882, that was progressively raised to \$8 in 1917. Also, in 1907 they introduced the first financial test, establishing that each individual must have \$25 (or \$50 per family). The same acts extended the classes of "excludable" immigrants, i.e. those who were prohibited to entry because they would have surely become a burden for the hosting society. Passenger acts in the US in the 19th century (fixing minimal standards to carry immigrants) is another policy that indirectly increased migration costs. Canada has also acted on costs in order to control the composition of the immigrant population. In 1870, a travel fund of C\$30 per adult (for Mennonites that agreed to build settlements) was introduced, while in 1910 migration was restricted by a tax of C\$50 per head (C\$200 per head for Asians). The first quota restriction in the United States came in 1921.

2. The model

We consider a world with two countries, a sending and a receiving one. We study the interaction between the workers in the sending country, who may decide to migrate, and the receiving country government, which sets the immigration policy.

2.1. The sending country. The sending country is populated by a mass n^* of heterogeneous workers, called foreigners. Each foreigner is endowed with some skill $\theta \in \{H, L\}$, where n_θ^* denotes the mass of workers with skill θ . A foreigner i with skill θ may work in the sending country for an exogenous wage w_θ^* .⁸ Alternatively, he can migrate and enjoy

$$w_\theta - (\gamma + \varepsilon_i),$$

where w_θ is the endogenous wage in the receiving country, and $(\gamma + \varepsilon_i)$ is the cost of migrating. The latter includes a monetary term γ , which is a common relocating cost, and a psychological cost ε_i , which may be influenced by a number of individual specific characteristics like age, family ties, access to networks at origin and destination countries.⁹ Specifically, ε_i is assumed to be a random variable following a log-concave cumulative distribution Π with continuous density π .¹⁰ This assumption implies that the ratio

$$(2.1) \quad \frac{\pi}{\Pi} \text{ is decreasing.}$$

We also assume that credit constraints may limit migration: immigrants have to incur the cost γ up-front, and the sending country has no credit market for them. Specifically, let each foreigner be endowed with some wealth, drawn by a distribution Ω_θ with continuous density ω_θ . If one interprets θ as an observable skill (like education), then it is natural to assume that the high skilled are on average wealthier than the low skilled (see for example Filmer and Pritchett, 1999 and Piketty, 2000). Formally, this means that the high skilled wealth distribution is more favorable than the low skilled one, in

⁸ We are interested in the determination of the immigration policy in the receiving country, hence we can keep wages in the sending country as exogenous. This need not be a realistic assumption: e.g. Mishra (2003) and Hanson (2005) document how Mexican wages are affected by emigration.

⁹ This interpretation can be justified to the extent that these elements are not systematically correlated with the type θ (see Schwartz, 1973, for a thoughtful discussion of the relation between education and migration costs).

¹⁰ This assumption captures the idea that a lower cost is associated with a higher probability of migration. It implies that the effect of a change in the returns from migration is higher on those with less incentive to migrate. Formally, log-concavity means that those at the tails of the distribution are not too sensitive to a change in the parameters, and a large class of distributions satisfies it (e.g. Uniform, Normal, Lognormal, Weibull, Exponential, Logistic, Laplace, Gamma, Chi-Squared, Pareto, see Bagnoli and Bergstrom, 2005).

the sense of conditional stochastic dominance.¹¹ That is, for every $\gamma \in \mathbb{R}_+$, we have

$$(2.2) \quad \frac{\omega_L}{1 - \Omega_L} \geq \frac{\omega_H}{1 - \Omega_H}.$$

2.2. The receiving country. The receiving country is populated by a mass n of workers, here called natives, who are also heterogeneous in skill $\theta \in \{H, L\}$. Let w_H and w_L denote the equilibrium wages in this country. We assume that such wages decrease with the amount of workers with the same skill, and increase in the amount of workers with different skills in the country. As such, equilibrium wages may be influenced by immigration. Specifically, let the production technology in this country be $F(N_H, N_L)$, where N_θ is the total mass of workers with skill θ , i.e. the sum of natives n_θ and immigrants x_θ with skill θ . We assume that the production function has constant returns to scale, the labor market is competitive, so that

$$(2.3) \quad w_\theta = F_\theta = \frac{\partial F(N_H, N_L)}{\partial N_\theta},$$

and that

$$(2.4) \quad F_{HH}, F_{LL} < 0 \text{ and } F_{HL} > 0.$$

In particular, for most of the analysis we simply let the production function be a Cobb-Douglas, that is

$$(2.5) \quad F(N_H, N_L) = N_H^\alpha N_L^{1-\alpha}.$$

We also assume that natives are only concerned about their wages, i.e. their utility is

$$U_\theta = w_\theta.$$

Accordingly, high skilled natives support any policy that increases the supply of low skilled workers, while low skilled natives support the opposite. The receiving country government is then interested in regulating the inflows of immigrants as these influence natives' utility. Its goal is to maximize the welfare function

$$W = \mu_H U_H + \mu_L U_L,$$

where μ_θ denotes the weight attached to group θ 's utility, as determined by the specific institutional setting.¹² Immigration policy acts on γ , which is the cost foreigners have

¹¹ Hazard rate, or conditional dominance is slightly stronger than first order stochastic dominance, but weaker than the standard assumption of monotone likelihood ratio (see Krishna, 2002, Appendix B).

¹² We will mostly consider a utilitarian function with $\mu_\theta = N_\theta$ or $\mu_\theta = n_\theta$. Alternatively, one could think of a majoritarian democracy where only the largest group of natives gets positive weight; lobbying or other political economy models (see Bianchi, 2006, for a discussion of how increasing immigration restrictions may be seen as the result of an increase in μ_L).

to incur to enter and work in the receiving country, so we write the government's program as

$$(2.6) \quad \max_{\gamma \in \mathbb{R}_+} \mu_H w_H(\gamma) + \mu_L w_L(\gamma).$$

Remarks.

The evidence supporting equation (2.4), i.e. immigrants are perfect substitutes for similarly skilled natives, is quite controversial. Several studies find a rather small impact of immigration on natives' wages (see Borjas, 1994; Friedberg and Hunt, 1995; and Card, 2005), while others (e.g. Borjas, 2003) document that immigrants compete with similarly skilled natives and significantly lower their equilibrium wages. For our purposes, it suffices that citizens' *beliefs*, rather than actual effects, are of this kind. Considering the centrality of the issue in past and current policy debates, and the evidence on attitudes towards immigration discussed in Section 4.3, this seems a much less controversial assumption. Some complications on the functioning of the labor market and on natives' preferences are considered in Section 4.3.

The government's program considered in equation (2.6) emphasizes that migration costs are partly endogenous, i.e. they can be significantly influenced by the receiving country policies.¹³ For example, governments impose direct fees, bureaucracies and other time consuming requirements that increase foregone earnings, or the expenses for consulting and legal services. This policy component need not be though only as an entry tax, and as such, we do not include its potential revenues in the welfare function. In fact, a significant part of migration costs depends on immigration restrictions, but they are not pocketed by the receiving country government. Further discussions on the policy space are in Section 4.2.

3. Analysis

We now show that, in order to set the optimal policy, the receiving country government has to predict the effects on immigrants' skill composition. This in turn requires an understanding of the forces driving the decision to migrate, i.e. of immigrants' self-selection.

3.1. The migration decision. A foreigner i with skill θ prefers to migrate if $w_\theta - (\gamma + \varepsilon_i) \geq w_\theta^*$, hence for each skill θ there exists a cut-off value $\varepsilon^\theta \equiv w_\theta - w_\theta^* - \gamma$ such that any individual with skill θ and a cost ε_i lower than ε^θ would like to migrate.

¹³ Of course, migration costs depend also on exogenous variables, like the distance between the two countries, and on partly exogenous ones, like transportation costs. However, notice that policies may become more and more relevant, given the historical trend of decreasing transportation costs and increasing immigration restrictions (Hatton and Williamson, *forthcoming*).

In addition, this individual must be sufficiently wealthy to incur the migration cost γ . Thus, the supply of migrants with skill θ is defined by

$$(3.1) \quad x_\theta = q_\theta n_\theta^*,$$

where

$$(3.2) \quad q_\theta = [1 - \Omega_\theta(\gamma)]\Pi[w_\theta - w_\theta^* - \gamma].$$

The term $\Pi[w_\theta - w_\theta^* - \gamma]$ represents those who have incentive to move, while the fraction $[1 - \Omega_\theta(\gamma)]$ are those who can afford to move. We define immigrants' skill composition as the ratio of high to low skilled migrants, i.e.

$$(3.3) \quad Q = \frac{q_H}{q_L}.$$

3.2. Optimal immigration restrictions. The receiving country government is interested in governing immigration flows as immigrants may change the supply of skills in the receiving country and then natives' equilibrium wages. Recall that the labor market is competitive and the production function is a Cobb-Douglas. Hence, as described in equations (2.3) and (2.5), equilibrium wages can be written as

$$w_H = \alpha R^{\alpha-1},$$

and

$$w_L = (1 - \alpha)R^\alpha,$$

where R is the ratio of high to low skilled workers N_H/N_L . This ratio varies with migration flows, and hence with the immigration policy γ , according to the equation

$$(3.4) \quad R = \frac{n_H + x_H}{n_L + x_L}.$$

We can then write the government's program in equation (2.6) as

$$(3.5) \quad \max_{\gamma \in \mathbb{R}_+} \mu_H \alpha R^{\alpha-1} + \mu_L (1 - \alpha) R^\alpha.$$

Obviously, the government preferred policy depends on the weights μ_θ . For now, we abstract from redistributive concerns or other political economy distortions, and consider a purely utilitarian setting in which each group is valued according to its size. In this case, if immigrants had the same influence as natives, then no immigration restrictions would be imposed. In fact, if

$$\mu_\theta = N_\theta,$$

then equation (3.5) is maximized when $\gamma = 0$, as the government is just interested in maximizing the total output. If instead the government cares only about natives, then

$$\mu_\theta = n_\theta,$$

and we have that

$$(3.6) \quad \frac{dW}{d\gamma} = \frac{dR}{d\gamma} \left[\frac{n_L x_H - n_H x_L}{n_H + x_H} \right].$$

In this case, the function W is convex and has a minimum at $R = n_H/n_L$. Efficiency gains from immigration are minimized when immigrants have the same skill composition as the native population, i.e. when $x_H/x_L = n_H/n_L$.¹⁴ Since the government maximizes efficiency, it aims at optimizing the skill ratio R . In particular, if immigrants are less skilled than natives, the optimal policy is the one preferred by high skilled natives, that is the one maximizing the ratio R ; and vice versa if immigrants are more skilled than natives.¹⁵

As it turns out, predicting how the policy affects the skill ratio R requires an understanding of the forces driving the migration flows. First, notice that

$$\frac{dR}{d\gamma} = \frac{1}{N_L^2} \left(N_L \frac{dx_H}{d\gamma} - N_H \frac{dx_L}{d\gamma} \right),$$

where

$$\frac{dx_\theta}{d\gamma} = \frac{\partial x_\theta}{\partial \gamma} + \frac{\partial x_\theta}{\partial w_\theta} \frac{\partial w_\theta}{\partial R} \frac{dR}{d\gamma}.$$

Hence, we have that

$$\frac{dR}{d\gamma} \left[N_L^2 - N_L \frac{\partial x_H}{\partial w_H} \frac{\partial w_H}{\partial R} + N_H \frac{\partial x_L}{\partial w_L} \frac{\partial w_L}{\partial R} \right] = \frac{\partial x_H}{\partial \gamma} N_L - \frac{\partial x_L}{\partial \gamma} N_H,$$

and, given that $\partial w_H/\partial R < 0$ and $\partial w_L/\partial R > 0$, the ratio R increases in γ if and only if

$$(3.7) \quad \frac{\partial x_H}{\partial \gamma} N_L \geq \frac{\partial x_L}{\partial \gamma} N_H.$$

Notice that $\partial x_\theta/\partial \gamma$ are *partial* derivatives, i.e. they describe the direct effect of immigration policy on immigration flows, abstracting from the effect on equilibrium wages. In order to highlight the role of self-selection, notice that, multiplying both sides of equation (3.7) by $x_L x_H$ and rearranging, the ratio R increases in γ if and only if

$$(3.8) \quad \frac{\partial x_H}{\partial \gamma} x_L \cdot \frac{x_H}{n_H + x_H} \geq \frac{\partial x_L}{\partial \gamma} x_H \cdot \frac{x_L}{n_L + x_L}.$$

Equation (3.8) can be decomposed as the product of two forces:

$$\frac{x_H}{n_H + x_H} \geq \frac{x_L}{n_L + x_L},$$

¹⁴ The result resembles a well known principle in international trade, where gains from trade are higher the greater the trading countries differ in their factor endowments. A similar point, in a more complicated setting, is made by Borjas (1995).

¹⁵ The intuition follows from the fact that the government is insensitive to the effects of the policy on immigrants, hence it maximizes the benefit for the group of workers where the proportion of natives is larger.

which is equivalent to

$$(3.9) \quad \frac{x_H}{x_L} \leq \frac{n_H}{n_L},$$

and

$$\frac{\partial x_H}{\partial \gamma} x_L \geq \frac{\partial x_L}{\partial \gamma} x_H,$$

which is equivalent to

$$(3.10) \quad \frac{\partial Q}{\partial \gamma} \geq 0.$$

Equation (3.9) describes a *size effect*, i.e. what happens to the skill ratio R when one admits more or less immigrants, while keeping their skill composition as fixed. If immigration restrictions had no other effect, then welfare would be maximized with free immigration. In fact, equation (3.9) tells that increasing the cost increases the ratio R if and only if immigrants are less skilled than natives. According to equation (3.6), we would then have

$$\frac{\partial W}{\partial \gamma} \leq 0.$$

However, any immigration policy also changes the average skill of immigrants, as described by equation (3.10). This represents a *composition effect*: higher restrictions increase the skill ratio R if and only if they increase immigrants' skill composition Q . We summarize these findings in the following Propositions.

PROPOSITION 9. *If immigrants' skill composition was taken as given, a utilitarian government would impose no immigration restrictions.*

PROPOSITION 10. *Any immigration policy affects not only the size but also the skill composition of the migration flow.*

3.3. The composition effect. Standard discussions about immigration policies abstract from the composition effect. However, as we now show, such abstraction may be misleading, since this effect may sometimes *reverse* the predictions based on the size effect only. In fact, there are situations where the tension between the two effects is inescapable, i.e.

$$(3.11) \quad \frac{\partial Q}{\partial \gamma} \geq 0 \text{ if and only if } \frac{x_H}{x_L} \geq \frac{n_H}{n_L},$$

and, in addition, the composition effect may be stronger.

To see this, it is useful to first rewrite condition (3.7) as

$$(3.12) \quad \frac{\partial q_H}{\partial \gamma} n_H^* n_L - \frac{\partial q_L}{\partial \gamma} n_L^* n_H + \frac{\partial q_H}{\partial \gamma} q_L n_H^* n_L^* - \frac{\partial q_L}{\partial \gamma} q_H n_L^* n_H^* \geq 0.$$

The composition effect is less likely to be an issue if the skill compositions of the two countries are very different. Suppose for example that the sending country has a very poor skill composition. All else equals, a more restrictive policy is likely to have a larger impact on low skilled foreigners, and so increase the ratio R . This effect being clear, we now consider the case where the skill composition between the sending and the receiving country is similar, and concentrate on selection issues. In particular, we let

$$(3.13) \quad n_H^* = n_H \text{ and } n_L^* = n_L,$$

and so we write condition (3.12) as

$$(3.14) \quad \frac{\partial q_H}{\partial \gamma} - \frac{\partial q_L}{\partial \gamma} + \frac{\partial q_H}{\partial \gamma} q_L - \frac{\partial q_L}{\partial \gamma} q_H \geq 0.$$

The last equation makes it clear that in order to predict the effect of the policy, it is necessary to know the response of low and high skilled foreigners. This requires an understanding of the forces behind immigrants' self-selection, as we now consider.

3.3.1. The simplest case: no wealth constraints. To illustrate our argument in the cleanest way, we first abstract from wealth constraints. Besides being simple, this way of modeling the migration decision is very similar to classical self-selection models like Borjas (1987). In this case, immigrants' skill composition in equation (3.3) writes simply as

$$Q = \frac{\Pi(w_H - w_H^* - \gamma)}{\Pi(w_L - w_L^* - \gamma)}.$$

Immigrants' self-selection is driven only by the incentives that foreigners face according to their skills. These incentives are greater for the high skilled if and only if the wage differential in the receiving country is higher than in the sending one, in which case immigration restrictions increase immigrants' skill composition. That is, denoting wage differentials as $\Delta w^* = w_H^* - w_L^*$ and $\Delta w = w_H - w_L$, we write that

$$(3.15) \quad \frac{\partial Q}{\partial \gamma} \geq 0 \iff \frac{\pi(\varepsilon^L)\Pi(\varepsilon^H) - \pi(\varepsilon^H)\Pi(\varepsilon^L)}{\Pi(\varepsilon^H)\Pi(\varepsilon^L)} \geq 0 \iff \Delta w \geq \Delta w^*.$$

Equation (3.15) describes an *incentive effect*. As expressed in condition (2.1), changing costs has a relatively higher impact on the people with lower gain from migration. If $\Delta w \geq \Delta w^*$, the low skilled are on average those with less incentives to migrate, thus they are the one more sensitive to an increase in costs. As a result, restrictions and skill composition are positively related.

To see how this effect matters for the receiving country, we simplify further our analysis by assuming that the psychological cost of migration is uniformly distributed

over the interval $[a, b]$. For $\varepsilon^\theta \in [a, b]$, $\pi(\varepsilon^L) = \pi(\varepsilon^H)$ and, substituting into equation (3.14), we see that

$$(3.16) \quad \frac{\partial R}{\partial \gamma} \geq 0 \iff \Pi(\varepsilon^H) - \Pi(\varepsilon^L) \geq 0 \iff \Delta w \geq \Delta w^*.$$

When instead one of the thresholds ε^L and ε^H lies outside the interval $[a, b]$, the sign of the derivative is reversed. Suppose for example that $\Delta w \leq \Delta w^*$ for all γ . Equation (3.16) tells that, despite immigrants are less skilled than natives, increasing immigration restrictions decreases the skill ratio in the receiving country. This is because the skill composition of the new immigrants gets even lower when restrictions become tighter. This is true as long as the cost is not too high, so that some high skilled have still incentive to move. Such threshold is implicitly defined by

$$\gamma \leq \bar{\gamma} \equiv w_H - w_H^* - a.$$

Beyond the cost $\bar{\gamma}$, only low skilled foreigners have incentive to migrate, so further restrictions just decrease the number of immigrants, without affecting their composition, and so the skill ratio R increases. Thus, in this case, the relation between R and γ is U-shaped, with a minimum at $\gamma = \bar{\gamma}$.

Even in such simple setting, we gain some fundamental insights on the relation between size and composition effects. First, the composition effect pushes the skill ratio in the opposite direction than the standard size effect. The reason is intuitive: the size effect is by definition random, so it hits a group of foreigners proportionally to their propensity to migrate, while the composition effect tends to be stronger on the least represented group. Second, the relation between the skill ratio, and hence natives' welfare, and immigration restrictions is non-monotonic. When restrictions are low, the composition effect is stronger, which has a number of counterintuitive implications. First, some natives may support further immigration restrictions even if immigrants are not harmful for them. For example, low skilled natives may be against immigration even if immigrants complement them, since further restrictions would improve immigrants' skill composition (see Section 4.3). Second, even a utilitarian government with no redistributive or political economy concerns may impose positive immigration restrictions in order to maximize or minimize the skill ratio (see Section 4.2). We summarize these findings with the following Proposition and Corollaries.

PROPOSITION 11. *When migration is driven only by incentives,*

- a):** *Immigrants are positively self selected if and only if $\Delta w \geq \Delta w^*$;*
- b):** *Immigration restrictions increase Q if and only if $\Delta w \geq \Delta w^*$;*
- c):** *Under condition (3.13), size and composition effects have the opposite direction;*

d): *The relation between R and γ may be non-monotonous. In particular, R is first increasing and then decreasing in γ when $\Delta w \geq \Delta w^*$; while the opposite occurs when $\Delta w < \Delta w^*$.*

COROLLARY 6. *When the composition effect prevails, some natives may support further restrictions even if immigrants are not harmful for them.*

COROLLARY 7. *Immigration restrictions are optimal even for a utilitarian government that cares only about natives' utility.*

3.3.2. *The general case: incentive and wealth effects.* We now explore how the previous insights carry through in a setting where potential migrants face also wealth constraints, and so self-selection is driven also by different ability to incur the migration cost. For our purposes, this implies that it may not be sufficient to know whether immigrants are positively or negatively self-selected, but one need to know also what drives self-selection. Those with the highest gain from migration, and hence the highest willingness to pay for it, are not necessarily the ones with the highest resources to pay for it.

Besides being a more general formulation of the migration decision, this setting matches better with the empirical evidence. As implied by equation (2.2), credit constraints are less severe for the high skilled. Hence, we have that

$$[1 - \Omega_H(\gamma)] \geq [1 - \Omega_L(\gamma)],$$

which tends to push towards positive self-selection in terms of observables. This implies that, immigrants may be positively self-selected even if returns to skills are higher in the source country, as recently documented in a number of studies.¹⁶

In this setting, we first notice that increasing immigration restrictions improves immigrant skill composition Q when

$$(3.17) \quad \frac{\pi(\varepsilon^L)\Pi(\varepsilon^H) - \pi(\varepsilon^H)\Pi(\varepsilon^L)}{\Pi(\varepsilon^H)\Pi(\varepsilon^L)} + \frac{\omega_L(1 - \Omega_H) - \omega_H(1 - \Omega_L)}{(1 - \Omega_H)(1 - \Omega_L)} \geq 0.$$

The first term is the same *incentive effect* described in the previous Section. The second term represents a *wealth effect*. By equation (2.2), this is always positive: increasing the

¹⁶ For example, Chiquiar and Hanson (2005) find that Mexican migrants to the US are positively self-selected despite returns to skills are higher in Mexico. This can be interpreted recognizing the role of migration costs in shaping immigrants' self-selection. The same is true for the more aggregate evidence in Hatton and Williamson (2004) and Brucker and Defoort (2006), where the gap in years of schooling between movers and stayers is positive; it increases with the distance between source and destination country and it decreases with source country per capita GDP.

cost one gets richer and more skilled migrants, and the strength of this term increases with the level of restrictions and with wealth inequality in the sending country.¹⁷

It is then clear that when only wealth constraints matter, migrants skill composition increases with migration costs, while if migration is driven mostly by incentives the same occurs if $\Delta w \geq \Delta w^*$. When instead $\Delta w < \Delta w^*$, the effect is ambiguous. For low levels of cost, the relation is negative, since the wealth effect is weak and incentives dominate. The shape of Q as costs increase depends on the strength of the two effects. Roughly, when $\Pi(\varepsilon^H)$ goes to zero faster than $(1 - \Omega_L)$, Q tends to zero as γ increases, since at some point a few high skilled are willing to migrate. When the opposite occurs, there exists a cost beyond which the wealth effect takes over, so the relation is U-shaped.

In this setting, size and composition effects need not go in opposite directions. It may be for example that an increase in restriction increases immigrant composition, and hence possibly the skill ratio in the receiving country, even if immigrants are negatively self-selected, since higher restrictions tend to reinforce the wealth effect. However, we are interested in the case where condition (3.11) holds, so there is a tension between size and composition effects. This happens whenever the relation between Q and γ is monotone, i.e. either $\Delta w \geq \Delta w^*$ or self-selection is driven only by wealth or by incentives. Given equations (3.11) and (3.13), a *sufficient* condition for the composition effect to prevail is that

$$(3.18) \quad x_H \geq x_L \text{ if and only if } \frac{\partial q_H}{\partial \gamma} \geq \frac{\partial q_L}{\partial \gamma}.$$

That is, R increases with γ despite immigrants are more skilled than natives whenever the least represented group is, in absolute terms, the most sensitive to a policy change.^{18,19} Moreover, as long as the flow of immigrants in the receiving country is relatively small, i.e. q_L and q_H are small, the condition is almost necessary.

Moreover, a *necessary* condition for the composition effect to prevail is that

$$(3.19) \quad \gamma \leq \gamma^{\max},$$

¹⁷ While we focus only on labor market interactions, immigrants' self-selection is important also for issues like the impact of migration on the source country, or assimilation, discrimination, crime, in the receiving country (see e.g. Butcher and Morrison Piehl, 2005). Bianchi (2006) provides some additional comparative statics on the relation between self-selection and the characteristics of the source country.

¹⁸ Notice that these derivatives are negative, hence the lower the derivative the more sensitive is a group of foreigners.

¹⁹ While not directly considered in the model, an example may be when immigrants are mostly low skilled and illegal. Immigrants depress R but immigration restrictions are likely to worsen the situation by discouraging high skilled migrants, without affecting low skilled (illegal) ones.

where the maximum cost depends on the functional forms Π and Ω_θ one assumes (see Bianchi, 2006, for some examples). Roughly, a low cost is needed to have a sufficiently heterogeneous population of migrants who responds to policy changes. If the cost is so high that only one group of foreigners migrates, being they the richest or the most motivated, then by definition there is no composition effect. Suppose for example migration is driven only by wealth constraints, and the cost is such that only high skilled can afford it. Immigrants are more skilled than natives, and Q increases in γ , so size and composition effects have opposite directions. However, at this point, the composition effect is very weak: a further increase in γ only prevents high skilled to migrate, and so it decreases R . In other words, as q_L and $\partial q_L/\partial\gamma$ are almost zero, equation (3.14) reduces to $\partial q_H/\partial\gamma$, which is negative.

It follows from equation (3.19) that the relation between R and γ need not be monotone. The composition effect may dominate for any $\gamma < \gamma^{\max}$, and the size effect dominates afterwards. Hence, natives may support a more restrictive policy despite immigrants are beneficial for them, up to the point in which the skill ratio R is maximized or minimized. Equation (3.19) also implies that the relation between natives' welfare and immigration restrictions need not be monotone. Hence, a utilitarian government, which solves the program in equation (3.5) with $\mu_\theta = n_\theta$, may optimally impose positive immigration costs. In this case, such restrictions are not due to distributional concerns, or other departures from pure efficiency, but they are a way to screen immigrants by affecting their self-selection. We summarize with the following Propositions.

PROPOSITION 12. *When immigration is driven only by wealth constraints, or when $\Delta w \geq \Delta w^*$, immigrants are positively self-selected. The statements (b)-(d) in Proposition 11 and Corollaries 6 and 7 still hold.*

PROPOSITION 13. *When $\Delta w < \Delta w^*$, the relation between Q and γ is ambiguous. Predicting the policy outcome in the receiving country requires an understanding of what drives immigrants' self-selection.*

PROPOSITION 14. *In general, the composition effect is stronger when those who tend to migrate less are the most sensitive to a change in immigration policy.*

4. Discussion and extensions

4.1. Migration costs and immigration policies. Taken literally, our model makes some important simplifications. First, migration costs are essentially money, and second, immigration restrictions only act on costs, unconditionally on skills. Since

these assumptions are obviously not realistic, we now see to what extent they can be relaxed without affecting the analysis.

On the first point, what migrants have to forsake is definitely not only money, and it is not even clear that these are always the most significant component of migration costs (see Massey, Arango, Hugo, Kouaouci, Pellegrino and Taylor, 1993). Consider for example insurance motives, which are typically more salient for low skilled, and so they may drive self-selection. Their effect, however, is ambiguous. Some literature considers these motives to explain positive self-selection: low skilled cannot migrate and give up the support of their family or community, in terms of access to credit (e.g. Banerjee and Newman, 1998; and Munshi and Rosenzweig, 2005) or unemployment insurance (e.g. Cuecuecha, 2005). Another stream of literature uses similar arguments to support negative self-selection: since low skilled cannot get formal insurance at home, they migrate and send remittances in order to smooth family consumption (e.g. Stark and Bloom, 1985; Taylor, 1986). Hence, at this level of generality, the lack of explicit formalization of network effects may not be a fundamental limitation for our analysis.

On the second point, the immigration policy space is typically more multifaceted than in our model. Receiving countries can act on many dimensions beside the monetary cost γ , and they can impose different restrictions on different types of immigrants. Of course, if the receiving country could perfectly contract on immigrant skills, it would directly select the desired size and type of immigration and the interaction between policy and skill composition would be trivial. However, this does not seem the case. There are many instances in which direct screening mechanisms are difficult to design and to implement. Moreover, many authors have stressed that they do not appear so powerful.²⁰ In such world, our formalization is on the one hand a way to highlight that even a policy independent on type has some screening power. On the other hand, it may be a starting point to complicate the policy space. For example, even if skills are not perfectly contractible, one can think of indirect ways to affect immigrant composition. For example, a country offering generous welfare benefits may attract lower skilled immigrants, as these benefits decrease the returns to skills. Alternatively, as considered below, a visa offered for a high fee and no bureaucracies is likely to attract a different type of applicants than one requiring no fees but many red tapes.

4.1.1. *Time and money.* Migration costs include also the time immigrants have to spend in filling out forms, queuing, and waiting for documents, and this time may be thought as foregone earnings and so as a monetary term. Moreover, since the value of time may differ according to skills, these procedures may affect self-selection. To keep

²⁰ See e.g. Jasso and Rosenzweig (1995) and Jasso, Rosenzweig and Smith (1998) on US immigrants, Miller (1999) on the Australian point system, Antecol, Cobb-Clark and Trejo (2003) and Jasso and Rosenzweig (2005) on Canada and Australia vs. the US.

the analysis simple, assume that each migrant has to invest some fixed amount of time β in bureaucracies, and this time is worth βw_θ^* . Hence, in absolute terms, bureaucracies are more harmful for high skilled.²¹ The average skills of migrants is then

$$Q^\beta = \frac{[1 - \Omega_H]}{[1 - \Omega_L]} \cdot \frac{\Pi[(w_L - w_L^*) + (\Delta w - \Delta w^*) - (\gamma + \beta w_H^*)]}{\Pi[w_L - w_L^* - \gamma - \beta w_L^*]}.$$

Notice first that the conditions for positive self-selection become harder to satisfy. When only incentives matter, we now need that $\Delta w > (1 + \beta)\Delta w^*$, i.e. differential returns to skill in the receiving country are sufficiently high to compensate also for the loss of time.

Increasing the time spent in bureaucracies affects immigrants' skill composition according to the relation

$$(4.1) \quad \frac{\partial Q^\beta}{\partial \beta} \geq 0 \Leftrightarrow w_L^* \pi(\varepsilon^L) \Pi(\varepsilon^H) - w_H^* \pi(\varepsilon^H) \Pi(\varepsilon^L) \geq 0.$$

Once again, without further assumptions, the sign of this relation is ambiguous. Roughly, we need the relative gain to be larger for high skilled than for low skilled, while in the basic framework we had the same condition with absolute gains.²² With respect to the case of no bureaucracies, however, we can say that increasing the cost γ is now more likely to reduce migrants' skill composition. In this setting, it is necessary but no longer sufficient that migrants are positively self-selected in terms of incentives in order to get a positive relation between cost and skill composition.

4.2. Returns to Skills. We have seen that differential returns to skills in sending vs. destination countries are a central determinant of self-selection and of the relation between immigration restrictions and immigrants' skill composition. However, we have not assumed any general pattern on such differentials. In fact, as we now argue, any simple generalization is likely to be misleading.

On the theoretical side, with competitive labor markets, everything depends on the production functions one assumes for the two countries. Developed country should have higher returns to skills if one focuses on Total Factor Productivity (Lucas, 1990; Hall and Jones, 1999; Caselli, 2004), skill complementarities (Kremer, 1993), or skill biased technological change (Acemoglu, 1998; Caselli and Coleman, 2005). On the other hand, if one emphasizes that skills are scarce in the South and abundant in the North, or that some labor market institutions compress wages in advanced economies,

²¹ One could think that skills also affect efficiency in complying with bureaucracy: if the value of foregone earning is then smaller for high skilled, the analysis in the previous sections applies (see Chiswick, 1999).

²² To see this, let ε_i be uniformly distributed. Condition (4.1) requires that relative gains increase with skills, i.e. $w_L^*(w_H - w_H^* - \gamma) > w_H^*(w_L - w_L^* - \gamma)$.

then developing countries should have higher returns to skills (Blau and Kahn, 1996; and Leuven, Oosterbeek and van Ophem, 2004).

Going to the data, as a first approximation, returns to skills decrease with per capita GDP (Bils and Klenow, 2000; Freeman and Oostendorp, 2000; Psacharopoulos and Patrinos, 2002; Caselli and Coleman, 2005), which would imply that real gains from migration tend to be higher for low skilled. However, these differences are not huge, and general patterns appear weak.²³

In the immigration literature, accordingly, there is no consensus. Various models simply assume that a worker with skill s in country j gets a wage $k_j \cdot s$. Hence, by construction, returns to skills are higher in more developed countries, and South-North immigrants should be positively self-selected (Chiswick, 1999; Giannetti, 2003; Jasso and Rosenzweig, 2005). However, this formalization bypasses the issue of wage inequality, which has been stressed in several studies, following Borjas (1987). Indeed, these studies conclude that low skilled may have the greatest incentives to migrate.

Instead of looking for a general pattern, it appears that sensible insights may be derived from specific microanalysis. Chiquiar and Hanson (2005), for example, look at earnings by skills of Mexican resident vs. Mexican immigrants, rather than Mexican resident vs. US residents. Building counterfactual wages, they estimate that real wage premia decrease, in absolute terms, with education.²⁴ Similar estimates can be found in the analysis of Palestinian immigrants to Israel (Yashiv, 2004). These studies confirm our general point that self-selection is in general driven both by incentives and by constraints: those who can access migration are not necessarily those who have more to gain from it.

4.2.1. *Discrimination.* A second dimension of returns to skills concerns the mapping from skills to jobs. Even for given skills, immigrants need not access the same spectrum of jobs and wages as natives, hence the wage gap *per se* may not be fully informative. For example, there are many instances in which immigrants are locked into traditional low skilled occupations (see Munshi, 2003). Hence, whether accessing well-paid jobs requires personal connections or skills is an equally important determinant of self-selection.²⁵

²³ Banerjee and Duffo (2004) argue that the common wisdom that returns to skills are higher in developing country is an artifact of low quality data.

²⁴ They report that "real U.S. wage premium is \$4.07 per hour for an individual with 5-8 years of schooling, \$3.52 for an individual with 12 years of education, and \$2.60 for an individual with 16 or more years of education".

²⁵ To see this, assume that in the sending country high skilled have probability p_H^* to get a good job (and $1 - p_H^*$ to get a bad one). Their expected wage is $E(w_H^*) = p_H^* w_H^* + (1 - p_H^*) w_L^*$. In the same way, low skilled expect $E(w_L^*) = p_L^* w_H^* + (1 - p_L^*) w_L^*$, with $p_H^* \geq p_L^*$. The country is meritocratic the more $p_H^* \rightarrow 1$ and $p_L^* \rightarrow 0$. Similarly for the destination country. Now the requirement that returns to skills are higher in the receiving country means $E(w_H) - E(w_L) > E(w_H^*) - E(w_L^*)$, i.e.

These issues are difficult to measure and to compare across countries, and indeed they have been typically overlooked in this literature. However, they may be quite important for our purposes. If good jobs are harder to get for immigrants, incentives induce negative self-selection, and labor market competition hurts only low skilled natives. To see this, assume that barriers to entry are more severe for well-paid jobs, hence high skilled immigrants are (partially) prevented to access them. This may come from immigrants' inability to assimilate (language, country-specific skills) or from a discriminatory labor market. In our model, we can define a measure of the relation between immigrants' skill and wages as $\tau \in [0, 1]$. High skilled immigrants expect to earn $w_L + \tau\Delta w$, where $\tau = 0$ corresponds to full discrimination and $\tau = 1$ to full integration.²⁶ The skill composition of migrants is defined by:

$$Q^\tau = \frac{[1 - \Omega_H(\gamma)] \Pi[(w_L + \tau\Delta w) - (w_L^* + \Delta w^*) - \gamma]}{[1 - \Omega_L(\gamma)] \Pi[w_L - w_L^* - \gamma]},$$

while the resulting skill ratio is:

$$R^\tau = \frac{n_H + \tau x_H}{n_L + x_L + (1 - \tau)x_H}.$$

Consider for example the case of $\tau = 0$, i.e. immigrants are treated as an homogeneous group of low skilled and there is no opportunity for them to access good jobs. High skilled foreigners have now less incentive to migrate and migrants' skill composition decreases. Moreover, discrimination creates a negative spillover on the low skilled market, which is bad news for all foreigners and it decreases the migration flow. High skilled natives thus always gain from immigration, since immigrants are prevented to compete with them; while low skilled natives bear all the costs. Thus, labor market discrimination induces a positive correlation between natives' skills and pro-immigration attitudes (see Section 4.3).

4.3. Immigration policy preferences. Our results on the composition effect may have interesting implications also for the recent literature on individual preferences towards immigration. This literature documents that, in developed countries, there is a strong positive correlation between individual education and the support for a more liberal immigration policy. A number of studies have interpreted these findings stressing economic motivations, and in particular labor market competition (e.g.

$(p_H - p_L)\Delta w > (p_H^* - p_L^*)\Delta w^*$. For example, given that in Mexico wage inequality is higher but social mobility is lower than in the US (as reported e.g. by Dahan and Gaviria, 2001), who should be more likely to migrate?

²⁶ These effects, to my knowledge, have not been explored in a formal model before (as recognized by Borjas, 1994, footnote 30). Moreover discrimination can be thought partly as a policy variable (e.g. anti discrimination laws, recognition of foreign qualifications...), thus one may replicate our analysis substituting γ with τ .

Scheve and Slaughter, 2001; Mayda, 2004; O'Rourke and Sinnott, 2004; Hanson, Scheve and Slaughter, 2005). In support of this interpretation, they report that once one restricts the sample to people out of the labor force, the correlation between education and pro-immigration preferences disappears.²⁷

Our model is in line with these studies in that preferences over immigration policy are fully described by the effects on the skill ratio R , which determines equilibrium wages. However, as we now argue, focusing on the effects on R may be a useful way to describe a number of phenomena that we have abstracted from, like fiscal policy and political economy. On fiscal policy, one may argue that high skilled immigrants are always preferred since they pay higher taxes and receive less welfare benefits. Hence, high skilled natives would trade-off the reduction in wages with the fiscal benefit of accepting high skilled immigrants.²⁸ Only a few empirical studies have looked at this effect on preferences. In support of the fact that labor market concerns are of first order, Hanson, Scheve and Slaughter (2005) documents that, across U.S. states, high skilled are less pro-immigration in areas with high skilled immigrants.

On political economy issues, if immigrants gain political power in the receiving country, then natives may trade-off the effect on their wages with the one on the political equilibrium (like in Ortega, 2005). However, in our setting, the sign of this effect depends on whether immigrants would vote according to their skills, hence protecting their wages, or they would remain "loyal to their roots" and oppose restrictions to immigration anyway. Lowell, Bean and de la Garza (1986) and Goldin (1994) report that immigrants lobbied and voted for pro-immigration policies, and a number of survey studies (e.g. Espenshade and Hempstead, 1996 and Scheve and Slaughter, 2001) report that immigrant have more favorable attitudes towards immigration, irrespective of their economic condition. Thus, it seems that accounting for this long run political economy effect would strengthen the standard result, i.e. low skilled would oppose immigration even more.

5. Conclusion

The paper has developed a simple framework for analyzing the interaction between immigrants' self-selection and the determination of immigration policy. This approach has uncovered some new and possibly important mechanisms shaping the migration

²⁷ Other studies instead focus on factors like racism, anxiety, social and political alienation and other cultural values and beliefs (e.g. Espenshade and Hempstead, 1996; Citrin, Green, Muste and Wong 1997; Hainmueller and Hiscox, 2004).

²⁸ The simplest way to introduce fiscal issues here is to think that the government collects tw_θ and distributes the revenues with a lump sum transfer to every worker. Now high skilled utility is a convex combination (with weight t) of w_θ , which depends negatively on R , and the transfers, which depend positively on R .

decision, the preferences over immigration policy in the receiving country, and the corresponding policy design. With a large set of *caveat* in mind, one may use these results to draw some policy implications. A prime motivation for this discussion is that the skill composition of immigrants matters: while the literature on the impact of immigration is divided in almost any respect, a consensus is that, considering the effect both on the labor market and on fiscal spending, high skilled immigrants create net benefits for host countries.²⁹ Hence, improving the ability to screen would represent a clear gain for receiving countries.

As we have shown, migrants' self-selection implies that any policy, even if independent on type, affects different migrants in a different way, and so it has some indirect screening power. This can limit the effectiveness of the policy, but it may also be viewed as an additional dimension to exploit. In fact, since as we argued the efficiency of direct screening mechanisms is sometimes limited, an alternative route is to act on costs. This would influence the migration decision, so it would act on self-selection *ex-ante* rather than imposing restrictions *ex-post*. The model should not be viewed as delivering absolute policy prescriptions in this respect. Instead, we have stressed that things may change dramatically depending on whether migration is driven by constraints or incentives, and in general on the source country characteristics. If those who migrate are simply those who can afford it, issues like labor market discrimination or red tapes may not have a crucial effect on immigrants' skill composition. Increasing the migration cost, e.g. through a head tax on entry, is likely to increase immigrants' skill composition.³⁰ Instead, as economic incentives become the main argument of the migration decision, the effect of these policies depends on differential returns to skills, and thus it may be more difficult to predict. Discrimination and bureaucracies push towards negative self-selection and, in this case, a more restrictive policy is likely to lead to an even less skilled immigration.

The most general conclusion of our exploration is that self-selection matters, also for receiving countries. The forces shaping self-selection affect the way different potential migrants respond to policy changes, which is obviously a central element to consider when thinking about immigration policy. Nothing is terribly surprising in this statement. There is a huge and fundamental literature explaining how different agents

²⁹ See e.g. Borjas (1995), Lee and Miller (2000), Storesletten (2000), Chojnicki, Docquier and Ragot (2005), Chiswick (2005).

³⁰ Of course, one concern is that restrictions tend to encourage illegal immigration (that is more attractive for low skilled). However, the issue is common to any intervention (e.g. setting quotas or entry requirements) directed to regulate legal migration, and it reveals once again that restricting entry cannot be the only dimension of a sound immigration policy.

respond differently to changes in prices.³¹ For some reason, the literature on immigration policy has generally overlooked this issue, and, under this perspective, this paper may be a step towards filling the gap.

³¹ These agents being borrowers dealing with interest rates, workers with wages or policyholders with insurance premia (see e.g. Stiglitz and Weiss, 1981)

References

- Acemoglu, D. (1998), 'Why do new technologies complement skills? directed technical change and wage inequality', *Quarterly Journal of Economics* 113, 1055–1090.
- Akee, R. (2005), 'Venturing beyond the reef: Deciphering immigrant self-selection from a developing country to a developed country', Mimeo. Harvard University.
- Antecol, H., Cobb-Clark, D. A. and Trejo, S. J. (2003), 'Immigration policy and the skills of immigrants to australia, canada and the united states', *Journal of Human Resources* 38, 192–218.
- Bagnoli, M. and Bergstrom, T. (2005), 'Log-concave probability and its applications', *Economic Theory* 26(2), 445–469.
- Banerjee, A. and Duflo, E. (2004), 'Growth theory through the lens of development economics', forthcoming in *Handbook of Economic Growth* edited by P. Aghion and S. Durlauf.
- Banerjee, A. and Newman, A. (1998), 'Information, the dual economy, and development', *Review of Economic Studies* 65(4), 631–653.
- Bellettini, G. and Berti Ceroni, C. (2005), 'Immigration policy, immigrant selection and human capital accumulation', Mimeo: Universita' di Bologna.
- Benhabib, J. (1996), 'On the political economy of immigration', *European Economic Review* 40, 1737–43.
- Bianchi, M. (2006), 'Immigration policy and self-selecting migrants', SSE/EFI Working Paper Series in Economics and Finance No. 627.
- Bils, M. and Klenow, P. J. (2000), 'Does schooling cause growth?', *American Economic Review* 90(5), 1160–1183.
- Blau, F. D. and Kahn, L. M. (1996), 'International differences in male wage inequality: institutions versus market forces', *Journal of Political Economy* 104, 791–837.
- Borjas, G. J. (1987), 'Self-selection and the earnings of immigrants', *American Economic Review* 77(4), 531–53.
- Borjas, G. J. (1994), 'The economics of immigration', *Journal of Economic Literature* 32(4), 1667–1717.

- Borjas, G. J. (1995), 'The economic benefits from immigration', *Journal of Economic Perspectives* 9(2), 3–22.
- Borjas, G. J. (2003), 'The labor demand curve is downward sloping: Reexamining the impact of immigration on the labor market', *Quarterly Journal of Economics* 118(4), 1335–1374.
- Brücker, H. and Defoort, C. (2006), 'The (self-)selection of international migrants reconsidered: Theory and new evidence', Institute for the Study of Labor (IZA) Discussion Paper No. 2052.
- Butcher, K. and Morrison Piehl, A. (2005), 'Why are immigrants' incarceration rates so low? evidence on selective immigration, deterrence, and deportation', Federal Reserve Bank of Chicago WP-05-19.
- Card, D. (2005), 'Is the new immigration really so bad?', NBER Working Paper No. 11547.
- Carrington, W. J., Detragiache, E. and Vishwanath, T. (1996), 'Migration with endogenous moving costs', *American Economic Review* 86(4), 909–30.
- Caselli, F. (2004), 'Accounting for cross-country income differences', forthcoming in *Handbook of Economic Growth* edited by P. Aghion and S. Durlauf.
- Caselli, F. and Coleman, W. J. (2005), 'The world technology frontier', Mimeo: LSE and Duke University.
- Chiquiar, D. and Hanson, G. H. (2005), 'International migration, self-selection, and the distribution of wages: Evidence from Mexico and the United States', *Journal of Political Economy* 113(2), 239–281.
- Chiswick, B. R. (1999), 'Are immigrants favorably self-selected?', *American Economic Review* 89(2), 181–185.
- Chiswick, B. R. (2005), 'High skilled immigration in the international arena', Institute for the Study of Labor (IZA) Discussion Paper No. 1782.
- Chiswick, B. R., Lee, Y. L. and Miller, P. W. (2005), 'Immigrant earnings: A longitudinal analysis', Institute for the Study of Labor (IZA) Discussion Paper No. 1750.
- Chojnicki, X., Docquier, F. and Ragot, L. (2005), 'Should the U.S. have locked the heaven's door? reassessing the benefits of the postwar immigration', Institute for the Study of Labor (IZA) Discussion Paper No. 1676.
- Citrin, J., Green, D. P., Muste, C. and Wong, C. (1997), 'Public opinion toward immigration reform: The role of economic motivation', *Journal of Politics* 59(3), 858–881.
- Clark, X., Hatton, T. J. and Williamson, J. G. (2002), 'Where do U.S. immigrants come from, and why?', NBER Working Paper No. 8998.
- Cuecuecha, A. (2005), 'The immigration of educated Mexicans: The role of informal social insurance and migration costs', mimeo: Instituto Tecnológico Autónomo de México.

- Dahan, M. and Gaviria, A. (2001), ‘Sibling correlations and intergenerational mobility in latin america’, *Economic Development and Cultural Change* 49(3), 537–554.
- Epstein, G. S. and Nitzan, S. (2005), ‘The struggle over migration policy’, Institute for the Study of Labor (IZA) Discussion Paper No. 1533.
- Espenshade, T. and Hempstead, K. (1996), ‘Contemporary american attitudes toward u.s. immigration’, *International Migration Review* 30(2), 535–570.
- Facchini, G. and Willman, G. (2005), ‘The political economy of international factor mobility’, *Journal of International Economics* 67, 201–219.
- Filmer, D. and Pritchett, L. (1999), ‘The effect of household wealth on educational attainment: Evidence from 35 countries’, *Population and Development Review* 25(1), 85–120.
- Freeman, R. and Oostendorp, R. (2000), ‘Wages around the world: Pay across occupations and countries’, NBER Working Paper No. 8058.
- Friebel, G. and Guriev, S. (2004), ‘Smuggling humans: A theory of debt-financed migration’, Institute for the Study of Labor (IZA) Discussion Paper No. 1025.
- Friedberg, R. M. and Hunt, J. (1995), ‘The impact of immigrants on host country wages, employment and growth’, *Journal of Economic Perspectives* 9(2), 23–44.
- Giannetti, M. (2003), ‘On the mechanics of migration decisions: skill complementarities and endogenous price differentials’, *Journal of Development Economics* 71, 329–349.
- Goldin, C. (1994), ‘The political economy of immigration restriction in the united states, 1890 to 1921’, in C. Goldin and G. D. Libecap, eds, *The Regulated Economy: A Historical Approach to Political Economy*, University of Chicago Press, Chicago.
- Hainmueller, J. and Hiscox, M. J. (2004), ‘Educated preferences: Explaining attitudes toward immigration in europe’, Mimeo: Harvard University.
- Hall, R. and Jones, C. I. (1999), ‘Why do some countries produce so much more output per worker than others?’, *Quarterly Journal of Economics* 114(1), 83–116.
- Hanson, G. H. (2005), ‘Emigration, labor supply, and earnings in mexico’, NBER Working Paper No. 11412.
- Hanson, G. H., Scheve, K. and Slaughter, M. J. (2005), ‘Individual preferences over high-skilled immigration in the united states’, Mimeo.
- Hatton, T. J. and Williamson, J. G. (2004), ‘International migration in the long-run: Positive selection, negative selection and policy’, NBER Working Paper No. 10529.

- Hatton, T. J. and Williamson, J. G. (Forthcoming), *World Mass Migration: Two Centuries of Policy and Performance*, MIT Press.
- Jasso, G. and Rosenzweig, M. (2005), 'Selection criteria and the skill composition of immigrants: A comparative analysis of Australian and US employment immigration', mimeo: Harvard and New York University.
- Jasso, G. and Rosenzweig, M. R. (1995), 'Do immigrants screened for skills do better than family reunification immigrants?', *International Migration Review* 29(1), 85–111.
- Jasso, G., Rosenzweig, M. R. and Smith, J. P. (1998), 'The changing skills of new immigrants to the United States: Recent trends and their determinants', NBER Working Paper No. 6764.
- Kremer, M. (1993), 'The O-ring theory of economic development', *Quarterly Journal of Economics* 108(3), 551–575.
- Krishna, V. (2002), *Auction Theory*, Academic Press.
- Lee, R. D. and Miller, T. W. (2000), 'Immigration, social security and broader fiscal impacts', *American Economic Review* 90, 350–354.
- Legrain, P. (2007), 'Migrant tax would slash illegal entry into Europe', *Financial Times*, 24 April 2007.
- Leuven, E., Oosterbeek, H. and van Ophem, H. (2004), 'Explaining international differences in male skill wage differentials by differences in demand and supply of skills', *Economic Journal* 114, 466–486.
- Lopez, R. and Schiff, M. (1998), 'Migration and the skill composition of the labor force: The impact of trade liberalization in LDCs', *Canadian Journal of Economics* 31(2), 318–336.
- Lowell, B. L., Bean, F. and de la Garza, R. (1986), 'Undocumented immigration: An analysis of the 1984 Simpson-Mazzoli vote', *Social Science Quarterly* 67(1), 118–127.
- Lucas, R. E. J. (1990), 'Why doesn't capital flow from rich to poor countries?', *American Economic Review* 80(2), 92–96.
- Massey, D. S., Arango, J., Hugo, G., Kouaouci, A., Pellegrino, A. and Taylor, E. J. (1993), 'Theories of international migration: A review and appraisal', *Population and Development Review* 19(3), 431–466.
- Mayda, A. M. (2004), 'Who is against immigration? A cross-country investigation of individual attitudes toward immigrants', Working Paper: Georgetown University.
- Mayda, A. M. (2005), 'International migration: A panel data analysis of economic and non-economic determinants', Institute for the Study of Labor (IZA) Discussion Paper No. 1590.
- McKenzie, D. and Rapoport, H. (2005), 'Migrant networks, migration incentives and education inequality in rural Mexico', Mimeo: Stanford and Bar-Ilan University.

- Miller, P. W. (1999), 'Immigration policy and immigrant quality: The Australian point system', *American Economic Review* 89(2), 192–197.
- Mishra, P. (2003), 'Emigration and wages in source countries: Evidence from Mexico', Mimeo, Columbia University.
- Munshi, K. (2003), 'Networks in the modern economy: Mexican migrants in the U.S. labor market', *Quarterly Journal of Economics* 118(2), 549–599.
- Munshi, K. and Rosenzweig, M. (2005), 'Why is mobility in India so low? Social insurance, inequality, and growth', Mimeo: Brown and Harvard University.
- O'Rourke, K. H. and Sinnott, R. (2004), 'The determinants of individual attitudes towards immigration', Working Paper: Trinity College Dublin Economics Department.
- Ortega, F. (2005), 'Immigration quotas and skill upgrading in the labor force', *Journal of Public Economics* 89(9-10), 1841–1863.
- Piketty, T. (2000), 'Theories of persistent inequality and intergenerational mobility', in A. Atkinson and F. Bourguignon, eds, *Handbook of Income Distribution*, Elsevier, chapter 6.
- Psacharopoulos, G. and Patrinos, H. A. (2002), 'Returns to investment in education: A further update', The World Bank Policy Research Working Paper 2881.
- Scheve, K. F. and Slaughter, M. J. (2001), 'Labor market competition and individual preferences over immigration policy', *Review of Economics and Statistics* 83(1), 133–145.
- Schwartz, A. (1973), 'Interpreting the effect of distance on migration', *Journal of Political Economy* 81(5), 1153–1169.
- Sjaastad, L. A. (1962), 'The costs and returns of human migration', *Journal of Political Economy* 70(5), 80–93.
- Stark, O. and Bloom, D. E. (1985), 'The new economics of labor migration', *American Economic Review* 75, 173–178.
- Stiglitz, J. E. and Weiss, A. (1981), 'Credit rationing in markets with imperfect information', *American Economic Review* 71(3), 393–410.
- Storesletten, K. (2000), 'Sustaining fiscal policy through immigration', *Journal of Political Economy* 108(2), 300–323.
- Taylor, J. E. (1986), 'Differential migration, networks, information and risk', in O. Stark, ed., *Research in Human Capital and Development*, Vol. 4, *Migration, Human Capital, and Development*, Greenwich, Conn.: JAI Press, pp. 147–171.

Timmer, A. S. and Williamson, J. G. (1998), 'Immigration policy prior to the thirties: Labor markets, policy interaction, and globalization backlash', *Population and Development Review* 24(4), 739–771.

Yashiv, E. (2004), 'The self-selection of migrant workers revisited', Institute for the Study of Labor (IZA) Discussion Paper No. 1094.

Entrepreneurs

CHAPTER 4

Financial Development, Entrepreneurship, and Job Satisfaction

Milo Bianchi

ABSTRACT. In this paper, I show both theoretically and empirically that greater financial development increases the job satisfaction of the self-employed, relative to employees. Financial development favors both job creation and matching between talents and occupations. Hence, in more financially-developed countries, individuals choose to become self-employed because of their talent, rather than for a lack of better opportunities. In addition, the effects of financial development are not only monetary. Increasing financial development makes the self-employed relatively more satisfied mainly because it allows them to enjoy greater independence in their job.

Keywords: Financial development; entrepreneurship; job satisfaction.

JEL codes: G20, J23, J28.

1. Introduction

Several studies have documented that the self-employed tend to be more satisfied with their job, both in absolute terms and controlling for wages and hours worked (Blanchflower and Oswald, 1998; Hundley, 2001; Benz and Frey, 2004). Such systematic differences have been referred to as labor market rents.¹ A popular explanation for these rents relies on barriers to entry into self-employment: financial constraints prevent some people from becoming self-employed, so those who can afford it enjoy utility above market clearing (Blanchflower and Oswald, 1998). If this interpretation is correct, then rents should decrease with financial development.

This paper studies labor market rents in a large set of countries over two decades, and it correlates them with financial development. We report three main findings. First, rents are not a universal phenomenon. The self-employed report higher job satisfaction than employees only in *more* developed countries. Second, rents *increase* with

⁰ Sincere thanks to Cedric Argenton, Andrew Clark, Tore Ellingsen, Justina Fischer, Martin Flöden, Ola Granström, Giancarlo Spagnolo, and especially to Erik Lindqvist and Jörgen Weibull for very helpful discussions. Financial support from the the Knut and Alice Wallenberg Foundation is gratefully acknowledged.

¹ Somewhat differently from other branches of economics, the term "rent" here is not necessarily associated with some market imperfections. Indeed, the present paper shows that it is quite the opposite.

financial development, and this effect is stronger in less developed countries. Third, greater financial development increases rents *not* by making entrepreneurs richer, but by allowing them to enjoy a higher freedom in their work.

These results are consistent with an occupational choice model in which self-employment may be either a profitable choice or the last resort to avoid unemployment. In such model, people face financial constraints to enter self-employment, but these constraints are more likely to bind for those who want to set up a firm and hire employees, rather than running a business alone. Being self-employed gives profits, which depends on market competition and possibly on one's talent, and non-monetary benefits of control. These benefits may be higher in big firms, where there are more employees and greater scope for entrepreneurial action.

In this setting, financial development favors both job creation and a better matching between talent and occupational choice. In fact, as a country develops, the poor and talented entrepreneurs expand their business, and so enjoy higher utility. The rich and untalented instead leave self-employment, as the demand for employees increases. As a result, more people are self-employed because they are highly motivated, rather than because they lack better opportunities. In this way, the model predicts that entrepreneurial rents are associated with *high* levels of financial development.

Moreover, these effects are not only monetary. Indeed, financial development increases competition, so it may reduce profits, particularly for small entrepreneurs. However, it also allows entrepreneurs to grow their firm, and so enjoy higher non-monetary benefits of control. This effect seems a major channel between financial development and labor market rents.

1.1. Related literature. The literature provides several occupational choice models that analyze who becomes an entrepreneur (for a review, see Bianchi and Henrekson, 2005). In this paper, as in Lucas (1978), individuals differ in their talent as entrepreneurs and this affects the size of their firm and their profits. In such models, more developed countries have a lower share of entrepreneurs and firms of larger size, which is confirmed in our data and in a number of other studies (see Acs, Audretsch and Evans, 1994; Schneider and Enste, 2000; Gollin, 2007). We enrich Lucas' classic approach with three ingredients. First, potential entrepreneurs may be limited by credit constraints; second, entrepreneurs enjoy also non-monetary benefits; third, individuals may become entrepreneurs by choice or by necessity. All these aspects have been widely documented, and, to some extent, incorporated in formal models.

The role of personal wealth and credit constraints in determining the probability of becoming an entrepreneur is shown in many studies including Evans and Leighton

(1989), Evans and Jovanovic (1989), Holtz-Eakin, Joulfaian and Rosen (1994), Blanchflower and Oswald (1998), and Hurst and Lusardi (2004). This relation is explored formally, for example, in Banerjee and Newman (1993), who analyze occupational choices with credit market imperfections and show how a country's development path may be highly dependent on the initial distribution of wealth.

The fact that the self-employed value also non-monetary dimensions of their job, and in particular the possibility of being independent, is documented e.g. in Taylor (1996), Hamilton (2000), and Benz and Frey (2004). This literature reports higher level of job satisfaction for the self-employed, but the analysis so far is limited mostly to OECD countries.

The idea that some people are forced into self-employment by lack of better opportunities stems from a classic proposition, expressed especially in studies on the informal sector in developing countries, stating that labor markets are imperfect and workers in the formal sectors are paid above equilibrium wages (see Harris and Todaro, 1970; and Loayza, 1994). In support of this view, for example, Evans and Leighton (1989) find that those who are unemployed, have low wages, or have changed job frequently are more likely to enter self-employment; while Borjas (1986) discusses the role of discrimination in pushing minorities into self-employment. To my knowledge, no study explores this view formally. In existing models, the self-employed have chosen to be so and they could have become employees, while employees for some reason could not become self-employed. However, if this were the case, the self-employed would always be better off than employees, which is not true in our data.

Finally, the present paper may be viewed also as a contribution to the literature on financial development. A growing body of works investigates the mechanisms linking finance to the real economy, with an increasing attention to micro evidence (see Levine, 2005, for a recent survey). No study that I know considers how financial development affects individuals, and none uses subjective data in order to capture the effects on non-monetary dimensions of individual utility. Indeed, we will see that financial development acts also through non-monetary channels.

2. The Model

Consider an economy populated by a continuum n of risk-neutral individuals with identical preferences. Each individual is characterized by a type $\theta = (a, t)$, where a describes his initial wealth and t his managerial talent. Wealth is drawn from a cumulative distribution function F ; talent from a cumulative distribution function G , with support on the interval $[\underline{t}, \bar{t}]$. These draws are assumed to be statistically independent. In addition, each individual is endowed with one unit of labor, which he

can use as follows: he can either set up a firm, look for a job as employee of such firm, or run a one-man business.

2.1. Options. The way an individual employs his labor defines his occupation. As just mentioned, three options are available. First, an individual can set up a firm, and enjoy a profit that depends on his managerial talent and on the level of competition in the product market. In particular, we assume that each firm produces the same homogeneous good and it has the same size: it employs k units of capital and l workers.² A firm run by an individual with talent t produces tq_1 units of output.³ The profit is then

$$(2.1) \quad \pi_1 = ptq_1 - wl - rk,$$

where p denotes the price of the good, w denotes workers' wage, and r is the market interest rate. In addition, managing a firm gives utility b_1 , which is independent of the talent t . Hence, an individual who sets up a firm enjoys

$$(2.2) \quad U_1 = \pi_1 + b_1.$$

These individuals are called *big entrepreneurs*, and we denote their population share with x_1 .

As a second option, an individual can look for a job in such firm. If he is hired, he enjoys

$$U_2 = w,$$

where the wage w is independent of his managerial talent t . The population share of *workers* is denoted with x_2 . If instead he is not hired, an individual can turn to the last option, which we now describe.

The last option is to run a one-man business, which requires no capital investment, no employees, and it generates q_3 units of the same good produced by big firms. This output is independent of managerial talent, so the profit of one-man businesses writes simply as

$$\pi_3 = pq_3.$$

In these businesses, labor is less productive than in big firms. More precisely, we assume that $(1+l)$ one-man businesses produce less than one of the big firms considered above,

² As will become clear, the effects of financial development in our model would be amplified if the amount of capital invested and the number of employees were a function of one's talent.

³ This formalization of managerial talent follows Lucas (1978), and several subsequent occupational choice models (e.g. Gollin, 2007).

even when such firm is managed by the least talented individual.⁴ That is,

$$(2.3) \quad (1 + l)q_3 \leq tq_1.$$

In addition, running a one-man business gives utility b_3 , which is lower than the non-monetary benefit enjoyed in big firms, that is

$$(2.4) \quad b_3 < b_1.$$

In sum, individuals who run a one-man business enjoy

$$(2.5) \quad U_3 = \pi_3 + b_3,$$

they are called *small entrepreneurs*, and their population share is denoted with x_3 .

2.2. Markets. There are three markets in our economy: a labor market, a product market, and a credit market. In the labor market, the wage w is fixed and exogenous, which implies that such market may not clear. In case of excess supply, each applicant has the same probability of getting a job.⁵ The number of workers equals firms' demand, so we have

$$(2.6) \quad x_2 = lx_1.$$

The product market is described by a strictly decreasing inverse demand function

$$(2.7) \quad p = P(Q),$$

where Q denotes the total output produced in the economy. Entrepreneurs take the price p as given, and inelastically supply their output.

The financial market is competitive, the interest rate r is fixed and exogenous, and we normalize it to zero. Individuals can tap this market and ask for a loan $m = k - a$ in order to set up a firm. However, ex-post moral hazard limits the maximum size of such loan. Since, at cost c , the borrower can renege on his contract and run away with the money, the required repayment m cannot exceed c . Hence, only individuals with enough wealth can become entrepreneurs, and we define such lower bound on wealth as

$$(2.8) \quad a \geq a^* \equiv k - c.$$

⁴ This assumption seems natural if one considers that a firm employs one manager, l workers and k units of capital, while one-man businesses employ one worker and no capital. On the other hand, it rules out socially inefficient firm creation, which may be driven by the private benefit b_1 . This is an interesting case to consider, but it deserves a separate paper.

⁵ As we will see, there cannot be excess demand in our economy. More sophisticated reasons for non-market clearing wages are for example in Shapiro and Stiglitz (1984).

The threshold a^* decreases with c , which measures how easy it is to enforce loan contracts and so it is an indicator of financial development.⁶

2.3. Equilibrium. In equilibrium, each individual, given his type θ , chooses an option in order to maximize his expected utility; everyone is given one occupation, so

$$(2.9) \quad x_1 + x_2 + x_3 = 1;$$

and the markets function according to equations (2.6), (2.7) and (2.8).

We first consider the case of an individual with wealth lower than a^* . Since he cannot afford to set up a firm, he can look for a job as worker or become a small entrepreneur. To simplify our analysis, we assume that, for any price level, the utility of small entrepreneurs does not exceed the one of workers. The highest utility of a small entrepreneur occurs when the output Q is at its minimum, and so the price p at its maximum. The output writes

$$(2.10) \quad Q = nx_1\hat{t}q_1 + nx_3q_3,$$

where \hat{t} is the average talent of a big entrepreneur. Given condition (2.3), Q is minimized when everyone is a small entrepreneur, in which case the output is nq_3 and the price is $P(nq_3)$. Hence, the maximum that a small entrepreneur can get is

$$\bar{U}_3 \equiv P(nq_3)q_3 + b_3,$$

and we assume that

$$(2.11) \quad \bar{U}_3 \leq w.$$

This condition implies that individuals become small entrepreneurs as they cannot get a better job. Hence, the choice of an individual with wealth lower than a^* is pretty trivial. He applies for a job as employee, and, if he is not assigned one, he becomes a small entrepreneur.

A more interesting case is when an individual has wealth greater than a^* . Depending on his talent, he can decide whether to set up a firm or go through the same procedure as those with wealth lower than a^* . Specifically, an individual with wealth $a \geq a^*$ and talent t sets up a firm if and only if

$$(2.12) \quad tq_1p - wl + b_1 \geq \frac{x_2}{1-x_1}w + \frac{x_3}{1-x_1}(q_3p + b_3),$$

where the right hand side is the sum of the utility as worker, weighted by the probability to be hired, and as small entrepreneur, weighted by the probability to be forced to run

⁶ Our formalization of financial market imperfections is very similar to the one in Banerjee and Newman (1993). The fact that only sufficiently rich people get loans is a very common feature in financial markets, and it can be derived also in a model of moral hazard à la Holmstrom and Tirole (1997) or costly screening.

a one-man business. Equation (2.12) implicitly defines a lower bound on the talent of big entrepreneurs as

$$(2.13) \quad t \geq \frac{wl + x_3(q_3p + b_3) - (1 - x_1)b_1}{(1 - x_1)q_1p} \equiv t^*.$$

Hence, provided that an equilibrium exists, the share of big entrepreneurs x_1 is implicitly defined by

$$(2.14) \quad x_1 = \min\left\{\frac{1}{1+l}, [1 - F(a^*)][1 - G(t^*)]\right\}.$$

Given equations (2.6) and (2.9), equation (2.14) is sufficient to characterize also the equilibrium share of workers and of small entrepreneurs.

3. Analysis

We first show that an higher number of big entrepreneurs increases competition in the product market, and so it lowers the price of the good.⁷

LEMMA 7. *The price p is decreasing in x_1 .*

PROOF. By equation (2.7), p decreases with the output Q . Recall from equations (2.6) and (2.9) that $x_3 = 1 - x_1 - lx_1$. Hence, differentiating equation (2.10), we write

$$\frac{\partial Q}{\partial x_1} = ntq_1 - (1+l)nq_3.$$

By equation (2.3), the last equation is positive, so p decreases in x_1 . \square

Given Lemma 7, we see that the minimal talent needed to run profitably a firm increases with the share of big entrepreneurs x_1 . In fact, an higher x_1 increases the demand for workers, which lowers the incentive to set up a firm as a way to escape one-man businesses, and it increases competition, which also reduces the incentive to set up a firm. This is expressed in the next Lemma.

LEMMA 8. *The minimal talent t^* is increasing in x_1 .*

PROOF. With simple algebra, one can write

$$\frac{\partial t^*}{\partial x_1} = \frac{l(w - pq_3 - b_3)}{q_1p(1 - x_1)^2} - \frac{1}{q_1p^2(1 - x_1)} \frac{\partial P}{\partial x_1} [wl - (1 - x_1)b_1 + x_3b_3].$$

The first term is positive due to equation (2.11). The second term is also positive. In fact, given equation (2.13), we have $wl - (1 - x_1)b_1 + x_3b_3 = t^*(1 - x_1)q_1p - x_3q_3p$, which is positive since $1 - x_1 > x_3$ and, by equation (2.3), $t^*q_1 > q_3$. Hence, t^* increases in x_1 . \square

⁷ While in the real world there are also positive externalities among firms, we show that, even abstracting from them, entrepreneurs may get higher payoffs when more firms are created.

Lemma 8 ensures that an equilibrium in our economy exists and it is unique. In fact, notice that the right hand side of equation (2.14) is always non-negative and weakly decreasing in x_1 . Hence, this equation uniquely defines the share of big entrepreneurs x_1 , and, with equations (2.6) and (2.9), this characterizes our equilibrium. We summarize with the following Proposition.

PROPOSITION 15. *An equilibrium exists and it is unique. It is defined by equations (2.6), (2.9) and (2.14).*

The mechanics of our model depend on how x_1 varies with financial development. First, notice that, by relaxing wealth constraints, financial development allows more people to set up a firm. This increases competition, and it increases the minimal talent needed to run profitably a firm. Hence, financial development allows the poor with high talent to become entrepreneurs, and induces the rich with low talent to exit and work as employees. As shown in the next Lemma, the share of big entrepreneurs increases with financial development.

LEMMA 9. *The share of big entrepreneurs x_1 is weakly increasing in c .*

PROOF. If $[1 - F(a^*)][1 - G(t^*)](1 + l) < 1$, implicitly differentiating equation (2.14), we have

$$\frac{\partial x_1}{\partial c} = \frac{\frac{\partial F}{\partial a^*}[1 - G(t^*)]}{1 + [1 - F(a^*)]\frac{\partial G}{\partial t^*}\frac{\partial t^*}{\partial x_1}}.$$

The numerator measures the increment in people who can afford to become big entrepreneurs. The denominator tells how the mass of people who are sufficiently talented and hence willing to be big entrepreneurs changes as entry increases. Given Lemma 8, $\partial t^*/\partial x_1$ is positive and hence $\partial x_1/\partial c$ is also positive. If instead $[1 - F(a^*)][1 - G(t^*)](1 + l) \geq 1$, then $\partial x_1/\partial c = 0$. \square

Now, recall that, from equation (2.9), the share of small entrepreneurs decreases in x_1 , and hence in financial development, since having more firms increases the demand for employees. Hence, we can define the maximal level of financial development \bar{c} so that both big and small entrepreneurs are in the market. Such threshold is defined by

$$x_1(\bar{c}) = \frac{1}{1 + l}.$$

We first describe the effects of financial development when $c < \bar{c}$ and so $x_3 > 0$.⁸ We are interested in analyzing entrepreneurial "rents", i.e. systematic differences in utility

⁸ Obviously, we are only considering the case in which $c < k$, so indeed financial development has an effect.

of entrepreneurs vs. employees. The average utility of an entrepreneur is

$$(3.1) \quad \bar{U} = \frac{x_3(pq_3 + b_3) + x_1(p\hat{t}q_1 + b_1 - wl)}{x_1 + x_3},$$

where \hat{t} denotes the average talent of a big entrepreneur, as a function of the equilibrium share of big entrepreneurs. It is useful to decompose such utility as the sum of the average profit of entrepreneurs

$$\bar{\pi} = \frac{x_3pq_3 + x_1(p\hat{t}q_1 - wl)}{x_1 + x_3},$$

and their average non-monetary benefit

$$\bar{b} = \frac{x_3b_3 + x_1b_1}{x_1 + x_3}.$$

Entrepreneurial rents are defined as

$$(3.2) \quad R = \bar{\pi} + \bar{b} - w.$$

Notice first that entrepreneurial rents need not be positive. By construction, employees enjoy higher utility than small entrepreneurs. Hence, whether or not the labor market displays rents depends on the composition of entrepreneurs, i.e. on the share x_1 vs. x_3 .

We now look at the effects of financial development on such rents. For consistency with the empirical analysis in the next Section, we do not explicitly consider how the utility of a given entrepreneur changes as the country develops. Changes in rents are rather analyzed by comparing the group of entrepreneurs in a financially developed country with the one in a less developed country. Differentiating equation (3.2) with respect to c , we can see that the ratio of $\partial R/\partial c$ over $\partial x_1/\partial c$ writes as

$$(3.3) \quad \frac{x_3q_3}{x_1 + x_3} \frac{\partial P}{\partial x_1} + \frac{x_1q_1}{x_1 + x_3} \left(\frac{\partial \hat{t}}{\partial x_1} p + \frac{\partial P}{\partial x_1} \hat{t} \right) + \frac{(\hat{t}q_1 - q_3)p - wl + (b_1 - b_3)}{(x_1 + x_3)^2}.$$

As shown in Lemma 9, financial development plays a fundamental role as it relaxes wealth constraints, thereby increasing the share of big entrepreneurs. Equation (3.3) describes the resulting effects. The first term describes the effect on those who remain small entrepreneurs: their profit is reduced as they suffer from more competition from big entrepreneurs. The second term describes the effects on big entrepreneurs. They also suffer from increased competition, but this may be compensated by an increase in their average talent (see Lemma 8). Hence, the total effect on their profits is unclear. The last term is particularly important in our analysis. It tells that financial development allows more jobs to be created and some people to access a more desirable occupation, in terms both of profits and of non-monetary benefits. That is, the composition of entrepreneurs changes with financial development, as some poor but talented

entrepreneurs have the possibility to leave their one-man businesses and become big entrepreneurs.

It is also clear that these effects will not be at play when financial development is very high, so that the share of big entrepreneurs is large enough to have no small entrepreneurs in the market.⁹ That is, if $c \geq \bar{c}$, then $x_3 = 0$, $x_1 = 1/(1+l)$ and

$$\frac{\partial R}{\partial c} = q_1 \left(\frac{\partial \hat{t}}{\partial x_1} p + \frac{\partial P}{\partial x_1} \hat{t} \right) \frac{\partial x_1}{\partial c} = 0.$$

The following Proposition summarizes these predictions, which we test in the next Section.

PROPOSITION 16. *a. Entrepreneurial rents R occur only in countries with high financial development.*

b. The average profit $\bar{\pi}$ decreases with financial development, especially when financial development is low.

c. The average non-monetary benefit \bar{b} increases with financial development, especially when financial development is low.

4. Testing the model

We are interested in exploring the effects of financial development on the utility of entrepreneurs relative to workers. However, as shown in many studies, money is not the only argument to evaluate the returns from a job. We need a broader indicator of the utility U , and for this purpose, we use self-reported levels of job satisfaction. In our view, this measure includes both monetary and non-monetary dimensions.¹⁰

It is also important to notice from the outset that we are going to estimate the changes in job satisfaction within the group of entrepreneurs relative to the group of employees, but indeed the composition of these groups changes with financial development. In other words, we do not estimate the effects on the same individuals, but rather the effects on a representative individual within a group over time or across countries.

Moreover, in order to interpret our findings along the lines suggested by our model, we also look at the effects of financial development on the income of entrepreneurs relative to workers, and on the share of entrepreneurs in the labor force. We also test whether the effects of financial development depend on the country's stage of development, and whether they affect non-monetary components of job satisfaction.

⁹ If $\bar{c} > k$, then the share of small entrepreneurs is positive for any level of financial development.

¹⁰ This assumption will be validated empirically: We will see that income is a major determinant, but not the only determinant, of job satisfaction.

4.1. Data. In most of our analysis, the dependent variable will be the individual level of job satisfaction. This variable is taken from the World Value Surveys and it is a 1 to 10 index based on the answer to the question: "Overall, how satisfied or dissatisfied are you with your job?" Our sample covers 46 countries over the period 1981 – 2001. The surveys were conducted in three waves (in the early 80s, early 90s and late 90s) and not all countries were included in all waves.¹¹ We denote each country and year pair with k , indicating e.g. Austria 1990, Austria 1999, Belgium 1990, and so on. In total, our data set comprises 46607 individual observations for full time employees and 6888 for self-employed, divided into 91 country-year groups. For each individual, we have information on demographic characteristics, income, employment status, and several variables describing beliefs, personality and different dimensions of his or her job.

This data set presents three major advantages: first, it displays a significant cross-country and time-series variation, which allows to explore rents and financial development beyond the standard sample of OECD countries. Second, these data may capture the effects on small-scale business, which are usually considered quite sensitive to finance. In our dataset, among the self-employed, 11% have more than 10 employees, 26% have less than 10 employees, 18% are farmers, and the rest are running a one-man business. Third, subjective data may capture also non-monetary dimensions, which have proven a fundamental determinant of occupational choices.

As indicator of financial development in a given country-year k , we use the level of domestic credit to the private sector, as percentage of GDP. The variable is taken from the World Development Indicators, available from the World Bank website. In our sample, it displays a considerable variation both within and across countries, ranging from 1.68 (Poland, 1989), to 195.98 (Japan, 1990). This is the most commonly used indicator in the literature on finance and growth, and it seems well suited for our purposes as well. It reflects the availability of bank credit, which is a fundamental ingredient to ease the creation of new enterprises.

The other macroeconomic variables too are taken from the World Development Indicators. These include per capita GDP, GDP growth, unemployment, and market capitalization of listed firms. Summary statistics of all our variables can be found in the Appendix.

5. Empirical evidence

5.1. Rents are not everywhere. As suggested by our model, the self-employed need not enjoy greater utility than employees: in imperfect labor markets, self-employment can be the last resort to avoid unemployment. To get a first picture of where the status

¹¹ All the results of these surveys are available at the webpage <http://www.worldvaluessurvey.org/>.

of self-employed is a significant determinant of one's job satisfaction, we estimate the following equation separately for each country-year k :

$$(5.1) \quad U_i = \alpha + \beta X_i + \gamma E_i + \varepsilon_i.$$

The dependent variable U_i denotes the individual job satisfaction, X_i is a set of individual variables including gender, age, age-squared, education, marital status, and E_i is a dummy equal to one if i is self-employed. If in a given country in a given year the self-employed enjoy some rent, then the coefficient γ should be positive.

Table 1a reports the estimates of the coefficient γ for each country-year k . It is evident that labor market rents are not everywhere, but tend to be a feature of developed countries. In developing countries, the self-employed are not systematically more satisfied than employees.

Moreover, the results remain basically unchanged if income is included in the set of controls X_i . In fact, the set of countries-years displaying rents becomes slightly larger, which already suggests that income differentials are not the explanation behind differences in job satisfaction. If anything, controlling for income strengthen the existence of rents.

We can see these relationships more clearly by constructing the following variables. The variable RENTS is a dummy equal to one if γ is positive and significant at the 5% level. We also run a similar regression with income as dependent variable in equation (5.1). Given this regression, we construct the dummy RICH_SE, which equals one if γ is positive and significant at the 5% level, and the dummy POOR_SE, which equals one if γ is negative and significant at the 5% level.

As shown in Table 1b, the variable RENTS is positively correlated with financial development, GDP per capita and POOR_SE; and negatively correlated with share of self-employment and RICH_SE. In accordance with our model, rents occur in countries with high GDP per capita, high financial development and low self-employment. Moreover, in these countries, the self-employed tend to have a lower income than employees.

5.2. Rents and Financial Development. The previous results suggest that labor market rents are *not* due to financial market imperfections. We now explore this argument in further detail. We first estimate the equation

$$(5.2) \quad U_{i,k} = \alpha + \beta X_{i,k} + \gamma I_k + \delta F_k * E_{i,k} + \varepsilon_{i,k},$$

where $U_{i,k}$ denotes the reported job satisfaction for an individual i in a country-year k ; $X_{i,k}$ is a set of individual variables including gender, age, age-squared, education, marital status and employment status; I_k is a country-year dummy, F_k is the level

financial development and $E_{i,k}$ is an employment status dummy, equal to one if i is self-employed.

Equation (5.2) follows the spirit of Rajan and Zingales (1998), and it allows to estimate the effect of financial development on a particular set of individuals, the self-employed, after having controlled for the effect on the whole population and for country-year fixed effects. Our main interest is in the coefficient δ , which describes how financial development affects the job satisfaction of the self-employed *relative to* (full-time) employees.¹²

Table 2a reports our estimates on the full sample. The first column includes only the controls $X_{i,k}$. Self-employed, old, married and well-educated individuals tend to be more satisfied with their job. The second column includes the most basic specification, as in equation (5.2). The coefficient δ is positive and statistically significant. Financial development benefits the self-employed more than employees, hence financial development appears positively correlated with entrepreneurial rents.

When we add GDP per capita, interacted with the employment status dummy (column 3), the effect of financial development is slightly weaker, but still highly significant. Adding other macroeconomic variables like GDP growth and unemployment, always interacted with the employment dummy, does not change the estimate of δ (column 4). Hence, our preferred specification, which will serve as the baseline for the next analysis, is the one in column 3.

In columns 5 and 6, we see that rents tend to occur in places with a low share of self-employment, where, according to our model, entrepreneurship is mostly by choice rather than by necessity. Moreover, financial development seems to capture this mechanism, since the share of self-employment becomes insignificant once we control for financial development.

The second set of regressions estimates whether the effect of financial development depends on the country's stage of development. We divide the sample into countries-years with high and with low financial development, where such threshold is determined by the median value in our sample.¹³ Similar exercises are done splitting the sample in relation to GDP per capita. The results are in columns 1-4 of Table 2b: the effects of financial development on rents are positive and significant only in less developed countries.

¹² To ease the interpretation of our coefficients, part-time employees are excluded from the analysis. This exclusion does not change our results.

¹³ Splitting the sample according to the mean gives the same qualitative results. We prefer to use the median to have approximately equal numbers of observations and so potentially enough variation in both samples.

Our model suggests a possible explanation for this result. In less developed countries, many people become self-employed out of necessity. As these countries develop their financial systems, some of these people can afford to expand their business, and hence access better occupation, or they can leave self-employment, since additional salaried jobs are created. In more developed countries, instead, most people become self-employed by choice, so this effect on the composition of the self-employed is weaker. Indeed, we get similar findings if we split the sample according to the share of self-employed over employees, or to the level of unemployment. Financial development is positively associated with rents only in countries where the self-employed are many and unemployment is high.¹⁴

The same result appears if we use another widely used measure of financial development: the market capitalization of listed companies (as a percentage of GDP). The coefficient is positive and significant in less developed countries, and insignificant in more developed ones (columns 5 and 6).¹⁵ This suggests that there is nothing peculiar in our measure of financial development. Still, we prefer to keep our original indicator because, apart from being the most commonly used, banks rather than markets should have a bigger impact on the availability of credit for small entrepreneurs.

To explore the nonlinearity in the effects of financial development, column 7 includes the level of financial development squared and cube. The first appears to be negative and the second positive, and both are significant. These estimates are used to draw Figure 1, which plots the relation between rents and financial development. It is evident that rents occur only in countries with high financial development. In less developed countries, rents increase with financial development. In highly developed countries, approximately those above the sample median, the effect of financial development is U-shaped, and it appears insignificant if one applies a linear model.

5.3. Mechanisms. We now explore the mechanisms underlying the relation between financial development and rents. According to the model, financial development allows people to access better occupations, as it favors both job creation and a better matching between talents and occupational choices. Moreover, the model stresses that these mechanisms cannot be evaluated only in monetary terms.

We start by enriching the set of regressors in equation (5.2). First, we control for income, both in the full sample and separating countries-years according to their level of financial development. As shown in columns 2-4 of Table 3a, the results are

¹⁴ The median value for financial development is 71.78, for GDP per capita is 15252.62, for the share of self-employed is 0.1220557, and for unemployment is 8.2. Results for the share of self-employment and unemployment are not reported.

¹⁵ The median value for market capitalization, used to split the sample, is 32.27.

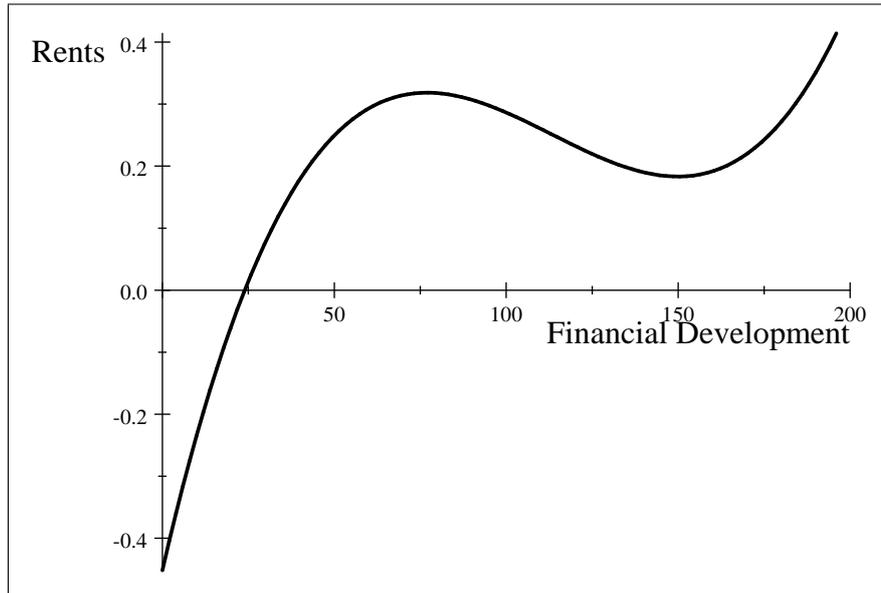


FIGURE 1. Rents and Financial Development (estimates obtained from Table 2b, Column 7).

basically unchanged. Income appears to be a major determinant of job satisfaction, but, as documented also in other works (e.g. Benz and Frey, 2004), higher income for entrepreneurs does not explain labor market rents. In addition, we document that the effects of financial development on job satisfaction are not only monetary.

Table 3b reports the estimates with income as the dependent variable. The overall effect is negative but not significant, but if we split the sample, we see that financial development decreases the income of the self-employed, relative to employees, in less developed countries, while its effect is not significant in more developed ones. This is consistent with our model in that financial development creates competition and this hurts disproportionately more the small entrepreneurs, who are the majority in less developed countries. For big entrepreneurs, instead, this may be counterbalanced by the fact that their average talent increases so their profits need not decrease.

Second, we explore whether financial development acts via non-monetary aspects of job satisfaction. We include variables like the degree of pride in the work, the satisfaction with job security, the degree of independence enjoyed in the job. We also control for work-related beliefs: how much work is important in life, the aspects of work one values more, the main reason why one works. None of these variables significantly affects our results, with the exception of independence, that is an indicator derived from the question: "How free are you to make decisions in your job?" Controlling for independence, the effect of financial development is considerably weaker, both in

magnitude and in significance (see Table 3a, column 7). The importance of this variable in explaining rents was already pointed out in Benz and Frey (2004), and indeed, also in our sample, being self-employed becomes negatively related to job satisfaction once one adds this control.

It is interesting to notice that part of the effects of financial development seems to work through this channel. This may be explained in terms of selection: financial development allows different people, and specifically those with higher value for independence, to become entrepreneur. Alternatively, it may be the case that financial development creates different working conditions, e.g. it gives entrepreneurs more freedom to experiment and innovate, and so higher utility from independence.

In order to explore further the issue of selection, we control for other aspects of personality, which are likely to be uncorrelated with financial development and so they would provide a strong case for selection arguments. These variables include the perceived freedom and control over the future, satisfaction with home life, personality traits like whether one welcomes major changes and innovations, enjoys taking responsibility, is confident of getting what he wants... However, while these aspects are generally an important determinant of job satisfaction, none of them appears to have a significant role in driving the effects of financial development on rents. Hence, selection along these dimensions does not seem to be a major player in our estimates.¹⁶

Last, we add occupation dummies, which classify workers according to whether they are employer/manager of a large or small firm, office workers or manual workers of different levels, agricultural workers, and so on. In all these occupations, workers can be either self-employed or employees.¹⁷ These dummies may capture the effect of financial development in allocating the self-employed into more desirable occupations, e.g. in increasing the fraction of big entrepreneurs among the self-employed. If this were the only effect, then we would expect financial development to become insignificant once occupation dummies are included. If instead financial development affected job satisfaction also for a given occupation, or if this classification of occupations were too coarse, then we would expect financial development to maintain some explanatory power. The data seem to validate the latter hypothesis: occupation dummies alone do

¹⁶ However, these questions were not asked in all countries in all waves, so we have some sample selection biases, and this makes it hard to assess whether such variables have indeed any affect. These results are not reported.

¹⁷ More precisely, occupations are divided as follows: employer/manager of an establishment with 500 or more employed; employer/manager of an establishment with 100 to 500 employed; employer/manager of an establishment with 10 to 100 employed, employer/manager of an establishment with less than 10 employed; professional worker; middle level non-manual office worker; supervisory non manual -office worker; junior level non manual; non manual -office worker; foreman and supervisor; skilled manual; semi-skilled manual worker; unskilled manual; farmer: has own farm; agricultural worker; member of armed forces; never had a job.

not change our estimates (see column 5). However, if we control also for independence, the effect of financial development becomes considerably smaller and statistically insignificant (column 6).

In sum, we interpret these results as suggesting that financial development increases labor market rents in developing countries as it allows the self-employed to access better occupations, especially in non-monetary terms. In addition, it allows them to enjoy higher utility within occupations, in particular because of increased freedom in taking decisions in their job.

6. Conclusion

In this paper, we have studied labor market rents in a large set of countries over two decades, and we have correlated them with financial development. We have shown that the self-employed report higher job satisfaction than employees only in more developed countries; that labor market rents increase with financial development, especially in less developed countries; and that financial development affects job satisfaction mostly along non-monetary dimensions, in particular independence. These findings have been interpreted with a simple occupational choice model in which financial development favors both job creation and a better matching between talents and occupations. Hence, in more developed countries, entrepreneurs display higher job satisfaction as they have chosen their occupation because of their talent, rather than for lack of better opportunities.

Financial development and job satisfaction are very broad concepts, and it is obviously difficult to pin down their mechanisms in a neat and indisputable way. While the results in this paper are preliminary in many respects, it appears that greater financial development has considerable effects on occupational choices, and that these effects are not only monetary. In our view, this is an interesting and highly unexplored avenue for future research.

7. Appendix

7.1. Variables.

Summary Statistics					
Variable	Observations	Mean	Std. Deviation	Min	Max
Job Satisfaction	53495	7.368427	2.137275	1	10
Financial Development	53495	66.01027	38.44095	1.68	195.98
GDP per capita	52941	13788.58	9755.539	317.15	43267.25
Unemployment	38045	8.447804	4.292947	0.6	16.4
Market Capitalization	38370	48.96034	51.08159	1.51	243.55
GDP growth	51967	1.666195	4.538714	-11.89	12.67
Share of Self-Employed	53495	0.1629402	0.1441684	0	0.80102
Self-Employed	53495	0.1287597	0.3349369	0	1
Gender (Male=1)	53443	1.386879	0.4870402	1	2
Age	53385	38.43434	12.09285	16	100
Married	53299	0.6778739	0.4672954	0	1
Education	49422	6.695905	2.733959	1	10
Income	46107	5.601297	2.424443	1	11

7.2. Results. The following tables report our main results. To ease the interpretation, Tables 2 and 3 report the estimates from OLS regressions. Results using ordered probit are qualitatively the same (see Ferrer-i-Carbonell and Frijters, 2004, for a methodological discussion).

Table 1a: Rents are not everywhere

Dependent variable: Job Satisfaction							
Country	Year	Self-Empl.	Std Error	Obs	Self-Empl.	Std Error	Obs
Argentina	1991	0.218498	0.1181881	356	0.2699054	0.139542	276
Austria	1990	0.0209013	0.1120682	689	0.0485767	0.1164346	663
Austria	1999	0.1391111	0.1168346	762	0.1815151	0.1438807	623
Belarus	2000	-0.2323577	0.2036998	636	-0.4506517	0.2084008	630
Belgium	1981	0.1887899	0.1593701	503	0.1042161	0.1849902	429
Belgium	1990	0.1450615	0.0795144	1282	0.0987844	0.1074882	825
Belgium	1999	0.1832029	0.1291225	836	0.317821	0.1458575	711
Brazil	1991	0.0726732	0.0965632	797	0.0924957	0.0999072	764
Bulgaria	1990	0.0800142	0.185855	414	-0.0186508	0.2561555	650
Bulgaria	1999	0.0775414	0.2435584	693	-0.0240338	0.1994808	397
Canada	1982	0.5367734	0.2491967	657	0.6687528	0.2771464	573
Canada	1990	0.3633602	0.1291664	1004	0.4820272	0.1430273	876
Chile	1990	-0.0251365	0.1090495	520	0.0115252	0.1096087	518
China	1990	-0.0007054	0.0967042	567	-0.0244881	0.0984042	560
Croatia	1999	-0.2486404	0.1799297	474	-0.3622906	0.1928684	459
Czech Republic	1990	0.1923959	0.269138	610	0.1579135	0.2703282	607
Czech Republic	1991	0.2870060	0.1416923	1371	0.2960847	0.1475212	1322
Czech Republic	1999	0.4087304	0.1141571	981	0.3788697	0.1208506	891
Denmark	1981	0.3384088	0.1682615	674	0.2992409	0.1794807	622
Denmark	1990	0.2796825	0.1396793	639	0.5643670	0.1803649	571
Denmark	1999	0.9144910	0.2072332	627	0.7675682	0.2184019	578
Estonia	1990	0.0730747	0.2963629	768	0.0614218	0.2967007	767
Estonia	1999	0.2926264	0.1674545	575	0.3125535	0.2182028	494
Finland	1990	0.1974547	0.1509961	424	-0.0218696	0.0305315	424
Finland	2000	0.1629761	0.1481705	490	0.2222935	0.1526546	449
France	1981	0.4995818	0.1486594	588	0.7115967	0.1979121	465
France	1990	0.2282116	0.2372401	453	0.176351	0.3375077	389
France	1999	0.2240496	0.2272016	736	0.1655217	0.2780597	610
Germany	1990	0.4758192	0.1186316	2124	0.4789135	0.1252535	2045
Germany	1999	0.0552915	0.1171329	862	0.1061966	0.1729893	607
Greece	1999	0.135569	0.1217611	644	0.0464116	0.1338035	563
Hungary	1991	0.062686	0.2834765	552	0.0590176	0.2833752	544
Hungary	1999	0.0118501	0.2869907	420	-0.1178635	0.3009352	409
Iceland	1984	0.1677834	0.1071644	679	0.174816	0.1118156	642
Iceland	1990	0.2014569	0.1065232	559	0.0463004	0.0221510	434
Iceland	1999	0.2725058	0.1004065	732	0.2937111	0.1046515	695
India	1990	0.0644648	0.0955979	580	0.0043798	0.0972911	573
Ireland	1981	0.1978731	0.1393969	598	0.0269862	0.2692462	222
Ireland	1990	0.0779317	0.135106	549	0.1616179	0.1481016	486
Ireland	1999	0.0860587	0.1271181	519	0.124438	0.1384172	461
Italy	1981	0.5217998	0.0921763	699	0.464150	0.1081868	541
Italy	1990	0.2923001	0.0921062	1030	0.2676279	0.1142673	782
Italy	1999	0.0922859	0.0815739	1031	0.101719	0.0929629	795

Results from ordered probit regression. Controls are gender, age, age squared, education, marital status. Columns 6-8 also include income. Coefficients significant at the 5% level are in bold.

Table 1a (continued): Rents are not everywhere

Dependent variable: Job Satisfaction							
Country	Year	Self-Empl.	Std Error	Obs	Self-Empl.	Std Error	Obs
Japan	1990	0.3220911	0.1274906	469	0.236718	0.1333606	429
Latvia	1999	-0.3273278	0.2978669	455	-0.4622407	0.3155974	430
Lithuania	1999	-0.5827697	0.2436269	513	-0.6629838	0.2908186	447
Luxembourg	1999	0.3633811	0.2197068	589	0.4690223	0.2849504	342
Malta	1983	0.5196789	0.3011161	205	1.0229910	0.4706113	147
Malta	1999	-0.0290905	0.1822961	464	-0.1450087	0.2061636	352
Mexico	1990	-0.1702467	0.105702	563	-0.1720934	0.10733	541
Netherlands	1981	0.2670397	0.2105276	480	0.0885858	0.2606471	350
Netherlands	1990	0.4447107	0.3047075	432	0.3888468	0.333793	358
Netherlands	1999	0.4887995	0.1304009	631	0.4953881	0.1384707	597
Nigeria	1990	0.0633401	0.1485335	226	-0.0287253	0.1563111	203
North Ireland	1981	0.670004	0.3710942	165	-0.2996089	0.1633143	36
North Ireland	1990	1.2420090	0.4946521	156	0.9445484	0.6738315	122
North Ireland	1999	0.1296344	0.1412417	464	0.1423546	0.1725758	340
Norway	1982	0.2632919	0.1187842	713	0.2995462	0.1267983	684
Norway	1990	0.2153681	0.1083191	845	0.3135317	0.1177507	741
Poland	1989	-0.1731505	0.1215037	477	-0.2318058	0.1226506	474
Poland	1990	-0.0753738	0.1160193	567	-0.1443434	0.1195364	548
Poland	1999	-0.4631074	0.1434599	506	-0.4752768	0.1440401	495
Portugal	1990	0.2235283	0.1285427	611	0.1880178	0.1320787	585
Portugal	1999	0.2374519	0.1411886	442	NA	NA	NA
Romania	1999	0.2960203	0.2269626	416	0.2847744	0.2375436	388
Russia	1999	0.1129042	0.1638472	1310	0.0913674	0.1756154	1235
Slovakia	1999	0.2435412	0.1654046	754	-0.0208156	0.1818599	707
Slovenia	1992	0.2700367	0.2015518	567	0.1919063	0.2044795	561
Slovenia	1999	-0.0854957	0.2082448	527	0.0801262	0.3321093	366
South Africa	1990	0.1922348	0.099104	1056	0.2063035	0.1038238	927
Spain	1981	0.2773977	0.0736411	984	0.2709970	0.0774734	897
Spain	1990	0.2931021	0.0671669	1624	0.2759815	0.0742207	1375
Spain	1999	0.2987274	0.1308793	478	0.0924025	0.1751715	319
Sweden	1982	0.3267322	0.1495769	651	0.3381471	0.1598292	619
Sweden	1990	0.4079321	0.1333923	770	NA	NA	NA
Sweden	1999	0.7091366	0.2334724	634	0.6261598	0.2399972	621
Turkey	2001	-0.0775182	0.1184038	369	-0.0575062	0.118804	369
UK	1981	0.7374404	0.2242033	678	0.5898634	0.261545	509
UK	1990	0.5916594	0.1292271	838	0.5926061	0.1527917	657
UK	1999	0.3124838	0.209776	434	0.145983	0.2437486	318
Ukraine	1999	-0.0973014	0.3300307	585	-0.1290457	0.3319256	571
US	1982	0.5060187	0.2303115	1317	0.5766977	0.2376557	1262
US	1990	0.3586575	0.1613992	1082	0.3210855	0.1680259	1016

Results from ordered probit regression. Controls are gender, age, age squared, education, marital status. Columns 6-8 also include income. Coefficients significant at the 5% level are in bold.

Table 1b: Partial Correlations

	RENTS	RICH SE	POOR SE	Financ Develop	GDP PC
RENTS	1				
RICH SE	-0.0275*	1			
POOR SE	0.0867*	-0.1671*	1		
Financ Develop	0.3296*	-0.1122*	0.0608*	1	
GDP PC	0.2983*	-0.1913*	0.0255*	0.6476*	1
Share of Self-Empl	-0.0323*	-0.1196*	-0.0057	-0.1708*	-0.1917*

Partial correlation coefficients. The star indicates significance at the 1% level.

Table 2a: Financial Development and Rents
Dependent Variable: Job Satisfaction (Full Sample)

Fin Dev*SE	0.0042804 (0.0008271)	0.0026134 (0.000977)	0.0027910 (0.0011157)	0.0040764 (0.0008514)
GDP pc*SE	0.0000095 (0.00000375)	0.0000112 (0.00000497)		
Unemploy*SE	0.0126092 (0.0088019)			
GDP growth*SE	-0.0088260 (0.0120389)			
Share SE*SE			-0.4288944 (0.2191341)	-0.1899596 (0.2254966)
Self-Employed	0.2988502 (0.0292735)	0.0202529 (0.0645355)	0.0056881 (0.0706861)	-0.1174733 (0.1486892)
Gender (Male=1)	0.0052085 (0.0194553)	0.0047338 (0.0194503)	0.0080404 (0.0195238)	-0.0022978 (0.0224946)
Age	0.0023248 (0.004653)	0.0024882 (0.0046506)	0.0024136 (0.0046622)	-0.0007280 (0.0053367)
Age squared	0.0001530 (0.0000544)	0.0001495 (0.0000543)	0.0001495 (0.0000545)	0.0001800 (0.0000619)
Married	0.1896052 (0.0222709)	0.1892803 (0.02227069)	0.1905596 (0.02227678)	0.1812794 (0.0257143)
Education	0.0355397 (0.0039674)	0.0351340 (0.0039667)	0.0352367 (0.0039746)	0.0358855 (0.0045765)
Observations	49101	49101	48648	35886
R-squared	0.0802	0.0808	0.0816	0.0881
			49101	49101
				0.0803
				0.0808

All regressions include country-year dummies. Robust standard errors in parenthesis. Coefficients significant at the 5% level in bold. SE means Self-Employed.

Table 2b: Financial Development and Rents

	Dependent Variable: Job Satisfaction (Sample splits)						
	Low Findev	High Findev	Low GDP	High GDP	Low Findev	High Findev	Full Sample
Fin Dev*SE	0.0097686 (0.0029152)	0.0009133 (0.0017962)	0.0035302 (0.0017837)	0.0015997 (0.0011241)			0.0241540 (0.0069934)
(Fin Dev*SE) ²							-0.0002375 (0.0000815)
(Fin Dev*SE) ³							0.0000007 (0.000000277)
Mkt Capitaliz*SE					0.0171855 (0.0080735)	0.0010193 (0.0008038)	
GDP _{pc} *SE	0.0000138 (0.00000555)	0.0000017 (0.00000525)	0.0000496 (0.0000115)	-0.0000038 (0.00000822)	0.0000120 (0.00000651)	0.0000170 (0.00000565)	0.0000075 (0.00000382)
Self-Employed	-0.2984839 (0.1163849)	0.2833401 (0.1646331)	-0.2497469 (0.1043552)	0.3497168 (0.1772009)	-0.1703375 (0.1269183)	-0.0944642 (0.1331187)	-0.4518678 (0.1640174)
Gender (Male=1)	0.0245148 (0.0284872)	-0.0097240 (0.0267184)	0.0087935 (0.0306634)	0.0052967 (0.024554)	0.0280590 (0.0335877)	0.0068214 (0.0314204)	0.0073872 (0.0195232)
Age	0.0033907 (0.0070569)	0.0014277 (0.006147)	-0.0063681 (0.0078771)	0.0093533 (0.0056197)	-0.0081005 (0.0090645)	-0.0140730 (0.007409)	0.0022860 (0.0046637)
Age squared	0.0001285 (0.0000831)	0.0001695 (0.0000711)	0.0002515 (0.0000935)	0.0000663 (0.0000649)	0.0002399 (0.0001088)	0.0003120 (0.0000846)	0.0001507 (0.0000545)
Married	0.1978516 (0.0334192)	0.1826483 (0.0309099)	0.2132834 (0.0367424)	0.1712164 (0.0280986)	0.1834348 (0.0399475)	0.2167752 (0.0351052)	0.1904068 (0.0227663)
Education	0.0444113 (0.0059072)	0.0267768 (0.0053484)	0.0595156 (0.0062113)	0.0124408 (0.0049987)	0.0416813 (0.0069007)	0.0390812 (0.0066912)	0.0352758 (0.0039741)
Observations	24399	24249	23704	24944	17188	17397	48648
R-squared	0.0915	0.0692	0.0646	0.0642	0.0714	0.0785	0.0819

All regressions include country-year dummies. Robust standard errors in parenthesis. Coefficients significant at the 5% level in bold.

Table 3a: Mechanisms
Dependent Variable: Job Satisfaction

	Full Sample	Full Sample	Low Finder	High Finder	Low Finder	Low Finder	Low Finder	Low Finder
Fin Dev*SE	0.0029554 (0.0011276)	0.0105339 (0.0031716)	-0.0009701 (0.0020137)	0.0058457 (0.0027549)	0.0107435 (0.0000058)	0.0047721 (0.0036372)		
GDP pc*SE	0.0000133 (0.00000407)	0.0000179 (0.00000593)	0.0000062 (0.00000574)	0.0000174 (0.0000053)	0.0000094 (0.0000058)	0.0000238 (0.0000060)		
Income	0.1058458 (0.0050712)	0.1054052 (0.0050838)	0.1172527 (0.0073951)	0.0941281 (0.006969)	0.3256878 (0.0059266)	0.3062039 (0.0068150)		
Freedom in job								
Self-Employed	0.2710708 (0.032662)	-0.0794165 (0.0762226)	-0.4117407 (0.1228474)	0.4063587 (0.1774827)	-0.8496402 (0.1109457)	-0.4160432 (0.1338845)	-0.7630571 (0.1353570)	
Gender (Male=1)	0.0029319 (0.0211411)	0.0061913 (0.0212347)	0.0283109 (0.0306485)	-0.0159328 (0.0293346)	0.1488681 (0.026247)	0.0394654 (0.0313340)	0.1546513 (0.0315293)	
Age	0.0004746 (0.0051649)	0.0004846 (0.0051752)	0.0001389 (0.0077473)	0.0009105 (0.006884)	-0.0196259 (0.0065276)	-0.0084715 (0.0073464)	-0.0214524 (0.0076354)	
Age squared	0.0001803 (0.0000607)	0.0001772 (0.0000608)	0.0001705 (0.0000917)	0.0001832 (0.0000801)	0.1438672 (0.0306824)	0.0002299 (0.0000864)	0.0003384 (0.0000903)	
Married	0.1080028 (0.0249903)	0.1092122 (0.0250734)	0.1202280 (0.036056)	0.1000381 (0.0347145)	0.0003154 (0.0000768)	0.1951896 (0.0349382)	0.1085367 (0.0352147)	
Education	0.0147632 (0.0044786)	0.0144140 (0.0044894)	0.0243810 (0.0065642)	0.0047208 (0.0061232)	-0.0103643 (0.0056281)	-0.0125758 (0.0072400)	-0.0412371 (0.0074339)	
Occupation	No	No	No	No	No	Yes	Yes	
Observations	42259	41809	21413	20396	24132	22147	19285	
R-squared	0.0912	0.0928	0.1026	0.0802	0.2307	0.1112	0.2331	

All regressions include country-year dummies. Robust standard errors in parenthesis. Coefficients significant at the 5% level are in bold.

Table 3b: The effects on income

Dependent Variable: Income			
	Full Sample	Low Findev	High Findev
Fin Dev*SE	-0.0018289 (0.0012567)	-0.0091418 (0.0035957)	0.0034655 (0.0026605)
GDP pc*SE	-0.0000104 (0.00000469)	-0.0000084 (0.00000731)	-0.0000095 (0.00000664)
Self-Employed	0.2967345 (0.0744278)	0.5926134 (0.1163334)	-0.2655474 (0.2144357)
Gender (Male=1)	-0.0525414 (0.0206253)	-0.0988691 (0.0287626)	0.0003247 (0.0295163)
Age	0.0119500 (0.0054366)	-0.0024892 (0.0074906)	0.0272718 (0.0079008)
Age squared	-0.0001460 (0.0000646)	-0.0000176 (0.0000895)	-0.0002823 (0.0000934)
Married	0.8743811 (0.0242663)	0.7233974 (0.0340043)	1.0336720 (0.0345314)
Education	0.2336714 (0.0042017)	0.2248508 (0.0059822)	0.2433377 (0.0058901)
Observations	41809	21413	20396
R-squared	0.3253	0.3456	0.294

All regressions include country-year dummies. Robust standard errors in parenthesis. Coefficients significant at the 5% level are in bold.

References

- Acs, Z. J., Audretsch, D. B. and Evans, D. S. (1994), 'Why does the self-employment rate vary across countries and over time?', CEPR Discussion Paper No. 871.
- Banerjee, A. V. and Newman, A. F. (1993), 'Occupational choice and the process of development', *Journal of Political Economy* 101(2), 274–98.
- Benz, M. and Frey, B. S. (2004), 'Being independent raises happiness at work', *Swedish Economic Policy Review* 11, 95–134.
- Bianchi, M. and Henrekson, M. (2005), 'Is neoclassical economics still entrepreneurless?', *Kyklos* 58(3), 353–377.
- Blanchflower, D. G. and Oswald, A. J. (1998), 'What makes an entrepreneur?', *Journal of Labor Economics* 16(1), 26–60.
- Borjas, G. J. (1986), 'The self-employment experience of immigrants', *Journal of Human Resources* 21(4), 485–506.
- Evans, D. S. and Jovanovic, B. (1989), 'An estimated model of entrepreneurial choice under liquidity constraints', *Journal of Political Economy* 97(4), 808–27.
- Evans, D. S. and Leighton, L. S. (1989), 'Some empirical aspects of entrepreneurship', *American Economic Review* 79(3), 519–35.
- Ferrer-i Carbonell, A. and Frijters, P. (2004), 'How important is methodology for the estimates of the determinants of happiness?', *Economic Journal* 114(497), 641–659.
- Gollin, D. (2007), 'Nobody's business but my own: Self employment and small enterprise in economic development', *Journal of Monetary Economics, Forthcoming*.
- Hamilton, B. H. (2000), 'Does entrepreneurship pay? an empirical analysis of the returns to self-employment', *Journal of Political Economy* 108(3), 604–631.
- Harris, J. R. and Todaro, M. P. (1970), 'Migration, unemployment and development: A two-sector analysis', *American Economic Review* 60(1), 126–42.
- Holmstrom, B. and Tirole, J. (1997), 'Financial intermediation, loanable funds, and the real sector', *Quarterly Journal of Economics* 112(3), 663–91.

- Holtz-Eakin, D., Joulfaian, D. and Rosen, H. S. (1994), 'Sticking it out: Entrepreneurial survival and liquidity constraints', *Journal of Political Economy* 102(1), 53–75.
- Hundley, G. (2001), 'Why and when are the self-employed more satisfied with their work?', *Industrial Relations* 40(2), 293–316.
- Hurst, E. and Lusardi, A. (2004), 'Liquidity constraints, household wealth, and entrepreneurship', *Journal of Political Economy* 112(2), 319–347.
- Levine, R. (2005), 'Finance and growth: Theory and evidence', in P. Aghion and S. Durlauf, eds, *Handbook of Economic Growth*, Elsevier, Vol. 1, chapter 12, pp. 865–934.
- Loayza, N. V. (1994), 'Labor regulations and the informal economy', Policy Research Working Paper Series 1335, The World Bank.
- Lucas, R. E. J. (1978), 'On the size distribution of business firms', *Bell Journal of Economics* 9(2), 508–523.
- Rajan, R. G. and Zingales, L. (1998), 'Financial dependence and growth', *American Economic Review* 88(3), 559–86.
- Schneider, F. and Enste, D. H. (2000), 'Shadow economies: Size, causes, and consequences', *Journal of Economic Literature* 38(1), 77–114.
- Shapiro, C. and Stiglitz, J. E. (1984), 'Equilibrium unemployment as a worker discipline device', *American Economic Review* 74(3), 433–44.
- Taylor, M. P. (1996), 'Earnings, independence or unemployment: Why become self-employed?', *Oxford Bulletin of Economics and Statistics* 58(2), 253–66.

CHAPTER 5

Is Neoclassical Economics still Entrepreneurless?

Milo Bianchi and Magnus Henrekson

ABSTRACT. We review and evaluate some recent contributions on modeling entrepreneurship within a neoclassical framework, analyzing how and to what extent the fundamental ingredients suggested in the social science literature were captured. We show how these approaches are important in stressing the main elements of a complex picture without being able to fully describe it. Each modeling attempt focuses only on one specific feature of entrepreneurship, and the entrepreneurial function broadly perceived eludes analytical tractability. As a consequence, the models can be useful in analyzing the effect of entrepreneurial behavior at an aggregate level, but not at explaining individual choices. From these observations, we highlight how a simplistic interpretation of the existing mainstream approaches incorporating entrepreneurship runs the risk of leading to distortionary policy interventions.

Keywords: Entrepreneurship; neoclassical modeling.

JEL codes: B41; D81; J23; L23; M13; O31.

1. Introduction

‘The theoretical firm is entrepreneurless – the Prince of Denmark has been expunged from the discussion of *Hamlet*.’¹ This oft-quoted observation was made by William J. Baumol (1968, p. 68) almost four decades ago in the *American Economic Review*. The article was an urge to the economics profession to start paying serious attention to the role of entrepreneurship in economic development.

A decade or so later entrepreneurship and the entrepreneur began to loom large on the political agenda, and ever since the entrepreneur has from time to time been said to be *the* factor crucial for the furthering of innovation and prosperity. Even as distinguished a social scientist as Ronald Reagan became captivated by these trends, stating

⁰ Magnus Henrekson is at the Stockholm School of Economics and at the Research Institute of Industrial Economics. This article was published in *Kyklos*, Vol. 58 – 2005 – No. 3, 353-377, and it is reprinted with the kind permission of Blackwell Publishing Ltd. Both authors gratefully acknowledge financial support from the Jan Wallander and Tom Hedelius Research Foundation. We have benefited from useful comments and suggestions from Niclas Berggren, Dan Johansson, Pavel Pelikan and Paul Segerstrom.

¹ Schumpeter (1942, p. 86) uses a similar formulation: ‘[A] theoretical construction which neglects this essential element of the case neglects all that is most typically capitalist about it ... it is like *Hamlet* without the Danish prince.’

‘We have lived through the age of Big Industry and the age of the giant corporation. But I believe this is the age of the entrepreneur.’²

There is also systematic empirical evidence that entrepreneurship is important for economic growth (e.g. Audretsch and Thurik, 2000; Carree et al., 2002; Wennekers and Thurik, 2001) and for job creation and renewal (e.g. Acs, 1999). Still, many scholars have pointed out that there has been little room for the entrepreneurial element in theoretical mainstream economics (e.g. Baumol, 1993; Kirchoff, 1994; Kirzner, 1997; Swedberg, 2000). As documented by Johansson (2004) the terms entrepreneur and entrepreneurship are virtually nonexistent in the leading graduate textbooks in micro, macro and industrial organization.

The reason for this disregard of entrepreneurship is not a denial of its relevance for economic development and the organization of economic activity. The reasons are methodological: the entrepreneur and the entrepreneurial function largely elude analytical tractability. In recent years, however, several attempts have been made to include entrepreneurship in mainstream economic modeling, not least in growth models.³

The purpose of this paper is to provide a systematic survey of the different mainstream/neoclassical modeling techniques that have been used in order to capture the entrepreneurial function in the economy. Among other things, we will highlight how and to what extent restrictive assumptions were required in order to attain the equilibrium (as properly defined in the given setting). An important additional purpose is to examine to what extent it may be possible to model the entrepreneurial function more realistically within mainstream economics (“the neoclassical paradigm”).

It turns out that entrepreneurship is invariably defined narrowly and it cannot be said to capture the wide-ranging and complex functions suggested outside mainstream economics. However, from this observation it does not follow that formal modeling of entrepreneurship in mainstream economics is of no value. In fact, while not being able to represent the whole picture, models help highlighting and exploring its main components. Hence, one has to be aware of the limitations of any given approach before drawing policy conclusions.

The paper is organized as follows. In Section 2, we briefly characterize the enormous and highly diverse body of research on entrepreneurship in the social sciences. In Section 3, we review some recent and influential approaches in modeling entrepreneurship, highlighting their main ideas, assumptions, results and possible critiques. In Section 4, we evaluate these modeling attempts in order to assess to what extent they

² Quoted in Brown, Hamilton and Medoff (1990, p. 1).

³ E.g. Romer (1990), Segerstrom, Anant and Dinopoulos (1990), Aghion and Howitt (1992), Helpman (1992) and Acemoglu, Aghion and Zilibotti (2003).

can be said to capture essential features of entrepreneurship and thereby contribute to our understanding of the workings of a modern market economy. In Section 5 we conclude, discussing the usefulness of (neoclassical) modeling and related policy issues.

2. Entrepreneurship in the social sciences

Considering the whole gamut of approaches to the study of entrepreneurship in the social sciences shows that the diversity of approaches and the number of contributions is extraordinary. It is sufficient to go through the recent handbooks and collections to see this (Acs and Audretsch, 2003; Shane, 2002; Westhead and Wright, 2000; Sexton and Landstrom, 2000). In all disciplines from social anthropology to highly abstract economic theory, entrepreneurship is studied, and the focus could be on personality, opportunity, ability, motivation, environment, organization, coordination, policy and finance. Hence, there is no way to summarize entrepreneurship research in a few pages. Fortunately, that is not necessary given the purpose of this paper.

Entrepreneurship research by non-economists is in most cases more descriptive in nature, and the theorizing is usually shaped by the empirical research (Swedberg, 2000, p. 24). But what does this boil down to? Well, a claim such that ‘the entrepreneur is the single most important player in a modern economy’ (Lazear, 2002, p. 1) must necessarily imply that it is about individuals and organizations (be they new, old, large or small) that actively contribute to renewal and change in the economy. It also implies that it is a function, but a function that is carried by specific individuals – individuals who can by their own volition decide whether to supply this function and, given that they choose to do so, do it in ways that are productive, unproductive or destructive from a social perspective (Baumol, 1990; Murphy et al., 1991).

Hence, entrepreneurship is not management, but it also involves more than just alertness and boldness. Its importance hinges on the fact that the real world is replete with Knightian uncertainty (uncalculable risk; Knight, 1921), knowledge is highly decentralized, particularized, largely noncodifiable (Hayek, 1945), human cognitive abilities are severely limited relative to the available information (Martens, 2004), and human action is an open-ended process rather than the result of dynamic optimization (Kirzner, 1997).

Moreover, as emphasized by Casson (1982/2003, p. 23) the entrepreneur is someone who ‘specializes in taking judgmental decisions about the coordination of scarce resources’. But because of the characteristics of the real world listed above, entrepreneurs will differ in decisions and actions. The essence of entrepreneurship is therefore about doing things differently. As long as opportunities for improved coordination

and innovation exist, entrepreneurial services will be socially valuable. However, getting entrepreneurial insights does not ascertain that these insights are translated into entrepreneurial action.

Trying to synthesize all this, we will settle on the following definition of entrepreneurship (partly adapted from Wennekers and Thurik, 1999): the ability and willingness of individuals, both on their own and within organizations to: (i) *innovate*, i.e. perceive and create new economic opportunities; (ii) *face uncertainty*, i.e. introduce their ideas in the market, by making decisions on location, form and the use of resources and institutions; and (iii) *manage their business* by competing with others for a share of that market. These three attributes will also accompany us in the next section, where we will analyze in some detail how and to what extent they have been captured in formal modeling.

What is already clear is how formidable is the challenge to model entrepreneurship analytically. As noted, Baumol was very pessimistic in 1968 about the possibility of integrating it within the neoclassical theory of the firm (as developed in the 1930s by scholars like John Hicks and Paul Samuelson): ‘there is no way in which they [entrepreneurs] can fit into the model’. Some two decades later Barreto (1989) presented a comprehensive analysis of the exclusion of the entrepreneur from the theory of the firm. He claims that it is indeed an infeasible task (p. 115, 141): ‘The confrontation between the basic axioms and the entrepreneur leaves two possibilities: to accept the entrepreneur and reject the modern theory of the firm, or to reject the entrepreneur and maintain allegiance to the modern theory of the firm... Simply put, entrepreneurship is above ‘formalization’ – it cannot be neatly packaged within a mechanistic, deterministic model. Importantly, the choice is an ‘either-or’ proposition; there is no happy medium. The corner solution which economic theory has chosen is consistency and for this reason the entrepreneur disappeared from microeconomic theory’.

However, some significant attempts have been made in recent years to model entrepreneurship in mainstream economics. Are these attempts largely aborted, or do they contribute to our understanding of the real world? This is the question to which we now turn.

3. Modeling entrepreneurship in neoclassical economics

If you say mainstream or neoclassical economics, basically every academic economist knows what you are referring to, at least as long as you do not ask for a formal definition. Still it is fair to say that the core of the neoclassical research paradigm is characterized by the study of the allocation of scarce resources, optimization, rationality, focus on marginal tradeoffs and relative prices, methodological individualism, the

use of calculus and a general equilibrium conception of the economy (e.g., Aspromourgos, 1987; Colander, 2000), or more colloquially, greed, rationality and equilibrium. However, as recently noted by numerous scholars (e.g. Solow, 1997; Colander, 2000, 2003; Schiffman, 2004), contemporary mainstream (micro) economics is far less dogmatic and is almost wholly defined by its method. It is about building models that are tested or that can at least be tested in principle. Hence, it is no surprise that ideas that cannot be modeled formally tend to be ignored in economics. It is then also logical, given the enormous interest in the entrepreneur and entrepreneurship in the public discussion, that there are a number of recent attempts to model entrepreneurship formally while remaining within the confines of mainstream economics.

In our review, we will not attempt to account for every author that has somehow addressed the issue, but rather to provide a systematic survey of the more representative contributions of a given idea of entrepreneurship.⁴ In our survey, we have identified three crucial attributes of entrepreneurs that have been stressed in the more substantive modeling efforts: (i) they are (generally speaking) “more talented”, (ii) they have greater tolerance towards risk-bearing and (iii) they are innovators (either as individuals or as firms involved in R&D activities). Within these broad strands we look more closely at the most influential works, analyzing how and to what extent the entrepreneurial function can be said to be captured in the models, their driving assumptions and main results (with particular focus on the existence and properties of the equilibrium).

The analysis will be conceptual rather than technical in order to stimulate some reflections on the concept of entrepreneurship and on the possibility to model it in a neoclassical framework.

3.1. Talented entrepreneurs. The model in Lucas (1978) focuses on the coordinating function of the entrepreneur. This basic feature goes back to Say (1845) and commonly characterizes the entrepreneur in economic theory. Each firm, irrespective of the size of its workforce, must have an entrepreneur (or equivalently, in Lucas’ terminology, a manager) in order to produce.⁵ He will organize the production employing the optimal level of labor and capital and will be remunerated by residual profits (i.e. by the value of output after having compensated the factors of production).⁶ In the model, workers are assumed to be completely homogeneous with respect to their productivity

⁴ A recent and comprehensive survey of the ‘the economics of entrepreneurship’ can be found in Parker (2004).

⁵ This indivisibility between output and entrepreneurial labor is typical when the entrepreneur is modeled as a coordinator of production. Technically, the nonconvexities created in this formulation do not constitute a problem for the existence of an equilibrium, since it is usually assumed that the economy is populated by a continuum of individuals (Aumann, 1966).

⁶ Notice however that there is no role for uncertainty in this model; everything is deterministic and agents are assumed to be risk neutral. Also, there’s no separation between ownership and control.

as employees, while they are endowed with different “talents for managing”. This variable will play a crucial role in the model, determining the choice between becoming a worker or an entrepreneur and the allocation of the workforce among entrepreneurs.

For a given standard constant return to scale production technology $f(l, k)$, the output actually produced y will depend on the talent x of the manager according to the following formulation

$$y = x \cdot g[f(l, k)],$$

where $g[\cdot]$ is increasing and concave.

We may already note how “entrepreneurship” is considered as a factor of production, as in Schultz (1980), where entrepreneurship enters the production function as any input whose price is determined by the interaction of supply and demand.

Although very convenient from an analytical point of view, and thus generally embraced in neoclassical approaches, the formulation is somehow at odds with the message in fundamental contributions in the field such as Knight (1921), Schumpeter (1934) or Kirzner (1997). They in fact rejected the idea of reducing entrepreneurship to a mere factor of production comparable to physical capital or “normal” labor. Instead, they insisted on the impossibility to quantify/identify its specific contribution to final output. This is necessary in the neoclassical approach, so that one can solve the maximization problem by using the customary tools (Barreto, 1989).

Entering the production function simply as a multiplier, talent has a huge effect on output. The effect is mitigated by the function $g[\cdot]$, which introduces diseconomies of scale in managing. This ensures that there exists an optimal size of the firm for any given x , and hence the most talented individual cannot control everything in equilibrium.

Regarding the first basic occupational choice, the optimal rule will identify a cut-off level x^* (increasing in factor prices and decreasing in $g[\cdot]$) such that everyone given a talent larger than x^* will become an entrepreneur and the others will be employees. The intuition behind the result is straightforward and robust to different specifications of the model; all that is required is that productivity differs among otherwise homogeneous individuals (with respect to a number of relevant parameters like initial wealth or risk aversion) and income for entrepreneurs is more sensitive to individual characteristics than for employees. Laussel and Le Breton (1995), for example, derive the same result in a slightly different general equilibrium model, where imperfect information on the quality of workers creates a “lemon effect” on the labor market and induces the most efficient workers to become entrepreneurs.

Lucas’s contribution also addresses the issue of the optimal distribution of factors of production across managers/entrepreneurs. Adding some more structure to the model,

and in particular imposing independence between firm growth, in terms of labor and assets employed, and size (Gibrat's Law), Lucas is able to derive the competitive equilibrium in the economy, characterized by a famous result: the most talented individual will manage the largest firm and earn the highest profits. This applies keeping fixed the distribution of talent (that intuitively gives an idea of the level of competition in talent required to manage a firm of given size) and factor prices (that determine the minimum threshold x^* needed to become an entrepreneur).⁷

As Lucas recognizes, the model does not specify exactly what managers do, but simply assumes that 'whatever managers do, some do it better than others'. The model thus does not really explain its crucial variable, "talent", and is incapable of addressing more realistic issues relating for example to the investment in one's own "talent" or to the dynamics of occupational choice.

The model draws attention to the fundamental issue of the relationship between skills and occupational choice. From an empirical perspective, for the reason just mentioned, its predictions are hardly testable, unless we assume something on what the variable "talent" refers to. Loosely speaking, it is quite unintuitive to think about the topic in such a clear-cut and monotonic way, where entrepreneurs are simply more talented than employed workers. In fact, it has been debated in many disciplines whether there is positive or negative self-selection into entrepreneurship, and empirical studies are not conclusive. For example, Evans and Leighton (1989) find that entrepreneurs are more likely to have experienced low wages and unemployment (they are the "misfits"), while Robinson and Sexton (1994) report that entrepreneurs are on average more educated than salary workers and Bates (1995) documents how these results vary greatly across industries.

Hence, much depends, on how one defines (or tries to proxy for) entrepreneurial "talent". An important contribution towards the understanding of this complex concept was recently made by Lazear (2002). He stresses the point that becoming an entrepreneur (as opposed to a specialist) requires having a more balanced talent that spans a number of different skills. To model the occupational choice in this fashion, it is assumed that, for given skills x_1 and x_2 , a specialist can earn $\max[x_1, x_2]$ while an entrepreneur will earn

$$\lambda \cdot \min[x_1, x_2],$$

where λ denotes the market value of entrepreneurial talent. Then the closer you get to the space where x_1 equals x_2 , the more likely that the agent becomes an entrepreneur.

⁷ More testable implications can be derived with further assumptions. In particular, assuming an elasticity of substitution in production less than unity will imply an average firm size increasing in per capita wealth in the country.

This probability is also affected by the parameter λ , which accounts for a number of technological (e.g. economies of scale or agency costs) and market (supply and demand of entrepreneurs) variables.

The empirical validity of the theory is still quite preliminary, but some support has been given in Lazear (2002) and Wagner (2003). The key factors in promoting entrepreneurship here reside on the possibility to train these “jack-of-all-trades” attitudes and to act on institutional factors comprised in the market return for entrepreneurship λ . The first point is currently being investigated empirically (Baumol, 2004 and Silva, 2004) and there seems to be little scope for policy intervention on this side: entrepreneurial talent appears to be mostly innate and difficult to teach. The second aspect is fundamental and it would clearly require further specifications. One interesting connection, that we can just mention here, leads to the recent literature on how different incentives determine the occupational choice of the most talented people in a country, whose growth rate will heavily depend on the choice of these people to devote their talent to productive vs. rent-seeking activities (Baumol, 1990; Murphy, Shleifer and Vishny, 1991; Holmes and Schmitz, 2001).

3.2. Risk-bearing entrepreneurs. The influential work of Kihlstrom and Laffont (1979) adds the element of risk bearing to a neoclassical general equilibrium model. Entrepreneurs not only provide organizational skills, but “take the responsibility of enterprise”, i.e. bear the risk associated with production.

The model is quite similar to that of Lucas (1978), although more refined: Individuals are homogeneous (in particular, they have the same ability in performing entrepreneurial and employee functions) except for their degree of risk aversion. Their basic decision is whether to be an employee or an entrepreneur. The latter foregoes a fixed wage in exchange for risky profits and decides also how many workers to employ.

Each firm is required to have one unit of entrepreneurship in order to operate and production is a function depending on labor input and a stochastic factor with an objective (i.e. known to everybody) and common (i.e. independent of characteristics of the firm or the entrepreneur) probability distribution.⁸ The equilibrium defines the partition of the population (i.e. decides who becomes worker and who is entrepreneur, based on agents’ expected utility maximization) and the resulting prices (wages) so that the labor market clears.

⁸ This assumption leaves no room for interactions between individual capacities and final outcome. Thus, situations where more able individuals have more accurate inferences or are less affected by global shocks are ruled out. However, it does not imply that shocks are perfectly correlated across firms, i.e. that the realization of the random variable will be the same for everybody.

Under standard assumptions on the utility function (everywhere continuous in the degree of risk aversion and strictly concave in income) and the production function (decreasing return to scale in labor), there exists an equilibrium where less risk averse individuals become entrepreneurs. More precisely there exists a unique cut-off level of risk aversion r^* such that all agents with risk aversion smaller than r^* will be entrepreneurs and the rest will be workers.⁹

In a fashion very similar to Lucas (1978), the less risk averse entrepreneurs will hire more workers in equilibrium, even if each firm faces the same decreasing return to scale production technology (and thus from an efficiency point of view should produce the same amount).¹⁰ This will hold if uncertainty enters the production technology multiplicatively or in general if in good states of the world workers are more productive.¹¹ Intuitively, this guarantees that the expected effect of a shock unambiguously determines the amount of risk to bear in order to run a firm of a given size.

The paper has turned out to be a fundamental contribution in many respects. It was probably the first study to extend the results regarding the existence and stability of equilibrium to a model comprising (some type of) entrepreneurship. Here the comforting point is that no departure from rational choice, well-defined production functions or complete information is required to invoke a need for entrepreneurs. Some heterogeneity in preferences is enough to justify a role for the entrepreneur in equilibrium, as long as risk cannot be insured.

Second, it provides a formal treatment of the theory of the entrepreneur as risk bearer, which is definitely one of the most plausible aspects commonly associated with entrepreneurship. At a somewhat deeper level, however, one may question whether bearing risk *per se* is a defining characteristic of entrepreneurs, or whether it emanates from external constraints (e.g. capital market imperfections). We will return to this issue shortly, trying to get some hints from the empirical literature.

Finally, it provides a plausible justification for a positive relationship between initial wealth and entrepreneurship. Differences in risk aversion may arise because individuals

⁹ It is also unique if we assume that the number of workers hired by an entrepreneur (with a given degree of risk aversion) decreases in the equilibrium wage. Furthermore, (Lyapunov) stability of the equilibrium is shown assuming a particular tâtonnement adjustment process and some more assumptions to assure that the utility function is “well-behaved” with respect to risk aversion.

¹⁰ Lucas himself points out (footnote 1) that the spirit of his paper can be replicated substituting the variable “talent for managing” with “attitude towards risk”.

¹¹ Formally, if x is the stochastic factor entering the production function $g(l, x)$, we say that x enters multiplicatively if we can write $g(l, x) = xh(l)$, as in Lucas’ formulation. What is required in the more general case is that $g(l, x)$ and $g_l(l, x)$ are both monotonic in x , i.e. a more favorable realization of x not only increases output but also the marginal product of labor. Under all these conditions it is also shown that a general increase in risk aversion, increasing the supply and decreasing the demand for workers, reduces the equilibrium wage.

with the same utility function differ in initial wealth levels. Under the widely accepted assumption that risk aversion decreases in wealth, the model implies that those who are initially wealthier will become entrepreneurs.

However, while there's little to criticize on the formal structure of the model, some concerns remain regarding the last two points. First, to be consistent with Knight's terminology, risk should be distinguished from uncertainty. Risk is a stochastic process with a known distribution, while uncertainty has to be handled with no information. And while Knight defines the entrepreneurial role as assuming responsibilities in an *uncertain* environment, the authors choose to model *risk*-bearing.¹² What (rational) decision making can deal with is in fact only risk, as one always needs an idea at least on some moments of the underlying probability distribution.¹³

Second, from a theoretical point of view, the recent literature on incentives and endogenous matching reveals the inconsistency of this interpretation of the Knightian theory of entrepreneurs as providers of insurance once we relax the assumption that the choice of occupation is the only institutional scheme of risk sharing. The fact that, in Kihlstrom and Laffont's model, there exists some institutional constraint to risk trading (basically the absence of a capital market) is really crucial for the result.

In fact, risk averse entrepreneurs would prefer to purchase some insurance, if a market for that would be opened. On the other hand, because of moral hazard effects perfect insurance would give entrepreneurs incentives not to supply the necessary effort. Moreover, if utility is separable in income and effort, wealthier agents will have to bear more risk in order to satisfy their incentive compatibility constraint for any given effort level. And here is the surprising result in Newman (1999): For a broad class of utility functions, the decrease of risk aversion in wealth is too slow compared to the increase in the level of risk to be borne. So the "increasing risk effect" dominates the "decreasing

¹² The distinction will be also important in order to reconcile Knight's and Schumpeter's theories, usually seen as contrasting (e.g. Evans and Jovanovic, 1989). Schumpeter makes a clear distinction between the risk bearer and the entrepreneur (who deals with uncertainty, i.e. directs his actions in new and unpredictable directions). Risk has to be borne by the capitalist, not necessarily by the entrepreneur, who handles uncertainty.

¹³ An interesting avenue in searching for foundations of the entrepreneurial choice might take into account subjective probability theory, where beliefs do not come from an exogenous and objective probability distribution (as in the von Neumann – Morgenstern theory) but where they are rather derived by individual preferences and thus will typically vary across individuals facing the same situation (see Kreps, 1988, for an introduction). A direct formalization of Knightian uncertainty can be found in Bewley (1986), where individuals possess a set of subjective probability distributions over states rather than a single (objective) one and make incomplete plans of action, that consider the possible occurrence of unexpected alternatives and state of the world. Unfortunately this literature has not inspired explicit applications to entrepreneurial choice and this seems definitely a worthwhile task.

risk aversion effect” and we obtain the empirically implausible prediction that wealthy agents will choose to be workers while the poor become entrepreneurs.¹⁴

The positive association between wealth and entrepreneurship can instead find an explanation recalling a basic result in incentive theory: Once the model is purged from risk-sharing considerations (i.e. when entrepreneurs are assumed to be risk neutral), higher wealth actually mitigates moral hazard problems in financial markets. In this setting greater wealth does not increase the likelihood of becoming an entrepreneur because of reduced risk aversion, but thanks to better access to the credit market.

Altogether, this casts some doubt on the view of entrepreneurs as risk bearers once an (imperfect) insurance market is opened. And these two points remain highly controversial, even when one moves on to the empirical side. On the role played by risk, for example, Parker (1996) and Ilmakunnas and Kanninen (2001) document that the degree of uninsurable risk that characterizes an economy drives individual occupational choice. Van Praag and Cramer (2001) stress the importance of risk aversion while Rosen and Willen (2001) report evidence that casts doubts on the “Knightian” interpretation that risk attitudes are a major determinant of the decision to become self-employed.

Similar debates can be found on the relationship between personal wealth and entrepreneurship. The crucial role of individual wealth has been largely recognized in many empirical studies, stressing the existence of liquidity constraints that restrict the possibilities for business start ups (e.g. Evans and Jovanovic, 1989; Evans and Leighton, 1989; Holtz-Eakin et al., 1994; Blanchflower and Oswald, 1998). On the other hand, recent studies (e.g. Dunn and Holtz-Eakin, 2000; and Hurst and Lusuardi, 2004) have reported no significant wealth effect and rejected the claim that these constraints play a key role in determining the propensity to enter self-employment.

3.3. Innovative entrepreneurs. Important contributions come from the acknowledgement of the centrality of innovation in the process of development. Most notably, recall Schumpeter’s theory of entrepreneurship as the engine of the capitalist economy. Implementing innovations (e.g. introducing new goods, new methods of production, new markets. . .) is regarded as the key feature of entrepreneurial activities, which are clearly separated from risk bearing or invention. We will review the most important attempts made to comprise this idea in formal models.

¹⁴ Technically, the total utility $U(y, e)$ can be written as: $U = u(y) - e$, where y denotes income and e effort. The result requires that $1/u'(y)$ is a strictly convex transformation of $u(y)$. Empirically, the fact that wealthy agents choose safer occupations than poor agents is not so implausible once these agency problems (and thus the endogeneity of matching between principals and agents) is taken into account. Akerberg and Botticini (2002), for example, find that in Medieval Tuscany wealthy farmers grew safer crops (cereals), while the poor peasants were dedicated to wine (riskier).

More recently, endogenous growth theory has developed some Schumpeterian ideas, focusing in particular on the process of innovation created by R&D activities. While not really consistent with the Schumpeterian idea of entrepreneurship (as we will argue in what follows), this literature has become the most prominent neoclassical attempt to model innovation, obsolescence and “creative destruction”.

3.3.1. *Entrepreneurs responding to recurrent profit opportunities.* Several schools of thought, in particular the Austrian School, have described the entrepreneur as somebody who discovers and responds to new economic opportunities. Holmes and Schmitz’s (1990) model refers explicitly to Baumol’s approach on the capacity to ‘respond to the opportunities for creating new products that arise from repeated technological breakthroughs’. This, and not risk bearing, is the crucial feature that defines an entrepreneur.

The model has two key features: individuals are continuously faced with opportunities for developing new products, i.e. the dynamics of growth (technical progress, demographic change) always create “disequilibria” and thus potential profits. Second, individuals differ in their ability to exploit these emerging opportunities. The first feature is modeled assuming an exogenous rate of technical progress and the second one by defining an ability parameter, θ , that mitigates the risk of creating a new business: the more talented agents are less likely to undertake an unsuccessful enterprise.

Specifically, businesses differ in their quality q : a unit of time allocated to managing a business of quality q will result in the production of q units of output. Given technical progress, the average quality of new businesses increases over time by some exogenous and constant parameter. Individual ability mitigates risk in the sense that for a given θ , q_t is a random variable with the cumulative distribution function $F(q, \theta) = Pr(q_t \leq q \mid \theta)$ and $F(q, \theta_1) > F(q, \theta_2)$, for $\theta_2 > \theta_1$; the higher the ability the greater the probability of starting up a high quality business.¹⁵

In the model, the occupational choice of individuals is derived in relation to their abilities and to the quality of the business that they are currently managing. They have to decide whether to manage their current business (and producing deterministically a consumption good) or starting-up a new one with stochastic quality q and incurring the fixed cost of one foregone period of production.¹⁶

¹⁵ Formally the distribution $F(., \theta_2)$ is strictly first-order stochastic dominant with respect to $F(., \theta_1)$, for $\theta_2 > \theta_1$.

¹⁶ It is interesting to note that individuals differ in their activities, not necessarily in their status of employee or self-employed. The model would not be different if the same “developing a new opportunity” vs. “managing an existing one” would be undertaken by an employee. Schumpeter (1934, p. 78) stated that ‘whatever the type, everyone is entrepreneur only when he actually carries out new combinations and loses that character as soon as he has built up his business, when he settles

Quite intuitively, in the unique competitive equilibrium people with low θ only manage already existing firms, people with high θ only set up new businesses, and people with intermediate θ either manage businesses they started or trade for higher quality businesses (dismissing their current one).

The major contribution consists in the simple and intuitive formalization of an idea that has been the central point of the analysis of authors like Schumpeter, Kirzner and Schultz, i.e. the existence of individuals with special abilities to innovate and capture profit opportunities.

The shortcomings concern the exogeneity of the technological progress and, once again, the focus on risk rather than Knightian uncertainty. The first aspect may not be crucial in a theory of entrepreneurship, once we stick to the Schumpeterian observation that ‘the inventor produces ideas, the entrepreneur gets things done’. In other words, the real focal point is not where “new opportunities” come from, but rather how some people respond to them and drive development.¹⁷ On the other hand, the recent and vast literature of endogenous growth and innovation (partly reviewed in the next subsection) may provide this missing ingredient, thus suggesting a possibly interesting combination of the two approaches.

On the second aspect, it seems that focusing solely on uncertainty would not allow the construction of an occupational choice model, where one can explain the motives for becoming an entrepreneur rather than a worker. As already mentioned, modeling individual choice requires either to somehow assign probabilities to different possible outcomes or to dismiss (standard) decision theory.

A different approach is proposed by Blanchflower and Oswald (1998). They create a simple and testable model of entrepreneurship, where people with different abilities to innovate act in a world of genuine uncertainty and where information asymmetries create imperfection in the capital market. Here entrepreneurs have not only to be gifted with entrepreneurial vision (in the model it is simply assumed that a fraction of the population is endowed with it), but must have enough capital or be “lucky” enough to receive an unsecured loan (again, there is an exogenous fraction of these “lucky” people in the population). Unsurprisingly, when there is a shortage of “entrepreneurial to running it as other people run their business’. The focus is therefore not really on a category of person, but rather on a function.

¹⁷ Formalizing how exactly entrepreneurs create new combinations is not an easy task. A possible approach is suggested in Olsson and Frey (2002). Entrepreneurs create growth through “normal” technological advance, i.e. they innovate by creating convex combinations of existing ideas within a set of technological possibility. The process will obviously terminate when the set will become convex, and we will have stagnation until an exogenous paradigm shift (extraordinary technological breakthrough) will reconfigure the technological possibilities, reintroducing non-convexities, new profit opportunities and ensuring long-run growth. Once again, “extraordinary” technological progress is not explained in the model.

people” with capital, they will earn a rent, i.e. they enjoy a higher utility than other workers do. The paper has mainly empirical concerns (and indeed finds evidence that liquidity constraints bar some potential entrepreneurs from becoming self-employed, and that self-employed persons report higher levels of satisfaction than employees), but it highlights an important point in our discussion. One can stick to the idea that returns from innovations are impossible to quantify (and thus incorporate uncertainty), but then any ambition to provide an explicit model of individual choice has to be abandoned.

3.3.2. *Innovation as a product of the R&D sector.* This recent and burgeoning literature builds on the crucial role of technological progress in the explanation of long-run growth. Innovation comes from purposive, profit-seeking investment in knowledge by firms involved in R&D activities, modeled following the patent race literature in industrial organization (see Tirole, 1998). Innovations typically emerge according to some stochastic process dependent on the amount of resources devoted to the sector.¹⁸ Moreover, it completely displaces previous products or leaders in an industry, allowing the winner to enjoy some monopolistic power until another innovation arrives. Thus there are winners and losers in the process; the obsolete products leave the market replaced by the new ones, in a Schumpeterian “creative destruction” fashion.¹⁹

What is not really Schumpeterian, however, is the way innovation takes place. There is no “new man” involved in novel and extraordinary activities, but rather a large scale, routinized, (almost) predictable path of innovation via investments in R&D.

Thus, a quite distinct conception of the innovative entrepreneur emerges. This conception does not find a proper justification in Schumpeter’s theory, but rather in one of his early critics, namely Solo (1951): ‘Innovation is more realistically analyzed as an ordinary business activity than as the extraordinary efforts of new firms and new men; invention and innovation are subject to costs and result in revenues like any other business activity; and both are carried out in competitive struggle by firms which are at once producers and innovators’.

¹⁸ The most common assumption is that the arrival of innovations is governed by a Poisson process with intensity proportional to the labor force in the R&D sector. While convenient analytically, the assumption is at odds with some stylized facts. First, the basic assumption underlying the Poisson process (the probability of two arrivals in a short period of time is practically zero) is not so intuitive given the (Schumpeterian) observation that innovations are often lumpy (they come in cycles). Second, the total amount of resources spent may be “adjusted” for by some institutional factor (see e.g. the evidence on Sweden vs. the U.S. presented in Henrekson and Rosenberg, 2001).

¹⁹ See for example Segerstrom et al. (1990); Aghion and Howitt (1992); Grossman and Helpman (1994); Dinopolous and Segerstrom (1999). We should also mention the pioneering efforts in modeling this kind of Schumpeterian competition by Futia (1980) and Grabowski and Vernon (1987). In many ways they anticipated the flavor of the recent research on the determinants of innovation.

It follows directly that if innovation is based on deliberate inventive efforts on the part of the firm, the resources used in this activity are a form of production factors. And if entrepreneurship becomes a factor of production, it may well be remunerated by the capitalist as other factors. Unsurprisingly then, in these models innovation is carried out by “highly skilled managers”, who receive fixed payments and thus bear no entrepreneurial risk (see e.g. Acemoglu et al., 2003).

The models were designed to study issues of growth, trade, technological convergence and there is not really much concern for providing an explicit formalization of entrepreneurship (modeling for example individual occupational choice and entrepreneurial effort). However, we find them relevant in our review since they impose quite strong assumptions on how innovations come about, and thus on the role of entrepreneurs in modern economies.

In particular, this approach challenges the traditional Schumpeterian distinction between inventor and entrepreneur that views R&D activities as mere providers of inventions (and look for the engine of development elsewhere). Indeed, in Schumpeter’s later writings (1942), one of the reasons for the decline of capitalism is its transformation into a “trustified” system, dominated by a small number of large firms, where ‘innovation itself is being reduced to routine. Technological progress is increasingly becoming the business of teams of trained specialists who turn out what is required and make it work in predictable ways’.

Assessing the relative importance of these theories (and thus also testing two different views on the actual “engine of growth”) is ultimately an empirical issue. But, as it has been a recurrent theme in this section, empirical studies are not conclusive. The relationship between innovative capacity and firm size or age, for example, is ambiguous and strongly relies on the specific definition of innovative activities adopted (as reported e.g. by Acs and Audretsch, 1988; Acs and Audretsch, 1995; Audretsch and Thurik, 1997).

At an intuitive level, one can say that small firms (or individual entrepreneurs) are more likely to introduce path-breaking ideas and “new combinations”, and that large firms will devote their activities (like R&D) to revise, refine and spread the innovation on a larger scale. And, instead of struggling within the debate of individual entrepreneurs vs. large firms it seems reasonable to embrace a “complementarity view”, where development requires both research effort to produce innovations and entrepreneurial/managerial skills to implement them, both the independent entrepreneurial innovations and routine R&D processes in large firms (Scherer, 1980; Baumol, 2002; Michelacci, 2003; Acs et al., 2004).

Table 1. *The Major Neoclassical Attempts to Model Entrepreneurship – A Summary*

Idea	Main contribution	Pros	Cons
“talent for managing” (<i>coordination role</i>)	Lucas (1978)	<ol style="list-style-type: none"> 1. formalization of a basic idea (J.B. Say) 2. model of occupational choice 	<ol style="list-style-type: none"> 1. no precise meaning of talent (not testable) 2. entrepreneurship as factor of production
“tolerance towards risk” (<i>mechanism to allocate risk</i>)	Kilbstrom and Laffont (1979)	<ol style="list-style-type: none"> 1. equilibrium analysis with entrepreneurship 2. formalization of Knight’s idea 3. explanation of wealth and entrepreneurship 	<ol style="list-style-type: none"> 1. no distinction between risk and uncertainty 2. not robust to the opening of capital market
“capacity to capture new profit opportunities” (<i>innovation</i>)	Holmes and Schmitz (1990)	<ol style="list-style-type: none"> 1. “disequilibrium” analysis 2. precise idea of entrepreneurial abilities 3. occupational choice 	<ol style="list-style-type: none"> 1. exogenous technological progress 2. no uncertainty
“working in the R&D sector” (<i>innovation</i>)	Aghion and Howitt (1992) Segenstrom et al. (1990)	<ol style="list-style-type: none"> 1. endogenous technological progress 2. formalization of “creative destruction” 	<ol style="list-style-type: none"> 1. no explicit modeling of entrepreneurs 2. the entrepreneur is defined as the inventor
Contribution			
Lucas (1978)	Production requires coordination	Those more talented in managing	
Kilbstrom and Laffont (1979)	Production is risky	The less risk averse	
Holmes and Schmitz (1990)	New profit opportunities always occur	Those talented in capturing opportunities	
Aghion and Howitt (1992)	Growth requires innovation	Firms involved in R&D activities	
Why entrepreneurs?			
Who is the entrepreneur?			

4. Summary and assessment of the contributions

In line with the definition of entrepreneurship given in Section 2, the previous section has identified three possible justifications for the role of entrepreneurs in the modern neoclassical theory of the firm. In general, they highlight some “special features” that distinguish entrepreneurs from other workers. The major modeling strategies are summarized in *Table 1*.

In Lucas (1978), the presence of entrepreneurs is explained by assuming that production requires special coordination skills that are embodied in particular individuals. Some people will become entrepreneurs since they are endowed with more “talent for managing”, i.e. they are able to guarantee a higher production for any given level of technology and factor employment. In Kihlstrom and Laffont (1979), production is risky, and the only way to allocate risk among people is through their occupational choice. Entrepreneurs emerge as those with more tolerance towards risk, those willing to provide income insurance to other workers in return for the right to the residual profits. In Holmes and Schmitz (1990), the world is always in disequilibrium, since technological, demographic and institutional changes continuously create new profit opportunities. Some people, who will be entrepreneurs, have a greater capacity to capture these opportunities, in the sense that they are more likely to start up a “high quality” enterprise. A somewhat different approach is taken in Aghion and Howitt (1992) and in much of the recent growth literature. Here innovation is still the engine of long-run growth, but its source is not the task of “extraordinary individuals”, but rather of firms that devote resources to large scale, routinized R&D activities.

All these contributions highlight and formally analyze *one* particular feature of entrepreneurship. If one wants a theory that captures the essentials of entrepreneurship at the level of the individual entrepreneur, these models are too restrictive, since the entrepreneurial function is far more complex. Rather than providing an exhaustive theory, these contributions should be viewed as pieces of a complex puzzle. In fact, it is not difficult to recognize that being a good coordinator without having an innovative mind, for example, is probably not enough to become a successful entrepreneur. It thus seems pointless to completely dismiss one approach in favor of another; they are both useful but fill different needs.

We have seen how the genuine uncertainty inherent in entrepreneurial activity is always missing. All risks are calculable, i.e. “known to be knowable”, so that the gathering and processing of information can be treated as a regular business cost amenable to optimization. Also, risk bearing does not appear as a fundamental attribute to define entrepreneurship, especially if we consider a more realistic setting with capital markets and agency conflicts.

We have reviewed what we deem to be the most influential papers concerned with innovations, either as a creative product of people with special talents or as a more predictable result of R&D activities. These two different approaches are somehow representative of a deeper debate, which leads directly to the foundation of traditional neoclassical analysis. Either we think of innovation as a perturbation of the system that will soon adapt and reach a new equilibrium. Or, we are willing to develop a more genuine “disequilibrium approach” that recognizes as a central argument of the analysis a world where possibilities of “new combinations” continuously arise and thus people are in fact acting “out of equilibrium”. In our view, a theory of entrepreneurship cannot abstract from this essential issue. If one believes that risk bearing and coordination are not crucial or not sufficient features to define entrepreneurship, then “in equilibrium” a role for entrepreneurs may be hard to find. At the end of the day, the “neoclassical” entrepreneur really needs a deep understanding of the dynamics of disequilibrium to become “really entrepreneurial”. And probably economic theory needs that too.

5. Useful models and useful policies

Can the entrepreneur at all be endogenized in an empirically meaningful way within an equilibrium framework? Or rather, can we model the entrepreneur as we would realistically like to see him or her within that theoretical framework? Rosen (1997), among others, is skeptical. The general equilibrium model is a powerful analytical tool. But analytical tools narrow the range of phenomena that can exist (within the model), and hence limit understanding to that restricted set of phenomena. Increased realism quickly implies loss of the powerful tools of calculus and game theory, and there may be no well-defined solution to the resource allocation problem. The entrepreneur, even though probably of critical importance for innovation and growth, ‘lacks operational definition’ and is too elusive to ever fit into the neoclassical model.

The complex, serendipitous, open-ended entrepreneur continuously obligated to deal with genuine uncertainty cannot be modeled in neoclassical economics, not even in a probabilistic sense. This presupposes calculable outcomes, while entrepreneurship is a process of discovery where one thing leads to another, and the discovery and exploitation of opportunities in turn create previously unknown opportunities, not even known to be knowable. In this sense, the neoclassical entrepreneur is not entrepreneurial.

But this is not necessarily a devastating blow to neoclassical modeling. While an individual entrepreneur cannot be modeled analytically, at the more aggregated level (the market, the industry) statistical “laws” may apply and expected regularities are likely to exist. A good model may summarize how a large number of economic agents

behave, and it can therefore be used to predict the likely change in the system resulting from a change in one of its parameters (a tax rate, a regulatory measure etc.). This could be the case despite the fact that the model is useless in describing an individual entrepreneur.²⁰

One way of understanding this is to start from Baumol (2002), where the capacity of the free-market economy ‘to produce a stream of applied innovations’ and a rapid rate of growth is examined. He stresses the combined and highly powerful effect of entrepreneurial innovation and routine/systematic innovation in incumbent firms, in particular the large firms in oligopolized industries. Capitalism is unrivalled when it comes to innovation and economic growth, but this presupposes both individual entrepreneurs and large oligopolistic firms. In most cases major innovations emanate from the ingenuity and serendipity of individual entrepreneurs (Scherer, 1980), but routinized step-by-step improvements of the original innovation by large (or rapidly growing) firms is then required to reap the full benefits of the capitalist organization of the economy.

Neoclassical modeling efforts of entrepreneurship capture important characteristics of these latter activities, even if they do not provide a useful description of the behavior of individual entrepreneurs. These models may adequately capture the dynamics of the whole population of entrepreneurial firms conditional upon the institutional framework within which they operate. The number of firms in mature industries is small, the operations of each firm is highly predictable; and despite the creative destruction of entrepreneurs, economic systems do not explode or implode, and no individual firm ever becomes infinitely large. At the same time, there is an enormous amount of churning and disequilibrium at lower levels of aggregation (Davis, Haltiwanger and Schuh, 1996; Kirchoff, 1994), but in a well-designed economy there exist entrepreneurs that equilibrate rather than disequilibrate. Or, if somebody disequilibrates by introducing a new technology there will, given favorable circumstances, be many entrepreneurs ready to exploit the new opportunities arising when the structure of the economy is adjusted to the new technology.

Equilibrium could still be relevant, although not in the traditional sense of a system that is stable or changing in the same way period after period. Economic systems are complex, humanly devised systems continuously evolving and adapting both to changing external circumstances and to internal changes and disruptions brought about in unpredictable ways by entrepreneurs (Colander, 2003; Pelikan, 2004). The “equilibrium” of a certain economy is in many ways unique in terms of rate of turnover of firms,

²⁰ An analogy can be drawn to molecular physics, where it may be impossible to describe the process by which the individual molecules move, and still one may be able to construct a model that captures very well how the whole system of molecules behaves.

industry structure, aggregate growth rate, rate of employment, degree of entrepreneurial activity etc. This can be expected to be largely a function of the institutional setup.²¹ Sometimes a great deal of entrepreneurship is needed to overcome the detrimental effects of dysfunctional institutions, sometimes favorable institutions spur a great deal of productive entrepreneurship that exploits opportunities that arise. In other, less propitious, circumstances what is profitable entrepreneurship from an individual perspective is wasteful rent-seeking from a social perspective (Acemoglu et al., 2004).²²

Few observers question the importance of entrepreneurial effort for economic development and increased prosperity. But, given the alleged importance it is probably still true that net wealth accumulation by entrepreneurs is not that large. Although one can always point to enormously wealthy entrepreneurs *ex post*, most of them fail or operate on a small scale. In fact, as recently estimated by Nordhaus (2004), the bulk of the social surplus from (Schumpeterian) entrepreneurial activity – a stunning 97.8 percent according to his calculations – flows to consumers in the form of cheaper and better products (in the U.S.).

Nordhaus's calculations suggest that the expected average return from entrepreneurship need not be so high to induce individual entrepreneurial effort, but the flip side of this suggestion is that misdirected policies inadvertently encouraging rent-seeking entrepreneurship are likely to be very costly. So, attempts to derive concrete policy recommendations based narrowly on any of the modeling efforts is hazardous, since the new opportunities for entrepreneurship arising from a policy change must be tied to activities that yield economic growth. Any attempt to create entrepreneurship, by trying to act directly on individual motives and behaviors, becomes extremely difficult and risky.

Acknowledging the fact that the different modeling efforts can pick up the effects of entrepreneurship at the aggregate level, they may still be very rough guides for policy design. Heeding the risk that the models are taken too literally is particularly important. For example, policymakers may think they have no meaningful role, if entrepreneurship is driven simply by “ability” or “attitudes toward risk”. Or, which is even worse, they may be inclined to adopt policies that promote the wrong actors (e.g. subsidizing R&D activities in large incumbent firms), or they may consider strategies

²¹ See Rodrik et al. (2004) for a recent and authoritative empirical study highlighting the crucial importance of institutional factors as determinants of economic growth.

²² Entrepreneurial self-employment may be partly pursued in search for independence, and fraudulent rent seeking may also be entrepreneurial (Baumol, 1990). It may also be noted that most self-employment is likely to be non-entrepreneurial (Henrekson, 2004).

for picking winners.²³ But, given that entrepreneurship is a generic trait in human societies, all programs of this kind are changing entrepreneurial incentives on one or several dimensions: making rent seeking relatively more profitable, expanding opportunities for political entrepreneurship, shifting the focus of limited entrepreneurial attention towards satisfying policymakers and bureaucrats rather than customers in the market place, and so forth. In a real economy where genuine uncertainty and path dependence play large roles, this is likely to have negative effects that tend to be ignored in standard neoclassical models, even when an element of entrepreneurship is highlighted.²⁴

We conclude that an individual real-world entrepreneur, even if highly stylized, cannot at present be modeled in mainstream economics, since he or she does elude analytical tractability. In this sense, the neoclassical entrepreneur is (still) not entrepreneurial. This is not to say that efforts to introduce certain analytically tractable facets of entrepreneurship in economic models are in vain. After all, any predictions about effects of policy proposals and institutional reforms changing the incentive structure vis-à-vis productive entrepreneurship presuppose a specific economic model. The introduction of entrepreneurship in a model, while abstracting from the political economy aspects, is likely to be particularly hazardous when the models are used as a guide for policy advice.

²³ Perhaps the most common example is support to existing entrepreneurs by artificially lowering the cost of capital for “entrepreneurial” firms, e.g. by providing seed capital or soft loans to firms selected through a bureaucratic procedure.

²⁴ Bergström (2000) finds that in the long-run firms that received government support performed worse than other comparable firms. See also Secrieru and Vigneault (2004) for a recent example.

References

- Acemoglu, D., Johnson, S. and J. Robinson (2004), 'Institutions as the Fundamental Cause of Long-run Growth', NBER Working Paper 10481.
- Acemoglu, D., Aghion, P. and Zilibotti, F. (2003), 'Distance to Frontier, Selection and Growth', *Journal of the European Economic Association* 1(2), 630–38.
- Acs, Z. J., ed. (1999), *Are Small Firms Important? Their Role and Impact*. Dordrecht: Kluwer.
- Acs, Z. J. and Audretsch, D. B. (1988), 'Innovation in Large and Small Firms: An Empirical Analysis', *American Economic Review*, 78(4), 678-90.
- Acs, Z. J. and Audretsch, D. B. eds. (2003), *Handbook of Entrepreneurship Research*. Dordrecht: Kluwer.
- Acs, Z. J., Audretsch, D. B., Braunerhjelm, P. and Carlsson, B. (2004), 'The Missing Link: The Knowledge Filter and Entrepreneurship in Endogenous Growth', Mimeo.
- Aspromourgos, T. (1987), 'Neoclassical', In *The New Palgrave Dictionary of Economics*. London: MacMillan.
- Audretsch, D. B. (1995), *Innovation and Industry Evolution*. Cambridge, MA: MIT Press
- Audretsch, D. B. and Thurik, A. R. (1997), 'Sources of Growth: The Entrepreneurial Versus the Managed Economy', Centre for Economic Policy Research, Discussion Paper No. 1710.
- Audretsch, D. B. and Thurik, A. R. (2000), 'Capitalism and Democracy in the 21st Century: from the Managed to the Entrepreneurial Economy', *Journal of Evolutionary Economics* 10(1), 17–34.
- Aumann, R. J. (1966), 'Existence of Competitive Equilibria in Markets with a Continuum of Traders', *Econometrica* 34(1), 1–17.
- Akerberg, D. and Botticini, M. (2002), 'Endogenous Matching and the Empirical Determinants of Contract Form', *Journal of Political Economy* 110(3), 564–591.
- Aghion, P. and Howitt, P. (1992), 'A Model of Growth through Creative Destruction', *Econometrica* 60(2), 323–351.
- Barreto, H. (1989), *The Entrepreneur in Micro-Economic Theory: Disappearance and Explanation*. London: Routledge.

- Baumol, W.J. (1968), 'Entrepreneurship in Economic Theory', *American Economic Review* 58(2), 64–71.
- Baumol, W.J. (1990), 'Entrepreneurship: Productive, Unproductive and Destructive', *Journal of Political Economy* 98(5), 893–921.
- Baumol, W.J. (1993), *Entrepreneurship, Management, and the Structure of Payoffs*. Cambridge, MA: MIT Press.
- Baumol, W.J. (2002), *The Free-Market Innovation Machine – Analyzing the Growth Miracle of Capitalism*. Princeton and Oxford: Princeton University Press.
- Baumol, W. J. (2004), 'Education for Innovation: Entrepreneurial Breakthroughs vs. Corporate Incremental Improvements', NBER Working Paper 10578.
- Bates, T. M. (1995), 'Self-Employment Entry Across Industries' *Journal of Business Venturing* 10 (6): 143-56.
- Bergström, F. (2000), 'Capital Subsidies and the Performance of Firms', *Small Business Economics* 14(3), 83–93.
- Bewley, T. (1986), 'Knightian Decision Theory: Part I', Cowles Foundation Discussion Paper No. 807. Yale University.
- Blanchflower, D. G. and Oswald, A. J. (1998), 'What Makes an Entrepreneur?' *Journal of Labor Economics* 16(1), 26–60.
- Bresnahan, T. F. and Trajtenberg, M. (1995), 'General Purpose Technologies 'Engines of Growth?'' *Journal of Econometrics* 65(1), 83–108.
- Brown, C., Hamilton, J. and Medoff, J. (1990), *Employers Large and Small*. Cambridge, MA: Harvard University Press.
- Carree, M., van Stel, A., Thurik, A. R. and Wennekers, S. (2002), 'Economic Development and Business Ownership: An Analysis Using Data of 23 OECD Countries in the Period 1976–1996', *Small Business Economics* 19(3), 271–290.
- Casson, M. (1982/2003), *The Entrepreneur – An Economic Theory*. 2nd ed. Cheltenham, UK: Edward Elgar.
- Colander, D. (2000), 'The Death of Neoclassical Economics', *Journal of the History of Economic Thought* 22(2), 127–143.
- Colander, D. (2003), 'The Complexity Revolution and the Future of Economics', Middlebury College Economics Discussion Paper No. 03-19. Department of Economics, Middlebury College.

- Davis, S. J., Haltiwanger, J. and Schuh, S. (1996), *Job Creation and Destruction*. Cambridge, MA: MIT Press.
- Dinopoulos, E. and Segerstrom, P. S. (1999), 'A Schumpeterian Model of Protection and Relative Wages', *American Economic Review* 89(3) 450–472.
- Dunn, T. and Holtz-Eakin, D. (2000), 'Financial Capital, Human Capital, and the Transition to Self-Employment: Evidence from Intergenerational Links', *Journal of Labor Economics*, 18(2), 282–305.
- Eliasson, G. (1984), 'Micro Heterogeneity of Firms and the Stability of Industrial Growth', *Journal of Economic Behavior & Organization* 5(3–4), 249–274.
- Evans, D. S. and Jovanovic, B. (1989), 'An Estimated Model of Entrepreneurial Choice under Liquidity Constraints', *Journal of Political Economy* 97(4), 808–27.
- Evans, D. S. and Leighton, L. S. (1989), 'Some Empirical aspects of Entrepreneurship', *American Economic Review* 79(3), 519–535.
- Futia, C. A. (1980), 'Schumpeterian Competition', *Quarterly Journal of Economics* 94(4), 675–695.
- Glancey, K. S. and McQuaid, R. W. (2000), *Entrepreneurial Economics*. New York: Palgrave.
- Grabowski, H. G. and Vernon, J. M. (1987), 'Pioneers, Imitators, and Generics – A Simulation Model of Schumpeterian Competition', *Quarterly Journal of Economics* 102(3), 491–526.
- Grossman, G. M. and Helpman, E. (1994), 'Endogenous Innovation in the Theory of Growth', *The Journal of Economic Perspectives*, 8(1), 23–44.
- Hayek, F. A. (1945), 'The Use of Knowledge in Society', *American Economic Review* 35(4), 21–30.
- Helpman, E. (1992), 'Endogenous Macroeconomic Growth Theory', *European Economic Review* 36(2), 237–267.
- Henrekson, M. (2004), 'Comment on David G. Blanchflower: Is More Self-Employment Good or Bad?', *Swedish Economic Policy Review* 11(2), *forthcoming*
- Henrekson, M. and Rosenberg, N. (2001), 'Designing Efficient Institutions for Science-Based Entrepreneurship: Lessons from the US and Sweden', *Journal of Technology Transfer* 26(3), 207–231.
- Holmes, T. and Schmitz, J. A. (1990), 'A Theory of Entrepreneurship and Its Application to the Study of Business Transfers', *Journal of Political Economy* 98(2), 265–294.
- Holmes, T. and Schmitz, J. A. (2001), 'A Gain from Trade: From Unproductive to Productive Entrepreneurship', *Journal of Monetary Economics* 47(3), 417–446.
- Holtz-Eakin, D., Joulfaian, D. and Rosen, H. S. (1994), 'Sticking it Out: Entrepreneurial Survival and Liquidity Constraints', *Journal of Political Economy* 102(1), 53–75.

- Hurst, E. and Lusardi, A. (2004), 'Liquidity Constraints, Household Wealth, and Entrepreneurship', *Journal of Political Economy* 112(2), 319-347.
- Ilmakunnas, P. and Kannianen, V. (2001), 'Entrepreneurship, Economic Risks, and Risk Insurance in the Welfare State: Results with OECD Data 1978-93', *German Economic Review* 2(3), 195-218.
- Johansson, D. (2004), 'Economics without Entrepreneurship or Institutions: A Vocabulary Analysis of Graduate Textbooks', *Econ Journal Watch* 3(1), 515-538.
- Kihlstrom, R. E. and Laffont, J. J. (1979), 'A General Equilibrium Entrepreneurial Theory of Firm Formation Based on Risk Aversion', *Journal of Political Economy* 87(3), 719-748.
- Kirchhoff, B. A. (1994), *Entrepreneurship and Dynamic Capitalism*. London: Praeger.
- Kirzner, I. M. (1997), 'Entrepreneurial Discovery and the Competitive Market Process: An Austrian Approach', *Journal of Economic Literature* 35(1), 60-85.
- Knight, F. H. (1921), *Risk, Uncertainty and Profit*. New York: Houghton Mifflin.
- Kreps, D. M. (1988), *Notes on the Theory of Choice*, Boulder, CO: Westview Press.
- Laussel, D. and Le Breton, M. (1995), 'A General Equilibrium Theory of Firm Formation Based on Individual Unobservable Skills', *European Economic Review* 39(7), 1303-1319.
- Lazear, E. M. (2002), 'Entrepreneurship', NBER Working Paper No. 9109.
- Libecap, G. D. (1993), 'Entrepreneurship, Property Rights and Economic Growth', In *Advances in the Study of Entrepreneurship, Innovation and Economic Growth: A Research Annual*. Vol. 6. Greenwich, CT: JAI Press.
- Lucas, R. E., Jr. (1978), 'On the Size Distribution of Business Firms', *Bell Journal of Economics* 9(3), 508-523.
- Martens, B. (2004), *The Cognitive Mechanics of Economic Development*. London: Routledge.
- Michelacci, C. (2003), 'Low Returns in R&D due to Lack of Entrepreneurial Skills', *The Economic Journal* 113, 207-225.
- Murphy, K. M., Shleifer, A. and Vishny, R. (1991), 'The Allocation of Talent: Implications for Growth', *Quarterly Journal of Economics* 106(2), 503-530.
- Newman, A. (1999), 'Risk-Bearing, Entrepreneurship and the Theory of Moral Hazard', Mimeo. University College London.
- Nordhaus, W. D. (2004), 'Schumpeterian Profits in the American Economy: Theory and Measurement', NBER Working Paper No. 10433.

- Olsson, O. and Frey, B. S. (2002), 'Entrepreneurship as Recombinant Growth', *Small Business Economics* 19(1), 69–80.
- Parker, S. C. (1996), 'A time series model of self-employment under uncertainty', *Economica*, 63, 459-75
- Parker, S. C. (2004), *The Economics of Self-employment and Entrepreneurship*. Cambridge: Cambridge University Press
- Pelikan, P. (2004), 'Interconnecting Evolutionary, Institutional and Cognitive Economics: Six Steps towards Understanding the Six Links', Ratio Working Paper No. 48.
- van Praag, C. M. and Cramer, J. S. (2001), 'The Roots of Entrepreneurship and Labour Demand: Individual Ability and Low Risk Aversion', *Economica*, 68(127), 45-62.
- Robinson, P. B. and Sexton, E. A. (1994), 'The Effects of Education and Experience on Self-employment Success' *Journal of Business Venturing*, 9(3), 141-156.
- Rodrik, D., Subramanian, A. and Trebbi, F. (2004), 'Institutions Rule: The Primacy of Institutions over Geography and Integration in Economic Development', *Journal of Economic Growth* 9(2), 131–165.
- Roine, J. (2004), 'The Political Economics of Not Paying Taxes', *Public Choice*, forthcoming.
- Romer, P. M. (1990), 'Endogenous Technological Change', *Journal of Political Economy* 98(5), 71–102.
- Rosen, H. and Willen, P. (2002), 'Risk, Return and Self-Employment', Mimeo, Princeton and Chicago University.
- Rosen, S. (1997), 'Austrian and Neoclassical Economics: Any Gains From Trade?', *Journal of Economic Perspectives* 11(4), 139–152.
- Say, J. B. (1845, 1964), *A Treatise on Political Economy: Or, the Production, Distribution and Consumption of Wealth*. New York: A.M. Kelley.
- Scherer, Frederic M. (1980), *Industrial Market Structure and Economic Performance*. Chicago: Rand McNally.
- Schiffman, D.A. (2004), 'Mainstream Economics, Heterodoxy and Academic Exclusion: A Review Essay', *European Journal of Political Economy* 20(4), 1079–1095.
- Schmitz, J. A. (1989), 'Imitation, Entrepreneurship, and Long-Run Growth', *Journal of Political Economy* 97(3), 721–739.
- Schultz, T. W. (1980), 'Investment in Entrepreneurial Ability', *Scandinavian Journal of Economics* 82(4), 437–48.

- Schumpeter, J. A. (1934), *The Theory of Economic Development*. Cambridge, MA: Harvard University Press.
- Schumpeter, J. A. (1942), *Capitalism, Socialism and Democracy*. New York: George Allen & Unwin.
- Secrieru, O. and Vigneault, M. (2004), 'Public Venture Capital, Occupational Choice, and Entrepreneurship', *Topics in Economic Analysis & Policy* 4(1), Article 25.
- Segerstrom, P. S., Anant, T.C.A. and Dinopoulos, E. (1990), 'A Schumpeterian Model of the Product Life Cycle', *American Economic Review* 80(5), 1077–1091.
- Sexton, D. L. and Landstrom, H. (2000), *The Blackwell Handbook of Entrepreneurship*. Oxford: Blackwell.
- Shane, S., ed. (2002), *The Foundations of Entrepreneurship*. Cheltenham, UK: Edward Elgar.
- Silva, O. (2004), 'Entrepreneurship: Can the Jack-of-All-Trades Attitude be Acquired?', CEP Discussion Paper 665, London School of Economics and Political Science.
- Solo, C. S. (1951), 'Innovation in the Capitalist Process: A Critique of the Schumpeterian Theory', *Quarterly Journal of Economics* 65(3), 417–428.
- Solow, R. M. (1997), 'How Did Economics Get That Way and What Way Did It Get?' *Daedalus* 126, 139.
- Swedberg, R. (2000), *Entrepreneurship: The Social Science View*. Oxford: Oxford University Press.
- Tirole, J. (1998), *The Theory of Industrial Organization*. Cambridge, MA: MIT Press.
- Wagner, J. (2003), 'Testing Lazear's jack-of-all-trades view of entrepreneurship with German micro data', *Applied Economics Letters*, 10, 687-89
- Wennekers, S. and Thurik, A. R. (1999), 'Linking Entrepreneurship and Economic Growth', *Small Business Economics* 13(1), 27–55.
- Westhead, P. and Wright, M. (2000), *Advances in Entrepreneurship*. Cheltenham, UK: Edward Elgar.

EFI, The Economic Research Institute

Published in the language indicated by the title. A complete publication list can be found at www.hhs.se/efi Books and dissertations can be ordered from EFI via e-mail: EFI.Publications@hhs.se

Reports since 2003

2007

Books

Wijkström, Filip och Torbjörn Einarsson. *Analysmodell för sektorsöverskridande statistik. Fallet vård och omsorg.*

Dissertations

Ahlersten, Krister. *Empirical Asset Pricing and Investment Strategies.*

Alexius, Susanna. *Regelmotståndarna – om konsten att undkomma regler.*

Berg, Bengt Åke. *Volatility, Integration and Grain Bank. Studies in Harvests, Rye Prices and Institutional Development of the Parish Magasins in Sweden in the 18th and 19th Centuries.*

Brodin, Karolina. *Consuming the Commercial Break: An Ethnographic Study of the Potential Audiences for Television Advertising.*

Elger, Max. *Three essays on investment-specific technical change.*

Hagberg, Axel. *Bankkrishantering: Aktörer, marknad och stat.*

Hjalmarson, Hanna. *En växande marknad: Studie av nöjdheten med konsumtionsrelaterade livsområden bland unga konsumenter.*

Hjelström, Tomas. *The Closed-End Investment Company Premium Puzzle – Model development and empirical tests on Swedish and British data.*

Kraus, Kalle. *Sven, inter-organisational relationships and control – a case study of domestic care of the elderly.*

Melian, Catharina. *Progressive Open Source*

Valiente, Pablo. *Re-innovating the Existing. A study of Wireless IS Capabilities to Support Mobile Workforces.*

2006

Books

Lundeberg, Mats, Pär Mårtensson and Magnus Mähring (eds) *IT & Business Performance. A Dynamic Relationship. EFI Yearbook 2006.* EFI/Studentlitteratur.

Thodenius, Björn. *Organisering av kunskap. En studie av Wallenberg Consortium North.*

Wijkström, Filip och Torbjörn Einarsson. *Från nationalstat till näringsliv? Det civila samhällets organisationsliv i förändring.*

Wijkström, Filip, Einarsson, Stefan och Larsson, Ola. *Staten och det civila samhället.*

Idétraditioner och tankemodeller i den statliga bidragsgivningen till ideella organisationer.

Östman, Lars. *Lysande ögonblick och finansiella kriser. Dramaten under ett sekel.*

Dissertations

Argenton, Cedric. *Quality provision in duopoly.*

Beckerman, Carina. *The Clinical Eye. Construction and Computerizing an Anesthesia Patient Record.*

Borglund, Tommy. *Aktievärden i fokus – internationell påverkan på intressentrelationer vid förvärv och fusion.*

Breman, Anna. *The Economics of altruism, paternalism and self-control.*

- Edquist, Harald. *Technological breakthroughs and productivity growth*.
- Eklund, Jana. *Essays on Forecasting and Bayesian Model Averaging*.
- Frostenson, Magnus. *Legitimitetskontrollen – en studie av etiska värderingars roll i gränsöverskridande förvärv och fusioner*.
- Gaspar, Raquel M. *Credit Risk and Forward Price Models*.
- Gustafsson, Peter. *Essays on Trade and Technological Change*.
- Hopkins, Elisabeth. *Is a higher degree of local currency pricing associated with lower exchange rate pass-through? A study of import pricing in 51 Swedish industries*.
- Kling, Ragnar. *Developing Product Development in Times of Brutal Change*.
- Langenskiöld, Sophie. *Peer Influence on Smoking: Causation or Correlation?*
- Lychnell, Lars-Olof. "Och fungerar det inte, gör vi på något annat sätt" – En klinisk fallstudie av IT-relaterat förändringsarbete i småföretag
- Meitz, Mika. *Five Contributions to Econometric Theory and the Econometrics of Ultra-High-Frequency Data*.
- Mendicino, Caterina. *Financial Market Imperfections, Business Cycle Fluctuations and Economic Growth*.
- Ovanfors, Anna. *Essays on Nonlinear Time Series Analysis and Health Economics*.
- Paltseva, Elena. *Essays on Commitment and Inefficiency in Political Economy*.
- Rogberg, Martin. *Den modeföljande organisationen. Om acceptansen av TQM och andra populära managementmodeller*.
- Silvennoinen, Annastiina. *Essays on Autoregressive Conditional Heteroskedasticity*.
- Sjögren, Ebba. *Reasonable Drugs. Making Decisions with Ambiguous Knowledge*.
- Slinko, Irina. *Essays in Option Pricing and Interest Rate Models*.
- Wilander, Fredrik. *Essays on Exchange Rates and Prices*.

2005

Books

- Andersson, Per, Susanne Hertz and Susanne Sweet (eds). *Perspectives on market networks – boundaries and new connections*.
- Charpentier, Claes. *IT inom omsorgen. Förväntade effekter av införande av IT-system för utförarna inom äldre- och handikappomsorgen*.
- Dembrower, Maria. *Entreprenörskap i industriella nätverk*.
- Lind, Johnny och Göran Nilsson (red). *Ekonomistyrningens metoder, sammanhang och utveckling – En vänbok till Lars A Samuelson*.
- Samuelson, Lars A. *Organizational governance and control – a summary of research in the Swedish society*.
- Svedberg Nilsson, Karin, Roger Henning och Karin Fernler (eds). *En illusion av frihet? Företag och organisationer i regelsamhället. EFIs Årsbok 2005*. EFI/Studentlitteratur.

Dissertations

- Andersson, Martin. *Making a Difference – Project Result Improvement in Organizations*.
- Arvidsson, Per. *Styrning med belöningsystem – Två fallstudier om effekter av belöningsystem som styrmedel*.
- Bems, Rudolfs. *Essays in International Macroeconomics*.
- Berg-Suurwee, Ulrika. *Nya trender, nya nämnder – effekter av en stadsdelsnämndsreform inom kultur och fritid*.
- Björkman, Hans. *Learning from members – Tools for strategic positioning and service innovation in trade unions*.
- Bodnaruk, Andriy. *Essays in Empirical Corporate Finance and Portfolio Choice*.
- Clapham, Eric. *Essays in Real Estate Finance*.
- Dareblom, Jeanette. *Prat, politik och praktik – Om individers möten med strukturer i en kommunal satsning på kvinnors företagande*.

- Fromm, Jana. *Risk Denial and Neglect: Studies in Risk Perception*.
- Hjelström, Anja. *Understanding International Accounting Standard Setting – A Case Study of IAS 12, Accounting for Deferred Tax*.
- Hortlund, Per. *Studies on Swedish Banking 1870-2001*.
- Lindahl, Therese. *Strategic and Environmental Uncertainty in Social Dilemmas*.
- Linnarsson, Håkan. *Alliance for Innovation. A structural perspective on new business development in cooperative ventures*.
- Madestam, Andreas. *Developing Credit Markets*.
- Nilsson, Roland. *The Market Impact of Short-Sale Constraints*.
- Nordfält, Jens. *Is consumer decision-making out of control? Non-conscious influences on the consumer decision-making process for fast moving consumer goods*.
- Nordin, Fredrik. *Externalising Services – Walking a Tightrope between Industrial and Service Logics*.
- Parmler, Johan. *Essays in Empirical Asset Pricing*.
- Simbanegavi, Witness. *Price Discrimination, Advertising and Competition*.
- Thodenius, Björn. *Användning av ledningsinformationssystem: en longitudinell studie av svenska storföretag*.
- Tolis, Christofer. *Framing the Business – Business Modelling for Business Development*.
- Östberg, Per. *Corporate Disclosure and Investor Recognition*.

2004

Books

- Ahrne, Göran och Nils Brunsson (red). *Regelexplosionen*.
- Lind, Johnny. *Strategi och ekonomistyrning. En studie av sambanden mellan koncernstrategi, affärsstrategi och ekonomistyrning*.
- Lind, Johnny och Walter Schuster (red). *Redovisningens teori, praktik och pedagogik. En vänbok till Lars Östman*.
- Sevón, Guje och Lennart Sjöberg (red). *Emotioner och värderingar i näringslivet. EFIs Årsbok 2004*.
- Wijkström, Filip and Stefan Einarsson. *Foundations in Sweden – Their scope, roles and visions*.

Dissertations

- Anderson, Anders. *Essays in Behavioral Finance*.
- Balsvik, Gudrun. *Information Technology Users: Studies of Self-Efficacy and Creativity among Swedish Newspaper Journalists*.
- Blix, Magnus. *Essays in Mathematical Finance – Modelling the Futures Price*.
- González Gómez, Andrés. *Nonlinear dynamics and smooth transition models*.
- Grönqvist, Erik. *Selection and Moral Hazard in Health Insurance: Taking Contract Theory to the Data*.
- Ivarsson Westerberg, Anders. *Papperspolisen – varför ökar administrationen i moderna organisationer*.
- Jutterström, Mats. *Att påverka beslut – företag i EUs regelsättande*.
- Jönsson, Kristian. *Macroeconomic Aspects of Capital Flows to Small Open Economies in Transition*.
- Larsson, Pär. *Förändringens villkor. En studie av organisatoriskt lärande och förändring inom skolan*.
- Lagerwall, Björn. *Empirical Studies of Portfolio Choice and Asset Prices*.
- Malmsten, Hans. *Properties and Evaluation of Volatility Models*.
- Marshall, Cassandra. *Dating for Innovation. Creating and Recognizing Opportunities through Collaborative Interorganizational Relationships in Fluid Environments*.

- Mattsson, Susanna. *På gränsen mellan ordning och oordning – tingens betydelse vid marknadsombildningar. En studie av svenska postväsendets ombildning under 1990-talet.*
- Nilsson, Charlotte. *Studies in Environmental Economics: Numerical Analysis of Greenhouse Gas Policies.*
- Nilsson, Hans. *Medborgaren i styrsystemet – beskrivning av VAD och HUR i styrning av kommunal verksamhet.*
- Nystedt, Jens. *Competition, Regulation and Integration in International Financial Markets.*
- Pajuste, Anete. *Corporate Governance and Controlling Shareholders.*
- Richtnér, Anders. *Balancing Knowledge Creation. Organizational Slack and Knowledge Creation in Product Development.*
- Salabasis, Mickael. *Bayesian Time Series and Panel Models – Unit Roots, Dynamics and Random Effects.*
- Sandberg, Rickard. *Testing the Unit Root Hypothesis in Nonlinear Time Series and Panel Models.*
- Skallsjö, Sven. *Essays on Term Structure and Monetary Policy.*
- Strikholm, Birgit. *Essays on Nonlinear Time Series Modelling and Hypothesis Testing.*
- Söderström, John. *Från Produkt till Tjänst. Utveckling av affärs- och miljöstrategier i produktorienterade företag.*
- Talia, Krim. *The Scandinavian Currency Union, 1873–1924 – Studies in Monetary Integration and Disintegration.*

2003

Books

- Lundahl, Mats (ed). *Globalization and Its Enemies. EFIs Årsbok 2003.*
- Sundgren, Bo, Pär Mårtensson, Magnus Mähring and Kristina Nilsson (eds). *Exploring Patterns in Information Management. Concepts and Perspectives for Understanding IT-Related Change.*

Dissertations

- Andersson, Henrik. *Valuation and Hedging of Long-Term Asset-Linked Contracts.*
- Bergman, Malin. *Essays on Human Capital and Wage Formation.*
- Damsgaard, Niclas. *Deregulation and Regulation of Electricity Markets.*
- Eklund, Bruno. *Four Contributions to Statistical Inference in Econometrics.*
- Hakkala, Katariina. *Essays on Restructuring and Production Decisions in Multi-Plant Firms.*
- Holgersson, Charlotte. *Rekrytering av företagsledare. En studie i homosocialitet.*
- Ivaschenko, Iryna. *Essays on Corporate Risk, U.S. Business Cycles, International Spillovers of Stock Returns, and Dual Listing.*
- Lange, Fredrik. *Brand Choice in Goal-derived Categories – What are the Determinants?*
- Le Coq, Chloé. *Quantity Choices and Market Power in Electricity Market.*
- Magnusson, Peter R. *Customer-Oriented Product Development – Experiments Involving Users in Service Innovation.*
- Meisiek, Stefan. *Beyond the Emotional Work Event Social Sharing of Emotion in Organizations.*
- Mårtensson, Anders. *Managing Mission-Critical IT in the Financial Industry.*
- Nilsson, Göran. *Processorientering och styrning – Regler, mål eller värderingar?*
- Sandberg, Robert. *Corporate Consulting for Customer Solutions Bridging Diverging Business Logics.*
- Sturluson, Jon Thor. *Topics in the Industrial Organization of Electricity Markets.*
- Tillberg, Ulrika. *Ledarskap och samarbete – En jämförande fallstudie i tre skolor.*
- Waldenström, Daniel. *Essays in Historical Finance.*

Wallén, Ulrika. *Effektivitet i grundskolan i anslutning till en stadsdelsnämndsreform.*
Ögren, Anders. *Empirical Studies in Money, Credit and Banking – The Swedish Credit Market
in Transition under the Silver and the Gold Standards, 1834–1913.*

