Essays in Dynamic General Equilibrium

by

Dan Cao

B.S., M.A., Ecole Polytechnique (2005)

Submitted to the Department of Economics
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETT INSTITUTE OF TECHNOLOGY

June 2010

Signature of Author ....
Department of Economics
May 10, 2010

Certified by ....
Daron Acemoglu
Charles P. Kindleberger Professor of Applied Economics
Thesis Supervisor

Certified by ....
Ivan Werning
Professor of Ecomics
Thesis Supervisor

Accepted by ....
Esther Duflo
Abdul Latif Jameel Professor of Poverty Alleviation and Development Economics
Chair, Department Committee on Graduate Studies

Essays in Dynamic General Equilibrium

by

Dan Cao

Submitted to the Department of Economics
on May 10, 2010, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

This thesis consists of three chapters studying dynamic economies in general equilibrium. The first chapter considers an economy in business cycles with potentially imperfect financial markets. The second chapter investigates an economy in its balanced growth path with heterogenous firms. The third chapter analyzes dynamic competitions that these firms are potentially engaged in.

The first chapter, "Asset Price and Real Investment Volatility with Heterogeneous Beliefs," sheds light on the role of imperfect financial markets on the economic and financial crisis 2007-2008. This crisis highlights the role of financial markets in allowing economic agents, including prominent banks, to speculate on the future returns of different financial assets, such as mortgage-backed securities. I introduce a dynamic general equilibrium model with aggregate shocks, potentially incomplete markets and heterogeneous agents to investigate this role of financial markets. In addition to their risk aversion and endowments, agents differ in their beliefs about the future aggregate states of the economy. The difference in beliefs induces them to take large bets under frictionless complete financial markets, which enable agents to leverage their future wealth. Consequently, as hypothesized by Friedman (1953), under complete markets, agents with incorrect beliefs will eventually be driven out of the markets. In this case, they also have no influence on asset prices and real investment in the long run. In contrast, I show that under incomplete markets generated by collateral constraints, agents with heterogeneous (potentially incorrect) beliefs survive in the long run and their speculative activities drive up asset price volatility and real investment volatility permanently. I also show that collateral constraints are always binding even if the supply of collateralizable assets endogenously responds to their price. I use this framework to study the effects of different types of regulations and the distribution of endowments on leverage, asset price volatility and investment. Lastly, the analytical tools developed in this framework enable me to prove the existence of the recursive equilibrium in Krusell and Smith (1998) with a finite number of types. This has been an open question in the literature.

The second chapter, "Innovation from Incumbents and Entrants," is a joint work with Daron Acemoglu. We propose a simple modification of the basic Schumpeterian endogenous growth models, by allowing incumbents to undertake innovations to improve their products. This model provides a tractable framework for a simultaneous analysis of entry of new firms and the

expansion of existing firms, as well as the decomposition of productivity growth between continuing establishments and new entrants. One lesson we learn from this analysis is that, unlike in the basic Schumpeterian models, taxes or entry barriers on potential entrants might *increase* economic growth. It is the outcome of the greater productivity improvements by incumbents in response to reduced entry, which outweighs the negative effect of the reduction in creative destruction. As the model features entry of new firms and expansion and exit of existing firms, it also generates an equilibrium firm size distribution. We show that the stationary firm size distribution is Pareto with an exponent approximately equal to one (the so-called "Zipf distribution").

The third chapter, "Racing: when should we handicap the advantaged competitor?" studies dynamic competitions, for example R&D competitions used in the second chapters. Two competitors with different abilities engage in a winner-take-all race; should we handicap the advantaged competitor in order to reduce the expected completion time of the race? I show that if the discouragement effect is strong, i.e., both competitors are discouraged from exerting effort when it becomes more certain who will win the race, we should handicap the advantaged. We can handicap him either by reducing his ability or by offering him a lower reward if he wins. Doing so induces higher effort not only from the disadvantaged competitor because of his higher incentive from a higher chance of winning the race but also from the advantaged competitor because of their strategic interactions. Therefore, the expected completion time is strictly shortened. To prove the existence and uniqueness of the equilibria (including symmetric and asymmetric equilibria) that leads to the conclusion, I use a boundary value problem formulation which is novel to the dynamic competition literature. In some cases, I obtain closed-form solutions of the equilibria.

Thesis Supervisor: Daron Acemoglu
Title: Charles P. Kindleberger Professor of Applied Economics

Thesis Supervisor: Ivan Werning
Title: Professor of Ecomics

To my grandmother, my parents, and Nhan

Acknowledgements

I am infinitely indebted to my thesis advisors, Daron Acemoglu and Ivan Werning for their guidance and infinite support during my entire time at MIT. Ivan made me decide to do research in macroeconomics. In the first summer after taking the general exams in industrial organization and econometrics and having started doing research in these two fields, I realized my intellectual curiosity was on something else. I thus tried it out in macroeconomics by working as a research assistant for Ivan on his paper on the Friedman rule. I discovered that macroeconomics fitted the best my research and personal interests. Since then frequent conversations, online and offline :), with Ivan have confirmed my choice. He has always been enthusiastic to share with me his passion for macroeconomics. Right after that summer, I sat in Daron's Advanced Topic in Endogenous Growth class and witnessed the beauty and richness of this subfield of macroeconomics. There were many open and exciting questions in this field. One of the questions led to the second chapter in this thesis. Monthly meetings with Daron for the last three years have been a great experience. After each meeting with him, I learned a new thing on doing research in economics. Daron was also very patient with me with my immature research questions at the beginning of my research career. Above all, I would not have been able to write my job market paper with the support of my two advisors. Daron constantly pushed me forward with setting up the general equilibrium framework. Ivan made excellent suggestions and comments on how to apply that framework to specific questions. I still remember vividly that during the two weeks before my internal job market seminar last October, Ivan spent at least two hours per day working with me on the paper. At some point we talked on Gchat until 4 a.m. and started talking again at 10 a.m. in the same morning. I also benefited a lot from talking with my other two advisors, Robert Townsend and Guido Lorenzoni since the beginning of the project leading to my job market paper. I also enjoyed and learned from my conversations with other faculty members including George-Marios Angeletos, Abhijit Banerjee, Olivier Blanchard, Ricardo Caballero, Arnaud Costinot, Dave Donaldson, Jerry Hausman, Bengt Holmstrom, Roberto Rigobon, Peter Temin, and Jean Tirole.

The five years at M.I.T. have not only been about economics. I have enjoyed the graduate student life within the Boston and Cambridge community. I would thus like to thank my friends who have made this experience much pleasant and memorable. Thanks to my former and current roommates David Cesarini, Jenifer La'o, Ngoc Huy Nguyen, Maria Polyakova, Daniel Rees, Bradley Shapiro, to my current and former officemates, Suman Basu, Jean-Paul L'Huillier, Monica Martinez-Bravo, Michael Powell, Florian Scheuer, Paul Schrimpf, Alp Simsek, as well as to my other classmates and non-economists friends, especially Dinh Thi Theu.

Finally, this thesis is dedicated to my grandmother, Dang Thi Huong, my father, Cao Hoa Binh, my mother, Vo Thi Thu, my brother, Cao Vu Nhan, and my cousin, Nguyen Chau Thanh, whose most earnest preoccupation was my happiness and my education. Without them this endeavor would not have been possible.

Contents

Chapter 1

Asset Price and Real Investment Volatility with Heterogeneous Beliefs

1.1 Introduction

The events leading to the financial crisis 2007-2008 have highlighted the importance of belief heterogeneity and how financial markets also create opportunities for agents with different beliefs to leverage up and speculate. Several investment and commercial banks invested heavily in mortgage-backed securities, which subsequently suffered large declines in value. At the same time, some hedge funds profited from the securities by short-selling them.

One reason for why there has been relatively little attention, in economic theory, paid to heterogeneity of beliefs and how these interact with financial markets is the *market selection hypothesis.* The hypothesis, originally formulated by Friedman (1953), claims that in the long run, there should be limited differences in beliefs because agents with incorrect beliefs will be taken advantage of and eventually be driven out the markets by those with the correct belief. Therefore, agents with incorrect beliefs will have no influence on economic activity in the long run. This hypothesis has recently been formalized and extended in recent work by Blume and Easley (2006) and Sandroni (2000). However these papers assume financial markets are complete and this assumption plays a central role in allowing agents to pledge all their wealth.

In this paper, I present a dynamic general equilibrium framework in which agents differ in their beliefs but markets are endogenously incomplete because of collateral constraints. Collateral constraints limit the extent to which agents can pledge their future wealth and ensure that agents with incorrect beliefs never lose so much as to be driven out of the market. Consequently all agents, regardless of their beliefs, survive in the long run and continue to trade on the basis of those heterogeneous beliefs. This leads to additional leverage and asset price volatility (relative to a model with homogeneous beliefs or relative to the limit of the complete markets economy).

The framework introduced in this paper also enables a comprehensive study of how the survival of heterogeneous beliefs and the structure of financial markets affect investment in the long run. I also use this framework for studying the impact of different types of regulations on welfare, asset price volatility and investment. The dynamic general equilibrium approach adopted here is central for many of these investigations. Since it permits the use of well specified collateral constraints, it enables me to look at whether agents with incorrect beliefs will be eventually driven out of the market. It allows leverage and endogenous investment (supply of assets) and it enables me to characterize the effects of different types of policies on welfare and economic fluctuations.

The dynamic stochastic general equilibrium model with incomplete markets I present in this paper is not only useful for the analysis of the effects of heterogeneity in the survival of agents with different beliefs, but also nests well-known models as special cases, including recent models, such as those in Kubler and Schmedders (2003), Fostel and Geanakoplos (2008) and Geanakoplos (2009), as well as more classic models including those in Kiyotaki and Moore (1997) and Krusell and Smith (1998). For instance, this model allows for capital accumulation with adjustment costs in the same model in Krusell and Smith (1998) and shows the existence of a recursive equilibrium. This equilibrium existence has been an open question in the literature. The generality is useful in making this framework eventually applicable to a range of questions on the interaction between financial markets, heterogeneity, investment and aggregate activity.

More specifically, I study an economy in dynamic general equilibrium with aggregate shocks and heterogeneous, infinitely-lived agents. Aggregate shocks follow a Markov process. Consumers differ in terms of their beliefs on the transition matrix of the Markov process (for simplicity, these beliefs differences are never updated as there is no learning; in other words agents in this economy agree to disagree).[1] There is a unique final good used for consumption and investment, and several real and financial assets. There are two classes of real assets: one class of assets, which I call trees, are in fixed supply and the other class of assets are in elastic supply. Only assets in elastic supply can be produced using the final good. The total quantity of final good used in the production of real assets is the aggregate *real investment*. I assume that agents cannot short sell either type of the assets. Assets in elastic supply are important to model real investment and also to show that collateral constraints do not arise because of artificially limited supply of assets.

Incomplete (financial) markets are introduced by assuming that all loans have to use financial assets as collateralized promises as in Geanakoplos and Zame (2002). Selling a financial asset is equivalent to borrowing and in this case agents need to put up some real assets as collateral. Loans are non-recourse and there is no penalty for defaulting. Consequently, whenever the face value of the security is higher than the value of its collateral, the seller of the security can choose to default without further consequences. In this case, security buyer seizes the collateral instead of receiving the face value of the security. I refer to equilibria of the economy with these financial assets as *incomplete markets equilibria* since the presence of collateral constrains introduces endogenous incomplete markets. Several key results involve the comparison of incomplete markets equilibria to the standard competitive equilibrium with complete markets.

Households (consumers) can differ in many aspects, such as risk-aversion and endowments. Most importantly they differ in their beliefs concerning the transition matrix governing transitions across aggregate states. Given the consumers' subjective expectations, they choose their consumption and real and financial asset holdings to maximize their intertemporal expected utility. In particular, the consumers' perceptions about the future value of each unit of real asset, including future rental prices and future resale value, determine the consumers' demand for new units of real assets. This demand, in turn, determines how many new units of real assets

[1] Alternatively, one could assume that even though agents differ with respect to their initial beliefs, they partially update them. In this case, similar results would apply provided that the learning process is sufficiently slow (which will be the case when individuals start with relatively firm priors)

are produced. Hence, demand determines real investment in a fashion similar to the neoclassical Tobin's Q theory of investment.

The framework delivers several results. The first set of results, already mentioned above, is related to the survival of agents with incorrect beliefs. As in Blume and Easley (2006) and Sandroni (2000), with perfect complete markets, in the long run, only agents with correct beliefs survive. Their consumption is bounded from below by a strictly positive number. Agents with incorrect beliefs see their consumption go to zero, as uncertainties realize. However, in any incomplete markets equilibrium, every agent survives because of no-default-penalty condition. When agents lose their bets, they can just simply walk away from their collateral while keeping their current and future endowments. They cannot do so under complete markets because they can commit to delivering all their future endowments.

More importantly, the survival or disappearance of agents with incorrect beliefs affects asset price volatility. To focus on asset price volatility, I consider economies with only trees as real assets. Under complete markets, agents with incorrect beliefs will eventually be driven out of the markets in the long run. The economies converge to economies with homogeneous beliefs, i.e., the correct beliefs. Markets completeness then implies that asset prices in these economies are independent of past realizations of aggregate shocks. In addition, asset prices are the net present discounted values of the dividend processes, with appropriate discount factors. As a result, asset price volatility is proportional to the volatility of dividends if the aggregate endowment, or equivalently the equilibrium stochastic discount factor, only varies by a limited amount over time and across states. These properties no longer hold under incomplete markets. Given that agents with incorrect beliefs survive in the long run, they exert permanent influence on asset prices. Asset prices are not only determined by the aggregate shocks as in the complete markets case, but also by the evolution of the wealth distribution across agents. This also implies that asset prices are history-dependent as the realizations of past aggregate shocks affect the current wealth distribution. The additional dependence on the wealth distribution raises asset price volatility under incomplete markets above the volatility level under complete markets.

I establish this result more formally using a special case in which the aggregate endowment is constant and the dividend processes are I.I.D. Under complete markets, asset prices are asymptotically constant. In contrast, asset price volatility, therefore, goes to zero in the long run. Asset price volatility stays well above zero under incomplete markets as the wealth distribution changes constantly, and asset price depends on the wealth distribution. Although this example is extreme, numerical simulations show that its insight carries over to less special cases. In general, long-run asset price volatility is higher under incomplete markets than under complete markets.

The volatility comparison is different in the short run, however. Depending on the distribution of endowments, short run asset price volatility can be greater or smaller under complete or incomplete markets. This happens because the wealth distribution matters for asset prices under both complete markets and incomplete markets in the short run. This formulation also helps clarify the long-run volatility comparison. In the long run, under complete markets, the wealth distribution becomes degenerate as it concentrates only on agents with correct beliefs. In contrast, under incomplete markets, the wealth distribution remains non-degenerate in the long run and affects asset price volatility permanently. However, the wealth of agents with incorrect beliefs may remain low as they tend to lose their bets. Strikingly, under incomplete markets and when the set of actively traded financial assets is endogenous, the poorer the agents with incorrect beliefs are, the more they leverage to buy assets. High leverage generates large fluctuations in their wealth, and as a consequence, in asset prices.

The results concerning volatility of asset prices also translate into volatility of real investment. Consequently, real investment under incomplete markets exhibits higher volatility than under complete markets. To illustrate this result, I choose a special case in which the aggregate endowment and productivity are constant over time. Under complete markets, as economies converge to economies with homogeneous beliefs, capital levels converge to their steady-state levels. Investments are therefore approximately constant; investment volatility is approximately zero. In contrast, under incomplete markets investment volatility remains strictly positive because it depends on the wealth distribution and the wealth distribution constantly changes as aggregate shocks hit the economies.

It is also useful to highlight the role of dynamic general equilibrium for some results mentioned above. In particular, the infinite horizon nature of the framework allows a comprehensive analysis of short-run and long-run behavior of asset price volatility. Such an analysis is not possible in finite horizon economies, including Geanakoplos's important study on the effects of heterogeneous beliefs on leverage and crises. For example, in page 35 of Geanakoplos (2009), he observes similar volatility as the economy moves from incomplete to complete markets. In my model, the first set of results described above shows that the similarity holds only in the short run. The long run dynamics of asset price volatility totally differs from complete to incomplete markets. In my model, the results are also based on insights in Blume and Easley (2006) and Sandroni (2000) regarding the disappearance of agents with incorrect beliefs. However, these authors do not focus on the effect of their disappearance on asset price or asset price volatility.

The second set of results that follow from this framework concerns collateral shortages. I show that collateral constraints will eventually be binding for every agent in complete markets equilibrium provided that the face values of the financial assets with collateral span the complete set of state-contingent Arrow-Debreu securities. Intuitively, if this was not the case, the unconstrained asset holdings would imply arbitrarily low levels of consumption at some state of the world for every agent, contradicting the result that consumption is bounded from below. In other words, there are always shortages of collateral even if I allow for an elastic supply of collateral. This result sharply contrasts with those obtained when agents have homogenous beliefs but still have reasons to trade due to differences in endowments or utility functions. In these cases, if the economy has enough collateral, or can produce it, then collateral constraints may not bind and the complete markets allocation is achieved. Heterogeneous beliefs, therefore, guarantee collateral shortages.

Another immediate implication of these results concerns Pareto inefficiency of incomplete markets equilibria. Incomplete markets equilibria are Pareto-suboptimal whenever agents strictly differ in their beliefs. This can be seen for the results that under complete markets equilibria, some agent's consumption will come arbitrarily close to zero while this never happens under incomplete markets. Intuitively, under complete markets agents pledged their future income, while collateral constraints put limits on such transactions. While allocations in which some agents experience very low levels of consumption may not be attractive according to some social welfare criteria, the equilibrium under complete markets is Pareto optimal under the subjective expectations of the agents. This result also implies that there is the possibility for Pareto improving regulations. However, given that this result is about unconstrained Pareto-efficiency, Pareto improving regulations might involve altering the incomplete markets structure.[2]

The above mentioned results are derived under the presumption that incomplete markets equilibria exist. However, establishing existence of incomplete markets equilibria is generally a challenging task. The third set of results establishes the existence of incomplete markets equilibria with a stationary structure. In their seminal paper, Geanakoplos and Zame (2002) shows that, with collateral constraints, the standard existence proof a la Debreu (1959) applies. Kubler and Schmedders (2003) extends the existence proof to infinite horizon economies. I use the insights from these works to show the existence of incomplete markets equilibria in finite and infinite horizon economies with production and capital accumulation. Following Kubler and Schmedders (2003), I look for *Markov equilibria*, i.e., in which equilibrium prices and quantities depend only on the distribution of *normalized financial wealth* and the total quantities of assets with elastic supply. I show the existence of the equilibria under standard assumptions. I also develop an algorithm, based on the algorithm in Kubler and Schmedders (2003), to compute these equilibria. The same algorithm can be used to compute the complete markets equilibrium benchmark. One direct corollary of the existence theorem is that the recursive equilibrium in Krusell and Smith (1998) exists.

The fourth set of results attempts to answer some normative questions in this framework. Simple and extreme forms of financial regulations such as shutting down financial markets are not beneficial. Using the algorithm described above, I provide numerical results illustrating that these regulations fail to reduce asset price volatility and moreover they may also reduces the welfare of all agents because of the restrictions they impose on mutually beneficial trades. In particular, the intuition for the greater volatility under such regulations is that, when the collateral constraints are binding, regulations restrict the demand for assets. Therefore asset prices are lower than they are in unregulated economies. Agents, however, will eventually save their way out of the constrained regime, at which point, asset prices will become comparable to the unregulated levels. Movements between constrained and unconstrained regimes create high asset price volatility. These results suggest that Pareto-improving or volatility reducing regulations must be sophisticated, for example, incorporating state-dependent regulations.

[2]For a two-period version of my model, the concept of constrained Pareto-inefficiency due to Geanakoplos and Polemarchakis (1986) can be checked. In some cases, the economy can be constrained inefficient in this sense, due to pecuniary externalities.

This paper is related to the growing literature studying collateral constraints, started with a series of paper by John Geanakoplos. The dynamic analysis of incomplete markets is closely related to Kubler and Schmedders (2003). They pioneer the introduction of financial markets with collateral constraints into a dynamic general equilibrium model with aggregate shocks and heterogeneous agents. There are two main technical contributions of this paper relative to Kubler and Schmedders (2003). The first is to introduce heterogeneous beliefs using Radner (1972) rational expectations equilibrium concept: even though agents assign different probabilities to the aggregate shocks, they agree on the equilibrium outcomes, including prices and quantities, once a shock is realized. This rational expectations concept differs from the standard rational expectation concept, such as the one used in Lucas and Prescott (1971), in which subjective probabilities should coincide with the true conditional probabilities given all the available information. The second is to introduce capital accumulation and production in a tractable way. Capital accumulation or real investment is modelled through intermediate asset producers with convex adjustment costs that convert old units of assets into new units of assets using final good.[3] The analysis of efficiency is related to Kilenthong (2009) and Kilenthong and Townsend (2009). They examine a similar but static environment.

My paper is also related to the literature on the effect of heterogeneous beliefs on asset prices studied in Xiong and Yan (2009) and Cogley and Sargent (2008). These authors, however, consider only complete markets. The survival of irrational traders is studied Long, Shleifer, Summers, and Waldmann (1990) and Long, Shleifer, Summers, and Waldmann (1991) but they do not have a fully dynamic framework to study the long run survival of the traders. Simsek (2009b) also studies the effects of belief heterogeneity on asset prices. He assumes exogenous wealth distributions to investigate the question which forms of heterogeneous beliefs affect asset prices. In contrast, I study the effects of the endogenous wealth distribution on asset prices as well as asset price volatility. Simsek (2009a) focuses on consumption volatility. He shows that as markets become more complete, consumption becomes more volatile as agents can speculate more. My first set of results suggests that this comparative statics only holds in the short run. In the long run, the reverse statement holds due to market selection.

[3] Lorenzoni and Walentin (2009) models capital accumulation with adjustment cost using used capital markets. Through asset producers, I assume markets for both used and new capital.

Related to the survival of agents with incorrect beliefs Coury and Sciubba (2005) and Beker and Chattopadhyay (2009) suggest a mechanism for agents' survival based on explicit debt constraints as in Magill and Quinzii (1994). These authors do not consider the effects of the agents' survival on asset prices. My framework is tractable enough for a simultaneous analysis of survival and its effects on asset prices and investment. Beker and Espino (2010) has a similar survival mechanism to mine based on the limited commitment framework in Alvarez and Jermann (2000). However, my approach to asset pricing is different because asset prices are computed explicitly as function of wealth distribution. Moreover, my approach also allows a comprehensive study of asset-specific leverage. Kogan, Ross, Wang, and Westerfield (2006) explore yet another survival mechanism but use complete markets instead.

The model in this paper is a generalization of Krusell and Smith (1998) with financial markets and adjustment costs. In particular, the existence theorem 1.2 shows that a recursive equilibrium in Krusell and Smith (1998) exists. Krusell and Smith (1998) derives numerically such an equilibrium, but they do not formally show its existence. My paper is also related to Kiyotaki and Moore (1997), although I provide a microfoundation for the financial constraint (3) in their paper using the endogeneity of the set of actively traded financial assets.

The rest of the paper proceeds as follow. In section 2, I present the model in its most general form and preliminary analysis of survival, asset price volatility and investment volatility under the complete markets benchmark as well as under incomplete markets. In section 3, I define and show the existence of incomplete markets equilibria under the form of Markov equilibria. In this section, I also prove important properties of Markov equilibria in this model. In section 4, I derive a general numerical algorithm to compute Markov and competitive equilibria. Section 5 focuses on assets in fixed supply with an example of only one asset to illustrate the ideas in sections 2 and 3. Section 6 concludes with potential applications of the framework in this paper. Lengthy proofs and constructions are in the Appendix.

1.2 General model

In this general model there are heterogeneous agents who differ in their beliefs about the future streams of dividend or about future productivities. There are also different types of assets (for examples trees, land, housing and machines) that differ in their adjustment costs, associated production technologies and collateral value.

1.2.1 The environment

There are H types of consumers, $h \in \mathcal{H} = \{1, 2, \ldots, H\}$ in the economy (there is a continuum of measure 1 of identical consumers in each type) with potentially different instantaneous preferences $U_h(c)$,discount rates β_h,endowments of good e_h and of labor L_h. They might also differ in their belief of the evolution of the aggregate productivities and of the aggregate dividend streams. In each period, there are S states of the world: $s \in \mathcal{S} = \{1, 2, \ldots, S\}$. Histories are denoted by

$$s^t = (s_0, s_1, \ldots, s_t),$$

the series of realizations of shocks up to time t. Notice that the space S can be chosen large enough to encompass both aggregate shocks, such as shocks to the productivity of aggregate production functions, to aggregate dividends, and idiosyncratic shocks, such as labor income shocks.[4]

There is only one final good in this economy. It can be consumed by consumer and can be used for the production of new units of assets. It is produced by final good producers specified below.

Real Assets: There are A types $a \in \mathcal{A} = \{1, 2, \ldots, A\}$ of physical assets.

Adjustment cost: There are two types of assets, one with elastic supply, $a \in \mathcal{A}_0$ and the other ones with fixed supply, $a \in \mathcal{A}_1$, associated with adjustment cost functions. Let A_0, A_1 respectively denote the numbers of assets with elastic and fixed supply.

We can think of assets with fixed supply, $a \in \mathcal{A}_1$, as having infinite adjustment costs, however for the rigorousness of the model, I treat them differently from the assets with elastic supply.

For each asset with elastic supply, $a \in \mathcal{A}_0$, in each period, k_a^n new units of asset a can be produced using k_a^o old units of asset a and $\Psi_a(k_a^n, k_a^o)$ units of the final good. The k_a^n new units are used for production in the next period. Let $q_{a,t}$ denote the ex-dividend price of each old

[4]See Krusell and Smith (1998) for a similar framework with incomplete market with both aggregate shocks and idyosyncratic shocks.

unit of asset a, and $q^*_{a,t}$ denote the price of each new unit of asset a. Notice that $\Psi_a(k^n_a, k^o_a)$ is the final good investment associated to asset a. One example typically used in macroeconomics, representing perfectly flexible investment, is

$$\Psi_a(k^n_a, k^o_a) = k^n_a - (1-\delta_a)k^o_a. \tag{1.1}$$

Another example with nonlinearity is the one used in Lorenzoni and Walentin (2009)

$$\Psi_a(k^n_a, k^o_a) = k^n_a - (1-\delta_a)k^o_a + \frac{\xi_a}{2}\frac{(k^n_a - k^o_a)^2}{k^o_a},$$

in which $0 < \xi_a < \min\{2(1-\delta_a), 1\}$.

We can also rewrite the adjustment cost under a more familiar form

$$k^n_a = (1-\delta_a)k^o_a + k^o_a\Phi\left(\frac{i_a}{k^o_a}\right), \tag{1.2}$$

in which i_a is real investment in terms of final good. $\Phi(.)$ is strictly increasing and weakly concave. Perfectly investment case(1.1) corresponds to $\Phi(x) = x$.

I make the following standard assumption on the adjustment cost function. This assumption ensures that the profit maximization of each asset producer yields upper-hemicontinous and convex solutions.

Assumption 1.1 *The adjustment cost function Ψ_a is homogeneous of degree* 1 *and convex in* $(k^n_a, -k^o_a)$. *Moreover, Ψ_a is strictly increasing in k^n_a and strictly decreasing in k^o_a.*

Production: Assets with fixed supply, $a \in \mathcal{A}_1$ generate a state-dependent stream of dividend $d_a(s)$. Asset with elastic supply can be used in production function with state-dependent production functions $F_a(K_a, L_a, s)$, in which K_a are units of assets of type a and L_a is labor of the type associated to the asset.

Similarly to the adjustment cost, I make the following standard assumption to ensure that the profit maximization of each final good producer yields upper-hemicontinous and convex solutions.

Assumption 1.2 *The production function $F_a(K_a, L_a, s)$ is homogeneous of degree* 1 *and con-*

cave in (K_a, L_a) and strictly increasing in both parameters.

One example is the standard Cobb-Douglas production function with state-dependent productivity used in the RBC literature

$$F_a(K_a, L_a, s) = A(s) K_a^{\alpha_a} L_a^{1-\alpha_a}.$$

Financial Assets: In each history s^t, there are also financial assets,

$$j \in \mathcal{J}_t = \{1, 2, \ldots, J_t\}.$$

The set of financial assets may depend on event nodes. Asset j traded at that node promises pay-off $b_j(s^{t+1}) = b_j(s_{t+1}) > 0$ in term of final good at the successor nodes $s^{t+1} = (s^t, s_{t+1})$. Agents can only sell the financial asset j if they hold shares of real assets as collateral. We associate j with an $A-$dimensional vector $k^j \geq 0$ of collateral requirements. If an agent sells one unit of security j, she is required to to hold k_a^j units of asset $a = 1, 2, \ldots, A$ as collateral. If an asset a can be used as collateral for different financial securities, the agent is required to invest k_a^j in each asset a for each $j = 1, \ldots, J$.

Since there are no penalties for default, a seller of the financial asset defaults at a node s^{t+1} whenever the total value of collateral assets falls below the promise at that state. By individual rationality, the actual pay-off of security j at node s^t is therefore always given by

$$f_{j,t+1}(s^{t+1}) = \min \left\{ b_j(s_{t+1}), \sum_{a=1}^{A} k_a^j \left(q_a(s^{t+1}) + d_a(s^{t+1})\right) \right\} \tag{1.3}$$

Let $p_{j,t}(s^t)$ denote price of security j at node s^t.

I allow k^j to depend on the current aggregate state as well as current and future prices. But I impose a lower bound on k^j to ensure that the supply of the financial assets are endogenously bounded in equilibrium. I also impose a upper bound on k^j to obtain a upper bound on prices of these financial assets in equilibrium. The lower and upper bounds can be chosen such that they are not binding in equilibrium.

Assumption 1.3 *There exist $\overline{k}$ and $\underline{k}$ strictly positive such that*

$$\underline{k} < k_a^j\left(s_t, d_t, q_t, q_{t+1}\right) < \overline{k}.$$

for all $a, j, s_t, d_t, q_t, q_{t+1}$.

By allowing k_a^j to depend on current and future prices, I want to capture the case

$$k_{a,t}^j = \max_{s^{t+1}|s^t}\left\{\frac{b_j\left(s_{t+1}\right)}{q_a\left(s^{t+1}\right) + d_a\left(s^{t+1}\right)}\right\}. \tag{1.4}$$

$k_{a,t}^j$ is the minimum collateral level that ensures no default. Therefore

$$f_{j,t+1}\left(s^{t+1}\right) = b_j\left(s_{t+1}\right).$$

This constraint captures the situation in Kiyotaki and Moore (1997) in which agents can borrow only up to the minimum across future states of the future value of their land[5]. With $S = 2$, and state non-contingent debts, i.e., $b_j\left(s_{t+1}\right) = b_j$, Geanakoplos (2009) argues that even if we allow for a wide range of collateral level, that is the unique collateral level that prevails in equilibrium. This statement for two future states still holds in this context of infinitely-lived agents as proved later in subsection 1.5. However, this might not be true if we have more than 2 future states.

Beside the group of consumers, there are two other groups of agents in this economy: the asset producers and the final good producers. These producers live only for one period, therefore they do not have to make inter-temporal decisions.

Asset Producers: In each state, there are A_0 representative asset producers. Asset producer $a \in \mathcal{A}_0$ produces $K_{a,t}^n$ unit of new asset from $K_{a,t-1}^o$ old units of old assets and $\Psi_a\left(K_{a,t}^n, K_{a,t-1}^o\right)$ units of final good. The producers take prices $q_{a,t}^*$ and $q_{a,t}$ as given to maximize their profit

[5] Of course, the collateral level in (1.4) does not satisfy Assumption 1.3. However, we can use an alternative collateral level $\overline{k}_a^j = \max\left(k_a^j, \epsilon\right)$ and show that in equilibrium $\overline{k}_a^j = k_a^j$ if we choose ϵ small enough.

$$\pi_t^a = \max_{\substack{K_{a,t}^n, K_{a,t-1}^o \geq 0 \\ \psi_{a,t} \geq \Psi_a\left(K_{a,t}^n, K_{a,t-1}^o\right)}} q_{a,t}^* K_{a,t}^n - \psi_{a,t} - q_{a,t} K_{a,t-1}^o. \tag{1.5}$$

Final Good Producers: In each state there is also A_0 representative final good producers. Producer $a \in \mathcal{A}_0$ produces $F_a(K_a, L_a, s)$ units of final good from K_a units of asset a and L_a units of labor associated to the asset[6]. The producers take rental prices $d_{a,t}$ and wages $w_{a,t}$ as given to maximize their profit

$$\pi_t^{f,a} = \max_{\substack{K_{a,t}^f, L_{a,t}, y_{a,t} \geq 0 \\ y_{a,t} \leq F_a\left(K_{a,t}^f, L_{a,t}, s_t\right)}} y_{a,t} - d_{a,t} K_{a,t}^f - w_t L_{a,t} \tag{1.6}$$

The consumers are the main actors in this economy, they make consumption saving and investment decisions based on their own assessment of the future prospects of the economy.

Consumers: In each state s^t, each consumer is endowed with $e_t^h = e^h(s_t)$ units of final good. I suppose there is a strictly positive lower bound on these endowments. This lower bound guarantees a lower bound on consumption, if a consumer decides to default on all her debt and withdraw from the financial markets.

Assumption 1.4 *There exists an* $\underline{e} > 0$ *such that* $e_h(s) > \underline{e}$ *for all* h *and* s*.*

For example, commercial banks receive deposits from their retail branches while these banks also have trading desks that trade independently in the financial markets.

She is also endowed with a vector of labor

$$L_h = \left(L_{h,a}(s_t)\right)_{a \in \mathcal{A}_0},$$

$L_{h,a}$ corresponds to labor associated with asset a.

The consumer maximizes her intertemporal expected utility with the per period utility function $U_h(.) : \mathbf{R}^+ \longrightarrow \mathbf{R}$ satisfies

Assumption 1.5 U_h *is concave and strictly increasing.*

[6]In an alternative model, assets use the same type of labor. That model is similar to the one presented here.

Notice that I do not require U_h to be strictly concave. This assumption captures linear utility functions in Geanakoplos (2009) and Harrison and Kreps (1978).

Consumer h takes sequences of prices as given and solves[7]

$$\max E_0^h \left[\sum_{t=0}^{\infty} \beta_h^t U_h\left(c_t^h\right)\right]$$

and in each history s^t, she is subject to the budget constraint

$$\begin{aligned}
& c_t^h + \sum_{a\in\mathcal{A}_1} q_{a,t}k_{a,t}^h + \sum_{a\in\mathcal{A}_0} q_{a,t}^* k_{a,t}^h + \sum_{j=1}^{J} p_{j,t}\phi_{j,t}^h \\
\leq\ & e_t^h + \sum_{a\in\mathcal{A}_0} w_{a,t}l_{a,t}^h + \sum_{j=1}^{J} f_{j,t}\phi_{j,t-1}^h + \\
& + \sum_{a\in\mathcal{A}_0} \left(q_{a,t} + d_{a,t}\right) k_{a,t-1}^h + \sum_{a\in\mathcal{A}_0} \Pi_a^h + \sum_{a\in\mathcal{A}_0} \Pi_a^{f,h} \\
& + \sum_{a\in\mathcal{A}_1} \left(q_{a,t} + d_{a,t}\right) k_{a,t-1}^h
\end{aligned} \tag{1.7}$$

the collateral constraints

$$k_{a,t}^h + \sum_{j:\phi_{j,t}<0} \phi_{j,t}^h k_{a,t}^j \geq 0 \tag{1.8}$$

One implicit condition from the assumption on utility functions is that consumptions are positive, i.e., $c_t^h \geq 0$. In the constraint (1.8), if the consumer does not use asset a as collateral to sell any financial security, then the constraint becomes the no-short sale constraint

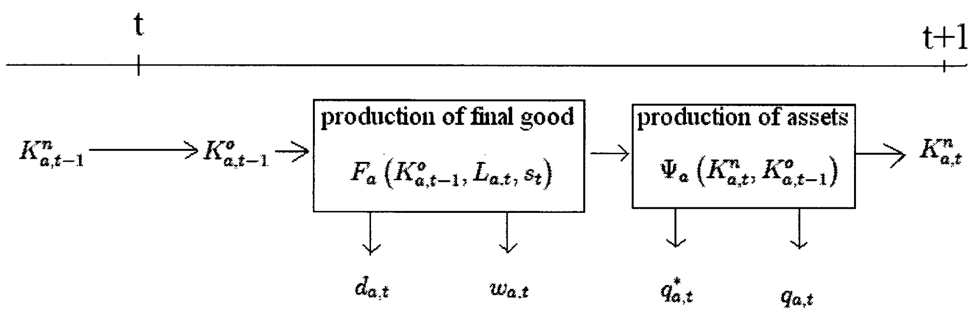
$$k_{a,t}^h \geq 0. \tag{1.9}$$

In the budget constraint (1.7), e_t^h is her endowment that can depend on the aggregate state s_t. Entering period t, the agent holds $k_{a,t-1}^h$ old units of real asset a and $\phi_{j,t-1}^h$ units of financial asset j. She can trade old units of real asset a at price $q_{a,t}$, buy new units of asset $k_{a,t}^h$ for time $t+1$ at price $q_{a,t}^*$. She can also buy and sell financial securities $\phi_{j,t}^h$ at price $p_{j,t}$. If she sell financial securities she is subject to collateral requirement (1.8). Finally, she works at the

[7] We can also introduce the disutility of labor in order to study employment in this environment. The existence of equilibria for finite horizon allows for labor choice decision.

wage $w_{a,t}$ in each production sector a. She also receives her shares of profit from the asset producer and final good producer at time t, Π_a^h and $\Pi_a^{f,h}$. However, given the homogeneity of the production functions, these profits should be zero in equilibrium.

Within a period, timing of decisions and actions taken by the agents are summarized in the following figure:



A number of features is worth noting in this setup: The demand of the consumers for new assets is similar to Tobin's Q theory of investment. They weigh the perceived marginal benefit of one additional unit of an asset a: future rental price, $d_{a,t+1}$, and future resale value $q_{a,t+1}$, against the marginal cost of buying one new unit of that asset at price $q^*_{a,t}$. The total demand for new units of asset a from the consumers is decreasing in price $q^*_{a,t}$ and the supply of the asset from the asset producers is increasing in $q^*_{a,t}$. In equilibrium both $q^*_{a,t}$ and $K^n_{a,t}$ are determined simultaneously. For instance, if the consumers expect low future resale price of an asset, they will demand less for new units of the asset. This low demand leads to low current price and low investment in the asset.

In this environment, I define an equilibrium as follows

Definition 1.1 *An **incomplete markets equilibrium** for an economy with initial asset holdings*

$$\left\{k^h_{a,0}\right\}_{h\in\{1,2,\ldots,H\}}$$

and initial shock s_0 is a collection

$$
\begin{aligned}
&(\left\{c_t^h\left(s^t\right), l_{a,t}^h\left(s^t\right), k_{a,t}^h\left(s^t\right), \phi_{j,t}^h\left(s^t\right)\right\}_{h\in\{1,2,\ldots,H\}} \\
&\left\{K_{a,t}^n\left(s^t\right), K_{a,t}^o\left(s^t\right), \psi_{a,t}\left(s^t\right)\right\}_{a\in\mathcal{A}_0} \\
&\left\{K_{a,t}^f\left(s^t\right), L_{a,t}\left(s^t\right), y_{a,t}\left(s^t\right)\right\}_{a\in\mathcal{A}_0} \\
&\left\{q_{a,t}^*\left(s^t\right), q_{a,t}\left(s^t\right), d_{a,t}\left(s^t\right), w_{a,t}\left(s^t\right)\right\}_{a\in\mathcal{A}_0} \\
&\left\{q_{a,t}\left(s^t\right)\right\}_{a\in\mathcal{A}_1}, \left\{p_{j,t}\left(s^t\right)\right\}_{j\in J_t(s^t)})
\end{aligned}
$$

satisfying the following conditions

i) Asset markets, labor market for each asset with elastic supply $a \in \mathcal{A}_0$ in each period clears: Demand by the consumers for new units of assets a equals supply of new units by the asset a producer:

$$\sum_{h=1}^{H} k_{a,t}^h\left(s^t\right) = K_{a,t}^n\left(s^t\right),$$

Demand by the asset a producer for old units of assets a equal supply of old units by the consumers:

$$K_{a,t}^o\left(s^t\right) = \sum_{h=1}^{H} k_{a,t-1}^h\left(s^t\right),$$

Demand by the asset a final good producer for old units of assets a equal supply of old units by the consumers:

$$K_{a,t}^f\left(s^t\right) = \sum_{h=1}^{H} k_{a,t-1}^h\left(s^t\right).$$

Labor demand by the asset a final good producer equal total labor supply by the consumers:

$$L_{a,t} = \sum_{h=1}^{H} L_{a,t}^h\left(s^t\right).$$

Market for each financial asset j clears:

$$\sum_{h=1}^{H} \phi_{j,t}^h\left(s^t\right) = 0.$$

ii) For each consumer h, $\left\{c_t^h\left(s^t\right), k_{a,t}^h\left(s^t\right), \phi_{j,t}^h\left(s^t\right)\right\}$ *solves the individual maximization problem subject to the budget constraint,* (1.7)*, and the collateral constraint,* (1.8)*. Asset producers and final good producers maximize their profit as in* (1.5) *and* (1.6)*.*

Notice that by setting the set of financial securities J_t is empty in each event node, we obtain a model with no financial markets, agents are only allowed to trade in real assets, but they cannot short-sell these assets. There are two important special cases of such a model. The first one is the case in which there are only asset in fixed-supply, i.e., $\mathcal{A}_0$ is empty. This case corresponds to Lucas (1978a)'s model with several trees and heterogeneous agents. The second one is the case in which there is only one asset with perfect elastic supply, i.e., adjustment cost described in (1.1). This case corresponds to Krusell and Smith (1998)'s model if we expand the set of aggregate shocks to incorporate idiosyncratic shocks of each individual and allow for a large number of agents. Therefore we can apply the existence proof in section 1.3 to show the existence of the recursive equilibrium in their original paper.

As benchmark I also study equilibrium with complete financial markets. Consumers and borrow and lend freely by buying and selling Arrow-Debreu state contingent securities, only subject to the no-Ponzi condition. In each node s^t, there are S financial securities. Financial security s deliver one unit of final good if state s happens at time $t+1$ and zero otherwise. Let $p_{s,t}$ denote time t price and let $\phi_{s,t}^h\left(s^t\right)$ denote consumer h's holding of this security. The budget constraint (1.7) of consumer h becomes

$$\begin{aligned}
& c_t^h + \sum_{a\in\mathcal{A}_1} q_{a,t}k_{a,t}^h + \sum_{a\in\mathcal{A}_0} q_{a,t}^* k_{a,t}^h + \sum_{s\in\mathcal{S}} p_{s,t}\phi_{s,t}^h \\
\leq\ & e_t^h + \sum_{a\in\mathcal{A}_0} w_{a,t}l_{a,t}^h + \phi_{s_t,t-1}^h + \\
& + \sum_{a\in\mathcal{A}_0}\left(q_{a,t}+d_{a,t}\right)k_{a,t-1}^h + \sum_{a\in\mathcal{A}_0}\Pi_a^h + \sum_{a\in\mathcal{A}_0}\Pi_a^{f,h} \\
& + \sum_{a\in\mathcal{A}_1}\left(q_{a,t}+d_{a,t}\right)k_{a,t-1}^h \qquad (1.10)
\end{aligned}$$

Definition 1.2 *A* ***complete markets equilibrium*** *is defined similarly to incomplete markets equilibrium except that each consumer solves her individual maximization problem subject to the budget constraint* (1.10) *and the no-Ponzi condition, instead of the collateral constraint* (1.8)*.*

In the next subsection, I establish some properties of incomplete markets equilibrium. I compare each of these properties to the one of complete markets equilibrium.

1.2.2 General properties of incomplete and complete markets equilibria

First, I restrict myself to studying equilibria in which the total quantities of assets with elastic supply are bounded, i.e., for each $a \in \mathcal{A}_0$, there exists a upper bound $\overline{K}_a$ such that $K_{a,t}\left(s^t\right) \leq \overline{K}_a$ for all t, s^t. Given this restriction we can show easily that total supply of final good in each period is bounded by a constant $\overline{e}$. Indeed in each period, total supply of final good is

$$\begin{aligned} & \sum_{h\in\mathcal{H}} e^h + \sum_{a\in\mathcal{A}_1} d_a K_a + \sum_{a\in\mathcal{A}_0} F\left(K_a^o, \sum_h L_h^a\right) - \sum_{a\in\mathcal{A}_0} \Psi_a\left(K_a^n, K_a^o\right) \\ \leq & \sum_{h\in\mathcal{H}} e^h + \sum_{a\in\mathcal{A}_1} d_a K_a + \sum_{a\in\mathcal{A}_0} F\left(\overline{K}_a, \sum_h L_h^a\right) - \sum_{a\in\mathcal{A}_0} \Psi_a\left(0, \overline{K}_a\right) < \overline{e}. \end{aligned} \tag{1.11}$$

The first term is the total final good endowment of each individual. The second is total dividends from fixed-supply assets. The third and forth terms are the maximum amount of final good that can be produced using elastic supply assets. Given that $\mathcal{S}$ is finite; we can choose an upper bound $\overline{e}$ of total final good over aggregate states $s \in \mathcal{S}$. In incomplete or complete markets equilibria, the market clearing condition for final good implies that total consumption is bounded from above by $\overline{e}$. Given that consumption of every agent is always positive, consumption of each agent is bounded from above by $\overline{e}$, i.e.,

$$c_{h,t}\left(s^t\right) \leq \overline{e}\ \forall t, s^t. \tag{1.12}$$

Under the boundedness of total quantities of assets, we can show that in any incomplete markets equilibrium, consumption of consumers is bounded from below by a strictly positive constant $\underline{c}$. Two assumptions are important for this result. First, no-default penalty allow consumers, at any moment in time, to walk away from their past debts and only lose their collateral assets. After defaulting, they can always keep their non-financial wealth (inequality (1.14) below). Second, increasingly large speculation by postponing current consumption is not an equilibrium strategy, because in equilibrium consumption is bounded by $\overline{e}$ (inequality (1.15) below). This assumptions prevent agents from constantly postpone their consumption to buy assets. Formally, we have the following proposition

Theorem 1.1 *Suppose that in a incomplete markets equilibrium, there is an upper bound on total quantities of assets with elastic supply. Moreover, there exists $\underline{c}$ such that*

$$U_h(\underline{c}) < \frac{1}{1-\beta} U_h(\underline{e}) - \frac{\beta}{1-\beta} U_h(\overline{e}), \tag{1.13}$$

where $\overline{e}$ is defined in (1.11).*Then in a incomplete markets equilibrium, consumption of each consumer in each history node always exceeds $\underline{c}$.*

Proof. As in (1.12) we can find an upper bound for consumption of each consumer. In each period one of the feasible strategies of consumer h is to default on all her past debts and consume her endowment from the current period on, therefore

$$U_h(c_{h,t}) + E_t^h \left[\sum_{r=1}^{\infty} \beta^r U_h(c_{h,t+r}) \right] \geq \frac{1}{1-\beta} U_h(\underline{e}). \tag{1.14}$$

Notice that in equilibrium, $\sum_h c_{h,t+r} \leq \overline{e}$ therefore $c_{h,t+r} \leq \overline{e}$. So

$$U_h(c_h) + \frac{\beta}{1-\beta} U_h(\overline{e}) \geq \frac{1}{1-\beta} U_h(\underline{e}) \tag{1.15}$$

This implies

$$U_h(c_h) \geq \frac{1}{1-\beta} U_h(\underline{e}) - \frac{\beta}{1-\beta} U_h(\overline{e}) > U_h(\underline{c})$$

■

Two remarks can be made here. First, condition (1.13) is automatically satisfied if

$$\lim_{c \longrightarrow 0} U_h(c) = -\infty,$$

for example, with log utility or CRRA utility with CRRA constant exceeds 1. Second, the lower bound of consumption, $\underline{c}$, is decreasing in $\overline{e}$. Therefore, the more the total available final good, the more profitable speculative activities are and the more incentives consumers have to defer current consumption to engage into these activities.

One immediate corollary of this proposition is that, every consumer survives in equilibrium. Therefore, incomplete markets equilibrium differs from complete markets equilibrium when consumers differ in their beliefs. The proposition below shows that in a complete markets equilibrium, with strict difference in beliefs, consumption of certain consumer will come arbitrarily close to 0 at some event node. The intuition for this result is that, if an agent believes that the likelihood of a state is much smaller than what other agents believe, the agent will want to exchange his consumption in that state for consumption in other states. Complete markets allow her to do so but, in incomplete markets equilibrium, collateral constraint limits the amount of consumption that she can sell in each state.

Proposition 1.1 *Suppose there is an upper bound on total quantities of assets with elastic supply and consumers have strictly heterogeneous beliefs. Moreover, the utility functions satisfy the Inada-condition*

$$\lim_{c \longrightarrow 0} U_h'(c) = +\infty.$$

Then, in a competitive equilibrium with complete markets, consumption of some agent comes arbitrarily close to zero at some state of the world. Formally

$$\inf_{h,s^t} c_h\left(s^t\right) = 0.$$

Proof. From the first-order condition

$$\left(\prod_{r=0}^{t-1} p_{s_{r+1}}\left(s^r\right)\right) U_h'\left(c_{h,0}\right) = P_h\left(s^t|s_0\right) U_h'\left(c_t\left(s^t\right)\right)$$

Therefore for h, h'

$$\frac{U_h'\left(c_h\left(s^t\right)\right)}{U_{h'}'\left(c_{h'}\left(s^t\right)\right)} = \frac{P_{h'}\left(s^t|s_0\right) U_{h'}\left(c_{h'}\left(0\right)\right)}{P_h\left(s^t|s_0\right) U_h'\left(c_h\left(s_0\right)\right)} \tag{1.16}$$

From inequality (1.12), we have $U_{h'}'\left(c_{h'}\left(s^t\right)\right) > U_{h'}'\left(\bar{e}\right)$, therefore

$$U_h'\left(c_h\left(s^t\right)\right) > \frac{P_{h'}\left(s^t|s_0\right) U_{h'}\left(c_{h'}\left(0\right)\right)}{P_h\left(s^t|s_0\right) U_h'\left(c_h\left(s_0\right)\right)} U_{h'}'\left(\bar{e}\right).$$

But given heterogeneity in belief, we can find s^t and h, h' such that $\frac{P_{h'}\left(s^t|s_0\right)}{P_h\left(s^t|s_0\right)}$ gets arbitrarily large. So $c_h(s^t)$ goes to zero as $\frac{P_{h'}(s^t|s_0)}{P_h(s^t|s_0)}$ goes to infinity. ■

Blume and Easley (2006) and Sandroni (2000) show an even stronger result: Under some agent's belief, with probability one, consumption of agents whose beliefs strictly differ from hers goes to zero at infinity. Their proofs use difficult results from probability theory, however the first-order conditions 1.16 play the main role in the proofs.

The survival mechanism in Theorem 1.1 is similar to the one in Beker and Espino (2010) which is again based on Alvarez and Jermann (2000). The idea is that agents have limited ability to pledge their future income, for example labor income. As the result they can always default and keep their future income. This limited commitment is even stronger in my setting than in Alvarez and Jermann (2000) and Beker and Espino (2010) because after defaulting agents can always come back and trade in the financial markets by buying new physical assets. This survival mechanism also shows that agents can disappear in Blume and Easley (2006) and Sandroni (2000) because they can perfectly commit to pay their creditor using their future income. They can do so using short-term debts and keep rolling over their debts while using their present income to pay the interests.[8]

Due to different conclusions about agents' survival, the following corollary asserts that complete and incomplete markets allocations strictly differ when some agents strictly differ in their beliefs.

Corollary 1.1 *Suppose that conditions in Theorem 1.1 and Proposition 1.1 are satisfied and some agents strictly differ in their beliefs. Then, an incomplete markets equilibrium never yields an allocation that can be supported by a complete markets equilibrium. By the Second Welfare Theorem, incomplete markets equilibrium allocations are Pareto-inefficient.*

Proof. In a incomplete markets equilibrium, consumptions are bounded away from 0, but

[8] The following story of the founder of Long Term Capital Management show that traders in the financial markets often have limited commitment: John Meriwether worked as a bond trader at Salomon Brothers. At Salomon, Meriwether rose to become the head of the domestic fixed income arbitrage group in the early 1980s and vice-chairman of the company in 1988. In 1991, after Salomon was caught in a Treasury securities trading scandal Meriwether decided to leave the company. Meriwether founded the Long-Term Capital Management hedge fund in Greenwich, Connecticut in 1994. Long-Term Capital Management spectacularly collapsed in 1998. A year after LTCM's collapse, in 1999, Meriwether founded JWM Partners LLC. The Greenwich, Connecticut hedge fund opened with $250 million under management in 1999 and by 2007 had approximately $3 billion. The Financial crisis of 2007-2009 badly battered Meriwether's firm. From September 2007 to February 2009, his main fund lost 44 percent. On July 8, 2009, Meriwether closed the fund.

in a complete markets equilibrium, consumptions of some agents will approach 0. Therefore, the two sets of allocations never intersect. ■

Using this corollary, we can formalize and show the shortages of collateral assets.

Proposition 1.2 (Collateral Shortages) *Suppose that J_t includes complete set of state-contingent Arrow-Debreu securities. Then, for any given time t, the collateral constraints must be binding for some agent after time t, despite the fact that collateral assets can be produced.*[9]

Proof. We prove this corollary by contradiction. Suppose none of the collateral constraints are binding after certain date. Then we can take the first-order condition with respect to the state-contingent securities. This leads to consumption of some agent approaches zero at infinity, as shown in the proof of Proposition 1.1. This contradicts the conclusion of Theorem 1.1. ■

Araujo, Kubler, and Schommer (2009) argue that when there are enough collateral we might reach the Pareto optimal allocation. However, in the complete markets case, there will never be enough collateral. Moreover, this conclusion holds even if we allow for elastic supply of collateralizable assets. Collateral shortages in this context mean that at some point in time some agent only hold the assets for collateral purposes but not for investment and saving purposes.

We also emphasize here the difference between belief heterogeneity and other forms of heterogeneity such as heterogeneity in endowments or in risk-aversion. The following proposition, in the same form Theorem 5 in Geanakoplos and Zame (2007), shows that if consumers share the same belief and discount rate, there exist endowment profiles with which, collateral equilibria attain the first-best allocations.

Proposition 1.3 *If consumers share the same belief and discount factor, there is an open set of endowment profiles with the properties that the competitive equilibrium can be supported by a financial market equilibrium.*

Proof. We start with an allocation such that there is no trade in the complete markets equilibrium, then as we move to a neighborhood of that allocation, all trade can be collateralized. ■

[9] *It can also be shown that, at any moment of time, for* every *agent, the collateral constraint must be binding some time in future.*

Lastly, we go back to the complete markets benchmark to study the behavior of asset price volatility. We will compare this volatility with the one in the collateralized economy and show that, in general, in the long run, asset price more volatile in a incomplete markets equilibrium than it is in a complete markets equilibrium.

Proposition 1.4 *Suppose that there are some agent with the correct belief and that there are no elastic supply assets. Then in the complete markets equilibrium, asset prices are independent of past realizations of the aggregate shocks in the long run.*

Proof. Blume and Easley (2006) shows that in the long run, only agents with correct belief survive. Therefore, in the long run, we fall back to the case with homogeneous belief. Given markets completeness, there exists a representative agent with instantaneous utility function U_{Rep}, and her marginal utility evaluated at the total endowment determines asset prices

$$\begin{aligned} q_a\left(s^t\right) U'_{\text{Rep}}\left(e\left(s_t\right)\right) &= \beta E_t^{\text{Rep}}\left\{\left(q_a\left(s^{t+1}\right) + d_a\left(s_{t+1}\right)\right) U'_{\text{Rep}}\left(e\left(s_{t+1}\right)\right)\right\} \\ &= E_t^{\text{Rep}}\left\{\sum_{r=1}^{\infty} d_a\left(s_{t+r}\right) \beta^r U'_{\text{Rep}}\left(e\left(s_{t+r}\right)\right)\right\} \end{aligned} \tag{1.17}$$

in which $e\left(s\right)$ is the aggregate endowment in the aggregate state s. We can see easily from this expression that $q_a\left(s^t\right)$ is history-independent. ∎

When there are assets with elastic supply, this proposition should be modified as, controlling for total quantities of assets with elastic supply, asset prices are independent of past realizations of the aggregate shocks.

In contrast to complete markets equilibrium, in the next section we will show that, in incomplete markets equilibrium, asset prices can be history-dependent, as past realizations of aggregate shocks affect the wealth distribution, which in turns affects asset prices.

One issue might arise when one tries to interpret Proposition 1.4 is that, in some economy, there might not be any consumer whose belief coincides with the truth. For example, in Scheinkman and Xiong (2003), all agents can be wrong all the time, except they constantly switch from over-optimistic to over-pessimistic. To avoid this issue, I use the language in Blume and Easley (2006) and Sandroni (2000). I reformulate the results above using the subjective belief of each consumer.

Proposition 1.5 *Suppose that there are no assets with elastic supply, and condition* (1.13) *is satisfied. Then each agents believes that:*

1) In complete markets equilibrium, only her and consumers sharing her belief survive in the long run. However, in incomplete markets equilibrium, everyone survives in the long run.

2) In complete markets equilibrium, asset prices are history-independent. However, in incomplete markets equilibrium, asset price can be history-dependent.

The properties in this section are established under the presumption that incomplete markets equilibria exists. The next section is devoted to show the existence of these equilibria with a stationary structure. The next two sections follow closely the organization in Kubler and Schmedders (2003). The first shows the existence and the second presents an algorithm to compute the equilibria.

1.3 Markov Equilibrium

1.3.1 The state space

I define the financial wealth of each agent by

$$\omega_t^h = \frac{\sum_a (q_{a,t} + d_{a,t})\, k_{a,t-1}^h + \sum_j \phi_{j,t}^h f_{j,t-1}}{\sum_a (q_{a,t} + d_{a,t})\, K_{a,t-1}^o}.$$

Let $\omega\left(s^t\right) = \left(\omega^1\left(s^t\right), ..., \omega^H\left(s^t\right)\right)$. Then in equilibrium $\omega\left(s^t\right)$ always lies in the (H-1)-dimensional simplex Ω, i.e., $\omega^h \geq 0$ and $\sum_{h=1}^H \omega^h = 1$. ω^h's are positive because of the collateral constraint (1.8) that requires the value of each agents' asset holdings to exceed the liabilities from their past financial assets holdings. And the sum of ω^h equals 1 because of the asset market clearing and financial market clearing conditions.

I will show that, under conditions detailed in Subsection 1.3.3 below, there exists a Markov equilibrium over a compact state space. I look for an equilibrium in which equilibrium prices and allocations depend only on the states $\left(s_t, \omega_t, K_{t-1}^o\right) \in \mathcal{S} \times \Omega \times E$, in which

$$E = \prod_{a \in \mathcal{A}_0} \left[0, \overline{K}_a\right].$$

$K_a^o \in [0, \overline{K}_a]$ are the total old units of assets with elastic supply at the beginning of a period.

Let the state space $\mathcal{X}$ consist of all exogenous and endogenous variables that occur in the economy at some node σ, i.e., $\mathcal{X} = \mathcal{S} \times \mathcal{V}$, where $\mathcal{S}$ is the finite set of exogenous shocks and $\mathcal{V}$ is the set of all possible endogenous variables.

In each node σ, an element $v(\sigma) \in \mathcal{V}$ includes: the normalized wealth distribution $(\omega_h(\sigma))_{h\in\mathcal{H}} \in \Omega$, the total old units of assets with elastic supply $(K_a^o)_{a\in\mathcal{A}_0} \in E$; together with consumers' decisions: consumption, $H + HA_0$ current consumption and labor supply $\left(c^h(\sigma), l_a^h(\sigma)\right)_{h\in\mathcal{H}}$, $HA+HJ$ real and financial asset holdings $\left(k_a^h(\sigma), \phi_j^h(\sigma)\right)_{h\in\mathcal{H}}$. It also includes the $4A_0$ current prices of new units of elastic supply assets, the prices of old units of these assets, the rental prices and wages associated with these assets

$$\left(q_a^*(\sigma), q_a(\sigma), w_a(\sigma), d_a(\sigma)\right)_{a\in\mathcal{A}_0},$$

and A_1 prices of assets with fixed supply $(q_a(\sigma))_{a\in\mathcal{A}_1}$. Finally it includes J prices of the financial assets $(p_j(\sigma))_{j\in J}$. Therefore $\mathcal{V} = \Omega \times E \times \widehat{\mathcal{V}}$ with

$$\widehat{\mathcal{V}} = \mathbf{R}_+^H \times \mathbf{R}_+^{HA_0} \times \mathbf{R}_+^{AH} \times \mathbf{R}^{JH} \times \mathbf{R}_+^{4A_0} \times \mathbf{R}_+^{A_1} \times \mathbf{R}_+^J \tag{1.18}$$

the set of endogenous variables other than the wealth distribution and total old quantities of assets with elastic supply.

Finally, let $\mathcal{X} \subset \mathcal{V}$ denote the set of vectors of all the endogenous variables that satisfy: 1) financial markets clears, 2) producers maximize their profit and 3) the budget constraints of consumers bind. Formally,

$$\sum_h \phi_j^h = 0$$

and

$$l_a^h = L_{h,a}.$$

In addition, for each $a \in \mathcal{A}_0$, given $K_a^n = \sum_h k_a^h$ and $L_a = \sum_h l_a^h$ we have

$$(K_a^n, K_a^o, \psi_a) \in \underset{\substack{\widetilde{K}_a^n, \widetilde{K}_a^o \geq 0 \\ \widetilde{\psi}_a \geq \Psi_a\left(\widetilde{K}_a^n, \widetilde{K}_a^o\right)}}{\arg\max} \quad q_a^* \widetilde{K}_a^n - \widetilde{\psi}_a - q_a \widetilde{K}_a^o \tag{1.19}$$

and

$$(K_a^o, L_a, y_a) \in \underset{\substack{\widetilde{K}_a^f, \widetilde{L}_a, \widetilde{y}_a \geq 0 \\ \widetilde{y}_a \leq F_a\left(\widetilde{K}_a^f, \widetilde{L}_a, s\right)}}{\arg\max} \quad \widetilde{y}_a - d_a \widetilde{K}_a^f - w_a \widetilde{y}_a \tag{1.20}$$

and consumers' budget constraints hold with equality[10][11]

$$c^h = e^h + w \cdot l + \omega^h (q + d) \cdot K^o - q^* \cdot k - p \cdot \phi. \tag{1.21}$$

Notice that profit maximizations $(1.19), (1.20)$ and binding budget constraints imply that good market clears

$$\sum_h c^h + \sum_{a \in \mathcal{A}_0} \Psi_a (K_a, K_a^o) = \sum_h e^h + \sum_{a \in \mathcal{A}_0} F_a (K_a^o, L_a, s).$$

1.3.2 Markov Equilibrium Definition

In order to define a Markov equilibrium, I use the following definition of expectation correspondence. Given a state $(s, v) \in \mathcal{X}$, the 'expectation correspondence'

$$g : \mathcal{X} \rightrightarrows \mathcal{V}^S$$

describes all next period states that are consistent with market clearing and agents' first-order conditions. A vector of endogenous variables

$$\left(v_1^+, v_2^+, \ldots, v_S^+\right) \in g(x)$$

[10] With some abuse of notation, we use $q_a^* = q_a$ for $a \in \mathcal{A}_1$.

[11] Profit maximization conditions (1.19) and (1.20) imply zero profits from the producers, hence the absence of these profits in the consumers' budget constraint.

and $(s, v_s^+) \in \mathcal{X}$ for each $s \in \mathcal{S}$ if for all households $h \in \mathcal{H}$ the following conditions holds

a) For all $s \in \mathcal{S}$

$$\omega_s^{h+} = \frac{k^h \cdot (q_s^+ + d_s^+) + \sum_{j \in \mathcal{J}} \phi_j^h \min \left\{ b_j (s), \sum_{a \in A} k_a^j (q^+ + d^+) \right\}}{\sum_a (q^+ + d^+) \cdot K^+}.$$

b) There exist multipliers μ_a^h corresponding to collateral constraints such that

$$0 = \mu_a^h - q_a^* U_h' \left(c^h\right) + \beta_h E^h \left\{ \left(q_a^+ + d_a^+\right) U_h' \left(c^{h+}\right) \right\} \tag{1.22}$$
$$0 = \mu_a^h \left(k_a^h + \sum_{j \in \mathcal{J} : \phi_j^h < 0} k_a^j \phi_j^h \right)$$
$$0 \leq k_a^h + \sum_{j \in \mathcal{J} : \phi_j^h < 0} k_a^j \phi_j^h.$$

c) Define $\phi_j^h(-) = \max\left(0, -\phi_j^h\right)$ and $\phi_j^h(+) = \max\left(0, \phi_j^h\right)$, there exist multipliers $\eta_j^h(+)$ and $\eta_j^h(-) \in \mathbf{R}^+$ such that

$$\begin{aligned}
0 &= \sum_{a \in \mathcal{A}} \mu_a^h k_a^j - p_j U_h' \left(c^h\right) + \beta^h E^h \left\{ f_j^+ U_h' \left(c^{h+}\right) \right\} - \eta_j^h(-) \\
0 &= -p_j U_h' \left(c^h\right) + \beta^h E^h \left\{ f_j^+ U_h' \left(c^{h+}\right) \right\} + \eta_j^h(-) \\
0 &= \phi_j^h(+) \eta_j^h(+) \\
0 &= \phi_j^h(-) \eta_j^h(-).
\end{aligned}$$

Notice that for the case of assets with fixed supply, $q_a^* = q_a$.

Definition 1.3 *A Markov equilibrium consists of a (non-empty valued) 'policy correspondence', P, and a transition function F*

$$P : \mathcal{S} \times \Omega \times E \rightrightarrows \widehat{\mathcal{V}}$$

and

$$F : graph(P) \longrightarrow \mathcal{V}^{\mathcal{S}}$$

such that $graph(P) \subset \mathcal{X}$ and for all $x \in graph(P)$ and all $s \in \mathcal{S}$ we have $F(x) \subset g(x)$ and $(s, F_s(x)) \in graph(P)$.

Lemma 1.1 *A Markov equilibrium is a incomplete markets equilibrium according to Definition 1.1.*

Proof. This result is similar to the one in Duffie, Geanakoplos, Mas-Colell, and McLennan (1994). We only need to show that the first order conditions as represented by Lagrange multipliers are sufficient to ensure the optimal solution of the consumers. This holds because the optimization each consumer faces is a convex maximization problem. ■

Before continue, let me briefly discuss asset prices and real investment in a Markov equilibrium.

Notice that in (1.22), $q^*_{a,t} = q_{a,t}$ for assets in fixed-supply. We can rewrite that first-order condition with respect to asset holding (1.22) as

$$q_a U'_h\left(c^h\right) = \mu^h_a + \beta_h E^h\left\{\left(q^+_a + d^+_a\right) U'_h\left(c^{h+}\right)\right\} \geq \beta_h E^h\left\{\left(q^+_a + d^+_a\right) U'_h\left(c^{h+}\right)\right\}.$$

By re-iterating this inequality we obtain

$$q_{a,t} \geq E^h_t\left\{\sum_{r=1}^{\infty} \beta^r_h d_{t+r} \frac{U'_h\left(c^h_{t+r}\right)}{U'_h\left(c^h\right)}\right\}.$$

We have a strict inequality if there is a strict inequality $\mu^h_{a,t+r} > 0$ in future. So the asset price is higher than the discounted value of the stream of its dividend because in future it can be sold to other agents, as in Harrison and Kreps (1978) or it can be used as collateral to borrow as in Fostel and Geanakoplos (2008). Proposition 1.2 shows some conditions under which collateral constraints will eventually be binding for every agents when they strictly differ in their belief. As a results, asset price is strictly higher than the discounted value of dividends.

Equation (1.22) also shows that asset a will have collateral value when some $\mu^h_a > 0$, in addition to the asset's traditional pay-off value weighted at the appropriate discount factors. Unlike in Alvarez and Jermann (2000), attempts to find a pricing kernel which prices assets using their pay-off value might prove fruitless because assets with the same pay-offs but different collateral values will have different prices. This point is also emphasized in Geanakoplos's papers.

For asset with elastic supply, the same equation (1.22) implies that the value of one unit of capital is increasing in the collateral value associated to the multiplier on the collateral constraints and short-sale constraints μ_a^h. Moreover when we derive the first-order of the capital producer's optimization problem using the form (1.2) we have

$$q_{a,t}^* \Phi' \left(\frac{I_{a,t}}{K_{a,t}} \right) = 1.$$

Given the concavity of Φ, this equation shows that the aggregate real investment, $I_{a,t}$, in asset a is an increasing function of asset price $q_{a,t}^*$ as in Tobin's Q theory of investment.

1.3.3 Existence and Properties of Markov equilibrium

The existence proof is based on Kubler and Schmedders (2003) and Magill and Quinzii (1994): approximating the Markov equilibrium by a sequence of equilibria in finite horizon. There are three steps in the proof. First, using Kakutani's fixed point theorem to prove the existence proof of the truncated T-period economy. Second, show that all endogenous variables are bounded. And lastly, show that the limit as T goes to infinity is the equilibrium of the infinite horizon economy.

However, the most difficult part, including in the two related papers, is to prove the second step and this step involves the most of problem-specific economics intuitions. Basically showing that quantities are bounded is easy (with collateral constraint especially), but showing prices are bounded is more challenging. For example, what are the upper bounds of prices of long-lived assets? These prices may well exceed the current aggregate endowment. With capital investment, we also have to bound the total supply of elastic-supply capital. I get around this difficulty by using the usual assumption in the neoclassical growth model: assuming capital depreciates and strictly concave production functions, then combine it with the artificial compact boxes trick in Debreu (1959).

Lemma 1.2 *Consider a finite horizon economy that last $T+1$ periods $t = 0, 1, \ldots, T$, identical to infinite horizon economy excepts consumers maximize the expected utility over $T+1$ periods*

$$E_0^h \left[\sum_{t=0}^{T} \beta_h^t U_h \left(c_t \right) \right]$$

and in the last period $t = T$, there are no financial markets. In the first period, the budget constraint of agents h is

$$\begin{aligned} c_0 + \sum_{a\in\mathcal{A}_0} q^*_{a,0}k_{a,0} + \sum_{a\in\mathcal{A}_1} q_{a,0}k_{a,0} + \sum_{j=1}^{J} p_{j,0}\phi_{j,0} \ \leq \ & e^h_0 + \sum_{a\in\mathcal{A}_0} w_{a,0}l_{a,0} \\ & + \omega^h_0 \sum_{a\in\mathcal{A}} \left(q_{a,0} + d_{a,0}\right) K_{a,-1} \end{aligned} \tag{1.23}$$

instead of (1.7) *.An equilibrium exists given any initial condition*

$$\left(s_0 \in \mathcal{S}, \omega_0 \in \Omega, \{K_{a,-1}\}_{a\in\mathcal{A}}\right).$$

Instead of the usual budget constraints using in recursive equilibria, we use the condition that each consumer holds a share of the final total value of assets. This sharing can be implemented by assuming each agent h holds exactly ω^h_0 share of each asset $a \in \mathcal{A}$.

Proof. The proof follows the steps in Debreu (1959) using Kakutani's fixed point theorem and is presented in the Appendix. However it uses a different definition of attainable sets. Indeed, in Definition 1.4 in the Appendix, negative excess demand (instead of zero excess demand as in the original text) is enough to guarantee the boundedness of the equilibrium allocations. In addition, I will also show that prices are strictly positive. ■

To prove that the Markov equilibrium exists, we need to first show that there exists a compact set in which finite horizon equilibria lie. We need the following three additional assumptions:

Assumption 1.6 *There exists a $\overline{K}_a > 0$ for each $a \in \mathcal{A}_0$ such that*

$$\Psi_a\left(\overline{K}_a, \overline{K}_a\right) \geq \max_{s\in\mathcal{S}} \overline{e}_a\left(s\right), \tag{1.24}$$

where

$$\begin{aligned}\overline{e}_a(s) &= \sum_{h=1}^{H} e^h(s) + \sum_{a' \in \mathcal{A}_1} d_{a'}(s) K_{a,0} + \sum_{a' \in \mathcal{A}_0} F_{a'}\left(\overline{K}_{a'}, \sum_{h=1}^{H} L^h_{a'}, s\right) \\ &+ \sum_{a' \in \mathcal{A}_0 \setminus \{a\}} \Psi_{a'}\left(0, \overline{K}_{a'}\right)\end{aligned}$$

Assumption 1.7 *The first-derivative of Ψ_a are bounded over $\left[0, \overline{K}_a\right]^2$.*

The first assumption ensures that total quantities of elastic-supply assets are bounded. For example, when we have only one elastic-supply asset and its supply is perfectly elastic, i.e., adjustment cost function is given by the flexible investment function (1.1) and the associated production is Cobb-Douglas with $\alpha_a \in (0,1)$. Then inequality (1.24) is equivalent to

$$\delta_a \overline{K}_a > const + A\left(\overline{K}_a\right)^{\alpha_a} L_a^{1-\alpha_a}$$

which must be true for $\overline{K}_a$ large enough. This is also the way one obtains a upper bound for capital in a neoclassical growth model. The second assumption, ensures that prices of new and old assets are bounded in equilibrium as they correspond to the first-derivatives of Ψ_a. For example, (1.1) gives

$$\begin{aligned}\frac{\partial \Psi_a\left(K^n_a, K^o_a\right)}{\partial K^n} &= 1 \\ \frac{\partial \Psi_a\left(K^n_a, K^o_a\right)}{\partial K^o} &= -\left(1-\delta_a\right).\end{aligned}$$

Remind that $\overline{e}$ is defined in (1.12).

Assumption 1.8 *There exist $\overline{c}, \underline{c} > 0$ such that*

$$\begin{aligned}& U_h(\underline{c}) + \max\left\{\frac{\beta_h}{1-\beta_h} U_h(\overline{e}), 0\right\} \\ \leq\ & \min\left\{\frac{1}{1-\beta_h} \min_{s \in \mathcal{S}} U_h(\underline{e}), \min_{s \in \mathcal{S}} U_h(\underline{e})\right\}\ \forall h \in \mathcal{H}. \qquad (1.25)\end{aligned}$$

and

$$\begin{aligned} & U_h(\bar{c}) + \min\left\{\frac{\beta_h}{1-\beta_h} U_h(\underline{e}), 0\right\} \\ \geq \ & \max\left\{\frac{1}{1-\beta_h} U_h(\bar{e}), U_h(\bar{e})\right\} \ \forall h \in \mathcal{H}. \end{aligned} \tag{1.26}$$

The intuition for (1.25) is detailed in the proof of proposition 1.1; it ensures a lower bound for consumption. (1.26) ensures that prices of assets with fixed supply are bounded from above. When there are only assets with elastic supply, the second inequality (1.26) is not needed. Both inequalities are obviously satisfied by log utility.

Lemma 1.3 *Suppose Assumptions 1.6, 1.7 and 1.8 are satisfied then there is a compact set that contains the equilibrium endogenous variables constructed in Lemma 1.2 for every T and every initial condition lying inside the set.*

Proof. Appendix. ■

Theorem 1.2 *Under the same conditions, a Markov equilibrium exists.*

Proof. Appendix. As in Kubler and Schmedders (2003), we extract a limit from the T-finite horizon equilibria. Lemma 1.3 guarantees that equilibrium prices and quantities are bounded as T goes to infinity. ■

Corollary 1.2 *In a Markov equilibrium, every consumer survives.*

Proof. From the construction of the equilibrium $c^h(s^t) > \underline{c}$ for all h, t, s^t. ■

Corollary 1.3 *The Markov equilibrium is Pareto-inefficient if agents strictly differ in their beliefs.*

Proof. In Proposition 1.1 we show that under complete market, i.e. Pareto efficient allocation, consumption of some agents get arbitrarily close to zero in some history. Given the lower bound on consumption of each Markov equilibrium, an allocation corresponding to a Markov equilibrium is not a complete markets allocation. Therefore it is not a Pareto efficient allocation. ■

Proposition 1.6 *In contrast to the complete markets benchmark, in these Markov equilibria, asset prices can be history-dependent.*

Proof. The realizations of aggregate shocks determine the evolution of the wealth distribution which is one factor that determines asset prices. ■

Proposition 1.7 *When aggregate endowment and aggregate productivity are constant, and shocks are I.I.D, long run asset price volatility and investment volatility are higher under incomplete markets than they are under complete markets.*

Proof. In the long run, under complete markets, the economy converges to the one with homogenous beliefs because agents with incorrect beliefs will eventually be driven out of the markets. We can thus find a representative agent. Standard arguments for representative agent economy imply that asset prices are constant and levels of investment converge to their steady state levels. For example, suppose we only have assets in fixed-supply and the aggregate endowment is independent of states: $e(s_t) = e$. Then, in equation (1.17) for the long run representative agent, we can divide both side of that equation by $U'_{\text{Rep}}(e)$ to obtain

$$\begin{aligned} q_a\left(s^t\right) &= E_t^{\text{Rep}}\left\{\sum_{r=1}^{\infty} d_a\left(s_{t+r}\right)\beta^r\right\} \\ &= \frac{\beta}{1-\beta}E_t^{\text{Rep}}\left\{d_a\left(s_{t+1}\right)\right\}. \end{aligned}$$

From the first to the second line, I use the fact that the dividends process is I.I.D. The last equality implies that asset price $q_a\left(s^t\right)$ is independent of time and state.

When we have assets in elastic supply, but with constant productivity, as in the neoclassical growth model, the total quantity of an asset a in fixed supply should converge to the steady-state level K_a^* which is determined by the

$$\frac{\partial \Psi_a\left(K_a^*, K_a^*\right)}{\partial K_a^n} = \beta\left(-\frac{\partial \Psi_a\left(K_a^*, K_a^*\right)}{\partial K_a^o} + F_{a,K}\left(K_a^*, L_a\right)\right)$$

and therefore the investment associated to this asset converges to $I_a^* = \Psi_a\left(K_a^*, K_a^*\right)$.

Hence, under complete markets, asset price volatility and investment volatility converge to zero in the long run. Under incomplete markets, asset price volatility and investment volatility remain well above zero as aggregate shocks constantly change the wealth distribution, which, in turn, changes asset prices and investment. ■

There are two components of asset price volatility. The first one comes from volatility in the dividend process and aggregate endowment. The second one comes from wealth distribution, when agents strictly differ in their beliefs. However, the second component disappears under complete markets because only agents with correct beliefs survive in the long run. Whereas, under incomplete markets, this component persists. As a result, when we shut down the first component, asset price is more volatile under incomplete markets than it is under complete markets in the long run. In general, the same comparison holds or not depending on the long-run correlation between the first and the second volatility components under incomplete markets.

1.3.4 Relationship to recursive equilibria

When we do not have financial assets and there is only one real asset, then Markov equilibria are recursive equilibria. This is also true when initially agents hold the same fraction of each assets. However, in general, Markov equilibria are not recursive equilibria. But in Kubler and Schmedders (2003), subsection 4.4 shows that we can construct recursive equilibria from Markov equilibria if we can extract a continuous mapping from the policy correspondence.

As an important special case discussed after Definition 1.1 of incomplete markets equilibrium, the economy in Krusell and Smith (1998) corresponds to the economy here with one asset in perfectly elastic supply and without financial markets. The existence of a Markov equilibrium implies the existence of recursive equilibrium. Indeed, given that there is no financial markets and only one asset. The "normalized financial wealth" ω_t^h becomes $\frac{k^h_{a,t-1}}{K_{a,t-1}}$ the fractions of capital asset holdings. Together with the total quantity of capital, $K_{a,t}$, the state variables $(s_t, \omega_t, K_{a,t-1})$ is equivalent to $\left(s_t, \left(k^h_{a,t-1}\right)\right)$, the aggregate state and capital holdings of each agent in the definition of recursive equilibrium in page 874 of Krusell and Smith (1998). In a recent paper, Miao (2006) shows the existence of recursive equilibrium however he has to include future expected discounted utilities of agents in the state-space. In addition, he wrote in page 291, that the question whether a recursive equilibrium in Krusell and Smith (1998) exists remains an open question. The existence proof here provides a positive answer to that question. However, in this paper I only consider a finite number of types.

1.4 Numerical Method

In this section, I present an algorithm to compute Markov equilibria defined in the last section. This algorithm can also be used to compute complete markets equilibria.

1.4.1 General Algorithm

Suppose we need to find a function ρ defined over $S \times E$ on to a compact set $A \subset \mathbf{R}^N$, where S has finite element and E is convex and compact, and ρ satisfies the functional equation

$$\rho = f + T\rho$$

We then first discretize E by $\{e_1, e_2, \ldots, e_K\}$, and $\rho^n = (\rho_1^n, \rho_2^n, \ldots, \rho_S^n)$, each component is defined over $\{e_1, e_2, \ldots, e_K\}$. Let $\widetilde{\rho}_s^n$ be the extrapolation of ρ_s^n over E. Then

$$\rho_s^{n+1}(e_k) = \arg\min_{r \in A} \|r - \{f(e_k) + T\widetilde{\rho}_s^n(e_k)\}\| \tag{1.27}$$

If we have a fixed point $\rho^{n+1} = \rho^n$ and $f(e_k) + T\widetilde{\rho}_s^n(e_k) \in A$ then

$$\rho_s^n(e_k) = f(e_k) + T\widetilde{\rho}_s^n(e_k)$$

We present an implementation of this general algorithm to compute Markov Equilibria. We can also use the algorithm to compute competitive equilibria with complete markets. The state space in this case is the current consumption of each agent and the total supply of assets with elastic supply. The details are presented in the Appendix.

1.4.2 Algorithm to Compute Markov Equilibria

The construction of Markov equilibria in the last section also suggests an algorithm to compute them. The following algorithm is based on Kubler and Schmedders (2003). There are two differences of the algorithm here compared to the original algorithm. The more important difference is that the future wealth distributions are included into the current mapping instead of solving for them using a sub-fixed-point loops. This innovation reduces significantly the computing time, given solving for a fixed-point is time consuming in MATLAB. Relatedly, in section 1.5, as we seek to find the set of actively traded financial assets, we can include future asset prices as one of the components of the function ρ.The minor difference between the algorithm presented in Kubler and Schmedders (2003) and the one here is, for each iteration, I solve for a constrained optimization problem presented in (1.27) instead of solving for a zero point as in the original algorithm. This difference avoids the non-existence of zero points at the beginning of the loop when the initial guess ρ^0 is far away from the true solution.

We look for the following correspondence

$$\begin{aligned} \rho : \mathcal{S} \times \Omega \times E &\longrightarrow \widehat{\mathcal{V}} \times \Omega^S \times \mathcal{L} \\ (s, \omega, K_a) &\longmapsto \left(\widehat{v}, \omega_s^+, \mu, \eta\right) \end{aligned} \tag{1.28}$$

$\widehat{v} \in \widehat{\mathcal{V}}$ is the set of endogenous variables excluding the wealth distribution and total capital, as defined in (1.18). $(\omega_s^+)_{s\in\mathcal{S}}$ are the wealth distributions in the S future states and μ, η are Lagrange multipliers as defined in subsection 1.3.2.

From a given continuous initial mapping $\rho^0 = \left(\rho_1^0, \rho_2^0, \ldots, \rho_S^0\right)$, we construct the sequence of mappings $\{\rho^n = (\rho_1^n, \rho_2^n, \ldots, \rho_S^n)\}_{n=0}^{\infty}$ by induction. Suppose we have obtained ρ^n, for each state variable (s, ω, K_a), we look for

$$\rho_s^{n+1}(\omega, K_a) = \left(\widehat{v}_{n+1}, \omega_{s,n+1}^+, \mu_{n+1}, \eta_{n+1}\right) \tag{1.29}$$

that solves the forward equations presented in the Appendix.

We construct the sequence $\{\rho^n\}_{n=0}^{\infty}$ on a finite discretization of $\mathcal{S} \times \Omega \times E$. So from ρ^n to ρ^{n+1}, we will have to extrapolate the values of ρ^n to outside the grid using extrapolation methods in MATLAB. Fixing a precision δ, the algorithm stops when $\left\|\rho^{n+1} - \rho^n\right\| < \delta$.

1.5 Asset price volatility and leverage

This section uses the algorithm just described to compute incomplete and complete markets equilibria and study asset price and leverage. In order to focus on asset price, I only keep one real asset in fixed supply. Each financial asset corresponds to a leverage level. Suppose selling financial asset j requires k_j units of the fixed-supply asset as collateral and price of j is p_j. This operation is equivalent to buy k_j units of the real asset, at price q with p_j borrowed. Therefore, leverage as defined in Geanakoplos (2009) by the ratio between total value of the real asset over the down payment paid by the buyer:

$$L_j = \frac{k_j q}{k_j q - p_j} = \frac{q}{q - \frac{p_j}{k_j}}.$$

If in equilibrium, only one financial asset j is traded, the leverage level corresponding to the financial asset is called the leverage level of the economy.

To make the analysis as well as numerical procedure simple, I allow for only one asset and two types of agents: optimists and pessimists, each in measure 1 of identical agents. The general framework in Section 1.2 allows for wide range of financial assets with different promises and collateral requirements. However, given that the total quantity of collateral is exogenously bounded, in equilibrium, only certain financial assets are actively traded. I choose a specific setting based on Geanakoplos (2009), in which I can find exactly which assets are traded. The setting requires that promises are state-incontingent and in each aggregate state there are only two possible future aggregate states. The assets that are traded are the assets that allow maximum borrowing while keeping the payoff to lenders riskless. Endogenous financial assets interestingly generates the most volatility in the wealth distribution as agents borrow to the maximum and lose most of their wealth as they lose their bets but their wealth increases largely when they win. This volatility in the wealth distribution in turn feeds in to asset price volatility.

Endogenous set of traded assets also implies endogenous leverage which has been of the object of interest during the current financial crisis. In order to match the observed pattern of leverage, i.e., high in good states and low in bad states, I introduce the possibility for changing types of uncertainty from one aggregate state to others. This feature is introduced in Subsection 1.5.2.

To answer questions related to collateral requirements, in Subsection 1.5.2, I allow regulators to control the sets of financial assets that can be traded. Given the restricted set, the endogenous active assets can still be determined. One special case is the extreme regulation that shuts down financial markets. There are surprising consequences of these regulations on welfare of agents, on the equilibrium wealth distribution and on asset prices.

1.5.1 The model

There are two aggregate states $s = G$ or B and one single asset of which the dividend depends on the state s

$$d(G) > d(B).$$

The state follows a I.I.D process, with the probability of high dividends π unknown to agents in this economy. However the transition matrix is unknown to the agents in this economy. The supply of the asset is exogenous and normalized to 1. Let $q(s^t)$ denote the ex-dividend price of the asset at each history $s^t = (s_0, s_1, \ldots, s_t)$.

Financial Markets: At each history s^t, we consider the set of J of financial assets which promise state-independent pay-offs next period. I normalize these promises to $b_j = 1$. Asset j also requires k_j units of the real asset as collateral. The effective pay-off is therefore

$$f_{j,t+1}\left(s^{t+1}\right) = \min\left\{1, k_j\left(q\left(s^{t+1}\right) + d\left(s^{t+1}\right)\right)\right\}$$

Fostel and Geanakoplos (2008), Geanakoplos (2009) and recently Simsek (2009b) argue that if we allow for the set J to be dense enough that contains the complete set of collateral requirements, then in equilibrium the only financial asset is traded is the one with the minimum collateral level $k^*(s^t)$ to avoid default:

$$k^*\left(s^t\right) = \max_{s^{t+1}|s^t}\left\{\frac{1}{q\left(s^{t+1}\right) + d\left(s^{t+1}\right)}\right\}.$$

This statement applies for my general set up under the condition that in each history node, there are only two future aggregate states. The following proposition makes it clear.

Proposition 1.8 *Suppose in each event node s^t, there are only two possible future aggregate states s_{t+1}. Given the set J, there is no more than one actively traded asset with collateral requirement less than or equal to $k^*(s^t)$. There is also no more than one actively traded asset with collateral requirement greater than or equal to $k^*(s^t)$.*

Proof. The proof of the first part requires an analysis of portfolio choice of the sellers of these securities and is detailed in the Appendix. For the second part, notice that all securities with collateral greater than or equal to $k^*(s^t)$ is riskless to the buyers, i.e. deliver 1 units of final good regardless of the future states. Hence, these securities are sold at the same price. In addition, the sellers of the securities prefer selling securities with the least level collateral requirement to save their collateral. Therefore in equilibrium, only one security, with the collateral requirement the smallest above $k^*(s^t)$, is traded. ■

Imagine that the set J includes all collateral requirements $k_j \in \mathbf{R}^+, k_j > 0.$[12]. Proposition 1.8 says that only securities with collateral requirement exactly equals to $k^*(s^t)$ are traded in equilibrium. Therefore the only actively traded financial asset is riskless to its buyers. Let $p(s^t)$ denote the price of this financial asset. The endogenous interest rate is therefore

$$r\left(s^t\right) = \frac{1}{p\left(s^t\right)} - 1.$$

Consumers: There are two types agents in this economy, optimists, O, and pessimists, P, each in measure one of identical agents. They have the same utility function

$$\sum_{t=0}^{\infty} \beta^t U\left(c_t\right), \tag{1.30}$$

and endowment e in each period. But they differ in their belief about the transition matrix of the aggregate state s. Suppose agent $h \in \{O, P\}$ estimates the probability of high dividends as $\pi_G^h = 1 - \pi_B^h$.We suppose $\pi_G^O > \pi_G^P$, i.e. optimists always think that good states are more likely than the pessimists think they are.

So each agent maximizes the inter-temporal utility (1.30) given their belief of the evolution

[12] To apply the existence theorem 1.2 I need J to be finite. But we can think of J as a fine enough grid.

of the aggregate state, they are subject to

$$c_t + q_t\theta_t + p_t\phi_t \leq e_t + (q_t + d_t)\,\theta_{t-1} + f_t\phi_{t-1} \tag{1.31}$$

no short-sale

$$\theta_t \geq 0 \tag{1.32}$$

and collateral constraint

$$\theta_t + \phi_t k^* \geq 0, \tag{1.33}$$

for each $h \in \{O, P\}$. At time t, each agent choose to buy θ_t units of real asset at price q_t and ϕ_t units of financial asset at price p_t. Moreover, Proposition 1.8 allows us to focus on only one level of collateral requirement k^*.

Given prices q and p, this program yields solution $c_t^h\left(s^t\right), \theta_t^h\left(s^t\right), \phi_t^h\left(s^t\right)$. In equilibrium prices $\left\{q_t\left(s^t\right)\right\}$ and $\left\{p_t\left(s^t\right)\right\}$ are such that asset and financial markets clear, i.e.,

$$\begin{aligned}
\theta_t^O + \theta_t^P &= 1 \\
\phi_t^O + \phi_t^P &= 0
\end{aligned}$$

for each history s^t.

I define the financial wealth of each agent at the beginning of each period as

$$\omega_t^h = \frac{(q_t + d_t)\,\theta_{t-1}^h + f_t\phi_{t-1}^h}{q_t + d_t}.$$

Due to the collateral constraint, in equilibrium, ω_t^h must always be positive and

$$\omega_t^O + \omega_t^P = 1.$$

The pay-off relevant state space

$$\left\{\left(\omega_t^O, s_t\right) : \omega_t^O \in [0,1] \text{ and } s_t \in \{G, B\}\right\}$$

is compact. I look for Markov equilibria in which prices and allocations depend solely on that state. In Sections 1.3 and 1.4, I show the existence of such a Markov equilibrium and develop an algorithm that computes the equilibrium.

1.5.2 Numerical Results

Numerical example

$$\begin{aligned} \beta &= 0.5 \\ d(G) &= 1 > d(B) = 0.2 \\ U(c) &= \log(c) \end{aligned}$$

And the beliefs are $\pi^O = 0.9 > \pi^P = 0.5$. I will vary the endowments of the optimists and the pessimists, e^O and e^P respectively, in different numerical exercises.

Asset Prices

Given that the main demand for the asset comes from the optimists, when their endowment is small, their demand is more elastic with respect to "normalized financial wealth". To investigate that relationship, I fix the endowment of the pessimists at

$$e^P = \begin{bmatrix} 10 & 10.8 \end{bmatrix}$$

and vary the endowment of the optimist

$$e^O = \begin{bmatrix} e & e \end{bmatrix}$$

I keep the aggregate endowment constant by choosing the pessimists' endowment to be state dependent.

Incomplete Markets Equilibrium: I rewrite the budget constraint of the optimists (1.31) using the normalized financial wealth, ω_t^O,

$$c_t + q_t\theta_t + p_t\phi_t \leq e^O + (q_t + d_t)\,\omega^O.$$

Therefore, their total wealth $e^O + (q_t + d_t)\,\omega^O$ affects their demand for the asset. If non-financial endowment e^O of the optimists is small relative to price of the asset, their demand for asset is more elastic with respect to their financial wealth $(q_t + d_t)\,\omega^O$. I compute Markov equilibria for two values of the optimists' wealth $e = 1$ and 10. Figure 1-1 plots price of the asset as function of the optimists' normalized financial wealth ω^O. The dashed line corresponds to the high "non-financial" wealth of the optimists: $e^O = 10$; the solid line corresponds to low the low "non-financial" wealth of the optimists: $e^O = 1$. The figure shows that the elasticity of price with respect to ω^O increases as we reduce the non-financial wealth of the optimists from $e^O = 10$ to $e^O = 1$.

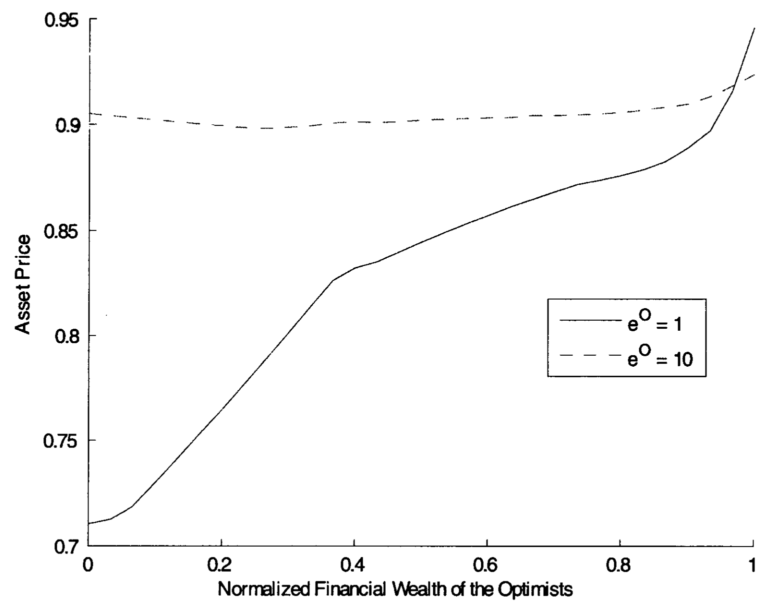


Figure 1-1: Asset Price Under Incomplete Markets

There are two main factors that affect asset prices. The first factor is the aggregate state. Aggregate states affect prices through endowments of agents. Because their endowments determine their consumption, and thus determine the marginal utility at which they evaluate value of the asset. Aggregate states also affect asset prices through the evolution of future aggregate states, if these states are persistent. The second factor that I emphasize here is the financial wealth distribution, as it affects the budget constraints of different agents. The financial wealth distribution may vary significantly, especially when some agents have limited non-financial wealth. Figure 1-2 shows the evolution of the "normalized financial wealth" of the optimists, ω^O, when their non-financial endowment is relatively small with respect to the price of the asset: $e^O = 1$. The left panel corresponds to the current state $s = G$, and the right panel corresponds to the current state $s = B$. The solid lines represent next period normalized wealth of the optimists as function of the current normalized wealth, if good shock realizes next period. The dashed lines represent the same function when bad shocks realizes next period. I also plot the 45 degree lines for comparison. This figure shows that, in general, good shocks tend to increase and bad shocks tend to decrease the normalized wealth of the optimists.

When ω^O is close to zero, the optimists are highly leveraged to buy the asset. If a bad shocks hits in the next period, they have to sell off their asset holdings to pay off their debts. Their next period "financial wealth" plummets and contributes to the fall in asset price.

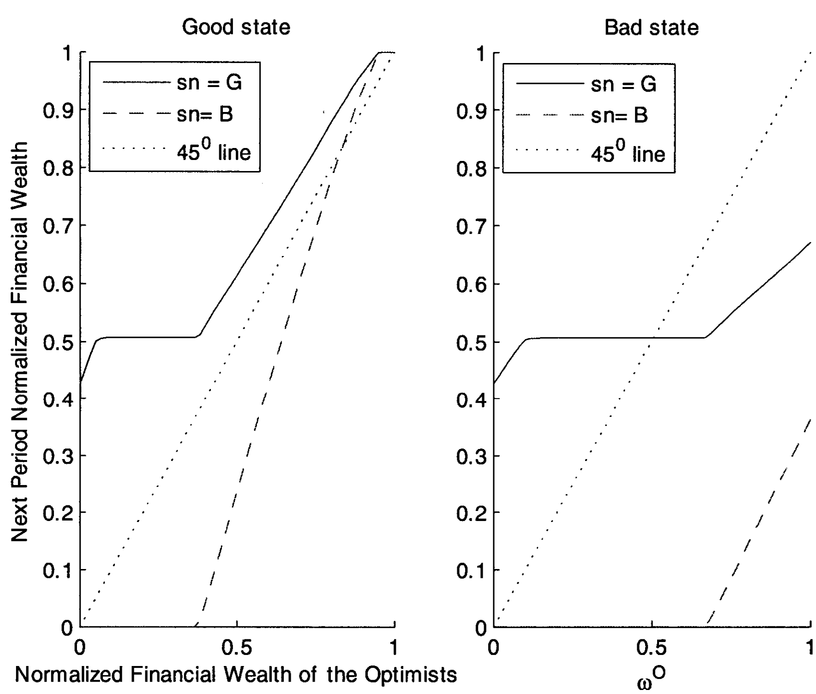


Figure 1-2: Dynamics of Wealth Distribution under Incomplete Markets

Complete Markets Equilibrium: In a complete markets equilibrium, as shown in the Appendix, remark 1.1, the state variable is the consumption of the optimists. However, there is a one-to-one mapping from this state variable to a more meaningful state variable which is the relative wealth of the optimists. Given that markets are complete, wealth of each consumer is defined as the current value of her current and future stream consumption

$$V_t^h = \sum_{r=0}^{\infty} p_t\left(s^{t+r}\right) c_{t+r}^h\left(s^{t+r}\right),$$

where $p_t\left(s^{t+r}\right)$ denotes the time t Arrow-Debreu price for a claim to a unit of consumption at date $t+r$ and sate s^{t+r}. Let

$$\widehat{\omega}_t^O = \frac{V_t^O}{V_t^O + V_t^P}$$

denote the relative wealth of the optimists with respect to the total wealth. Similar to the incomplete markets equilibrium, this variable determines asset price and constantly changes as aggregate shocks hit the economy. Figure 1-3 depicts the relationship between asset price and relative wealth. This figure is the counterpart of Figure 1-1 for complete markets.

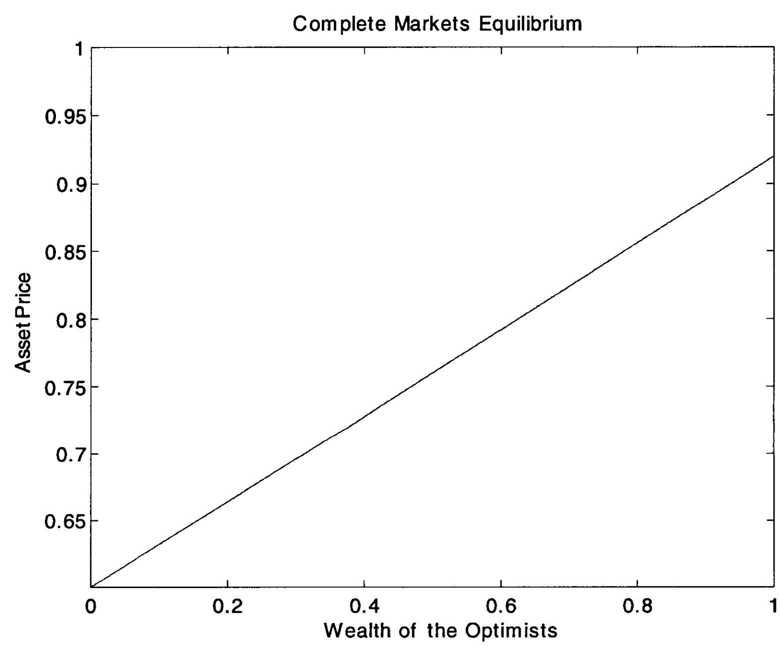


Figure 1-3: Asset Price under Complete Markets

Notice that at two extreme $\widehat{\omega}_t^O = 0$ (on the left of Figure 1-3) or 1 (on the right of Figure 1-

3), we go back to the representative agent economy in which there are either only the optimists or the pessimists. The representative consumer consumes all the aggregate endowment in each period. Asset price is determined by her marginal utility.

$$q(s)\,U'(e(s)) = \sum_{s'} \beta P(s,s')\,U'(e(s'))\,(q(s') + d(s'))$$

so, we can rewrite these equations as

$$\begin{bmatrix} q(G) \\ q(B) \end{bmatrix} = X^{-1}Y \begin{bmatrix} d(G) \\ d(B) \end{bmatrix} \tag{1.34}$$

where X and Y are matrices with elements that are functions of marginal utilities and transition probabilities. This formula also suggests that volatility of price of an asset is proportional to volatility of its dividends if $X^{-1}Y$ is state-independent.

Consider a special case when the aggregate endowments are constant across states and shocks are I.I.D, we have

$$q(G) = q(B) = \frac{\beta}{1-\beta}\,(P(G)\,d(G) + P(B)\,d(B)),$$

i.e., asset price is constant in the long run. When $\widehat{\omega}_t^O = 0$, asset price is the discounted value of average dividends evaluated at the pessimists' belief

$$q^P = \frac{\beta}{1-\beta}\left(\pi^P d(G) + \left(1-\pi^P\right) d(B)\right),$$

which is smaller than when $\widehat{\omega}_t^O = 1$, where asset price is the discounted value of average dividends evaluated at the optimists' belief

$$q^O = \frac{\beta}{1-\beta}\left(\pi^O d(G) + \left(1-\pi^O\right) d(B)\right) > q^P.$$

In the short-run, however, the wealth distribution constantly changes as shocks hit the economy. Figure 1-4 depicts the evolution of the relative wealth distribution that determines the evolution of asset price under complete markets. This figure is the counterpart of Figure

1-2 under complete markets. Given that the aggregate endowment is constant, the transition of the wealth distribution does not depend on current aggregate state, unlike under incomplete markets. The optimists buy more Arrow-Debreu assets that deliver in the good future states and buy less Arrow-Debreu assets that delivers in bad future states. Therefore, when a good shock hits, the relative wealth of the optimists increases (solid line) and vice versa when a bad shock hits (dashed line).

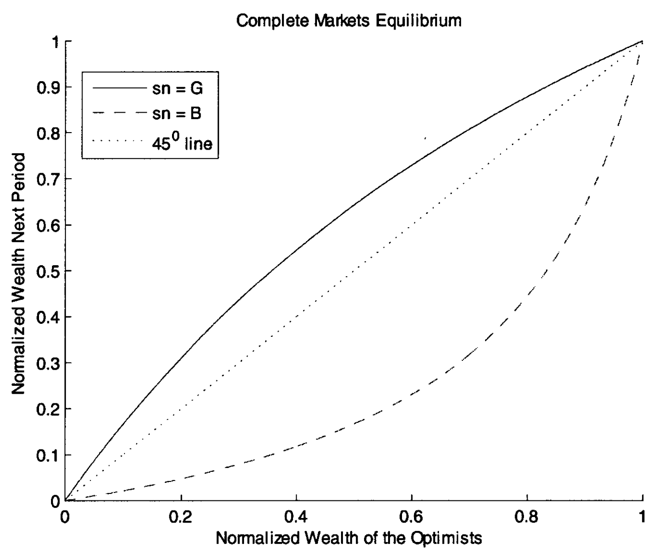


Figure 1-4: Dynamics of Wealth Distribution under Complete Markets

Asset Price Volatility

We compare asset prices, and asset price volatility of the Markov equilibrium with the complete markets benchmark. Consider first what happens with complete markets: Asset price does depend on the wealth distribution $\widehat{\omega}_t^O$ and its evolution. However, in the long run $\widehat{\omega}_t^O$ converges to 0 or to 1 depending on whether the pessimists or the optimists hold the correct belief. Therefore, in the long run, asset price only depends on the aggregate states.

In the case of Markov equilibrium, however, consumers with incorrect beliefs are protected by the no default penalty assumption. They always survive in equilibrium, and constantly speculate on asset prices. First, asset prices are not only state dependent but also depend on the wealth of the optimists. Second, their wealth undergoes large swings as they lose or win their bet after each period. The two components increase the volatility of asset price compared to the complete markets case.

I measure price volatility as one-period ahead standard deviation of price. This measure is the discrete time equivalence of the continuous instant volatility, see for example Xiong and Yan (2009). The following figure shows the evolution of asset price volatility under the assumption that the pessimists hold the correct belief. The figure shows that, in short run, asset price is more or less volatile in the complete markets equilibrium than in the incomplete markets economy depending on the relative non-financial wealth of agents. However, in the long run, as the optimists are driven out in the complete markets equilibrium, asset price is history independent and price volatility is proportional to dividend volatility. This property does not hold in the incomplete markets equilibrium, the overly optimistic agents constantly speculate on asset price using the same asset as collateral. Asset price becomes more volatile than in the complete markets equilibrium, given the wealth of the optimists constantly change as they win or loose their bets.

Strikingly, the smaller the non-financial wealth of the optimist is, the higher the short-run asset price volatility in the incomplete markets equilibrium but the lower the short-run asset price volatility in the complete markets equilibrium. This is because, incomplete markets, it takes less time to drive out the optimists if they have lower non-financial wealth. As we increase the non-financial wealth of the optimists, we increase the short-run volatility of asset price with complete markets and decrease the short-run volatility of asset price with incomplete markets. Figure 1-5 plots the average asset price volatility over time for complete markets (dashed lines) and for incomplete markets (solid lines) equilibria, with different levels of "non-financial wealth" of the optimists (low and high). This figure shows that, above some certain level of non-financial wealth of the optimists, in the short-run asset price is more volatile under complete markets. But in the long run, the reverse inequality holds (right panel).

The financial crisis 2007-2008

Geanakoplos (2009) argues that the introduction of CDS triggered the financial crisis 2007-2008. The reason is that the introduction of CDS moves the markets close to complete. CDS

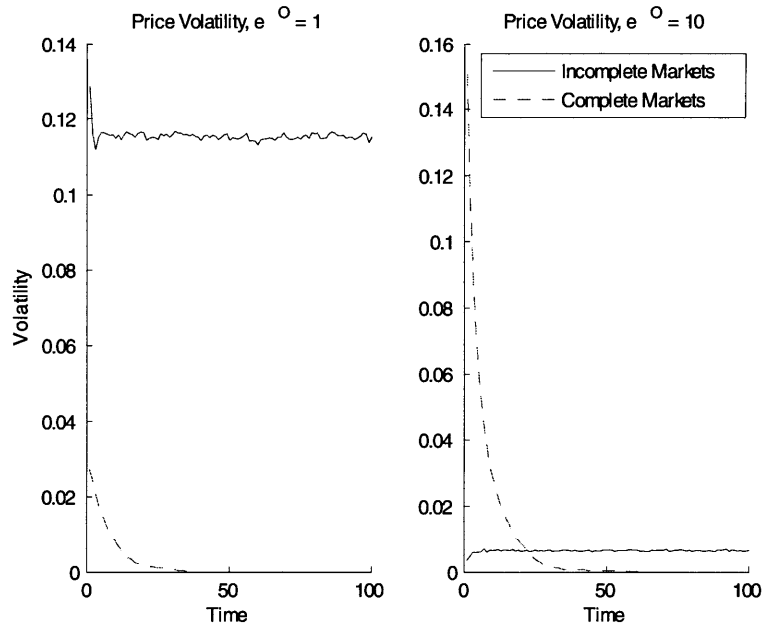


Figure 1-5: Asset Price Volatility Over Time

allow pessimists to leverage their pessimism about the assets. I do the same exercise here by simulating a financial markets equilibrium in its stationary state from time $t = 0$ until time $t = 50$.[13] At $t = 51$ markets suddenly become complete. In Figure 1-6 left panel plots asset price level and right panel plots asset price volatility over time. The simulation shows that, asset price decreases, but asset price volatility increases in the short run after the introduction of CDS. The reason for the fall in asset price is that the "pessimists can leverage their view". The reason for increasing in asset price volatility is the movement in the wealth distribution toward the long-run wealth distribution, which concentrates on pessimists. Asset price decreases because the pessimists can leverage their pessimism with complete markets.

Dynamic leverage cycles

Even though the example in Subsection 1.5.2 generates high asset price volatility, leverage is not consistent with what we observe in financial markets: high leverage in good times and low

[13]In order to generate high short-run asset price volatility, I choose a high level of the optimists' endowment $e^O = 10$.

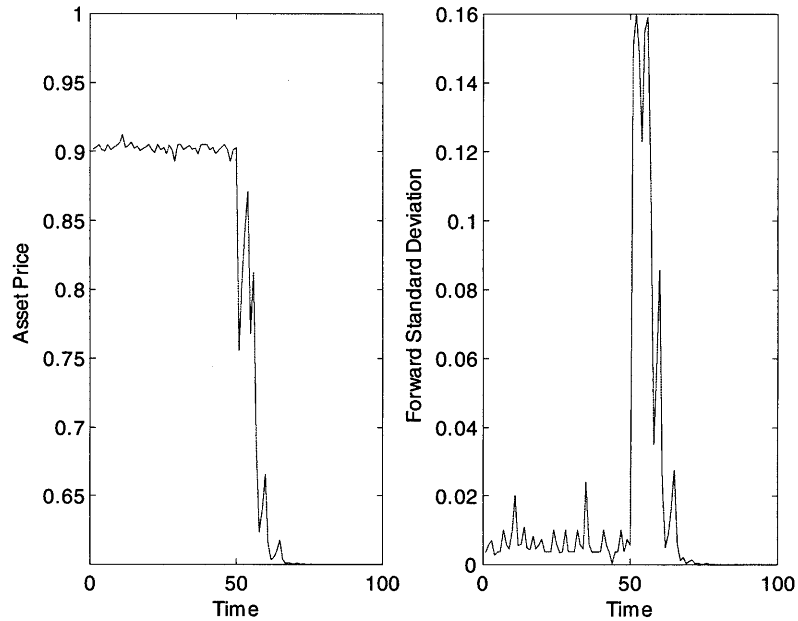


Figure 1-6: Financial Crisis 2007-2008

leverage in bad times, as documented in Geanakoplos (2009).

In order to generate the procyclicality of leverage, I use the insight from Geanakoplos (2009) regarding aggregate uncertainty: bad news must generate more uncertainty and more disagreement in order to reduce equilibrium leverage significantly. To formalize this type of news, I assume that after a series of good shocks, the first bad shock does not immediately reduce dividends. After this bad shock, however, dividends plunge if a second bad shock hits the economy. Therefore the first bad shock increases uncertainty regarding dividends. In a dynamic setting, the formulation translates to a dividend process that depends not only on current aggregate shock but also on last period aggregate shock. Therefore we need to use four aggregate states, instead of the two aggregate states in the last subsections:

$$s \in \{GG, GB, BG, BB\}.$$

Figure 1-7, left panel, shows that the initial bad shocks following a series of good shocks does not reduce dividends. However, the fall in dividends increases, falling to 0.2, if a second bad

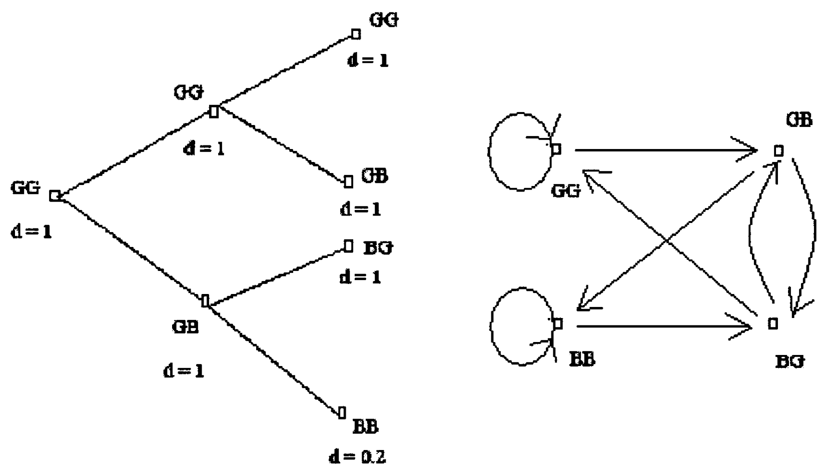


Figure 1-7: Evolution of the Aggregate States

shock hits the economy, i.e., the first bad shock increases uncertainty in dividends. The right panel of the figure shows the evolution through time of the aggregate states using Markov chain representation.

This aggregate uncertainty structure generates high leverage at good states GG and BG and low leverage in bad states GB and BB. Figure 1-8 shows this pattern of leverage. The dashed line represents leverage level in good states $s = GG$ or BG as a function of the normalized wealth distribution. The two solids lines represent leverage level in bad states $s = GB$ or BB. We see that leverage decreases dramatically from good states to bad states. However, in contrast to the static version in Geanakoplos (2009), changes in the wealth distribution do not amplify the decline in leverage from good states to bad states as leverage is insensitive to the wealth distribution in bad states.

Moreover, this version of dynamic leverage cycles generates a pattern of leverage build-up in good times. Good shocks increase leverage as they increase the wealth of the optimists relative to the wealth of the pessimists and leverage is increasing the wealth of the optimists. Figure 1-9 shows the evolution of the wealth distribution and leverage over time. The economy starts at good state and $\omega^O = 0$. It experiences 9 consecutive good shocks from $t = 1$ to 9 and two bad shocks at $t = 10, 11$ then another 9 good shocks from $t = 12$ to 20. This figure shows that,

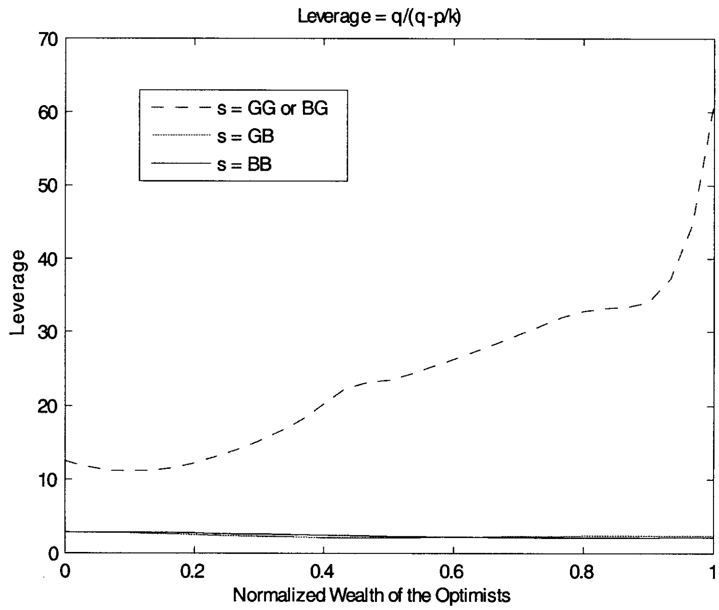


Figure 1-8: Leverage Cycles

in good states, both the wealth of the optimists and leverage increase. However their wealth and leverage plunge when bad shocks hit the economy.

Regulating Leverage

Subsection 1.5.2 shows that, in a incomplete markets equilibrium, when the non-financial wealth of the optimists is small relative to asset prices, variations in their wealth play an important role in driving up asset price volatility. It is then tempting to conclude that by restricting leverage, we can reduce the variation of wealth of the optimists, therefore reduce asset price volatility. However, this simple intuition is not always true by two reasons. First, restricting leverage limits the demand for asset of the optimists when their "financial wealth" is small, therefore drives down asset price. In contrast, when their "financial wealth" is large, restricting leverage does not affect the demand, thus does not affect asset price. The two channels create a potential for higher asset price volatility. Second, restricting leverage does reduce asset price in the short run when the optimists are poor, however in the long run they can accumulate the asset and become wealthier. High leverage requirements prevent them from falling back to the

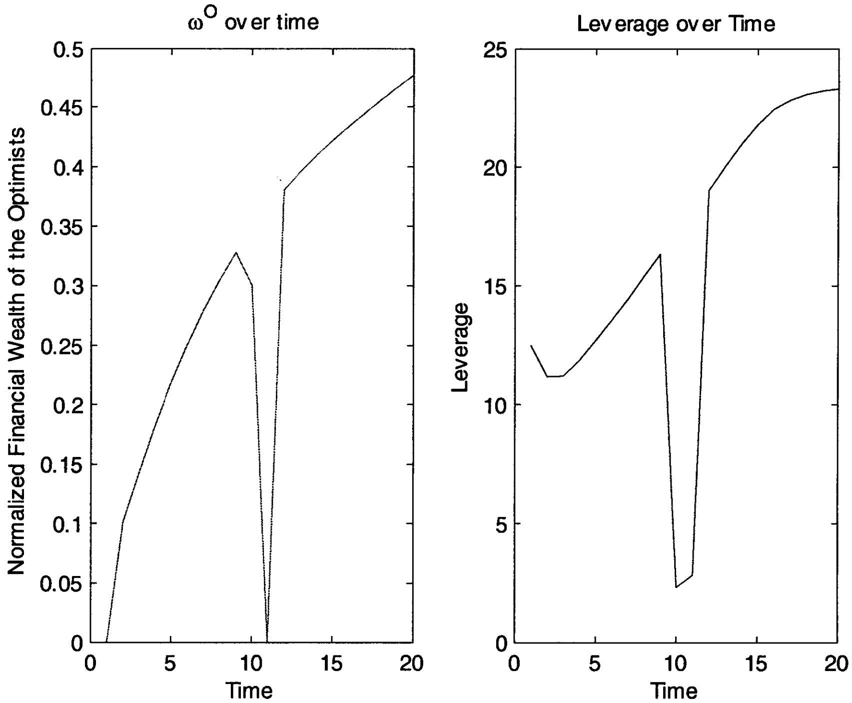


Figure 1-9: Leverage Cycles

low wealth region. So in the long run, restricting leverage drives up asset price volatility due to the first reason and high long run wealth of the optimists.

To show this statement, I go to the extreme case, when leverage is strictly forbidden, i.e., there are no financial assets. The Figure 1-10 plots the volatility of asset price as function of the wealth of the optimists in two cases, with financial markets and without financial markets[14]. We can see that, with financial markets, asset volatility is higher when the optimists are poor and lower when they are rich. The reverse holds without financial markets. The numerical solution also shows that, without financial markets, the optimists always accumulate assets to move up to the high wealth region. This dynamics makes asset price more volatile without financial markets then it is with financial markets.

Figure 1-11 shows the Monte-Carlo simulation for an economy starting in good state and $\omega^O = 0$. The figure plots the evolution of the average of the normalized financial wealth of the

[14] Without financial market, "financial wealth" is asset holding itself.

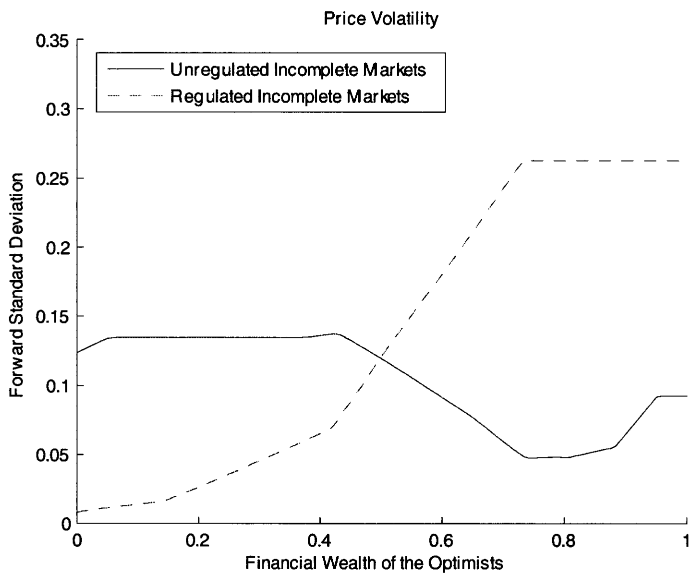


Figure 1-10: Asset Price Volatility in Unregulated and Regulated Economies

optimists, left panel, and asset price volatility, right panel, over time (the solid lines represent the unregulated economy and the dashed lines represent the regulated economy). As discussed above, the wealth of the optimists remains low in average in the unregulated economy but increases to a permanently high level under regulation. Thus, initially asset price volatility is higher in the unregulated economy than in the regulated economy. The reverse inequality holds, however, as over time, the wealth of the optimists increase more in the regulated economy than in the unregulated economy.

I conclude this part with two additional remarks. First, intermediate regulations can be computed using Proposition 1.8. If the regulator requires collateral $k \geq k_r$. Then the proposition shows that in equilibrium, only the leverage level $\max(k_t^*, k_r)$ prevails. Numerical solution for intermediate regulations, confirms the conclusion in the paragraphs above. Second, regulation not only fails to reduce asset price volatility, it also reduces welfare of both types of agents as it reduces trading possibilities.

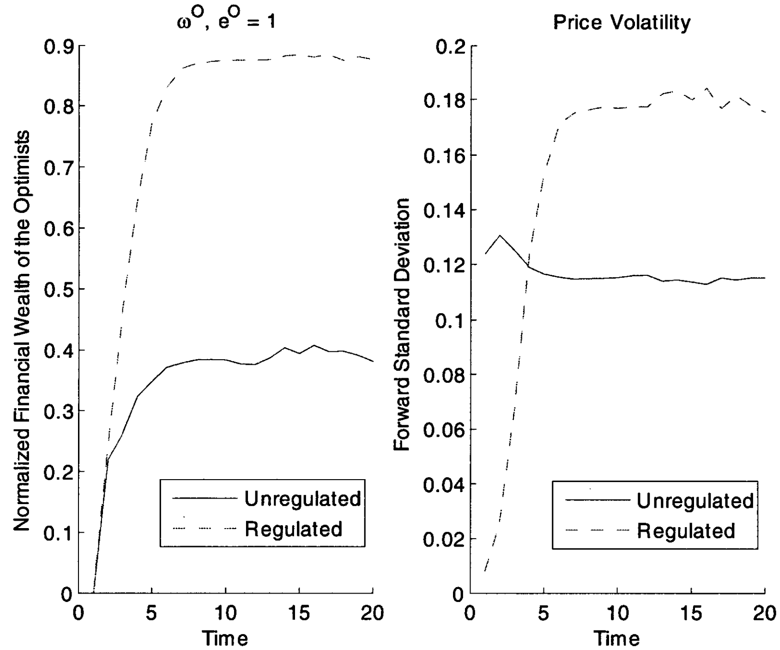


Figure 1-11: Wealth Distribution and Asset Price Volatility over Time

1.6 Conclusion

In this paper I develop a dynamic general equilibrium model to examine the effects of belief heterogeneity on the survival of agents and on asset price and investment volatility under different financial markets structures. I show that, when financial markets are endogenously incomplete, agents with incorrect beliefs survive in the long run. The survival of these agents leads to higher asset price and investment volatility. This result contrasts with the frictionless complete markets case, in which agents holding incorrect beliefs are eventually driven out and as a result, asset prices and investment exhibit lower volatility.

In addition, I show the existence of stationary Markov equilibria in this framework with incomplete financial markets and with general production and capital accumulation technology. I also develop an algorithm for computing the equilibria. As a result, the framework can be readily used to investigate questions about the interaction between financial markets and the macroeconomy. For instance, it would be interesting in future work to apply these methods in calibration exercises using more rigorous quantitative asset pricing techniques, such as in

Alvarez and Jermann (2001). This could be done by allowing for uncertainty in the growth rate of dividends rather than uncertainty in the levels, as modeled in this paper, in order to match the rate of return on stock markets and the growth rate of aggregate consumption. Such a model would provide a set of moment conditions that could be used to estimate relevant parameters using GMM as in Chien and Lustig (2009). A challenge in such work, however, is that finding the Markov equilibria is computationally demanding.

A second avenue for further research is to examine more normative questions in the framework developed in this paper. My results suggest, for example, that financial regulation aimed at reducing asset price and real investment volatility should be state-depedent, as conjectured by Geanakoplos (2009). It would also be interesting to consider the effects of other intervention policies, such as bail-out or monetary policies.

1.7 Appendix

To prove the existence of equilibrium in finite horizon, I allow utility to be dependent of labor decision. So per period utility of agent h is $U_h(c, L_h - l) : (\mathbf{R}^+)^2 \longrightarrow \mathbf{R}$ over consumption and leisure. I replace Assumption 1.5 by the following Assumption

Assumption 1.5b: $U_h(c, l)$ is strictly increasing in c, non-decreasing in l and concave in (c, l).

Definition 1.4 *An allocation*

$$\begin{pmatrix} c_t^h(s^t), k_{a,t}^h(s^t), l_{a,t}^h(s^t), \phi_{j,t}^h(s^t) \\ K_{a,t}^n(s^t), K_{a,t}^o(s^t), \psi_{a,t}(s^t) \\ K_{a,t}^f(s^t), L_{a,t}(s^t), y_{a,t}(s^t) \end{pmatrix}$$

together with the no default penalty defined in (1.3), *is attainable if consumptions, real asset holdings, labor decision from the consumers, new and old real assets decision from the real asset producers and capital and labor decisions of final good producers are positive. The resources constrained are satisfied*

$$L_a^h \geq l_{a,t}^h(s^t)$$
$$\psi_{a,t}(s^t) \geq \Psi_a(K_{a,t}^n(s^t), K_{a,t}^o(s^t))$$
$$F\left(K_{a,t}^f(s^t), L_{a,t}(s^t), s\right) \geq y_{a,t}(s^t)$$

($\psi_{a,T}(s^T) \geq \Psi_a\left(0, K_{a,T}^o(s^T)\right)$ *given* $K_{a,T+1}(s^T) = 0$) *and excess demands are negative: First, excess demands on the good markets are negative:*

$$\sum_{h=1}^{H} c_t^h + \sum_{a\in\mathcal{A}_0} \psi_{a,t} - \sum_{h=1}^{H} e_t^h - \sum_{a\in\mathcal{A}_1} d_{a,t} \sum_{h} k_{a,t-1}^h - \sum_{a\in\mathcal{A}_0} y_{a,t} - \sum_{\substack{j=1 \\ f_{j,t}=b_{j,t}}}^{J} \left(\sum_{h=1}^{H} \phi_{j,t-1}^h\right) b_{j,t} \leq 0$$

Second, for $a \in \mathcal{A}_0$

$$\sum_{h=1}^{H} k_{a,t}^h - K_{a,t}^n \leq 0$$

$$K^o_{a,t} - \sum_{h=1}^{H} k^h_{a,t-1} - \sum_{\substack{j=1 \\ f_{j,t}<b_{j,t}}}^{J} \left(\sum_{h=1}^{H} \phi^h_{j,t-1} \right) k^a_{j,t-1} \leq 0$$

$$K^f_{a,t} - \sum_{h=1}^{H} k^h_{a,t-1} - \sum_{\substack{j=1 \\ f_{j,t}<b_{j,t}}}^{J} \left(\sum_{h=1}^{H} \phi^h_{j,t-1} \right) k^a_{j,t-1} \leq 0$$

$$L_{a,t} - \sum_{h=1}^{H} l^h_{a,t} \leq 0$$

$$\sum_{h=1}^{H} \phi^h_{j,t} \leq 0. \tag{1.35}$$

for $a \in \mathcal{A}_1$

$$\sum_{h=1}^{H} k^h_{a,0} - K_{a,-1} \leq 0.$$

in each time-state t, s^t with $0 < t < T$. For the initial period there is no explicit initial debt and the aggregate supply of asset a is $K_{a,-1}$ so

$$\sum_{h=1}^{H} c^h_0 + \sum_{a \in \mathcal{A}_0} \psi_{a,0} - \sum_{h=1}^{H} e^h_0 - \sum_{a \in \mathcal{A}_1} d_{a,0} K_{a,-1} - \sum_{a \in \mathcal{A}_0} y_{a,0} \leq 0$$

For $a \in \mathcal{A}_1$

$$\sum_{h=1}^{H} k^h_{a,0} - K_{a,-1} \leq 0$$

$$K^o_{a,0} - K_{a,-1} \leq 0$$

$$K^f_{a,0} - K_{a,-1} \leq 0$$

$$L_{a,0} - \sum_{h=1}^{H} l^h_{a,0} \leq 0$$

$$\sum_{h=1}^{H} \phi^h_{j,0} \leq 0. \tag{1.36}$$

For $a \in \mathcal{A}_0$

$$\sum_{h=1}^{H} k^h_{a,0} - K^n_{a,0} \leq 0$$

For $t = T$, there is no financial assets that pay-off at $T+1$, so

$$\sum_{h=1}^{H} c_T^h + \sum_{a\in\mathcal{A}_0} \psi_{a,T} - \sum_{h=1}^{H} e_T^h - \sum_{a\in\mathcal{A}_1} d_{a,T} \sum_{h} k_{a,t-1}^h - \sum_{a\in\mathcal{A}_0} y_{a,T} - \sum_{\substack{j=1\\ f_{j,T}=b_{j,T}}}^{J} \left(\sum_{h=1}^{H} \phi_{j,T-1}^h\right) b_{j,T} \leq 0$$

For $a \in \mathcal{A}_1$

$$K_{a,T}^f - \sum_{h=1}^{H} k_{a,T-1}^h - \sum_{\substack{j=1\\ f_{j,T}<b_{j,T}}}^{J} \left(\sum_{h=1}^{H} \phi_{j,T-1}^h\right) k_{j,T-1}^a \leq 0$$

$$L_{a,T} - \sum_{h=1}^{H} l_{a,T}^h \leq 0. \tag{1.37}$$

For $a \in \mathcal{A}_0$

$$K_{a,T}^o - \sum_{h=1}^{H} k_{a,T-1}^h - \sum_{\substack{j=1\\ f_{j,T}<b_{j,T}}}^{J} \left(\sum_{h=1}^{H} \phi_{j,T-1}^h\right) k_{j,T-1}^a \leq 0$$

Lemma 1.4 *The set of attainable allocations is bounded.*

Proof. We prove this Lemma by induction in t.

Before all, notice that given

$$\sum_h k_{a,t}^h - \sum_h k_{a,t-1}^h - \sum_{f_{j,t}<b_{j,t}} k_{a,t}^j \sum_h \phi_{j,t-1}^h \leq 0$$

for each $a \in \mathcal{A}_1$, and $t \leq T$, we have

$$\begin{aligned}\sum_h k_{a,t}^h &\leq \sum_h k_{a,t-1}^h + \sum_{f_{j,t}<b_{j,t}} k_{a,t}^j \sum_h \phi_{j,t-1}^h\\ &\leq \sum_h k_{a,t-1}^h\\ &\leq ..\\ &\leq \sum_h k_{a,0}^h \leq K_{a,-1}\end{aligned}$$

Step 1 $t \longmapsto t+1$: Suppose there is an M_t such that for each attainable allocations associate with an economy that

$$M_t \geq c_t^h\left(s^t\right) \geq 0$$
$$M_t \geq k_{a,t-1}^h\left(s^t\right) \geq 0$$
$$M_t \geq K_{a,t}^n\left(s^t\right) \geq 0$$
$$M_t \geq K_{a,t}^o\left(s^t\right) \geq 0$$
$$M_t \geq K_{a,t+1}^f\left(s^{t+1}\right) \geq 0$$
$$M_t \geq y_{a,t+1}\left(s^{t+1}\right) \geq 0$$
$$M_t \geq \left|\psi_{a,t}\left(s^t\right)\right|$$

we show that the statement holds at $t+1 \leq T$ by using the system of inequalities (1.35) and (1.37): For $a \in \mathcal{A}_0$ we have

$$K_{a,t+1}^o - \sum_{h=1}^H k_{a,t}^h - \sum_{\substack{j=1 \\ f_{j,t+1}<b_{j,t+1}}}^J \left(\sum_{h=1}^H \phi_{j,t}^h\right) k_{j,t}^a \leq 0$$

and

$$\sum_{h=1}^H \phi_{j,t}^h \leq 0,$$

therefore

$$K_{a,t+1}^o \leq HM_t = M_{t+1}^o.$$

Similarly

$$K_{a,t+1}^f - \sum_{h=1}^H k_{a,t}^h - \sum_{\substack{j=1 \\ f_{j,t+1}<b_{j,t+1}}}^J \left(\sum_{h=1}^H \phi_{j,t}^h\right) k_{j,t}^a \leq 0$$

therefore

$$K_{a,t+1}^f \leq M_{t+1}^o.$$

Besides,

$$\begin{aligned}\psi_{a,t+1} &\geq \Psi_a\left(K^n_{a,t+2}, K^o_{a,t+1}\right)\\ &\geq \Psi_a\left(0, M^o_{t+1}\right) = -M^{\psi-}_{a,t+1}.\end{aligned}$$

Second

$$\sum_{h=1}^{H} c^h_{t+1} + \sum_{a\in\mathcal{A}_0} \psi_{a,t+1} - \sum_{h=1}^{H} e^h_{t+1} - \sum_{a\in\mathcal{A}_1} d_{a,t+1} \sum_{h} k^h_{a,t-1} - \sum_{a\in\mathcal{A}_0} y_{a,t+1} - \sum_{\substack{j=1\\ f_{j,t+1}=b_{j,t+1}}}^{J} \left(\sum_{h=1}^{H} \phi^h_{j,t}\right) b_{j,t+1} \leq 0$$

and

$$\sum_{h=1}^{H} \phi^h_{j,t} \leq 0$$

implies

$$\sum_{h=1}^{H} c^h_{t+1} + \sum_{a\in\mathcal{A}_0} \psi_{a,t+1} \leq \sum_{h=1}^{H} e^h_{t+1} + \sum_{a\in\mathcal{A}_1} d_{a,t+1} K_{a,-1} + A_0 M_t.$$

Given that $c^h_{t+1} \geq 0$, for $a \in \mathcal{A}_0$

$$\psi_{a,t+1} \leq \max_{s^T} \sum_{h=1}^{H} e^h_{t+1} + \sum_{a\in\mathcal{A}_1} d_{a,t+1} K_{a,-1} + A_0 M_t + (A_0 - 1)\, M^{\psi-}_{a,t+1} = M^{\psi}_{t+1}.$$

Therefore,

$$\Psi_a\left(K^n_{a,t+1}, K^o_{a,t+1}\right) \leq M^{\psi}_{t+1}$$

since

$$\Psi_a\left(K^n_{a,t+1}, K^o_{a,t+1}\right) \leq \psi_{a,t+1}.$$

Also, given

$$K^o_{a,t+1} \leq M^o_{t+1},$$

and Ψ_a is decreasing in $K^o_{a,t+1}$, we have

$$\Psi_a\left(K^n_{a,t+1}, M^o_{t+1}\right) \leq M^{\psi}_{t+1}$$

so

$$K^n_{a,t+1} \leq \Psi^{-1}_{a,2}\left(M^o_{t+1}, M^\psi_{t+1}\right) = M^n_{a,t+1}.$$

Finally

$$\begin{aligned}
\sum_{h=1}^{H} c^h_{t+1} &\leq \sum_{h=1}^{H} e^h_{t+1} + \sum_{a\in\mathcal{A}_1} d_{a,t+1}K_{a,-1} + A_0 M_t - \sum_{a\in\mathcal{A}_0} \psi_{a,t+1} \\
&\leq \sum_{h=1}^{H} e^h_{t+1} + \sum_{a\in\mathcal{A}_1} d_{a,t+1}K_{a,-1} + A_0 M_t + \sum_{a\in\mathcal{A}_0} M^{\psi-}_{a,t+1} = M^c_{t+1}.
\end{aligned}$$

Lastly,

$$L_{a,t+2} \leq \sum_{h=1}^{H} L^h_a = \overline{L}_a$$

and

$$K^f_{a,t+2} \leq \sum_{h=1}^{H} k^h_{a,t+1} \leq K^n_{a,t+1} \leq M^n_{a,t+1}$$

therefore

$$y_{t+2} \leq \max_s F_a\left(M^n_{a,t+2}, \overline{L}_a, s\right) = M^f_{a,t+1}$$

Let

$$M_{t+1} = \max\left(M^c_{t+1}, M^o_{t+1}, M^n_{a,t+1}, M^f_{a,t+1}, M^\psi_{t+1}, M^{\psi-}_{a,t+1}\right).$$

we have

$$c^h_t, k^h_{a,t+2}, K^n_{a,t+1}, K^o_{a,t+1}, \psi_{a,t+1}, K^f_{a,t+2}, y_{a,t+2}$$

are bounded by M_{t+1}.

Step 2 $t = 0$: Similarly proof using (1.36). ■

Proof of Theorem 1.2. In this proof, we allow non-trivial labor choice decision, by supposing utility function of each consumer is concave over consumption and leisure $U_h(c, L-l)$. We restrict choices of produces and consumers to $[-2M_T, 2M_T]$,(keeping bond holding choices in $[-B, +B]$ and labor choices of final good producers in $\left[0, 2\overline{L}_a\right]$) constructed from above. To simplify the proof, we switch from the final good as numeraire to the following normalization:

Let Δ denote the set of prices $(p^c, q_a^*, q_a, d_a, w_a, p_j)$ such that

$$\begin{aligned} p^c, q_a^*, q_a, d_a, w_a, p_j &\geq 0 \\ p^c + \sum_{a\in\mathcal{A}_0} q_a^* + \sum_{a\in\mathcal{A}} q + \sum_{a\in\mathcal{A}_0} d_a + \sum_{a\in\mathcal{A}_0} w_a + \sum_{j\in\mathcal{J}} p_j &= 1 \end{aligned}$$

For each state s^t we normalize prices in each time-state pair such that

$$\left(p^c\left(s^t\right), q_a^*\left(s^t\right), q_a\left(s^t\right), d_a\left(s^t\right), w_a\left(s^t\right), p_j\left(s^t\right)\right) \in \Delta.$$

for $t \leq T-1$ and for the final date

$$(q, p^c, d_a, w_a, p) \in \Delta^f,$$

in which

$$\Delta^f = \left\{ (p^c, q_a, d_a, w_a) \geq 0 : p^c + \sum_{a\in\mathcal{A}_0} q_a + \sum_{a\in\mathcal{A}_0} d_a + \sum_{a\in\mathcal{A}_0} w_a = 1 \right\}.$$

Notice that the no-default constraint has become

$$f_{j,t+1}\left(s^{t+1}\right) = \min\left\{ p_t^c\left(s^t\right) b_j\left(s_{t+1}\right), \sum_{a=1}^{A} k_a^j\left(s^t\right)\left(q_a\left(s^{t+1}\right) + d_a\left(s^{t+1}\right)\right)\right\}$$

The optimal decisions of the capital producers yield $\left(K_{a,t}^n, K_{a,t}^o, \psi_{a,t}\right)_{s\in\Sigma^{T-1}}$, final good producers yields $\left(K_{a,t}^f, L_{a,t}, y_{a,t}\right)_{s\in\Sigma^T}$ and the decision of the consumer yields

$$\left(c_t^h, l_{a,t}^h, k_{a,t+1}^h, \phi_{j,t+1}^h\right)_{s\in\Sigma^{T-1}} \times \left(c_t^h, l_{a,t}^h\right)_{s\in\Sigma^T\setminus\Sigma^{T-1}}.$$

Let Z denote the correspondence that maps each set of prices

$$\left(p_t^c, q_t^*, q_t, d_t, w_t, p_{j,t}\right)_{s\in\Sigma^{T-1}} \times \left(q, p_t^c, d_t, w_t\right)_{s\in\Sigma^T\setminus\Sigma^{T-1}}$$

to the excess demand in each market in each time-state pair

$$Z : \Delta^{\left\|\Sigma^{T-1}\right\|} \times \left(\Delta^f\right)^{\left\|\Sigma^T\setminus\Sigma^{T-1}\right\|} \rightrightarrows \mathbf{R}^{(1+A_1+4A_0+J)\left\|\Sigma^{T-1}\right\|+(1+3A_0)\left\|\Sigma^T\setminus\Sigma^{T-1}\right\|}$$

$$p \in \Delta^{\|\Sigma^{T-1}\|} \times \left(\Delta^{f}\right)^{\|\Sigma^{T}\setminus\Sigma^{T-1}\|} \longmapsto z = \text{(excess demands)} \tag{1.38}$$

The component of the excess demand in each market corresponds to the component of the price system in that market. When $\sigma \in \|\Sigma^{T-1}\|$ there is one market for final good, A_1 markets for assets with fixed supply and $4A_0$ markets corresponding to new units, old units for asset production, old units for production and labor market for each asset with elastic supply, and finally J market for financial securities. When $\sigma \in \|\Sigma^{T}\|$ there are no market for financial securities nor new units of assets.

In Lemma 1.6, we establish that Z is upper hemi-continuous and compact, convex-valued. Given each individual choice is bounded, Z is bounded for example by a closed cube K of

$$\mathbf{R}^{(1+A_1+4A_0+J)\|\Sigma^{T-1}\|+(1+3A_0)\|\Sigma^{T}\setminus\Sigma^{T-1}\|}.$$

Consider the following correspondence

$$\digamma : \left(\Delta^{\|\Sigma^{T-1}\|} \times \left(\Delta^{f}\right)^{\|\Sigma^{T}\setminus\Sigma^{T-1}\|}\right) \times K \rightrightarrows \left(\Delta^{\|\Sigma^{T-1}\|} \times \left(\Delta^{f}\right)^{\|\Sigma^{T}\setminus\Sigma^{T-1}\|}\right) \times K$$

$$\begin{array}{c} \left\{p \in \Delta^{\|\Sigma^{T-1}\|} \times \left(\Delta^{f}\right)^{\|\Sigma^{T}\setminus\Sigma^{T-1}\|}, z \in K\right\} \\ \longmapsto \underset{\widetilde{p}\in\left(\Delta^{\|s^T\|\cdot(T-1)}\times\left(\Delta^{f}\right)^{\|s^T\|}\right)}{\arg\max} \left\{\widetilde{p}\cdot z\right\} \times Z\left(p\right). \end{array}$$

Since $\digamma$ is an upper hemi-continous correspondence, with non-empty, compact convex value. Kakutani's theorem guarantees that $\digamma$ has a fixed point

$$\overline{p} = \left(\left(\overline{p}^{c}_{t}, \overline{q}^{*}_{a,t}, \overline{q}_{a,t}, \overline{d}_{a,t}, \overline{w}_{a,t}, \overline{p}_{j,t}\right)_{s\in\Sigma^{T-1}} \times \left(\overline{p}^{c}_{T}, \overline{q}_{a,T}, \overline{d}_{a,T}, \overline{w}_{a,T}\right)_{s\in\Sigma^{T}\setminus\Sigma^{T-1}}, \overline{z}\right).$$

We simplify the notations by denoting

$$\begin{aligned} \overline{p}_t\left(s^t\right) &= \left(\overline{p}^{c}_{t}, \overline{q}^{*}_{a,t}, \overline{q}_{a,t}, \overline{d}_{a,t}, \overline{w}_{a,t}, \overline{p}_{j,t}\right) \text{ for } s^t \in \Sigma^{T-1} \\ \overline{p}_T\left(s^T\right) &= \left(\overline{p}^{c}_{T}, \overline{q}_{a,T}, \overline{d}_{a,T}, \overline{w}_{a,T}\right) \text{ for } s^T \in \Sigma^{T}\setminus\Sigma^{T-1} \end{aligned}$$

Notice that, by summing up over consumers' budget constraint as done in Lemma 1.5 we obtain the following inequalities

$$p_t^c Z_t^c\left(s^t\right) + \sum_{a\in\mathcal{A}_1} q_{a,t} Z_{a,t}^{Ko}\left(s^t\right) + \sum_{j=1}^{J} p_{j,t} Z_{j,t}\left(s^t\right) + \sum_{a\in\mathcal{A}_0} q_t^* Z_{a,t}^{Kn}\left(s^t\right) + \sum_{a\in\mathcal{A}_0} q_{a,t} Z_{a,t}^{K}\left(s^t\right)$$
$$+ \sum_{a\in\mathcal{A}_0} d_{a,t} Z_{a,t}^{Kf}\left(s^t\right) + \sum_{a\in\mathcal{A}_0} w_{a,t} Z_{a,t}^{L}\left(s^t\right) \leq 0 \tag{1.39}$$

for each $t \leq T-1$. Notice that for $t = 0$ $\sum_{h=1}^{H} \phi_{j,-1}^h = 0$ and $\sum_{h=1}^{H} k_{a,-1}^h = K_{a,-1}$. For the notations used above, we have

$$Z_t^c\left(s^t\right) = \sum_{h=1}^{H} c_t^h + \sum_{a\in\mathcal{A}_0} \psi_{a,t} - \sum_{h=1}^{H} e_t^h - \sum_{a\in\mathcal{A}_1} d_{a,t} \sum_h k_{a,t-1}^h - \sum_{a\in\mathcal{A}_0} y_{a,t} - \sum_{\substack{j=1 \\ f_{j,t}=p_t^c b_{j,t}}}^{J} \left(\sum_{h=1}^{H} \phi_{j,t-1}^h\right) b_{j,t}$$

$$Z_{a,t}^{Kn}\left(s^t\right) = \sum_{h=1}^{H} k_{a,t}^h - K_{a,t}^n$$

$$Z_{a,t}^{Ko}\left(s^t\right) = K_{a,t}^o - \sum_{h=1}^{H} k_{a,t-1}^h - \sum_{\substack{j=1 \\ f_{j,t}<b_{j,t}}}^{J} \left(\sum_{h=1}^{H} \phi_{j,t-1}^h\right) k_{j,t}^a$$

$$Z_{a,t}^{Kf}\left(s^t\right) = K_{a,t}^f - \sum_{h=1}^{H} k_{a,t-1}^h - \sum_{\substack{j=1 \\ f_{j,t}<b_{j,t}}}^{J} \left(\sum_{h=1}^{H} \phi_{j,t-1}^h\right) k_{j,t}^a$$

$$Z_{a,t}^{L}\left(s^t\right) = L_{a,t} - \sum_{h=1}^{H} l_{a,t}^h$$

For $a \in \mathcal{A}_0$

$$Z_{a,t}^{K}\left(s^t\right) = \sum_h k_{a,t}^h - \sum_h k_{a,t-1}^h - \sum_{f_{j,t}<p_t^c b_{j,t}} k_{a,t}^j \sum_h \phi_{j,t-1}^h$$

and for $j \in \mathcal{J}_t$

$$Z_{j,t}\left(s^t\right) = \sum_h \phi_{j,t}^h$$

For $t = T$

$$p_T^c Z_T^c\left(s^T\right) + \sum_{a\in\mathcal{A}_0} q_{a,T} Z_T^{Ko}\left(s^T\right) + \sum_{a\in\mathcal{A}_0} d_{a,T} Z_T^{Kf}\left(s^T\right) + \sum_{a\in\mathcal{A}_0} w_{a,T} Z_T^{L}\left(s^T\right) \leq 0 \tag{1.40}$$

with

$$Z_T^c\left(s^T\right) = \sum_{h=1}^{H} c_T^h + \sum_{a\in\mathcal{A}_0} \psi_{a,T} - \sum_{h=1}^{H} e_T^h - \sum_{a\in\mathcal{A}_1} d_{a,T} \sum_{h} k_{a,T-1}^h - \sum_{a\in\mathcal{A}_0} y_{a,T} - \sum_{\substack{j=1 \\ f_{j,T}=p_T^c b_{j,T}}}^{J} \left(\sum_{h=1}^{H} \phi_{j,T-1}^h\right) b_{j,T}$$

$$Z_{a,T}^{Ko}\left(s^T\right) = K_{a,T}^o - \sum_{h=1}^{H} k_{a,T}^h - \sum_{\substack{j=1 \\ f_{j,T}<b_{j,T}}}^{J} \left(\sum_{h=1}^{H} \phi_{j,T-1}^h\right) k_{j,T}^a$$

$$Z_{a,T}^{Kf}\left(s^T\right) = K_{a,T}^f - \sum_{h=1}^{H} k_{a,T}^h - \sum_{\substack{j=1 \\ f_{j,T}<b_{j,T}}}^{J} \left(\sum_{h=1}^{H} \phi_{j,T-1}^h\right) k_{j,TT}^a$$

$$Z_{a,T}^{L}\left(s^T\right) = L_{a,T} - \sum_{h=1}^{H} l_{a,T}^h$$

We re-write these inequalities compactly as

$$\overline{p}_t\left(s^t\right) \cdot \overline{z}_t\left(s^t\right) \leq 0 \ \forall t = 0, s^t.$$

Given

$$\overline{p}_t \in \arg\max_{\widetilde{p}\in\Delta} \left\{\widetilde{p} \cdot \overline{z}_t\right\}$$

we have (by choosing $\widetilde{p}$ in the corner of Δ or Δ_f depending on whether $t < T$) $\overline{z}_t \leq 0$ for each time-state pair t, s^t. In Lemma 1.4, the choices are bounded by M_T therefore the artificial bound $2M_T$ is not binding. Now we can show that prices are strictly positive:

1) $\overline{p}_t^c > 0$ otherwise $\overline{c}_t^h$ will reach the artificial bound $2M_T$, which contradicts the fact that the bound is not binding. Similarly

2) Given $\overline{p}_t^c > 0$, $\overline{d}_{a,t} > 0$ otherwise $\overline{K}_{a,t}^f$ will reach the artificial bound.

3) Given $\overline{d}_{a,t+1} > 0$, $\overline{q}_{a,t}^* > 0$ otherwise $\overline{k}_{a,t}^h$ will reach the artificial bound.

4) Given $\overline{p}_t^c > 0$, $\overline{q}_{a,t} > 0$ otherwise $\overline{K}_{a,t}^o$ will reach the artificial bound.

5) If $\overline{w}_{a,t} = 0$ then $\overline{L}_t = 2\overline{L}_a, \overline{l}_{a,t}^h \leq L_a^h$ which contradicts the negative excess demand in the labor markets, so $\overline{w}_{a,t} > 0$.

6) Finally if $\overline{p}_{j,t}^b = 0$ then $\overline{\phi}_{j,t+1}^h = B$ because $f_{j,t+1} > 0$, therefore $\sum_h \overline{\phi}_{j,t+1}^h = HB > 0$, which contradicts the negative excess demand in the financial market for asset j.

Therefore, we must have

$$\overline{p}_t^c, \overline{q}_{a,t}^*, \overline{q}_{a,t}, \overline{d}_{a,t}, \overline{w}_{a,t}, \overline{p}_{j,t}^b > 0.$$

$\overline{p}_t^c > 0$ also implies budget constraints, and therefore (1.39) and (1.40) hold with equality, so markets must clear. The collateral constraints (1.8) implies that if $\phi_{j,t} < 0$ then $-\phi_{jt} < \frac{M_T}{k_{j,a}}$, where $k_{j,a} = \min_{s \in \mathcal{S}} k_{j,a}(s) > \underline{k}$. Therefore if $\phi_{j,t}^h > 0$; $\phi_{j,t}^h < (H-1)\frac{M_T}{k_{j,a}}$. We can choose M_T independent of B, so we can choose B such that $B = (H-1)\frac{M_T}{k_{j,a}}$; this artificial constraint will not be binding. To conclude, observing that in this fixed point, all the artificial bounds are slack: we have thus found an equilibrium. ■

Lemma 1.5 ***(Walras' Law)*** *Given that consumers, firms optimize subject to their constraints, we obtain inequalities* (1.39) *and* (1.40).

Proof. We sum up the budget constraints (1.7) across all consumers

$$\begin{aligned}
& \sum_h p_t^c c_t^h + \sum_h \sum_{a \in \mathcal{A}_1} q_{a,t} k_{a,t}^h + \sum_h \sum_{a \in \mathcal{A}_0} q_{a,t}^* k_{a,t}^h + \sum_h \sum_{j=1}^{J} p_{j,t} \phi_{j,t}^h \\
\leq & \sum_h p_t^c e_t^h + \sum_h \sum_{a \in \mathcal{A}_0} w_{a,t} l_{a,t}^h + \sum_h \sum_{j=1}^{J} f_{j,t} \phi_{j,t-1}^h \\
& + \sum_h \sum_{a \in \mathcal{A}_0} (q_{a,t} + d_{a,t}) k_{a,t-1}^h + \sum_h \sum_{a \in \mathcal{A}_1} (q_{a,t} + d_{a,t}) k_{a,t-1}^h \\
& + \sum_a \Pi_t^f + \sum_a \Pi_t^a
\end{aligned}$$

So, moving endowment in final good e_t^h from the right hand side to the left hand side we obtain

$$\begin{aligned}
& p_t^c\left(\sum_h c_t^h-\sum_h e_t^h\right)+\sum_{a\in\mathcal{A}_1} q_{a,t}\left(\sum_h k_{a,t}^h\right) \\
& +\sum_{a\in\mathcal{A}_0} q_{a,t}^*\left(\sum_h k_{a,t}^h\right)+\sum_{j=1}^J p_{j,t}\left(\sum_h \phi_{j,t}^h\right) \\
\leq\ & \sum_h\sum_{a\in\mathcal{A}_0} w_{a,t}l_{a,t}^h+\sum_h\sum_{j=1}^J f_{j,t}\phi_{j,t-1}^h \\
& +\sum_h\sum_{a\in\mathcal{A}_0}\left(q_{a,t}+d_{a,t}\right)k_{a,t-1}^h+\sum_h\sum_{a\in\mathcal{A}_1}\left(q_{a,t}+p_t^c d_{a,t}\right)k_{a,t-1}^h \\
& +\sum_{a\in\mathcal{A}_0}\Pi_{a,t}^f+\sum_{a\in\mathcal{A}_0}\Pi_{a,t}
\end{aligned} \tag{1.41}$$

Notice that

$$\Pi_{a,t}^f=p_t^c y_{a,t}-d_{a,t}K_{a,t}^f-w_{a,t}L_{a,t}$$

and

$$\Pi_{a,t}=q_{a,t}^*K_{a,t}^n-p_t^c\psi_{a,t}-q_{a,t}K_{a,t}^o.$$

and if $f_{j,t}<p_t^c b_{j,t}$

$$f_{j,t}=\sum_{a=1}^A k_{a,t}^j\left(q_{a,t}+d_{a,t}\right).$$

Plugging these equalities into, (1.41) we obtain exactly the inequality(1.39). The inequality (1.40) is obtained similarly. ■

Lemma 1.6 *Z defined in* (1.38) *is upper hemi-continuous and compact, convex-valued.*

Proof. These properties are standard. ■

Proof of Lemma 1.3. Given any equilibrium, let $\mu_{a,t}$ denote the Lagrange multipliers associated to the collateral constraints, (1.8) in the consumers' optimization problem. First we show that consumptions are bounded from above and below: Market clearing condition implies $c_t^h\leq\bar{e}$. Second for each t one of the feasible strategies is to consume at least the endowment in each period therefore

$$\sum_{t'=t}^{T} \mathbf{P}_h \left(s^{t'}|s^t\right) \beta_h^{t'-t} U_h \left(c_t^h\right) \geq \sum_{t'=t}^{T} \mathbf{P}_h \left(s^{t'}|s^t\right) \beta_h^{t'-t} U_h \left(e_h \left(s_{t'}\right)\right)$$

Therefore

$$U_h \left(c_t^h\right) + \max\left\{\frac{\beta_h}{1-\beta_h} U_h \left(\overline{e}\right), 0\right\} \geq \min\left\{\min_{s\in\mathcal{S}} \frac{1}{1-\beta_h} U_h \left(e^h \left(s\right)\right), \min_{s\in\mathcal{S}} U_h \left(e^h \left(s\right)\right)\right\}$$

so

$$c_t^h \geq \underline{c}.$$

Second, we prove by induction that for each $a \in \mathcal{A}_0$, $K_{a,t+1} \leq \overline{K}_a$. Indeed, good market at time t clears implies

$$\Psi_a \left(K_{a,t+1}, K_{a,t}^o\right) \leq \overline{e}_a \left(s\right) \leq \Psi_a \left(\overline{K}_a, \overline{K}_a\right)$$

given

$$K_{a,t}^o = K_{a,t} \leq \overline{K}_a$$

and Ψ_a is decreasing in the second parameter, we have

$$\Psi_a \left(K_{a,t+1}, K_{a,t}^o\right) \geq \Psi_a \left(K_{a,t+1}, \overline{K}_a\right).$$

Therefore

$$\Psi_a \left(K_{a,t+1}, \overline{K}_a\right) \leq \Psi_a \left(\overline{K}_a, \overline{K}_a\right).$$

Since Ψ_a is increasing in the first parameters, we have

$$K_{a,t+1} \leq \overline{K}_a.$$

Now, the first-order condition of the asset producers implies

$$q_{a,t}^* = \frac{\partial \Psi_a \left(K_{a,t+1}, K_{a,t}^o\right)}{\partial K_{a,t+1}}.$$

Therefore

$$\underline{q}_a^* = \inf_{0 \leq K, K^o \leq \overline{K}_a} \frac{\partial \Psi_a (K, K^o)}{\partial K} \leq q_{a,t}^* \leq \sup_{0 \leq K, K^o \leq \overline{K}_a} \frac{\partial \Psi_a (K, K^o)}{\partial K} = \overline{q}_a^*.$$

Similarly

$$\underline{q}_a = \inf_{0 \leq K, K^o \leq \overline{K}_a} -\frac{\partial \Psi_a (K, K^o)}{\partial K^o} \leq q_{a,t} \leq \sup_{0 \leq K, K^o \leq \overline{K}_a} -\frac{\partial \Psi_a (K, K^o)}{\partial K^o} = \overline{q}_a.$$

The first-order condition with respect to k_{t+1}^h implies

$$\mu_{a,t}^h - q_{a,t}^* U_h' \left(c_t^h\right) + \beta_h E_t^h \left[\left(q_{a,t+1} + d_{a,t+1}\right) U_h' \left(c_{t+1}^h\right)\right] = 0$$

therefore

$$\begin{aligned} \overline{q}^* U_h' \left(c_t^h\right) &\geq \beta_h E_t^h \left[\left(q_{t+1} + d_{a,t+1}\right) U_h' \left(c_{t+1}^h\right)\right] \\ &> \beta_h E_t^h \left[d_{a,t+1} U_h' \left(c_{t+1}^h\right)\right] \end{aligned}$$

so $d_{a,t+1}$ is bounded by

$$\frac{\overline{q}^* U_h' (\overline{e})}{\beta_h \Pr_h (s_{t+1} | s^t) U_h' (\underline{c})} = \overline{d}_a.$$

Given

$$d_{a,t+1} = F_{a,K} (K_{a,t+1}, L_a)$$

$K_{a,t+1}$ is bounded from below by $\underline{K}_a > 0$. Also $d_{a,t+1} = F_{a,K} (K_{a,t+1}, L_a) > F_{a,K} \left(\overline{K}_a, L_a\right) = \underline{d}_a$ Similarly we have bounds $\overline{w}$ and $\underline{w}$ for $w_{a,t+1}$.

For $a \in \mathcal{A}_1$

$$q_{a,t} \leq \frac{H}{K_{a,-1}} \overline{c} = \overline{q}_a$$

otherwise there will be a consumer that holds at least $\frac{K_{a,-1}}{H}$ units of asset a at s^t after paying-off her debt. This consumer can sell part of her holding to pay-off debt and consume the rest of the sale. This strategy would give her more expected utility than her current one. This contradicts the optimally of her current choice. More formally, given

$$\sum_h \left(k^h_{a,t-1} + \sum_j \phi^h_{j,t-1} k^j_{a,t-2} \right) = K_{a,-1},$$

there must exist a consumer h such that

$$k^h_{a,t-1} + \sum_j \phi^h_{j,t-1} k^j_{a,t-2} \geq \frac{K_{a,-1}}{H}.$$

Therefore her budget at the beginning of period t will exceed

$$\begin{aligned}
& e^h_t + \sum_{a\in\mathcal{A}_0} w_{a,t} l^h_{a,t} + \sum_{j=1}^{J} f_{j,t}\phi^h_{j,t-1} + \sum_{a\in\mathcal{A}} \left(q_{a,t} + d_{a,t}\right) k^h_{a,t-1} \\
\geq\ & \left(q_{a,t} + d_{a,t}\right) \left(k^h_{a,t-1} + \sum_j \phi^h_{j,t-1} k^j_{a,t-2} \right) \\
\geq\ & \left(q_{a,t} + d_{a,t}\right) \frac{K_{a,-1}}{H} \\
>\ & \bar{c}.
\end{aligned}$$

To continue, the first-order condition

$$\mu^h_{a,t} - q_{a,t} U'_h\left(c^h_t\right) + \beta_h E^h_t \left[\left(q_{a,t+1} + d_{a,t+1}\right) U'_h\left(c^h_{t+1}\right) \right] = 0$$

yields

$$q_{a,t} \geq \max_h \frac{\beta_h \min_{s\in\mathcal{S}} d_a(s)\, U'_h(\bar{e})}{U'_h(\underline{c})} = \underline{q}_a.$$

The first-order condition with respect to $\phi_{j,t+1}$ implies for an agent h with $\phi^h_{j,t+1} \geq 0$ (check the deviation $\phi^h_{j,t+1} + \delta\phi$)

$$-p_{j,t} U'_h\left(c^h_t\right) + \beta_h E_h \left[f_{j,t+1} U'_h\left(c^h_{t+1}\right) \right] \leq 0.$$

So

$$\begin{aligned} p_{j,t} &\geq \frac{\beta_h E_h \left[f_{j,t+1} U_h' \left(c_{t+1}^h \right) \right]}{U_h' \left(c_t^h \right)} \\ &\geq \frac{\beta_h \min \left(b_j, \underline{k} \left(\underline{q} + \underline{d} \right) \right) U_h' \left(\overline{e} \right)}{U_h' \left(\underline{c} \right)} = \underline{p}_j. \end{aligned}$$

Moreover we should have

$$\begin{aligned} p_{j,t} &\leq \sum_{a \in \mathcal{A}_0} \overline{q}_a^* k_a^j + \sum_{a \in \mathcal{A}_1} \overline{q}_a k_a^j \\ &\leq \left(\sum_{a \in \mathcal{A}_0} \overline{q}_a^* + \sum_{a \in \mathcal{A}_1} \overline{q}_a \right) \overline{k} = \overline{p}_j, \end{aligned}$$

otherwise it is more than enough to simultaneously buy assets and sell security j, the aggregate demand of ϕ_j will be strictly negative. Also because of the market clearing condition we have $0 \leq k_{a,t+1}^h \leq \overline{K}_a$. Because of the collateral constraint

$$\begin{aligned} \phi_{j,t}^h &\geq \max_a \left(-\frac{\overline{K}_a}{k_{a,t}^j} \right) \\ &\geq \max_a \left(-\frac{\overline{K}_a}{\underline{k}} \right) = \underline{\phi}_j \end{aligned}$$

therefore

$$\phi_{j,t}^h \leq -\left(H - 1 \right) \underline{\phi}_j = \overline{\phi}_j.$$

■

Proof of Theorem 1.2. Let the compact set $\mathcal{T} \subset \widehat{\mathcal{V}}$ denote the set over which the equilibrium endogenous variables of the finite horizon economies lie and E is defined such that the set of equilibrium total units of assets always lie in E as well. For each correspondence $V : \mathcal{S} \times \Omega \times E \rightrightarrows \mathcal{T}$ define an operator that maps the correspondence to a new correspondence $W : \mathcal{S} \times \Omega \times E \rightrightarrows \widehat{\mathcal{V}}$ such that

$$W\left(s, \omega, K\right) = \left\{ \begin{array}{c} \widehat{v} \in \mathcal{T} \text{ such that } \left(s, \omega, K, v\right) \in \mathcal{X}\text{: } \exists \left(v_s\right)_{s \in \mathcal{S}} \in g\left(s, \omega, K, v\right) \\ \text{such that } \forall s' \ \widehat{v}_{s'} \in V\left(s', \omega'\right) \text{ in which } v_{s'} = \left(\omega', K', \widehat{v}_{s'}\right) \end{array} \right\}$$

Let $V^0 = \mathcal{T}$ and $V^{n+1} = G_{\mathcal{T}}(V^n)$. In Lemma 1.7 below, we show that V^{n+1} is a non-empty correspondence for all $n \geq 0$. We have $W(s, \omega, K)$ is not empty and $W(s, \omega, K) \subset V^0 = \mathcal{T}$ .It is also easy to show that

$$V(s, \omega, K) \subset V'(s, \omega, K)$$

for all $(s, \omega, K) \in \mathcal{S} \times \Omega \times E$ (denote $V \subset V'$) then the same inclusion holds for W and W'. By definition $V^1 \subset V^0$ so by induction we can show $V^{n+1} \subset V^n$.Therefore we have obtained a sequence of decreasing compact sets. Let

$$V^*(s, \omega, K) = \bigcap_{n=1}^{\infty} V^n(s, \omega, K)$$

Then V^* is a non-empty correspondence and $G_{\mathcal{T}}(V^*) \subset V^*$. Since graph of g is closed, we have that $G_{\mathcal{T}}(V^*)$ is non-empty as well. Let V^* be the 'policy correspondence' and

$$\mathcal{F}^*(s, \omega, K, v) = \left\{ \begin{array}{c} (v_s)_{s \in \mathcal{S}} \in g(s, \omega, K, v) \text{ such that} \\ \forall s' \; \widehat{v}_{s'} \in V^*(s', \omega') \text{ where } v_{s'} = (\omega', K', \widehat{v}_{s'}) \end{array} \right\}.$$

Then $(V^*, \mathcal{F}^*)$ is a Markov equilibrium. ■

Lemma 1.7 *V^{n+1} is a non-empty correspondence for all $n \geq 0$.*

Proof. For each n let consider the equilibrium constructed in Lemma 1.2 for the initial condition (s, ω, K) it is easy to show that the resulting allocation at time 0 belong to $V^n(s, \omega, K)$. For example, for $n = 0$: We use the equilibrium constructed in For each $s_1 \in \mathcal{S}$ Let v_{s_1} is defined by

$$\begin{aligned} q_a^* &= 0 \\ p_j &= 0 \end{aligned}$$

and q_a, w_a, d_a are defined as in that construction. We also add $k_a^h = 0$, $K_a = 0$ and $\phi_j^h = 0$ the other allocations are defined in the construction as well. Then $(s_1, v_{s_1}) \in \mathcal{X}$. It easy to see that $(v_s)_{s \in \mathcal{S}} \in g(s, \omega, K, v)$. Also $\widehat{v}_{s_1} \in V(s_1, \omega_1)$ by definition.

■

Algorithm 1.1 *Computing Complete Market Equilibria: The state space should be*

$$\left((c_h)_{h\in\mathcal{H}}, (K_a)_{a\in\mathcal{A}_0}\right).$$

We find the mapping ρ from that state space into the set of current prices and investment levels $\{q_a, q_a^*, w_a, d_a, K_a^n\}_{a\in\mathcal{A}_0}, \{q_a\}_{a\in\mathcal{A}_1}$, future consumptions $\left\{(c_h^+)_{h\in\mathcal{H}}\right\}_{s\in\mathcal{S}}$, and $\{p_s\}_{s\mathcal{S}}$ the Arrow-Debreu state prices. There are therefore $5A_0 + A_1 + SH + S$ unknowns.
First, notice that $l_{h,a} = L_a^h$.
For each $a \in \mathcal{A}_0$, from the first order condition for the asset producers and final good producers, we obtain.

$$\begin{aligned} q_a^* &= \frac{\partial \Psi_a\left(K_a^n, K_a\right)}{\partial K_a^n} \\ q_a &= -\frac{\partial \Psi_a\left(K_a^n, K_a\right)}{\partial K_a} \\ d_a &= F_K\left(K_a, \sum_a l_a^h, s\right) \\ w_a &= F_L\left(K_a, \sum_a l_a^h, s\right) \end{aligned}$$

which give $4A_0$ equations. From the non-arbitrage equations, it should be that

$$q_a^* = \sum_s p_s\left(q_a^+ + d_a^+\right)$$

this gives another A_0 equations.
For each $a \in \mathcal{A}_1$ we also have A_1 equations

$$q_a = \sum_s p_s\left(q_a^+ + d_a^+\right)$$

Regarding p_s, the inter-temporal Euler equation implies

$$p_s = \frac{U_h'\left(c_h^+\right)}{U_h'\left(c_h\right)}$$

that give SH equations and finally

$$\sum_h c_h^+ + \sum_a \Psi_a\left(K_a^{n+}, K_a^n\right) = \sum_h e_h + \sum_{a\in\mathcal{A}0} F\left(K_a^n, \sum_a l_a^h\right) + \sum_{a\in\mathcal{A}_1} e_a K_{a,-1}$$

which give another S equations. With these $5A_0 + A_1 + SH + S$ equations, we can solve for the $5A_0 + A_1 + SH + S$ unknowns. That solution determines the mapping $T\rho$.

In order to find an equilibrium corresponding to an initial asset holdings $(\theta_{h,a})_{h\in\mathcal{H},a\in\mathcal{A}}$ we find the value of stream of consumption and endowment of each consumers

$$V_c^h = c_h + \sum_{s\in\mathcal{S}} p_s V_c^{h+}(s)$$

and

$$V_e^h = e_h + \sum_{s\in\mathcal{S}} p_s V_e^{h+}(s)$$

Then we solve for H unknowns $(c_h)_{h\in\mathcal{H}}$ using H equations

$$V_c^h = V_e^h + \sum_{a\in\mathcal{A}} \theta_{h,a} q_a.$$

Remark 1.1 *When there are no assets with elastic supply, calculation is easier: The state space should be $\left((c_h)_{h\in\mathcal{H}-1}\right)$ We find the mapping ρ from that state space into $\left\{\left(c_h^+\right)_{h\in\mathcal{H}-1}\right\}_{s\in\mathcal{S}}$ future consumptions and $\{p_s\}_{s\mathcal{S}}$ the Arrow-Debreu state prices. In total we have HS unknowns. Notice that we need to keep track of the consumption of only $H-1$ consumers. The consumption of the remaining consumer is determined by the market clearing condition*

$$c_H(s) = e_a(s) - \sum_{h=1}^{H-1} c_h(s).$$

he intertemporal Euler equation implies

$$p_s = \frac{U_h'\left(c_h^+\right)}{U_h'(c_h)}$$

that give HS equations. From these HS equations we can solve for the HS unknowns. When *we have CRRA utility function, we can solve for closed form solution of p_s and c_h^+.*

Algorithm 1.2 ***Computing Incomplete Markets Equilibria:*** *We look for the equilibrium mapping defined in* (1.28), *for each iteration, given ρ^n,*

$$\rho_s^{n+1}(\omega, K_a) = \left(\widehat{v}_{n+1}, \omega_{s,n+1}^+, \mu_{n+1}, \eta_{n+1}\right)$$

is determined to satisfied the following equations

$$\begin{aligned}
0 &= \mu_{a,n+1}^h - q_{a,n+1}^* U_h'\left(c_{n+1}^h\right) \\
&\quad + \beta_h E^h\left\{\left(q_a^+ + d_a^+\right) U_h'\left(c^{h+}\right)\right\} \\
0 &= \mu_{a,n+1}^h \left(k_{a,n+1}^h + \sum_{j\in\mathcal{J}:\phi_j^h<0} k_{a,j}^j \phi_{j,n+1}^h \right) \\
0 &\leq k_{a,n+1}^h + \sum_{j\in\mathcal{J}:\phi_j^h<0} k_a^j \phi_{j,n+1}^h .
\end{aligned}$$

The variables with superscript $+$,$q_a^+, d_a^+, c^{h+}, l^{h+}$ *are determined using the mapping* ρ^n *on the state variables* $(s, \omega_s^+, K_{a,n+1})$ *where*

$$K_{a,n+1} = \begin{cases} \sum_{h\in\mathcal{H}} k_{a,n+1}^h & \text{if } a \in \mathcal{A}_0 \\ K_{a,-1} & \text{if } a \notin \mathcal{A}_0 \end{cases}$$

We also require

$$\begin{aligned}
0 &= \sum_{a\in\mathcal{A}} \mu_{a,n+1}^h k_a^j - p_{j,n+1} U_h'\left(c_{n+1}^h\right) \\
&\quad + \beta^h E^h\left\{f_j^+ U_h'\left(c^{h+}\right)\right\} - \eta_{j,n+1}^h(-) \\
0 &= -p_{j,n+1} U_h'\left(c_{n+1}^h\right) + \beta^h E^h\left\{f_j^+ U_h'\left(c^{h+}\right)\right\} + \eta_{j,n+1}^h(-) \\
0 &= \phi_{j,n+1}^h(+)\, \eta_{j,n+1}^h(+) \\
0 &= \phi_{j,n+1}^h(-)\, \eta_{j,n+1}^h(-) .
\end{aligned}$$

The budget constraints of the consumers hold with equality

$$c_{n+1}^{h} = e^{h}(s) + \omega^{h}(q_{n+1} + d_{n+1}) \cdot K - q_{n+1}^{*} \cdot k_{n+1}^{h} + w_{n+1} \cdot l_{n+1}^{h} - p_{n+1} \cdot \phi_{n+1}^{h}$$

where, for each $a \in \mathcal{A}_0$

$$\begin{aligned} q_{a,n+1}^{*} &= \frac{\partial \Psi_a (K_{a,n+1}, K_a^o)}{\partial K} \\ q_{a,n+1} &= -\frac{\partial \Psi_a (K_{a,n+1}, K_a^o)}{\partial K^o} \\ d_{a,n+1} &= \frac{\partial F_a (K_a^o, L_{a,n+1})}{\partial K^o} \\ w_{a,n+1} &= \frac{\partial F_a (K_a^o, L_{a,n+1})}{\partial L} \end{aligned}$$

with $l_{a,n+1}^{h} = L_{h,a}$ *and* $L_{a,n+1} = \sum_{h \in \mathcal{H}} l_{a,n+1}^{h}$. *Finally, the future wealth distributions are consistent with current asset holdings and future prices*

$$\omega_s^{h+} = \frac{k_{n+1}^{h} \cdot (q_s^{+} + d_s^{+}) + \sum_{j \in \mathcal{J}} \phi_{j,n+1}^{h} \min \left\{ b_j(s), \sum_{a \in A} k_{a,n+1}^{j} (q^{+} + d^{+}) \right\}}{\sum_a (q^{+} + d^{+}) \cdot K_{n+1}}$$

again the variables with superscript $+$, q_s^{+}, d_s^{+}, *are determined using the mapping* ρ^n.

Proof of Proposition 1.8. Since there are only to future states, let u denote the higher return

$$u = \max_{s^{t+1}|s^t} \left(q\left(s^{t+1}\right) + d\left(s_{t+1}\right) \right)$$

and d denote the lower return

$$d = \min_{s^{t+1}|s^t} \left(q\left(s^{t+1}\right) + d\left(s_{t+1}\right) \right)$$

We are considering the set of debt assets that promise 1 in both states and requires k unit of the real asset as collateral. The price of such an asset is p_k

1. $k < \frac{1}{u}$; then this asset is essentially the real asset because its effective pay-off is (ku, kd)

2.$\frac{1}{d} \geq k \geq \frac{1}{u}$. Then the pay-off to the borrower of the asset is

$$(ku - 1, 0)$$

and he has to pay $kq - p_k$: she buys k real asset but she get p_k from selling the financial asset. So the borrowers only choose k such that

$$\frac{ku-1}{kq-p_k}$$

is maximum among $k \in \left[\frac{1}{u}, \frac{1}{d}\right]$. So only assets belonging to

$$\arg\max_{k\in\left[\frac{1}{u},\frac{1}{d}\right]} \frac{ku-1}{kq-p_k} \tag{1.42}$$

will be chosen by borrowers in equilibrium.

Consider an actively traded financial asset with collateral level k^* belong to the argmax set above. If another asset with $k < k^*$ is also actively traded, price of this asset, p_k, will be strictly less than

$$\frac{ku-1}{k^*u-1}p_{k^*} + \frac{k^*-k}{k^*u-1}q.$$

Otherwise, due to collateral value of the real asset, buyers of this asset will strictly prefer the portfolio $\frac{ku-1}{k^*u-1}$ of asset k^* and $\frac{k^*-k}{k^*u-1}$ of the real asset. This portfolio gives the same payoff value as buying one unit of asset k as

$$\begin{bmatrix} 1 \\ kd \end{bmatrix} = \frac{ku-1}{k^*u-1}\begin{bmatrix} 1 \\ k^*d \end{bmatrix} + \frac{k^*-k}{k^*u-1}\begin{bmatrix} u \\ d \end{bmatrix},$$

on top of that it gives the buyer an additional collateral value from holding the real asset. Therefore

$$\frac{ku-1}{kq-p_k} < \frac{k^*u-1}{k^*q-p_{k^*}}.$$

Thus every seller of this asset k will strictly prefer selling asset k^*. ■

Chapter 2

Innovation by Entrants and Incumbents

2.1 Introduction

The endogenous technological change literature provides a coherent and attractive framework for modeling productivity growth at the industry and the aggregate level. It also enables a study of how economic growth responds to incentives, policies, and market structure. A key aspect of the growth process is the interplay between innovations and productivity improvements by existing firms on the one hand and entry by more productive, new firms on the other. Existing evidence suggests that this interplay is important for productivity growth. For example, Bartelsman and Doms (2000) and Foster and Krizan (2000), among others, document that entry of new establishments (plants) accounts for about 25% of average TFP growth at the industry level, with the remaining productivity improvements accounted for by continuing establishments. These issues are difficult to address with either of the two leading frameworks of endogenous technological change, the expanding variety models, e.g., Romer (1990), Grossman and Helpman (1991a), Jones (1995), and the Schumpeterian quality ladder models, e.g., Segerstrom and Dinopoloulos (1990), Aghion and Howitt (1992), Grossman and Helpman (1991b). The expanding variety models do not provide a framework for directly addressing these questions.[1] The Schumpeterian models are potentially better suited to studying the interplay between incumbents and entrants, since they focus on the process of creative destruction and entry. Nevertheless, because of Arrow's replacement effect (Arrow 1962), these baseline models predict that all innovation should be undertaken by new firms and thus do not provide a framework for the analysis of over 75% of the productivity growth due to innovation by existing establishments.[2]

This paper provides a simple framework that involves simultaneous innovation by new and existing establishments.[3] The model is a tractable (and minimal) extension of the multisector Schumpeterian growth model. It generates endogenous growth in a manner similar to the standard endogenous technological change models, but the contribution of incumbent (continuing plants) and new firms to growth is determined in equilibrium. Although the parameters necessary for a careful calibration of the model are not currently available, plausible choices of parameters generates numbers consistent with 75% of productivity growth being driven by continuing plants. Since existing plants are involved in innovation and expand their sizes as they increase their productivity, the model also generates an endogenous distribution of firm sizes. In particular, the available evidence suggests that firm growth can be approximated by "Gibrat's Law," whereby the rate of growth of a particular firm is independent of its size. In addition, the distribution of firms can be approximated by a Pareto distribution with a coefficient close to one (i.e., the so-called "Zipf's distribution," where the fraction of firms with size greater than S is proportional to $1/S$).[4] In the model here, there is entry and exit into an industry, and conditional on not exiting, firms grow on average. Consistent with Gibrat's Law,

[1] In the expanding variety models, the identity of the firms that are undertaking the innovation does not matter, so one could assume that it is the existing producers that are inventing new varieties, though this will be essentially determining the distribution of productivity improvements across firms by assumption.

[2] Models of step-by-step innovation, such as Aghion, Harris, Howitt, and Vickers (2001), Aghion, Bloom, Blundell, Griffith, and Howitt (2005), and Acemoglu and Akcigit (2006), allow innovation by incumbents, but fix the set of firms competing within an industry, thus do not feature entry. Aghion, Burgess, Redding, and Zilibotti (2005) consider an extension of these models in which there is entry, but focus on how the threat of entry may induce incumbents to innovate.

[3] In the model, each firm will consist of a single plant, thus the terms "establishment," "plant" and "firm" can be used interchangeably.

[4] For evidence on firm growth, see, among others, Sutton (1997) and Sutton (1998). For patterns of firm entry and exit, see Dunne and Samuelson (1988), Dunne and Samuelson (1989) and Klepper (1996). For evidence on firm size distribution, see the classic paper by Simon and Bonini (1958) and the recent evidence in Axtell (2001). For the size distribution of cities, see, among others, Dobkins and Ioannides (1998), Gabaix (1999)and Eeckhout and Jovanovic (2002).

both the growth rates conditional on not existing and the unconditional growth rates are independent of firm size. With an additional assumption (allowing a minimal amount of imitation), the model also generates a stationary equilibrium distribution of firm sizes that is Pareto with an exponent approximately equal to one (the so-called Zipf's distribution, e.g., Lucas 1978b, Gabaix 1999), which appears to be a very good approximation to the distribution of firm sizes in the US data (Axtell 2001).

The model consists of a given number of sectors producing inputs (machines) for the unique final good of the economy. In each sector, there is a quality ladder, and at any point in time, a single firm has access to the highest quality input (machine). This firm has incentives to increase its quality continuously by undertaking R&D in order to increase profits. These R&D investments are the source of productivity growth by continuing firms. At the same time, potential entrants undertake R&D in order to create a better input and replace the incumbent. This is the source of productivity growth by entrants. The building blocks of the model, that incumbents engage in continuous productivity-increasing research, while potential entrants seek more radical innovations, are consistent with the case study evidence on the nature of innovation. Freeman (1982), Pennings and Buitendam (1987),Tushman and Anderson (1986) and Scherer (1984) document how established firms involved in innovations that improve existing products, while new firms invest in more radical and "original" innovations (see also the discussion in Arrow 1974). In the model, this pattern arises because Arrow's replacement effect implies that incumbents do not wish to pursue radical innovations, but they have access to a technology for improving the quality of their machines/products and have the incentives to do so.

The dynamic equilibrium of this economy can be characterized in closed form and leads to a number of interesting comparative static results. Despite the Schumpeterian character of the model, there is an endogenous negative relationship between the rate of entry of new firms and the rate of productivity growth (the higher is entry in equilibrium, the lower is growth). This reflects the importance of productivity growth by incumbents. In particular, reduced entry makes incumbents more profitable, and they respond by undertaking more R&D and increasing productivity growth. This same economic mechanism also leads to a surprising result. While taxes on existing firms unambiguously reduce productivity growth, taxes or entry barriers on potential entrants *increase* economic growth. This is a rather paradoxical result, since the underlying model is only a small variant of the baseline Schumpeterian growth, where entry by new firms drives growth entirely. It is the outcome of the greater productivity improvements by incumbents in response to reduced entry, which outweighs the negative effect of the reduction in creative destruction. This result does not necessarily imply that entry barriers would be growth-enhancing in practice, but isolates a new effect of entry barriers on productivity growth.[5]

This paper is most closely related to Klette and Kortum (2004). Klette and Kortum construct a rich model of firm and aggregate innovation dynamics. Their model is one of expanding product varieties and the firm size distribution is driven by differences in the number of products that a particular firm produces. Klette and Kortum assume that firms with more products have an advantage in discovering more new products. With this assumption, their model generates the same patterns as here and also matches additional facts about propensity to patent and differential survival probability of firms by size. One disadvantage of this approach is that the firm size distribution is not driven by the dynamics of continuing plants (in fact, if new products are interpreted as new plants or establishments, the Klette-Kortum model predicts that all productivity growth will be driven by entry of new plants, though this may be an extreme interpretation, since some new products are produced in existing plants). The current model is best viewed as complementary to Klette and Kortum (2004), and focuses on innovations by existing firm in the same line of business instead of the introduction of new products. In practice, both types of innovations appear to be important and it is plausible that existing large firms might be more successful in locating new product opportunities.[6] Nevertheless, both qualitative and some recent quantitative evidence suggests that innovation by firms and existing lines of products are most important. Abernathy (1980), Lieberman (1984), and Scherer (1984), among others, provide various case studies documenting the importance of innovations by existing firms and establishments in the same line of business for productivity growth (for example, the role of innovations by General Motors and Ford in the car industry). Empirical work by Bartelsman and Doms (2000) and Foster and Krizan (2000) is also consistent with the

[5]In some way, the current model may be viewed as combining two of Schumpeter's important ideas; the process of creative destruction and the importance of large (here continuing) firms in innovation (see Schumpeter 1934, and Schumpeter 1942).

[6]Scherer (1984), for example, emphasizes both the importance of innovation by continuing firms (and establishments) and documents that larger firms produce more products.

importance of productivity growth by continuing establishments rather than the importance of new products, while Broda and Weinstein (1996) provide empirical evidence on the importance of improvements in the quality of products in international trade. An additional advantage of the model presented here relative to that of Klette and Kortum (2004) is its relative simplicity and tractability, which makes it particularly useful for policy analysis. The model is a slight variant of the standard Schumpeterian framework, thus constitutes a minimal departure from a "textbook" model. Other related papers include Lentz and Mortensen (2008), Klepper (1996) and Atkeson and Burstein (2007). Lentz and Mortensen (2008) extend Klette and Kortum's model by introducing additional sources of heterogeneity and estimate the model on Danish data. Klepper (1996) documents various facts about the firm size, entry and exit decisions and innovation, and provides a simple descriptive model that can account for these facts. The recent paper by Atkeson and Burstein (2007) also incorporates innovations by existing firms, but focuses on the implications of these innovation dynamics for the relationship between trade opening and productivity growth. None of these papers consider a Schumpeterian growth with innovation both by incumbents and entrants that can be easily mapped to decomposing the contribution of new and continuing plants (firms) to productivity growth.

Another set of related papers include Jovanovic (1982), Hopenhayn (1992), Ericson and Pakes (1995), Melitz (2003),Rossi-Hansberg and Wright (2007b), Rossi-Hansberg and Wright (2007a), Lagos (2001), and Luttmer (2004), Luttmer (2007), which analyze firm dynamics. Many of these papers generate realistic firm size distributions based on heterogeneity of productivity (combined with fixed costs of operation). Nevertheless, these papers typically take the stochastic productivity growth process of firms as exogenous, whereas my focus here is to understand how R&D decisions of firms shape the endogenous process of productivity growth. Another noteworthy contribution is that although the current model is not explicitly designed for studying firm size distributions, the equilibrium distribution of firm sizes approximates a Pareto distribution with an exponent of one (i.e., the "Zipf distribution"), such as in Luttmer (2007) for firm sizes or Gabaix (1999) for cities.

The rest of the paper is organized as follows. Section 2.2 presents the basic environment and characterizes the equilibrium. This section also shows that under some plausible parameterizations the model generates a large fraction of productivity growth driven by incumbents. Section

2.3 looks at the effects of policy on equilibrium growth and shows the paradoxical result that entry barriers and taxes on entrants increase economic growth. We also briefly characterizes the Pareto optimal allocation in this economy and compares it to the equilibrium. In Section 2.4 we show that in the original economy, there does not exist an equilibrium stationary firm size distribution. However by adding another type of entry we obtain a balanced growth path equilibrium with a stationary firm size distribution. We show that the stationary distribution approximates a Pareto distribution with an exponent close to one. Section 2.5 concludes.

2.2 Model

2.2.1 Environment

The economy is in continuous time and admits a representative household with the standard constant relative risk aversion (CRRA) preferences

$$\int_0^{\infty} e^{-\rho t} \frac{C(t)^{1-\theta} - 1}{1-\theta} dt,$$

where $\theta \geq 1$ is the coefficient of relative risk aversion or the inverse of the intertemporal elasticity of substitution. Population is constant at L and labor is supplied inelastically. The resource constraint at time t takes the form

$$C(t) + X(t) + Z(t) \leq Y(t), \tag{2.1}$$

where $C(t)$ is consumption, $X(t)$ is aggregate spending on machines, and $Z(t)$ is total expenditure on R&D at time t.

There is a continuum of machines (inputs) normalized to 1 used in the production of a unique final good. Each machine line is denoted by $\nu \in [0,1]$. The production function of the unique final good is given by:

$$Y(t) = \frac{1}{1-\beta} \left[\int_0^1 q(\nu,t)^{\beta} x(\nu,t|q)^{1-\beta} d\nu \right] L^{\beta}, \tag{2.2}$$

where $x(\nu,t|q)$ is the quantity of the machine of type v of quality q used in the production process. This production process implicitly imposes that only the highest quality machine will be used in production for each type of machine $\nu \in [0,1]$.

Throughout, the price of the final good at each point is normalized to 1 (relative prices of final goods across different periods being determined by the interest rate).

The engine of economic growth here will be process innovations that lead to quality improvements. This will be driven by two types of innovations:

1. Innovation by incumbents

2. Creative destruction by entrants

Let $q(\nu,t)$ be the quality of machine line v at time t. We assume the following "quality ladder" for each machine type:

$$q(\nu,t) = \lambda^{n(\nu,t)} q(\nu,s) \text{ for all } v \text{ and } t,$$

where $\lambda > 1$ and $n(\nu,t)$ denotes the number of incremental innovations on this machine line between time $s \leq t$ and time t, where time s is the date at which this particular machine type was first invented and $q(\nu,s)$ refers to its quality at that point. The incumbent has a fully enforced patent on the machines that it has developed (though this patent does not prevent entrants leapfrogging the incumbent's quality). We assume that at time $t=0$, each machine line starts with some quality $q(\nu,0) > 0$ owned by an incumbent with a fully-enforced patent on this initial machine quality.

Incremental innovations can only be performed by the incumbent producer. So we can think of those as "tinkering" innovations that improve the quality of the machine. If the current incumbent spends an amount $z(\nu,t)\,q(\nu,t)$ of the final good for this type of innovation on a machine of current quality $q(\nu,t)$, it has a flow rate of innovation equals to $\phi(z(\nu,t))$ for $\phi(z)$ strictly increasing, concave in z and satisfies the following Inada-type assumptions[7]

$$\phi(0) = 0 \text{ and } \phi'(0) = \infty. \tag{2.3}$$

[7]More formally, this implies that for any interval $\Delta t > 0$, the probability of one incremental innovation is $\phi z(\nu,t)\,\Delta t$ and the probability of more than one incremental innovation is $o(\Delta t)$ with $o(\Delta t)/\Delta t \to 0$ as $\Delta t \to 0$.

Recall that such an innovation results in a proportional improvement in quality and the resulting new machine will have quality $\lambda q(\nu, t)$.

The alternative to incremental innovations are *radical innovations.* A new firm (entrant) can undertake R&D to innovate over the existing machines in machine line v at time t[8]. If the current quality of machine is $q(\nu, t)$, then by spending one unit of the final good, this new firm has a flow rate of innovation equal to $\frac{\eta(\widehat{z}(\nu,t))}{q(\nu,t)}$, where $\eta(.)$ is a strictly decreasing, continuously differentiable function, and $\widehat{z}(\nu, t)$ is total amount of R&D by new entrants towards machine line ν at time t. The presence of the strictly decreasing function η captures the fact that when many firms are undertaking R&D to replace the same machine line, they are likely to try similar ideas, thus there will be some amount of "external" diminishing returns (new entrants will be "fishing out of the same pond"). Since each entrant attempting R&D on this line is potentially small, they will all take $\widehat{z}(\nu, t)$ as given. Throughout we assume that $z\eta(z)$ is *strictly increasing* in z so that greater aggregate R&D towards a particular machine line increases the overall probability of discovering a superior machine. We also suppose that $\eta(z)$ satisfies the following Inada-type assumptions

$$\lim_{z \longrightarrow 0} \eta(z) = 0 \text{ and } \lim_{z \longrightarrow \infty} \eta(z) = \infty. \tag{2.4}$$

An innovation by an entrant leads to a new machine of quality $\kappa q(\nu, t)$, where $\kappa > \lambda$. Therefore, innovation by entrants are more "radical" than those of incumbents. Existing empirical evidence from studies of innovation support the notion that innovations by new entrants are more significant or radical than those of incumbents[9]. We assume that whether the entrant was a previous incumbent or not does not matter for its technology of innovation or for the outcome of its innovation activities.

A simple example of functions $\phi(.)$ and $\eta(.)$ that satisfy the requirements above are

$$\phi(z) = Az^{1-\alpha} \text{ and } \eta(z) = Bz^{-\gamma}, \tag{2.5}$$

[8] Incumbents could also access the technology for radical innovations, but would choose not to. Arrow's replacement effect implies that since entrants make nonpositive profits from this technology (because of free entry), the profits of incumbents, who would be replacing their own product, would be negative. Incumbents will still find it profitable to use the technology for incremental innovations, which is not available to entrants.

[9] However, it may take a while for the successful entrants to realize the full productivity gains from these innovations (e.g., Freeman 1982). we are abstracting from this aspect.

with $\alpha, \gamma \in (0,1)$. We will use this functional form to derive some simple quantitative implications from the model in Subsection 2.2.5. For the rest of the analysis, there is no reason to assume a specific functional form.

Now we turn to describing the production technology. Once a particular machine of quality $q(\nu, t)$ has been invented, any quantity of this machine can be produced at constant marginal cost ψ. We normalize $\psi = 1 - \beta$ without loss of any generality, which simplify the algebra below. This implies that the total amount of expenditure on the production of intermediate goods at time t is

$$X(t) = (1-\beta) \int_0^1 x(\nu, t)\, d\nu, \tag{2.6}$$

where $x(\nu, t)$ is the quantity of this machine used in final good production. Similarly, the total expenditure on R&D is

$$Z(t) = \int_0^1 [z(\nu, t) + \widehat{z}(\nu, t)]\, q(\nu, t)\, d\nu, \tag{2.7}$$

where $q(\nu, t)$ refers to the highest quality of the machine of type ν at time t. Notice also that total R&D is the sum of R&D by incumbents and entrants ($z(\nu, t)$ and $\widehat{z}(\nu, t)$ respectively). Finally, define $p^x(\nu, t|q)$ as the price of machine type ν of quality $q(\nu, t)$ at time t. This expression stands for $p^x(\nu, t|q(\nu, t))$, but there should be no confusion in this notation since it is clear that q refers to $q(\nu, t)$, and we will use this notation for other variables as well.

2.2.2 Equilibrium Definitions

In this Subsection, we define the equilibrium of the economy described in the previous section.

An allocation in this economy consists of time paths of consumption levels, aggregate spending on machines, and aggregate R&D expenditure $[C(t), X(t), Z(t)]_{t=0}^{\infty}$, time paths for R&D expenditure by incumbents and entrants $[z(\nu, t), \widehat{z}(\nu, t)]_{v\in[0,1],t=0}^{\infty}$, time paths of prices and quantities of each machine an the net present discounted value of profits from that machine, $[p^x(\nu, t|q), x(\nu, t), V(\nu, t|q)]_{v\in[0,1],t=0}^{\infty}$, and time paths of interest rates and wage rates, $[r(t), w(t)]_{t=0}^{\infty}$. An equilibrium is given by an allocation in which R&D decisions by entrants maximize their net present discounted value, pricing, quantity and R&D decisions by incumbents maximize their net present discounted value, the representative household chooses the path of consumption and allocation of spending across machines and R&D optimally, and the labor market clears.

Let us start with the aggregate production function for the final good producers. Profit maximization by the final good sector implies that the demand for the highest available quality of machine $\nu \in [0,1]$ at time t is given by

$$x(\nu,t) = p^x(\nu,t|q)^{-1/\beta} q(\nu,t) L \text{ for all } \nu \in [0,1] \text{ and all } t. \tag{2.8}$$

The price $p^x(\nu,t|q)$ will be determined by the profit maximization of the monopolist holding the patent for machine of type ν and quality $q(\nu,t)$. Note that the demand from the final good sector for machines in (2.8) is iso-elastic, so the unconstrained monopoly price is given by the usual formula with a constant markup over marginal cost. Throughout, we assume that

$$\kappa \geq \left(\frac{1}{1-\beta}\right)^{\frac{1-\beta}{\beta}}, \tag{2.9}$$

so that after an innovation by an entrant, there will not be limit pricing. Instead, the entrant will be able to set the unconstrained profit-maximizing (monopoly) price. By implication, an entrant that innovates further after its own initial innovation will also be able to set the unconstrained monopoly prices[10]. Condition (2.9) also implies that, when the highest quality machine is sold at the monopoly price, the final good sector will only use this machine type and thus justifies the way we wrote the final good production function, (2.2), imposing that only the highest quality machines in each line will be used.

Since the demand for machines in (2.8) is iso-elastic and $\psi = 1-\beta$, the profit-maximizing monopoly price is

$$p^x(\nu,t|q) = 1. \tag{2.10}$$

Combining this with (2.8) implies

$$x(\nu,t|q) = qL. \tag{2.11}$$

Consequently, the flow profits of a firm with the monopoly rights on the machine of quality q

[10] Notice that given the Inada-condion (2.3) on ϕ, the incumbent which has recently been replaced always has incentives to innovate using the "tinkering" innovative technology to try to catch-up with the new entrant in a patent race a la (Aghion-Howitt). However, we can make appropriate assumptions so that it is always more profitable to invest in radical innovations. As a result, we can consider the incumbent as a potential entrant.

can be computed as:

$$\pi(\nu, t|q) = \beta q L. \tag{2.12}$$

Next, substituting (2.11) into (2.2), we obtain that total output is given by

$$Y(t) = \frac{1}{1-\beta} Q(t) L, \tag{2.13}$$

where

$$Q(t) = \int_0^1 q(\nu, t)\, d\nu \tag{2.14}$$

is the average total quality of machines and will be the only state variable in this economy. Since we have assumed that $q(\nu, 0) > 0\ \forall \nu$, (2.14) also implies $Q(0) > 0$ as the relevant initial condition of our economy.[11]

As a by product, we also obtain that the aggregate spending on machines is

$$X(t) = (1-\beta) Q(t) L. \tag{2.15}$$

Moreover, since the labor market is competitive, the wage rate at time t is

$$w(t) = \frac{\partial Y}{\partial L} = \frac{\beta}{1-\beta} Q(t). \tag{2.16}$$

To characterize the full equilibrium, we need to determine R&D effort levels by incumbents and entrants. To do this, let us write the net present value of a monopolist with the highest quality of machine q at time t in machine line ν:

$$V(v, t|q) = \mathbf{E}_t \left[\int_t^{T(\nu,t)} e^{-\int_t^{t+s} r(t+\tilde{s}) d\tilde{s}} \left(\pi(\nu, t+s|q) - z(\nu, t+s)\, q(t+s) \right) ds \right], \tag{2.17}$$

where the quality $q(\nu, t+s)$ follows a Poisson process that in each instant $q(\nu, t+s+ds) = \lambda q(\nu, t+s)$ with probability $\phi(z(\nu, t+s))\, ds$, and $T(\nu, t)$ is a stopping time where a new

[11] One might be worried about whether the average quality $Q(t)$ in (2.14) is well-defined, since we do not know how $q(\nu, t)$ will look like as a function of ν and the function $q(\cdot, t)$ may not be integrable. This is not a problem in the current context, however. Since the index ν has no intrinsic meaning, we can rank the ν's such that $\nu \mapsto q(\nu, t)$ is nondecreasing. Then the average in (2.14) exists when defined as a Lebesgue integral.

entrant enters into the sector ν. So if the R&D of the entrants into the sector is $\widehat{z}(\nu, t+s_1)$, then the distribution of $T(\nu, t)$ is

$$\Pr(T(\nu,t) \geq t+s) = \mathbf{E}_t\left[e^{-\int_0^s \widehat{z}(\nu,t+s_1)\eta(\widehat{z}(\nu,t+s_1))ds_1}\right].$$

Under optimal R&D choice of the incumbents, their value function $V(\nu, t|q)$ defined in (2.17) satisfies the standard Hamilton-Jacobi-Bellman equation:

$$\begin{aligned} r(t)V(\nu,t|q) - \dot{V}(\nu,t|q) = \max_{z(\nu,t)\geq 0} \{&\pi(\nu,t|q) - z(\nu,t)q(\nu,t) \\ &+\phi(z(\nu,t))(V(\nu,t|\lambda q) - V(\nu,t|q)) - \widehat{z}(\nu,t)\eta(\widehat{z}(\nu,t))V(\nu,t|q)\}. \end{aligned} \tag{2.18}$$

where $\widehat{z}(\nu,t)\eta(\widehat{z}(\nu,t))$ is the rate at which radical innovations by entrants occur in sector ν at time t and $\phi(z(\nu,t))$ is the rate at which the incumbent improves its technology. The first term in (3.7) is $\pi(\nu,t)$ flow of profit given by , while the second term is the expenditure of the incumbent for improving the quality of its machine. The second and third line include changes in the value of the incumbent due to innovation either by itself (at the rate $\phi(z(v,t))$, the quality of its product increases from q to λq) or by an entrant (at the rate $\widehat{z}(\nu,t)\eta(\widehat{z}(\nu,t))$, the incumbent is replaced and receives zero value from then on).[12] The value function is written with a maximum on the right hand side, since $z(\nu,t)$ is a choice variable for the incumbent.

Free entry by entrants implies that we must have:[13]

$$\begin{aligned} \eta(\widehat{z}(\nu,t))V(\nu,t|\kappa q(\nu,t)) &\leq q(\nu,t), \text{ and} \\ \eta(\widehat{z}(\nu,t))V(\nu,t|\kappa q(\nu,t)) &= q(\nu,t) \text{ if } \widehat{z}(\nu,t) > 0, \end{aligned} \tag{2.19}$$

which takes into account that by spending an amount $q(\nu,t)$, the entrant generates a flow rate of innovation of $\eta(\widehat{z})$, and if this innovation is successful (flow rate $\eta(\widehat{z}(\nu,t))$), then the

[12] The fact that the incumbent receives a zero value from then on follows from the assumption that a previous incumbent has no advantage relative to other entrants in competing for another round of innovations.

[13] Since there is a continuum of machines $\nu \in [0,1]$, all optimality conditions should be more formally stated as "for all $\nu \in [0,1]$ except subsets of $[0,1]$ of zero Lebesgue measure" or as "almost everywhere". We will not add this qualification to simplify the notation and the exposition.

entrant will end up with a machine of quality κq, thus earning the (net present discounted) value $V(\nu, t|\kappa q)$. The free entry condition is written in complementary slackness form, since it is possible that in equilibrium there will be no innovation by entrants.

Finally, maximization by the representative household implies the familiar Euler equation,

$$\frac{\dot{C}(t)}{C(t)} = \frac{r(t) - \rho}{\theta}, \tag{2.20}$$

and the transversality condition takes the form

$$\lim_{t \longrightarrow \infty} e^{-\int_0^t r(s)ds} \left[\int_0^1 V(\nu, t|q) \, d\nu \right] = 0. \tag{2.21}$$

This transversality condition follows because the total value of corporate assets is $\int_0^1 V(\nu, t|q) \, d\nu$. Even though the evolution of the quality of each machine is line is stochastic, the value of a machine of type ν of quality q at time t, $V(\nu, t|q)$ is non-stochastic. Either q is not the highest quality in this machine line, in which case the value function of the firm with a machine of quality q is 0, or alternatively, $V(\nu, t|q)$ is given by (2.17).

We summarize the conditions for an allocation to be an equilibrium in the following definition:

Definition 2.1 *Equilibrium is time paths of* $\{C(t), X(t), Z(t)\}_{t=0}^{\infty}$ *that satisfy (2.1), (2.7), (2.15) and (2.21); time paths for R&D expenditure by incumbents and entrants,* $\{z(\nu, t), \widehat{z}(\nu, t)\}_{\nu \in [0,1], t=0}^{\infty}$ *that satisfy (2.18) and (2.19); time paths of prices and quantities of each machine and the net present discounted value of profits,* $\{p^x(\nu, t|q), x(\nu, t|q), V(\nu, t|q)\}_{t=0}^{\infty}$ *given by (2.10), (2.11) and (2.17) or (2.18); and time paths of wage and interest rates,* $\{w(t), r(t)\}_{t=0}^{\infty}$ *that satisfy (2.16) and (2.20).*

In addition, we define BGP (balanced growth path) as an equilibrium path in which innovation, output and consumption grow at a constant rate. Notice that in BGP, aggregates grow at the constant rate, but there will be firm deaths and births, and the firm size distribution may change. We will discuss the firm size distribution in Section 2.4 and will refer to BGP equilibrium with a stationary (constant) distribution of normalized firm sizes as "a stationary BGP equilibrium". For now, we refer to allocation as a BGP regardless of whether the distribution of (normalized) firm sizes is stationary.

Definition 2.2 *A balanced growth path (hereafter BGP) is an equilibrium path in which innovation, output and consumption grow at a constant rate.*

2.2.3 Existence and Characterization of the BGP

The requirement that consumption grows at a constant rate in the BGP implies that $r(t) = r^*$, from (2.20). Moreover, in BGP, we must also have $z(\nu, t|q) = z(q)$ and $\widehat{z}(\nu, t|q) = \widehat{z}(q)$. These together imply that in BGP $\dot{V}(\nu, t|q) = 0$ and $V(\nu, t|q) = V(q)$. The following Proposition shows the existence of a linear BGP, in which the value function of incumbents is linear in the incumbents' product quality

Proposition 2.1 *Starting from any initial distribution of incumbent firms' product quality, there exists a unique linear BGP. Moreover, there are not any transitional dynamics. The economy jumps immediately on to a BGP.*

Indeed, from the optimal research decision of the incumbents:

$$rV(q) = \beta Lq + \max_{z} \phi(z)(V(\lambda q) - V(q)) - zq - \widehat{z}(q)\,\eta(\widehat{z}(q))\,V(q) \tag{2.22}$$

From the free-entry condition

$$\eta(\widehat{z}(q))\,V(\kappa q) = q$$

or

$$\widehat{z}(q) = \eta^{-1}\left(\frac{q}{V(\kappa q)}\right).$$

Since we focus on linear equilibria in which $V(q)$ is linear in q, we conjecture that

$$V(q) = vq$$

and look for v. Then

$$rv = \beta L + \max_{z} \phi(z)(\lambda - 1)v - z - \widehat{z}\eta(\widehat{z})\,v \tag{2.23}$$

and

$$\eta(\widehat{z})\,\kappa v = 1.$$

Let $z(v)$ denote $\arg\max_z \phi(z)(\lambda-1)v - z$, then $z(v)$ is strictly increasing in v given that $\phi(z)$ is strictly concave. And let $\widehat{z}(v)$ denote $\eta^{-1}\left(\frac{1}{\kappa v}\right)$ then $\widehat{z}(v)$ is strictly increasing in v as well given $\eta(z)$ is decreasing in z. Moreover, since $z\eta(z)$ is strictly increasing in z, $\widehat{z}(v)\,\eta(\widehat{z}(v))$ is strictly increasing in v.

From the Euler equation (2.20), we have

$$\begin{aligned}\frac{\dot{C}(t)}{C(t)} &= \frac{r-\rho}{\theta} \\ &= g,\end{aligned}$$

where g is the growth rate of consumption and output.

From (2.13) we have

$$\frac{\dot{Y}(t)}{Y(t)} = \frac{\dot{Q}(t)}{Q(t)}.$$

As noted above, in a BGP, for all machines, incumbents and entrants will undertake constant R&D z^* and $\widehat{z}^*$, respectively. Consequently, in a small interval of time Δt, there will be $\phi(z^*)\,\Delta t$ sectors that experience one innovation by the incumbent (increasing their productivity by λ) and $\widehat{z}^*\eta(\widehat{z}^*)\,\Delta t$ sectors that experience replacement by new entrants (increasing productivity by factor of κ). The probability that there will be two or more innovations of any kind within an interval of time Δt is $o(\Delta t)$. Therefore, we have

$$\begin{aligned}Q(t+\Delta t) &= (\lambda\phi(z^*)\,\Delta t)\,Q(t) + (\kappa\widehat{z}^*\eta(\widehat{z}^*)\,\Delta t)\,Q(t) \\ &\quad + (1-\phi(z^*)\,\Delta t - \widehat{z}^*\eta(\widehat{z}^*)\,\Delta t)\,Q(t) + o(\Delta t).\end{aligned}$$

Now substracting $Q(t)$ from both sides, dividing Δt and taking the limit as $\Delta t \longrightarrow 0$, we obtain

$$\frac{\dot{Q}(t)}{Q(t)} = (\lambda-1)\,\phi(z^*) + (\kappa-1)\,\widehat{z}^*\eta(\widehat{z}^*).$$

Thus

$$g = \phi(z^*)(\lambda - 1) + \widehat{z}^* \eta(\widehat{z}^*)(\kappa - 1). \tag{2.24}$$

We combine the equations (2.20), (2.23) and (2.24) and rearrange terms to establish the equation that determines v:

$$\beta L = \rho v + (\theta - 1)\phi(z(v))(\lambda - 1)v + z(v) + (\theta(\kappa - 1) + 1)\widehat{z}(v)\eta(\widehat{z}(v))v \tag{2.25}$$

If $\theta \geq 1$, since the right hand side is strictly increasing, it equals 0 at $v = 0$ and goes to $+\infty$ as v goes to $+\infty$. There exists a unique $v^* > 0$ such that the right hand side equals the left hand side. From the implied investment rate of the incumbents and entrants

$$\begin{aligned} z^* &= z^*(v^*) && (2.26) \\ \widehat{z}^* &= \widehat{z}^*(v^*). && (2.27) \end{aligned}$$

We can then recover the equilibrium growth rate

$$g^* = \phi(z^*)(\lambda - 1) + \widehat{z}^* \eta(\widehat{z}^*)(\kappa - 1) \tag{2.28}$$

and the equilibrium interest rate is determined from the consumer's Euler equation (2.20)

$$r^* = \rho + \theta g^*. \tag{2.29}$$

Lastly, we need to also verify that the transversality condition of the representative household, (2.21) is not violated. The condition for this is $r^* > g^*$ which is also satisfied if $\theta \geq 1$.

Another set of interesting implications of this model concern firm size dynamics. The size of a firm can be measured by its sales, which is equal to $x(\nu, t \mid q) = qL$ for all ν and t. We have seen that the quality of an incumbent firm increases at the flow rate $\phi(z^*)$, with z^* given by (2.26), while the firm is replaced at the flow rate $\hat{z}^* \eta(\hat{z}^*)$. Hence, for Δt sufficiently small, the stochastic process for the size of a particular firm is given by

$$x(\nu, t+\Delta t \mid q) = \begin{cases} \lambda x(\nu, t \mid q) & \text{with probability } \phi z^* \Delta t + o(\Delta t) \\ 0 & \text{with probability } \hat{z}^* \eta(\hat{z}^*) \Delta t + o(\Delta t) \\ x(\nu, t \mid q) & \text{with probability } (1 - \phi z^* \Delta t - \hat{z}^* \eta(\hat{z}^*) \Delta t) + o(\Delta t) \end{cases} \tag{2.30}$$

for all ν and t. Firms therefore have random growth, and surviving firms expand on average. However, firms also face a probability of bankruptcy (extinction). In particular, denoting the probability that a particular incumbent firm that started production in machine line ν at time s will be bankrupt by time $t \geq s$ by $P(t \mid s, \nu)$, we clearly have $\lim_{t\to\infty} P(t \mid s, \nu) = 1$, so that each firm will necessarily die eventually. The implications of equation (2.30) for the stationary firm size distribution will be discussed in Section 2.4. For now it suffices to say that this equation satisfies Gibrat's Law, which postulates that firm growth is independent of size (e.g., Sutton 1997, Gabaix 1999).[14]

2.2.4 Linear return to R&D by the incumbents

Consider the limiting case in which $\phi(z)$ is linear: $\phi(z) = \phi z$. We look for an interior equilibrium in which both incumbents and entrants undertake R&D.

Equation (2.22) implies

$$\phi(V(\lambda q) - V(q)) = 1, \tag{2.31}$$

otherwise the incumbents will undertake infinite amount of R&D or no R&D at all. Therefore, the value of an incumbent with quality q simplifies to

$$V(q) = \frac{q}{\phi(\lambda - 1)}. \tag{2.32}$$

Moreover, from the free entry condition (again holding as equality from the fact that the equilibrium is interior):

$$\eta(\hat{z}) V(\kappa q) = q.$$

[14] The most common form of Gibrat's Law involves firm sizes evolving according to the stochastic process $S_{t+1} = \gamma_t S_t + \varepsilon_t$, where γ_t is a random variable orthogonal to S_t and ε_t is another random variable with mean 0. The law of motion in (2.30) is a special case of this, with $\varepsilon_t \equiv 0$.

This equation implies that BGP R&D level by entrants $\widehat{z}^*$ is implicitly defined by

$$\widehat{z}(q) = \widehat{z}^* = \eta^{-1}\left(\frac{\phi(\lambda-1)}{\kappa}\right) \ \forall q > 0, \tag{2.33}$$

where η^{-1} is the inverse of the $\eta(z)$ function. Since $\eta(.)$ is strictly decreasing, so is $\eta^{-1}(.)$. In a linear BGP, the fact that $V(\nu, t|q) = vq \ \forall \nu, t$ and q together with (2.23) also implies

$$V(q) = \frac{\beta L q}{r^* + \widehat{z}\eta(\widehat{z})}. \tag{2.34}$$

Next, combining this equation with (2.32) we obtain the BGP interest rate as

$$r^* = \phi(\lambda - 1)\beta L - \widehat{z}^*\eta(\widehat{z}^*).$$

Therefore, the BGP growth rate of consumption and output is

$$g^* = \frac{1}{\theta}\left(\phi(\lambda-1)\beta L - \widehat{z}^*\eta(\widehat{z}^*) - \rho\right). \tag{2.35}$$

Equation (2.35) already has some interesting implications. In particular, it determines the relationship between the rate of innovation by entrant $\widehat{z}^*$ and the BGP growth rate g^*. In standard Schumpeterian models, this relationship is positive. In contrast, here we have:

Remark 2.1 *There is a negative relationship between $\widehat{z}^*$ and g^*.*

Proof. This follow immediately from (2.35) and the fact that $\widehat{z}\eta(\widehat{z})$ is strictly increasing in $\widehat{z}$. ∎

We will see in Subsection 2.3.1 that one of the implications of Remark 2.1 will be a positive relationship between entry barriers and growth.

2.2.5 Growth Decomposition

In this framework, we can calibrate how much of productivity growth is driven by creative destruction (innovation by entrants) and how much of it comes from productivity improvements by incumbents. To determine this, we use (2.24) which decomposes growth into the component coming from incumbent firms (the first term) and that coming from new entrants (the second term).

It would be informative to derive the quantitative implications of the model concerning the decomposition of productivity growth between incumbents and entrants. Unfortunately, however, some of the parameters of the current model are difficult to pin down with our current knowledge of the technology of R&D. Hence, instead of a careful calibration exercise, we will provide some suggestive numbers using plausible parameter values. The purpose of the exercise is to get a better sense for the range of values for the contribution of incumbents and entrants to innovation and productivity growth. We proceed as follow. First, we normalize population to $L = 1$ and choose the following standard numbers:

$$\begin{aligned} g^* &= 0.02 \\ \rho &= 0.01 \\ r^* &= 0.05 \\ \theta &= 2, \end{aligned}$$

where θ, the intertemporal elasticity of substitution, is pinned down by the choice of the other three numbers. The first three numbers refer to annual rates (implicitly defining $\Delta t = 1$ as one year). The remaining variables will be chosen so as to ensure that the equilibrium growth rate is indeed $g^* = 0.02$.

As a benchmark, we take

$$\beta = 2/3,$$

which implies that two thirds of national income accrues to labor and one third to profits. The requirement in (2.9) then implies that $\kappa > 1.7$. We will start with the benchmark value of $\kappa = 3$ so that entry by new firms is sufficiently "radical" as suggested by some of the qualitative accounts of the innovation process (e.g., Freeman, 1982, Scherer, 1984). Innovation by incumbents is taken to be correspondingly smaller

$$\lambda = 2$$

Parameter Values								$\hat{z}^*\eta(\hat{z}^*)$	$\phi(z^*)$	$\frac{(\kappa-1)\hat{z}^*\eta(\hat{z}^*)}{g^*}$
1.	$\kappa=3$	$\lambda=1.2$	$\beta=2/3$	$\alpha=0.9$	$\gamma=0.5$	$A=0.0977$	$B=0.0083$	0.0033	0.0667	0.333
2.	$\kappa=3$	$\lambda=1.2$	$\beta=2/3$	$\alpha=0.1$	$\gamma=0.5$	$A=0.3626$	$B=0.0094$	0.0033	0.0667	0.333
3.	$\kappa=3$	$\lambda=1.2$	$\beta=2/3$	$\alpha=0.1$	$\gamma=0.5$	$A=0.3500$	$B=0.0033$	0.0004	0.0958	0.0418
4.	$\kappa=4$	$\lambda=1.2$	$\beta=2/3$	$\alpha=0.1$	$\gamma=0.5$	$A=0.3500$	$B=0.0032$	0.0006	0.0909	0.0904
5.	$\kappa=2$	$\lambda=1.2$	$\beta=2/3$	$\alpha=0.1$	$\gamma=0.5$	$A=0.3500$	$B=0.0034$	0.0003	0.0983	0.0164

Table 2.1: Growth Decomposition

so that productivity gains from a radical innovation is about twice that of a standard "incremental" innovation by incumbents (i.e., $\frac{\kappa-1}{\lambda-1}=2$). We will then show how results change when the magnitudes of the radical and incremental innovations are varied. For the functions $\phi(z)$ and $\eta(z)$, we adopt the functional form in (2.5) and choose the benchmark values of $\alpha=0.1$ or 0.9 and $\gamma=0.5$. The remaining two parameters A and B will be chosen to ensure $g^*=0.02$ with two third coming from the innovations of the incumbents and one third coming from the entrants, i.e., the firm term in (2.24) $\phi(z^*)(\lambda-1)$ equals 0.0133 and the second term $\widehat{z}^*\eta(\widehat{z}^*)(\kappa-1)$ equals 0.0067. Given the value of κ, we obtain $\widehat{z}^*\eta(\widehat{z}^*)$ equals 0.0033 which implies that there is entry of a new firm (creative destruction) in each machine line on average once every 7.5 years (recall that $r^*=0.05$ as the annual interest rate so that $\frac{r^*}{\widehat{z}^*\eta(\widehat{z}^*)}\approx 7.46$). Similarly, we have $\phi(z^*)$ equals 0.0667, so that there are on average 1.2 incremental innovations per year by an incumbent in a particular machine line ($r^*/\phi(z^*)\approx 1.2$).

Table 2.1 shows how these numbers change as we vary the parameters β, κ, λ and α. The first five columns of the table give the choices of parameters. The sixth and seventh columns are the values of A and B that will lead to equilibrium growth $g^*=0.02$. The next two columns report the innovation rate by the entrants, $\widehat{z}^*\eta(\widehat{z}^*)$ and the incumbents, $\phi(z^*)$. The final column reports the fraction of total productivity growth accounted for by entrants, i.e., $\frac{(\kappa-1)\widehat{z}^*\eta(\widehat{z}^*)}{g^*}$.

2.3 The effects of Policy on Growth

2.3.1 Entry Barriers

Let now use this model to analyze the effects of different policies on equilibrium productivity growth and its decomposition between incumbents and entrants. Since the model has a Schumpeterian structure (with quality improvements as the engine of growth and creative destruction playing a major role), it may be conjectured that entry barriers (or taxes on potential entrants) will have the most negative effect on economic growth. To investigate whether this is the case, let us suppose that there is a tax (or an entry barrier) τ_e on R&D expenditures by entrants and a tax τ_i on R&D expenditure by incumbents. Tax revenues are not redistributed back to the representative household (for example, they finance an additive public good). Note also that τ_e, can be interpreted not only as a tax or an entry barrier, but also as a more strict patent policy. Nevertheless, to keep the analysis brief, we only focus on the case in which tax revenues are collected by the government rather than rebated back to incumbents as patent fees.

Repeating the analysis in Subsection 2.2.3 for the case of nonlinear return to R&D by the incumbents, we obtain the following equilibrium conditions

$$z_\tau(v) = \arg\max_z \phi(z)(\lambda - 1)v - (1 + \tau_i)z \tag{2.36}$$

and

$$\widehat{z}_\tau(v) = \eta^{-1}\left(\frac{1+\tau_e}{\kappa v}\right). \tag{2.37}$$

Plugging again in these two functions into (2.25) to obtain the equation that determines v

$$\begin{aligned} \beta L &= \rho v + (\theta - 1)\phi(z_\tau(v))(\lambda - 1)v + (1+\tau_i)z_\tau(v) \\ &\quad + (\theta(\kappa - 1) + 1)\widehat{z}_\tau(v)\eta(\widehat{z}_\tau(v))v \end{aligned} \tag{2.38}$$

In the case of linear return, we repeat the analysis in Subsection 2.2.4. We have:

$$\eta(\widehat{z}^*[\tau])V(\kappa q) = (1+\tau_e)q \text{ or } V(q) = \frac{q(1+\tau_e)}{\kappa\eta(\widehat{z}^*_\tau)}, \tag{2.39}$$

where $\widehat{z}^*[\tau]$ is explicitly conditioned on the vector of taxes, $\tau = (\tau_i, \tau_e)$. The equation that determines the optimal R&D decisions of incumbents, is also modified because of the tax rate τ_i and becomes

$$\phi(V(\lambda q) - V(q)) = (1+\tau_i)q. \tag{2.40}$$

Now combining (2.39) with (2.40), we obtain

$$\phi\left(\frac{(\lambda-1)(1+\tau_e)}{\kappa\eta(\widehat{z}^*[\tau])}\right)=1.$$

Consequently, the BGP R&D level by entrants $\widehat{z}^*[\tau]$, when their R&D is taxed at the rate τ_e, is given by

$$\widehat{z}^*[\tau]=\eta^{-1}\left(\frac{(\lambda-1)(1+\tau_e)}{\kappa(1+\tau_i)}\right). \tag{2.41}$$

Equation (2.34), in Subsection 2.2.4, which was derived from the value function (2.18) still applies, so that the BGP interest rate is $r^*[\tau]=(1+\tau_e)^{-1}\kappa\eta(\widehat{z}^*[\tau])\beta L-\widehat{z}^*[\tau]\eta(\widehat{z}^*[\tau])$, or substituting for (2.41),

$$r^*[\tau]=\frac{\phi(\lambda-1)\beta L}{1+\tau_i}-\widehat{z}^*[\tau]\eta(\widehat{z}^*[\tau]),$$

and the BGP growth rate is

$$g^*[\tau]=\frac{1}{\theta}\left(\frac{\phi(\lambda-1)\beta L}{1+\tau_i}-\widehat{z}^*[\tau]\eta(\widehat{z}^*[\tau])-\rho\right), \tag{2.42}$$

which $\widehat{z}^*[\tau]$ is given by (2.41). Armed with the equilibrium with introduction of entry barriers and R&D taxes, we can study the effect of those policies on the incentive of the entrants, incumbents and the aggregate growth. The following proposition summarizes the results:

Proposition 2.2 *An increase in R&D tax on incumbents increases the value of entrants, therefore induces higher entry. However, the disincentive effect of the tax on incumbents reduces their investment more than increases entry. As a result, aggregate growth is unambiguously decreasing in R&D tax on incumbents. Similarly, an increase in R&D tax or entry barrier on entrants increases the value of incumbents, therefore increases their R&D investment. The overall effect on entry and aggregate growth is in generally ambiguous. Only in the case of linear return, the growth rate of the economy is (strictly) increasing in the tax rate on entrants, i.e.,* $dg^*[\tau]/d\tau_e>0$.

Proof. To prove the proposition we will use the following results that can be obtained by applying the envelope theorem to (2.36):

$$\frac{\partial z_{\tau}(v)}{\partial \tau_i} = -z_{\tau}(v) \text{ and } \frac{\partial z_{\tau}(v)}{\partial v} = \phi(z_{\tau}(v))(\lambda - 1),$$

and the first order condition implies

$$\phi'(z_{\tau}(v))(\lambda - 1)v = 1 + \tau_i.$$

First, consider the case of an R&D tax on the incumbent, $\tau_i > 0$. The derivative of the right hand side of (2.38) with respect to τ_i is

$$\left((\theta - 1)\phi'(z_{\tau}(v))(\lambda - 1)v + 1\right)\frac{dz_{\tau}(v)}{d\tau_i} + z_{\tau}(v). \tag{2.43}$$

Using the fact that $\frac{dz_{\tau}(v)}{d\tau_i}$ equals $-z_{\tau}(v)$, (2.43) simplifies to

$$-(\theta - 1)\phi'(z_{\tau}(v))(\lambda - 1)vz_{\tau}(v) < 0.$$

This means the right hand side of (2.38) is also pushed downward as τ_i increases. This increase in v^* induces a higher level of entry given by (2.37). As we show next, the effect on $z_{\tau}(v)$ is negative enough to cancel the increasing entry. Indeed,

$$\begin{aligned}\frac{dz_{\tau}(v^*)}{d\tau_i} &= \frac{\partial z_{\tau}(v^*)}{\partial \tau_i} + \frac{\partial z_{\tau}(v^*)}{\partial v}\frac{dv^*}{d\tau_i} \\ &= -z_{\tau}(v^*) + \phi(z_{\tau}(v^*))(\lambda - 1)\frac{dv^*}{d\tau_i}.\end{aligned}$$

So

$$\begin{aligned}\frac{dg^*}{d\tau_i} &= (\lambda - 1)\phi'(z_{\tau}(v^*))\frac{dz_{\tau}(v^*)}{d\tau_i} \\ &\quad + (\kappa - 1)\frac{d}{dv}\left(\widehat{z}_{\tau}(v)\eta(\widehat{z}_{\tau}(v))\right)\frac{dv^*}{d\tau_i}.\end{aligned}$$

By the implicit function theorem

$$\frac{dv^*}{d\tau_i} = -\frac{\partial \Phi}{\partial \tau_i} / \frac{\partial \Phi}{\partial v}$$

in which

$$\Phi(\tau_i, v) = \rho v + (\theta - 1)\phi(z_\tau(v))(\lambda - 1)v + (1 + \tau_i)z_\tau(v) + (\theta(\kappa - 1) + 1)\widehat{z}_\tau(v)\eta(\widehat{z}_\tau(v))v.$$

Further detailed algebra in the Appendix shows that $\frac{dg^*}{d\tau_i}$ has the same sign as

$$\begin{aligned}
&-\frac{1+\tau_i}{v}\rho - \theta\phi(z_\tau(v))(\lambda - 1) - (1 + \tau_i)(\theta(\kappa - 1) + 1)\frac{d}{dv}(\widehat{z}_\tau(v)\eta(\widehat{z}_\tau(v))) \\
&-(1 + \tau_i)\widehat{z}_\tau(v)\eta(\widehat{z}_\tau(v)) - (\kappa - 1)\frac{d}{dv}(\widehat{z}_\tau(v)\eta(\widehat{z}_\tau(v))) < 0.
\end{aligned}$$

Second, consider the case of an entry barriers, $\tau_e > 0$. Plugging that tax in (2.37) and (2.38), the right hand side of (2.38) is pushed downward, therefore the equilibrium value of v increases. This increase in v^* induces a higher level of the investment from the incumbents since (2.36) implies $\frac{dz_\tau(v)}{dv} = \phi(z_\tau(v))(\lambda - 1) > 0$. This increase in v^* ,however, has opposite effects on the equilibrium value of $\widehat{z}$ since the direction of change of the ratio $\frac{1+\tau_e}{v^*}$ in (2.37) is ambiguous. The change in g^* is also ambiguous. In the case of linear return, we have a better estimates of the effect of policies on equilibrium growth. From (2.42), g^* does not directly depend on τ_e. Therefore,

$$\frac{dg^*[\tau]}{d\tau_e} = \frac{\partial g^*[\tau]}{\partial \widehat{z}^*}\frac{\partial \widehat{z}^*[\tau]}{\partial \tau_e}. \tag{2.44}$$

From Remark 2.1, $\frac{\partial g^*[\tau]}{\partial \widehat{z}^*} < 0$. (2.41) then shows that $\frac{\partial \widehat{z}^*[\tau]}{\partial \tau_e} < 0$, and thus $\frac{dg^*[\tau]}{d\tau_e} > 0$. With respect to τ_i, note that the indirect effect in (2.44) is now negative, since, from (2.41), $\frac{\partial \widehat{z}^*[\tau]}{\partial \tau_i} > 0$, and in addition τ_i is also has a negative effect on g^* as shown by (2.42). Therefore $\frac{dg^*[\tau]}{d\tau_i} < 0$. ■

The second result for the case of linear return is rather surprising (by continuity this results also holds for the case in which α is sufficiently small). In Schumpeterian models, making entry more difficult, either with entry barriers or by taxing R&D by entrants, has negative effects on economic growth. Despite the Schumpeterian nature of the current model, here blocking entry increases equilibrium growth. Moreover, since as we prove in Subsection 2.3.2 that there tends to be too much entry in the decentralized equilibrium, a tax on entry also tends to improve welfare in this model. The intuition for this result is related to the main departure of this model from the standard Schumpeterian models. The engine of growth is still quality improvements, but, in contrast to the textbook models, these are undertaken both by incumbents and entrants. Entry barriers and taxes on entrants, by protecting incumbents, increase their profitability and value, and greater value by incumbents encourages more R&D investments and faster productivity growth. Taxes on entrants or entry by barriers also further increase the contribution of incumbents to productivity growth.

Ambiguity of the effects of policies on aggregate growth comes when the return to R&D investment is sufficiently concave. In the case of the functional form (2.5), this corresponds to a sufficiently high α. The following example illustrates the results:

Example 2.1 *Consider the case in which the aggregate growth is* 2% *and the parameters* A, B *in* (2.5) *are chosen such that one third of growth comes from entrants and two third comes from incumbents. When* $\alpha = 0.1$*, an entry barrier equivalent to* 10% *R&D tax on entrants increases growth to* 2.004% *and when* $\alpha = 0.9$ *the barrier lowers growth to* 1.96%*.*

2.3.2 Pareto optimal allocation

We now briefly discuss the Pareto optimal allocation, which will maximize the utility of the representative household starting with some initial value of average quality of machines $Q(0) > 0$. As usual, we can think of this allocation as resulting from an optimal control problem by a social planer. There will be two differences between the decentralized equilibrium and the Pareto optimal allocation. The first is that the social planner will not charge a markup for machines. This will increase the value of machines and innovation to society. Second, the social planner will not respond to the same incentives in inducing entry (radical innovation). In particular, the social planner will not be affected by the "business stealing" effect, which makes entrants more aggressive because they wish to replace the current monopolist, and she will also internalize the negative externalities in radical research captured by the decreasing function η.

Let us first observe that the social planner will always "price" machines at marginal cost, thus in the Pareto optimal allocation, the quantities of machine used in final good production will be given by

$$\begin{aligned} x^S(v,t|q) &= \psi^{-\frac{1}{\beta}} qL \\ &= (1-\beta)^{-\frac{1}{\beta}} qL. \end{aligned}$$

Substituting this into (2.2), we obtain the amount of output in the Pareto optimal allocation as

$$Y^S(t) = (1-\beta)^{-\frac{1}{\beta}} Q^S(t) L,$$

where the superscript S refers to the social planner's allocation and $Q^S(t)$ is the average quality of machines at time t in this allocation. Part of this output will be spent on production machines and is thus not available for consumption or research. For this reason, it is useful to compute net output as

$$\begin{aligned} \overline{Y}^S(t) &= Y^S(t) - X^S(t) \\ &= (1-\beta)^{-\frac{1}{\beta}} Q^S(t) L - \psi(1-\beta)^{-\frac{1}{\beta}} Q^S(t) L \\ &= \beta(1-\beta)^{-\frac{1}{\beta}} Q^S(t) L. \end{aligned}$$

Given the specification of the innovation possibilities frontier above consisting of radical and incremental innovations, the evolution of average quality of machines is

$$\frac{\dot{Q}^S(t)}{Q^S(t)} = (\lambda - 1)\phi z^S(t) + (\kappa - 1)\hat{z}^S(t)\eta\left(\hat{z}^S(t)\right), \tag{2.45}$$

where $z^S(t)$ is the average rate of incumbent R&D and $\hat{z}^S(t)$ is the rate of entrant R&D chosen by the social planner. The total cost of R&D to the society is $\left(\phi z^S(t) + \hat{z}^S(t)\eta\left(\hat{z}^S(t)\right)\right) Q^S(t)$.[15]

The maximization of the social planner can then be written as

$$\max \int_0^\infty e^{-\rho t} \frac{C^S(t)^{1-\theta} - 1}{1-\theta} dt$$

[15]We assume here that the social planner invests into each sector proportionately to its highest quality. This can be proved using the convexity of the social planner maximization problem.

subject to (2.45) and the resource constraint, which can be written as

$$C^S(t) + \left(z^S(t) + \hat{z}^S(t)\right) Q^S(t) \leq \beta(1-\beta)^{-\frac{1}{\beta}} Q^S(t) L.$$

As we derive in the Appendix, the two equations that determine z^S and $\hat{z}^S$ are:

$$\begin{aligned} & \theta\left((\lambda-1)\phi\left(z^S\right) + (\kappa-1)\hat{z}^S\eta\left(\hat{z}^S\right)\right) + \rho \\ = & \left(\beta(1-\beta)^{-\frac{1}{\beta}} L - \left(z^S + \hat{z}^S\right)\right)(\lambda-1)\phi'\left(z^S\right) \\ & + \left((\lambda-1)\phi\left(z^S\right) + (\kappa-1)\hat{z}^S\eta\left(\hat{z}^S\right)\right) \end{aligned} \tag{2.46}$$

and

$$(\lambda-1)\phi'\left(z^S\right) = (\kappa-1)\left(\eta\left(\hat{z}^S\right) + \hat{z}^S\eta'\left(\hat{z}^S\right)\right). \tag{2.47}$$

In the case of linear return, (2.47) becomes

$$(\lambda-1)\phi = (\kappa-1)\left(\eta\left(\hat{z}^S\right) + \hat{z}^S\eta'\left(\hat{z}^S\right)\right) \tag{2.48}$$

that determines. Then (2.46) becomes

$$\begin{aligned} & \theta\left((\lambda-1)\phi z^S + (\kappa-1)\hat{z}^S\eta\left(\hat{z}^S\right)\right) + \rho \\ = & \left(\beta(1-\beta)^{-\frac{1}{\beta}} L - \hat{z}^S\right)(\lambda-1)\phi + (\kappa-1)\hat{z}^S\eta\left(\hat{z}^S\right) \end{aligned} \tag{2.49}$$

that determines z^S and

$$\begin{aligned} g^S &= (\lambda-1)\phi z^S + (\kappa-1)\hat{z}^S\eta\left(\hat{z}^S\right) \\ &= \frac{\left(\beta(1-\beta)^{-\frac{1}{\beta}} L - \hat{z}^S\right)(\lambda-1)\phi + (\kappa-1)\hat{z}^S\eta\left(\hat{z}^S\right) - \rho}{\theta}. \end{aligned} \tag{2.50}$$

Equations (2.47) and (2.48) shows that the trade-off between radical and incremental innovations for the social planner is different because she internalizes the negative effect that one more unit of R&D creates on the success probability of other firms performing radical R&D on the same machine line. This is reflected by the negative term $\hat{z}^S\eta'\left(\hat{z}^S\right)$ on the right-hand side of (2.47) and (2.48). This effect implies that the social planner will tend to do more incremental innovations than the decentralized equilibrium. Since z^S and $\hat{z}^S$ are constant, consumption growth rate is also constant in the optimal allocation (thus no transitional dynamics). This Pareto optimal consumption growth rate can not be directly compared to the equilibrium BGP growth rate, g^*, because there are two counteracting effects. One the one hand, the social planner uses machines more intensively (because she avoids the monopoly distortions), and this tends to increase g^S above g^* given in (2.28) or (2.35) (this can be seen by the fact that the first term in g^S, $(1-\beta)^{-\frac{1}{\beta}} L(\lambda-1)\phi$ is strictly greater than the first term in g^* in (2.35), since $(1-\beta)^{-\frac{1}{\beta}} > 1$). This same effect can also encourage radical R&D. On the other hand, the social planner also has a reason for choosing a lower rate of radical R&D because she internalizes the negative R&D externalities and the business stealing effect. One can construct examples in which the growth rate of the Pareto optimal allocation is greater or less than that of the decentralized equilibrium (though only in exceptional cases is the equilibrium rate of Pareto optimal allocation smaller than that of the decentralized equilibrium). The following proposition illustrates the intuition:

Proposition 2.3 *When return to R&D of the incumbents is linear then the growth rate of the Pareto optimal allocation is always greater than that of the decentralized equilibrium. However, generally we can choose parameters such that the Pareto optimal growth rate is strictly less the equilibrium growth rate.*

Proof. Replacing $(\lambda-1)\phi$ from (2.48) into (2.50) and simplify, we obtain

$$\begin{aligned} g^S &= \frac{\beta(1-\beta)^{-\frac{1}{\beta}} L(\lambda-1)\phi - (\kappa-1)\left(\hat{z}^S\right)^2 \eta'\left(\hat{z}^S\right) - \rho}{\theta} \\ &> \frac{\beta L(\lambda-1)\phi-\rho}{\theta} = g^*, \end{aligned}$$

in which the growth rate in the decentralized equilibrium, g^*, is defined in (2.35). Notice also that (2.48) implies

$$\eta\left(\hat{z}^S\right) > \frac{(\lambda-1)\phi}{\kappa-1} = \eta\left(\hat{z}^*\right).$$

Therefore, $\hat{z}^* > \hat{z}^S$, i.e., entry is too high in the decentralized equilibrium compared to the Pareto optimal level of entry.

■

We choose a set of parameters to show that the result does not hold generally.

Example 2.2 *Suppose we start with a set of parameters in which the innovations from the entrants contribute significantly more to the aggregate growth than that of the incumbents do. Since the social planner internalizes the business stealing effect of entry on incumbents' innovation incentive, he will invest less into radical innovations. However, given the large share of radicals innovation into growth, Pareto optimal growth rate is smaller than the competitive equilibrium growth rate.*

$$\begin{aligned} A &= 0.003 \\ B &= 0.012 \\ \alpha &= 0.9 \\ \gamma &= 0.9 \end{aligned}$$

Then $g^* = 0.02$ *with* 90% *contributions from entrants. In the Pareto allocation,* $g^S = 0.018$.

2.4 Stationary BGP equilibrium

We discussed at the end of Subsection 2.2.3 that the evolution of firm size follows the Gibrat's law, i.e., firm growth is independent of size. This process potentially gives rise to an equilibrium size distribution following the Zipf's distribution, as in Gabaix (1999). However, we show in Subsection 2.4.1 below that in the original model, a stationary distribution does not exist. In particular, as time goes, all firms have approximately zero size relative to the average size and a vanishingly small fraction of firms become arbitrarily large.

Subsection 2.4.2 shows that if we introduce one more type of entry, *entry by imitation*, which allows potential entrants at any moment of time to pay some cost to copy a technology with quality proportional to the current average quality in the economy. The entrants will then enter and replace any entrants with product quality relative to the average quality falling below a threshold. We thus impose a minimum size of existing firms in the economy. As a result, a stationary firm size distribution exists. We also show that by choosing appropriate parameter of the imitation technology, the equilibrium in this modified economy is "closed" to the equilibrium in the original economy.

2.4.1 Equilibrium firm size dispersion

The model generates a dynamic equilibrium in which the economy, and thus the size of average firm, as measure by sales $x(t)$, grows, but does so stochastically. Therefore, the equilibrium also generates a firm size distribution. An interesting question is whether this firm size distribution resembles the empirical distributions. To investigate this question, we first need to normalize firm sizes by the average size of firms in the economy[16], $X_1(t) = \int_0^1 x(v,t)\,dt$, given in (2.6) - so that the equilibrium has a possibility of generating a stationary distribution. In particular, let the normalized firm size be

$$\widetilde{x}(t) = \frac{x(t)}{X_1(t)}.$$

Notice that

$$\widetilde{x}(t) = \widetilde{q}(t) = \frac{q(t)}{Q(t)},$$

the product quality relative to the average product quality. We would like to determine the equilibrium law of motion of the normalized firm size $\widetilde{x}(t)$ and its stationary distribution. Since in equilibrium $\frac{\dot{X}_1(t)}{X_1(t)} = g^* > 0$, for Δt sufficiently small the law of motion for the normalized size of the leading firm in each industry can be written as

$$\widetilde{x}(t+\Delta t) = \begin{cases} \frac{\lambda}{1+g^*\Delta t}\widetilde{x}(t) & \text{with probability } \phi(z^*)\,\Delta t \\ \frac{\kappa}{1+g^*\Delta t}\widetilde{x}(t) & \text{with probability } \widehat{z}^*\eta(\widehat{z}^*)\,\Delta t \\ \frac{1}{1+g^*\Delta t}\widetilde{x}(t) & \text{with probability } 1-\phi(z^*)\,\Delta t - \widehat{z}^*\eta(\widehat{z}^*)\,\Delta t. \end{cases} \tag{2.51}$$

Notice that this expression does not refer to the growth rate of a single firm, but to the leading firm in a representative industry. In particular, when there is entry, this leads to an increase in size rather than extinction.

The following proposition shows that if a stationary distribution of (normalized) firm sizes exists, then it must take the form of the Pareto distribution with an exponent equal to 1. Recall that the Pareto distribution takes the form $\Pr[\widetilde{x} \le y] = 1 - \Gamma y^{-\chi}$ with $\Gamma > 0$ and $y \ge \Gamma$.

[16] Another way to interpret this normalization is to consider the sales relative to labor wage $w(t)$ which is proportional to $Q(t)$ and $X(t)$.

Proposition 2.4 *In the economy studied here, if a stationary distribution of (normalized) firm sizes exists, then it is a Pareto distribution with exponent equal to* 1*. However, in the economy studied here, a stationary distribution does not exist.*

Proof. The equation that determines the stationary distribution with cdf $F(y)$ (derived in the Appendix) is

$$0 = F_y(y)\, yg^* - \phi(z^*)\left(F(y) - F\left(\frac{y}{\lambda}\right)\right) - \widehat{z}^*\eta(\widehat{z}^*)\left(F(y) - F\left(\frac{y}{\kappa}\right)\right) \tag{2.52}$$

This yields $F(y) = 1 - \left(\frac{\Gamma}{y}\right)^{\chi}$, plugging in this distribution into (2.52) to obtain

$$\phi(z^*)(\lambda^{\chi} - 1) + \widehat{z}^*\eta(\widehat{z}^*)(\kappa^{\chi} - 1) - g^*\chi = 0.$$

However, given $\frac{\lambda^{\chi}-1}{\chi}$ and $\frac{\kappa^{\chi}-1}{\chi}$ are strictly increasing in χ and by definition of g^* we have equality at $\chi = 1$. Therefore $\chi = 1$ is the unique solution. In some ways, this result looks quite remarkable, since it generates a stationary firm size distribution given by a Pareto distribution with an exponent of one, in a much simpler manner than any existing approaches, and does so despite the fact that the model was not designed to study firm size distribution. Unfortunately, we can show a stationary distribution does not exist. Indeed, we have just shown that if a stationary distribution exists, it must take the form $\Pr[\widetilde{x} \leq y] = 1 - \frac{\Gamma}{y}$ with $\Gamma > 0$. But the Pareto distribution is only defined for all $y \geq \Gamma$, thus Γ should be the minimum normalized firm size. However (2.51) shows that it is possible for the normalized size of a firm $\widetilde{x}$ to tend to 0. Therefore Γ should be equal to 0, which implies that there does not exist a stationary firm size distribution. ■

The essence of Proposition 2.4 is that with the random growth process in (2.51), the distribution of firm sizes will continuously expand. The nature of the "limiting distribution" is therefore similar to the immiserization result for income distribution in Atkeson and Lucas's (1992) economy with dynamic hidden information; in the limit, all firms have approximately zero size relative to the average $X_1(t)$ and a vanishingly small fraction of firms become arbitrarily large (so that average firm size $X_1(t)$ remains large and continues to grow).

2.4.2 Epsilon Economy

A way to avoid the dispersion of the firm size distribution is to introduce a lower bound on firm size. To do this, let us introduce a third type of innovation, "imitation". A new firm can enter in sector $\nu \in [0,1]$ with a technology $q^e(\nu,t) = \omega Q(t)$, where $\omega \geq 0$ and $Q(t)$ is average quality of machines in the economy given by (2.14) - after entry, the firm can engage in incremental innovations as usual. The cost of this type of innovation is assumed to be $\mu_e \omega Q(t)$. The fact that the cost should be proportional to average quality is in line with the structure of the model so far. We call the economy with this "imitation" technology an *epsilon economy.*[17]

Given this cost of imitation, if a firm could enter into a particular sector, become the monopolist and obtain the BGP value (2.17) ($T(\nu,t)$ is now the stopping time where either a entrant or an imitator enters and replaces the monopolist), it would be happy to do so. More precisely, it would be indifferent between entering and not entering, and we suppose that it chooses to enter depending on the quality of the incumbent in the sector. However, since there is already an incumbent in the industry, the entrant may not be able to charge the monopoly price and may be forced to charge a limit price. Even if the entrant had to charge a limited price for a short time interval, its value would strictly less than that implied by (2.17), and entry by paying the cost $\mu_e \omega Q(t)$ would not be profitable. Recall from condition (2.9) that the higher-quality firm can charge the unconstrained monopoly price when its quality is greater than a fraction $(1-\beta)^{-(1-\beta)/\beta}$ of the next highest quality, and otherwise, it will be forced to charge a limit price (upon entry). This reasoning then implies that entry by imitating the average technology is profitable in machine line ν at time t only when

$$q(\nu,t) \leq \omega (1-\beta)^{(1-\beta)/\beta} Q(t).$$

Given the imitators are indifferent between entering and not entering, we will assume that there exists $\epsilon \leq \omega (1-\beta)^{(1-\beta)/\beta}$ such that the imitators enter into a sector ν if and only if $q(\nu,t) \leq \epsilon Q(t)$. This implies that there will be no firm with quality $q(\nu,t) < \epsilon Q(t)$, because

[17] *This imitation technology captures the knowledge spillover channel as in Romer (1990). However, there are other ways to introduce the epsilon economy. For example, in a companion paper we consider the case in which each firm has to pay an epsilon-fixed labor cost. Such an economy is similar to the one in Luttmer (2007) except firm growth is exogeneous in his paper and endogeneous in our paper. What matters is that the cost is proportional to the aggregate quality.*

they will be immediately replaced by imitators (the case where $\omega = \epsilon = 0$ is identical to the economy we studied so far). The problem with this modified model is that the possibility that there will be another type of entry, as a function of the gap between current and average quality, will affect both the value function of firms and their incremental innovation decisions. However as $\omega \longrightarrow 0$, the value function and the innovation decisions converge to those characterized in Subsection 2.2.3 and so does the equilibrium. Therefore for ω arbitrarily small, the equilibrium characterized in Section 2.2.3 provides an arbitrarily close approximation to the equilibrium of the economy with $\omega > 0$. However, once we have that $q(\nu, t) \geq \epsilon Q(t)$ for all ν and t, we obtain the result that a stationary firm size distribution exists and takes the form of an asymptotic Pareto distribution.

To prove the existence of a balanced growth path equilibrium of the epsilon economy. We need the following assumption on the technology for radical innovation:

Assumption 2.1 $\max_{z>0} \epsilon_\eta(z) \leq 1 - \frac{1}{\kappa^\theta}$, *where* $\epsilon_\eta(z) = -\frac{z\eta'(z)}{\eta(z)}$ *is the elasticity of the entry function* $\eta(z)$.

Remark 2.2 *Under functional form* (2.5), *Assumption 2.1 is equivalent to* $\eta \leq 1 - \kappa^{-\theta}$ *or the entry function* $\eta(z)$ *is sufficiently inelastic. This assumption is used in Lemma 2.2 in the Appendix to ensure boundedness and some limit behaviors of the value function of incumbent firms under the presence of the two types of entry. The condition is imposed on the elasticity of* η *because it is the source of non-monotonicity of the value of the incumbent firms. For example, high value to the incumbent firms will attract more entry by radical innovation which in turn depress present value to the incumbents. Lastly, when there is no entry by radical innovation, this assumption is automatically satisfied.*

If Assumption 2.1 is satisfied, we have the following theorem describing the epsilon economy:

Theorem 2.1 (Existence of the Epsilon Economy) *Suppose the equilibrium in the benchmark economy* r^*, v^*, g^* *in Subsection 2.2.3 and Assumption 2.1 is satisfied . There exists an interval* $(\underline{\mu}, \overline{\mu})$ *and* $\Delta > 0$, $\overline{\omega} > 0$ *such that given* $\mu_e \in (\underline{\mu}, \overline{\mu})$ *and for each* $\omega \in (0, \overline{\omega})$, *there is a BGP with the following properties :*

1) An imitator pays $\mu_e \omega Q(t)$ *to buy a product quality* $\omega Q(t)$ *and to enter into a sector* ν *if*

$q(\nu,t) \leq \epsilon(\omega) Q(t)$, *where* $0 < \epsilon(\omega) \leq \omega(1-\beta)^{\frac{1-\beta}{\beta}}$ *.The imitator can charge unlimited monopolistic price and it replaces the incumbent in the sector.*[18]

2) The equilibrium growth rate of the aggregate product quality is $g(\omega) \in (g^*, g^* + \Delta)$ *which satisfies*

$$\lim_{\omega \longrightarrow 0} g(\omega) = g^*.$$

This economy admits a stationary distribution of normalized firm size with the cdf $f(.)$ which is approximately Pareto. Formally:

Theorem 2.2 (Tail Index) *The stationary distribution has an approximate Pareto tail with the Pareto exponent* $\chi = \chi(\omega) > 1$ *such that:* $\forall \xi > 0$ *there exist* $\overline{B}$,$\underline{B}$ *and* y_0 *such that*

$$f(y) < 2\overline{B} y^{-(\chi-1-\xi)}, \forall y \geq y_0$$

and

$$f(y) > \frac{1}{2}\underline{B} y^{-(\chi-1+\xi)}, \forall y \geq y_0.$$

In other words, $f(y) = y^{-\chi-1}\varphi(y)$, *where* $\varphi(y)$ *is a slow-varying function. Moreover*

$$\lim_{\omega \longrightarrow 0} \chi(\omega) = 1.$$

Proof. This is a direct consequence of the Lemma 2.7 in the Appendix. ■

Remark 2.3 *This theorem regarding the stationary distribution differ from the literature on city and firm size distribution, for example Gabaix (1999) in two respects. First, Pareto tail is obtained using Gibrat's law which is assumed exogenous in his paper, but is a result of endogenous investment decisions of firms given that their innovation cost is proportional to the quality of their current technologies. Second, the endogenous growth rate of the economy pushes the Pareto tail towards one, which is the Zipf's law.*[19]

[18] *The incumbents and innovative entrants solve the net present value maximization problem as in (2.18), but they take into account the behavior of the imitators.*

[19] *See Edward Glaeser's comment http://economix.blogs.nytimes.com/2010/04/20/a-tale-of-many-cities/*

Sketch proof of Theorem 2.1. For $\mu_e \in (\underline{\mu}, \overline{\mu})$ and $\omega \in (0, \overline{\omega})$ we show that there exists an equilibrium growth rate function g^* that satisfies the condition of BGP in three steps:
Step 1: For each $g \in (g^*, g^* + \Delta)$, we show the existence of a value function of an incumbent in sector ν at time t under the form

$$V_g(\nu, t|q) = Q(t)\,\widehat{V}_g\left(\frac{q(\nu,t)}{Q(t)}\right), \tag{2.53}$$

and a threshold $\epsilon_g(\omega)$ that an imitator will pay the cost $\mu_e \omega Q_t$ to buy a technology with quality ωQ_t to enter into sector ν at time t and replace the incumbent if $q(\nu,t) \leq \epsilon_g(\omega)\,Q(t)$.[20] Value of the incumbent depends only on the current average quality, $Q(t)$, and the gap between the current quality and the average quality. Plugging (2.53) in (2.18) and notice that

$$\dot{V}_g(\nu, t|q) = gQ(t)\,\widehat{V}_g\left(\frac{q(\nu,t)}{Q(t)}\right) - gQ(t)\,\widehat{V}_g'\left(\frac{q(\nu,t)}{Q(t)}\right)$$

we have

$$\begin{aligned}(r-g)\,\widehat{V}_g(\widetilde{q}) - g\widehat{V}_g'(\widetilde{q}) &= \beta L\widetilde{q} + \max_{z(\nu,t)}\left\{\phi(z(\nu,t))\left(\widehat{V}_g(\lambda\widetilde{q}) - \widehat{V}_g(\widetilde{q})\right) - z(\nu,t)\,\widetilde{q}\right\} \\ &\quad -\widehat{z}(\nu,t)\,\eta(\widehat{z}(\nu,t))\,\widehat{V}_g(\widetilde{q}),\end{aligned} \tag{2.54}$$

in which we have $r = \rho + \theta g$ and $\widetilde{q}(\nu,t) = \frac{q(\nu,t)}{Q(t)}$. Free-entry condition (2.19) becomes

$$\eta(\widehat{z}(\nu,t))\,\widehat{V}_g(\kappa\widetilde{q}(\nu,t)) = \widetilde{q}(\nu,t).$$

Moreover, the free-entry condition of the imitators implies

$$\widehat{V}_g(\omega) = \mu_e \omega. \tag{2.55}$$

Given the imitators will replace the incumbent in sector ν at time t if $q(\nu,t) \leq \epsilon_g Q(t)$, we also have

$$\widehat{V}_g(\epsilon_g) = 0. \tag{2.56}$$

[20] Again, due to the Arrow replacement effects, the incumbent firm will never purchase this imitation technology.

We will show that there exists a solution $V_g(\widetilde{q})$ to the functional equation (2.54) with the pointwise condition (2.55) and (2.56). We will assume that μ_e belong to certain interval $(\underline{\mu}, \overline{\mu})$ such that the imitators enter in equilibrium but they do not enter too early, and when they enter they can charge unlimited monopoly price, i.e., $\epsilon \leq \omega(1-\beta)^{\frac{1-\beta}{\beta}}$.Moreover $V_g(\widetilde{q})$ is equicontinuous with respect to g. The functional equation (2.54) also implies the investment decision of the incumbents $z_g(\widetilde{q})$ and of the entry rate of the innovative entrants $\widehat{z}_g(\widetilde{q})$.[21]

Step 2: These investment decisions together with the entry rule of the imitators and the growth rate g of the average quality yields a stationary distribution over the normalized sizes $\widetilde{q}$ with the cdf and pdf $F(.)$ satisfying:

If $y \geq \omega$

$$0 = F_y(y)\, yg - \int_{\frac{y}{\lambda}}^{y} \phi(z(\widetilde{q}))\, dF(\widetilde{q}) - \int_{\frac{y}{\kappa}}^{y} \widehat{z}(\widetilde{q})\, \eta(\widehat{z}(\widetilde{q}))\, dF(\widetilde{q}). \tag{2.57}$$

If $y < \omega$

$$0 = F_y(y)\, yg - F_y(\epsilon)\, \epsilon g - \int_{\frac{y}{\lambda}}^{y} \phi(z(\widetilde{q}))\, dF(\widetilde{q}) - \int_{\frac{y}{\kappa}}^{y} \widehat{z}(\widetilde{q})\, \eta(\widehat{z}(\widetilde{q}))\, dF(\widetilde{q}). \tag{2.58}$$

Let F_g denote such a distribution.

The intuition for (2.57) and (2.58) is the following: Given $y > 0$ the mass of firms with size jumping out of (ϵ, y) consists of firms with size between $\left(\frac{y}{\lambda}, y\right)$ and experience tinkering innovation, $\int_{\frac{y}{\lambda}}^{y} \phi(z(\widetilde{q}))\, dF(\widetilde{q})$, and firms with size between $\left(\frac{y}{\kappa}, y\right)$ and experience radical innovation, $\int_{\frac{y}{\kappa}}^{y} \widehat{z}(\widetilde{q})\, \eta(\widehat{z}(\widetilde{q}))\, dF(\widetilde{q})$. When $y < \omega$, we must also add the mass of firms being replaced by imitators with relative quality ω. This mass consists of firms around ϵ and do not experience any innovation, therefore are drifted under ϵ due to the growth rate g of the average quality Q, which is $F_y(\epsilon)\, \epsilon g$. Because of the stationarity of the distribution, this mass of firms must be equal to the mass of firms going into the interval (ϵ, y). This mass consists of firms around y and do not experience any innovation, therefore are drifted inside due to the growth rate g of the average quality Q, which is $F_y(y)\, yg$.

[21] There are two difficulties associated with proving the existence of the value function. The first one is that this is a differential equation with deviating arguments given that the right hand-side depends on value of V evaluated at λq and κq. As a result, we cannot apply standard existence proofs used for ordinary differential equations. We use here instead bound function techniques used in monotone iterative solution methods, see for example Jankowski (2005). The second difficulty is that we want to show that the value function statisfies some properties at infinity. This non-standard boundary problem is solved as in Staikos and Tsamatos (1981).

Step 3: We obtain an implied growth rate of the average product quality $g' = \frac{\dot{Q}_t}{Q_t}$ from the investment decision $z_g(\widetilde{q})$, and $\widehat{z}_g(\widetilde{q})$, imitation decision ϵ_g and stationary distribution F_g.

$$g' = \frac{(\lambda - 1)\,\mathbf{E}_{F_g}\left[\phi\left(z\left(\widetilde{q}\right)\right)\widetilde{q}\right] + (\kappa - 1)\,\mathbf{E}_{F_g}\left[\widehat{z}_g\left(\widetilde{q}\right)\eta\left(\widehat{z}_g\left(\widetilde{q}\right)\right)\widetilde{q}\right]}{1 - \epsilon_g F_g'\left(\epsilon_g\right)\left(\omega - \epsilon_g\right).}. \tag{2.59}$$

This formula is similar to the decomposition of growth in (2.24). In particular the nominator consists of innovation from the incumbents

$$(\lambda - 1)\,\mathbf{E}_{F_g}\left[\phi\left(z\left(\widetilde{q}\right)\right)\widetilde{q}\right]$$

and entrants

$$(\kappa - 1)\,\mathbf{E}_{F_g}\left[\widehat{z}_g\left(\widetilde{q}\right)\eta\left(\widehat{z}_g\left(\widetilde{q}\right)\right)\widetilde{q}\right].$$

The denominator shows the contribution of imitation to growth. The higher the gap $\omega - \epsilon_g$ is, the more important this component. Finally the equilibrium growth rate $g^*(\omega)$ is solution of the equation

$$D(g) = g' - g = 0.$$

Notice that, by using Theorem 2.2, as g goes to g^*, the asymptotic tail index of the quality distribution goes to 1, as a result the mean quality goes to infinity. Therefore (2.59) shows that

$$\lim_{g \longrightarrow g^*} D(g) = +\infty.$$

It remains to show that there exists some Δ such that $D(g^* + \Delta) < 0$ and that $D(g)$ is continuous to show the existence of $g^*(\omega)$.

The details of these steps are given in Appendix. ∎

2.4.3 Simulations

In addition to the parameters in the first row of Table 2.1 in the calibration part. We pick the following parameters for the epsilon economy:

$$\mu_e = 15$$

$$\omega = 0.1.$$

The resulting growth rate is $g_q(\omega) = 0.0205 > g_q^* = 0.02$. The exit threshold for incumbent firm is $\varepsilon(\omega) = 0.045 < \omega(1-\beta)^{\frac{1-\beta}{\beta}}$, therefore the imitators can charge unlimited monopolist price.

Then, we have the following equilibrium tail of the station distribution

$$\gamma(\omega) = 1.12$$

The resulting rank distribution is:

$$G(\widehat{q}) = \int_{\widehat{q}}^{\infty} f(\widehat{q}_1) \, d\widehat{q}_1.$$

Figure 2-1 represents the following relationship, similar to the one in Gabaix (1999)

$$\log(rank) = C - \gamma \log(size)$$

Figure 2-2 presents the value functions of the incumbents in a sector with the imitation threat (solid line) and without (dashed line). Under entry by imitation, the value of the incumbents is zero if $\widehat{q}_t = \frac{q_t}{Q_t} \leq \varepsilon$. We see here that the value of incumbent firms without the imitation threat is greater than it is with the threat. However this might not be true in general, given that entry by radical innovation will react to the value of the incumbents. Figure 2-3 presents the contributions of the incumbents and the entrants to the aggregate growth of product quality. About two-third of the aggregate growth is attributed to the incumbents. Notice that, the incumbents with lower quality invest more because of the threat of entry by imitation. This threat also makes the innovative entry less profitable.

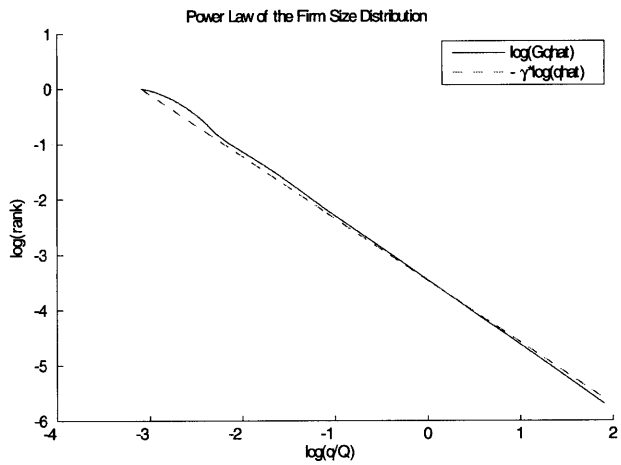


Figure 2-1: Stationary Distribution of Firm Size

2.5 Conclusion

A large fraction of US industry-level productivity growth is accounted for by existing firms and continuing establishments. Standard growth models either predict that most growth should be driven by new innovations brought about by entrants (and creative destruction) or do not provide a framework for decomposing the contribution of incumbents and entrants to productivity growth. In this paper, I proposed a simple modification of the basic Schumpeterian endogenous growth models that can address these questions. The main departure from the standard models is that incumbents have access to a technology to undertake incremental innovations and improve their existing machines (products). A different technology can then be used to generate more radical innovations. Arrow's replacement effect implies that only entrants will undertake R&D for radical innovations, while incumbents will invest in incremental innovations. This general pattern is in line with qualitative and quantitative evidence on the nature of innovation.

The model is not only consistent with the broad evidence but also provides a tractable framework for the analysis of productivity growth and of the entry of new firms and the expansion of existing firms. It yields a simple equation that decomposes productivity growth between continuing establishments and new entrants. Although the parameters to compute the

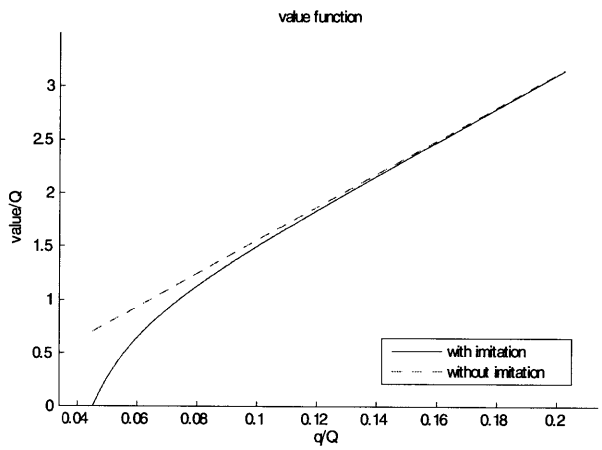


Figure 2-2: Value Function

exact contribution of different types of establishments to productivity growth have not yet been estimated, the use of plausible parameter values suggests that, in contrast to basic endogenous technological change models and consistent with the US data, a large fraction—but not all—of productivity growth is accounted by continuing establishments.

The comparative static results of this model are also very different from those of existing growth models. Most importantly, despite the presence of entry and creative destruction, the model shows that entry barriers or taxes on potential entrants increase the equilibrium growth rate of the economy. This is because, in addition to their direct negative effects, such taxes create a positive effect on productivity growth by making innovations by incumbents more profitable. In the current model, this indirect effect always dominates. This result is rather extreme in the model, because of the simplifying assumptions (in particular, because the technology of incremental innovations is linear). It should therefore *not* be interpreted as suggesting that entry barriers generally increase growth, but as highlighting that they not only create the well-understood negative effects by reducing creative destruction, but may also encourage further productivity-enhancing activities by incumbent producers. Which effect is more important in practice is an empirical question.

Finally, because the model features entry of new firms, and expansion and exit of existing firms, it also generates an equilibrium firm size distribution. Although the model has not been

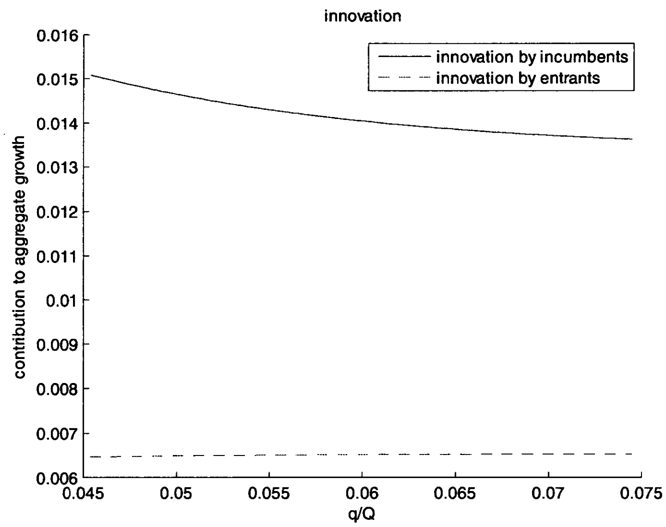


Figure 2-3: Innovations by Entrants and Incumbents.

designed to generate equilibrium firm size distributions, the resulting stationary distribution approximates the Pareto distribution with an exponent of one (the so-called "Zipf distribution") observed in US data (e.g., Axtell 2001).

The model presented in this paper should be viewed as a first step in developing tractable models with endogenous productivity processes for incumbents and entrants (which take place via innovation and other productivity-increasing investments). It contributes to the literature on endogenous technological change by incorporating additional industrial organization elements in the study of economic growth. An important advantage of the specific features emphasized here is that they generate predictions not only about the decomposition of productivity growth between incumbents and entrants, but also about the process of firm growth, entry and exit, and the equilibrium distribution of firm sizes. Nevertheless, the stochastic process for firm size is rather simple and does not incorporate rich firm dynamics that have been emphasized by other work, for example, by Klette and Kortum (2004), who allow firms to operate multiple products, or by Hopenhayn (1992), Melitz (2003) and Luttmer (2007), who introduce a non-trivial exit decision because of fixed costs of operation and also allow firms to learn about their productivity as they operate. Combining these rich aspects of firm entry and exit dynamics with innovation decisions that endogenize the stochastic processes of productivity growth of incumbents and entrants appears to be an important area for future theoretical research. Perhaps more important would be a more detailed empirical analysis of the predictions of these various approaches using data on productivity growth, exit and entry of firms. The relatively simple structure of the model presented in this paper should facilitate these types of empirical exercises. For example, a version of the current model, enriched with additional heterogeneity in firm growth, can be estimated using firm-level data on innovation (patents), sales, entry and exit.

2.6 Appendix

Algebraic Manipulation for Proposition 2.2.

$$\begin{aligned}
\frac{dg^*}{d\tau_i} &= (\lambda-1)\,\phi'\left(z_\tau\left(v^*\right)\right)\frac{dz_\tau\left(v^*\right)}{d\tau_i} \\
&\quad +(\kappa-1)\,\frac{d}{dv}\left(\widehat{z}_\tau\left(v\right)\eta\left(\widehat{z}_\tau\left(v\right)\right)\right)\frac{dv^*}{d\tau_i} \\
&= -z_\tau\left(v^*\right)(\lambda-1)\,\phi'\left(z_\tau\left(v^*\right)\right) \\
&\quad +(\lambda-1)\,\phi'\left(z_\tau\left(v^*\right)\right)\phi\left(z_\tau\left(v^*\right)\right)(\lambda-1)\,\frac{dv^*}{d\tau_i} \\
&\quad +(\kappa-1)\,\frac{d}{dv}\left(\widehat{z}_\tau\left(v\right)\eta\left(\widehat{z}_\tau\left(v\right)\right)\right)\frac{dv^*}{d\tau_i} \\
&= -z_\tau\left(v^*\right)(\lambda-1)\,\phi'\left(z_\tau\left(v^*\right)\right) \\
&\quad +\left(\begin{array}{c}(\lambda-1)\,\phi'\left(z_\tau\left(v^*\right)\right)\phi\left(z_\tau\left(v^*\right)\right)(\lambda-1) \\ +(\kappa-1)\,\frac{d}{dv}\left(\widehat{z}_\tau\left(v\right)\eta\left(\widehat{z}_\tau\left(v\right)\right)\right)\end{array}\right)\frac{dv^*}{d\tau_i}
\end{aligned} \tag{2.60}$$

From the implicit function theorem, (2.38) implies

$$\frac{dv^*}{d\tau_i} = -\frac{\partial\Phi}{\partial\tau_i}/\frac{\partial\Phi}{\partial v}, \tag{2.61}$$

We rewrite

$$\begin{aligned}
\frac{\partial\Phi}{\partial\tau_i} &= \left((\theta-1)\,\phi'\left(z_\tau\left(v\right)\right)(\lambda-1)\,v+(1+\tau_i)\right)\frac{\partial z_\tau\left(v\right)}{\partial\tau_i}+z_\tau\left(v\right) \\
&= -\left((\theta-1)\,\phi'\left(z_\tau\left(v\right)\right)(\lambda-1)\,v+(1+\tau_i)\right)z_\tau\left(v\right)+z_\tau\left(v\right) \\
&= -\left((\theta-1)\left(1+\tau_i\right)+(1+\tau_i)\right)z_\tau\left(v\right)+z_\tau\left(v\right) \\
&= -\left(\theta\left(1+\tau_i\right)-1\right)z_\tau\left(v\right).
\end{aligned}$$

since $\frac{\partial z_\tau(v)}{\partial \tau_i} = -z_\tau(v)$ and $\phi'(z_\tau(v))(\lambda-1)v = (1+\tau_i)$ from (2.36). Similarly

$$\begin{aligned}
\frac{\partial \Phi}{\partial v} &= \rho + (\theta-1)\phi(z_\tau(v))(\lambda-1) + \left((\theta-1)\phi'(z_\tau(v))(\lambda-1)v + (1+\tau_i)\right)\frac{dz_\tau(v)}{dv} \\
&\quad + (\theta(\kappa-1)+1)\frac{d}{dv}\left(\widehat{z}_\tau(v)\eta(\widehat{z}_\tau(v))v\right) \\
&= \rho + (\theta-1)\phi(z_\tau(v))(\lambda-1) + \left((\theta-1)\phi'(z_\tau(v))(\lambda-1)v + (1+\tau_i)\right)\phi(z_\tau(v))(\lambda-1) \\
&\quad + (\theta(\kappa-1)+1)\frac{d}{dv}\left(\widehat{z}_\tau(v)\eta(\widehat{z}_\tau(v))v\right) \\
&= \rho + (\theta(1+\tau_i) + (\theta-1))\phi(z_\tau(v))(\lambda-1) \\
&\quad + (\theta(\kappa-1)+1)\frac{d}{dv}\left(\widehat{z}_\tau(v)\eta(\widehat{z}_\tau(v))v\right).
\end{aligned}$$

Plugging the expression (2.61) into (2.60), we have

$$\begin{aligned}
\frac{dg^*}{d\tau_i} &= -z_\tau(v^*)(\lambda-1)\phi'(z_\tau(v^*)) \\
&\quad + \begin{pmatrix} (\lambda-1)\phi'(z_\tau(v^*))\phi(z_\tau(v^*))(\lambda-1) \\ +(\kappa-1)\frac{d}{dv}\left(\widehat{z}_\tau(v)\eta(\widehat{z}_\tau(v))\right) \end{pmatrix} \\
&\quad * \frac{(\theta(1+\tau_i)-1)z_\tau(v)}{\frac{\partial\Phi}{\partial v}}
\end{aligned}$$

Replacing $(\lambda-1)\phi'(z_\tau(v^*))$ by $\frac{1+\tau_i}{v}$, we just have to show

$$
\begin{aligned}
\frac{1+\tau_i}{v}\frac{\partial\Phi}{\partial v} &= \frac{1+\tau_i}{v}\left(\begin{array}{c}\rho+(\theta(1+\tau_i)+(\theta-1))\,\phi(z_\tau(v))\,(\lambda-1)\\ +(\theta(\kappa-1)+1)\frac{d}{dv}(\widehat{z}_\tau(v)\,\eta(\widehat{z}_\tau(v))\,v)\end{array}\right)\\
&> \left(\frac{1+\tau_i}{v}\phi(z_\tau(v^*))(\lambda-1)+(\kappa-1)\frac{d}{dv}(\widehat{z}_\tau(v)\,\eta(\widehat{z}_\tau(v)))\right)(\theta(1+\tau_i)-1)\\
&\iff\\
&\frac{1+\tau_i}{v}\rho+\theta\phi(z_\tau(v))(\lambda-1)\\
&+\frac{1+\tau_i}{v}(\theta(\kappa-1)+1)\frac{d}{dv}(\widehat{z}_\tau(v)\,\eta(\widehat{z}_\tau(v))\,v)\\
&= \frac{1+\tau_i}{v}\rho+\theta\phi(z_\tau(v))(\lambda-1)\\
&+(1+\tau_i)(\theta(\kappa-1)+1)\frac{d}{dv}(\widehat{z}_\tau(v)\,\eta(\widehat{z}_\tau(v)))\\
&+\frac{1+\tau_i}{v}(\theta(\kappa-1)+1)\,\widehat{z}_\tau(v)\,\eta(\widehat{z}_\tau(v))\,v\\
&> (\kappa-1)\frac{d}{dv}(\widehat{z}_\tau(v)\,\eta(\widehat{z}_\tau(v)))(\theta(1+\tau_i)-1)
\end{aligned}
$$

which is trivially true. ■

Solution to the social planning problem.

$$
\begin{aligned}
\widehat{H}\left(Q^S,z^S,\widehat{z}^S,\mu^S\right) &= \frac{\left(\beta(1-\beta)^{-\frac{1}{\beta}}Q^SL-\left(z^S+\widehat{z}^S\right)Q^S\right)^{1-\theta}-1}{1-\theta}\\
&+\mu^S\left((\lambda-1)\,\phi\left(z^S\right)+(\kappa-1)\,\widehat{z}^S\eta\left(\widehat{z}^S\right)\right)Q^S.
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial\widehat{H}}{\partial z^S} &= -Q^S\left(\beta(1-\beta)^{-\frac{1}{\beta}}Q^SL-\left(z^S+\widehat{z}^S\right)Q^S\right)^{-\theta}\\
&+\mu^S(\lambda-1)\,\phi'\left(z^S\right)Q^S\\
\frac{\partial\widehat{H}}{\partial\widehat{z}^S} &= -Q^S\left(\beta(1-\beta)^{-\frac{1}{\beta}}Q^SL-\left(z^S+\widehat{z}^S\right)Q^S\right)^{-\theta}\\
&+\mu^S(\kappa-1)\left(\eta\left(\widehat{z}^S\right)+\widehat{z}^S\eta'\left(\widehat{z}^S\right)\right)Q^S.
\end{aligned}
$$

Lastly

$$\begin{aligned}
\rho\mu^S - \dot{\mu}^S &= \frac{\partial \widehat{H}}{\partial Q^S} \\
&= \left(\beta(1-\beta)^{-\frac{1}{\beta}} L - \left(z^S + \widehat{z}^S\right)\right)\left(\beta(1-\beta)^{-\frac{1}{\beta}} Q^S L - \left(z^S + \widehat{z}^S\right) Q^S\right)^{-\theta} \\
&\quad + \mu^S\left((\lambda-1)\phi\left(z^S\right) + (\kappa-1)\widehat{z}^S \eta\left(\widehat{z}^S\right)\right)
\end{aligned}$$

Since

$$\begin{aligned}
\mu^S(\lambda-1)\phi'\left(z^S\right) &= -\left(\beta(1-\beta)^{-\frac{1}{\beta}} Q^S L - \left(z^S + \widehat{z}^S\right) Q^S\right)^{-\theta} \\
&= -\left(\beta(1-\beta)^{-\frac{1}{\beta}} L - \left(z^S + \widehat{z}^S\right)\right)^{-\theta}\left(Q^S\right)^{-\theta},
\end{aligned}$$

by differentiating both sides with respect to t, we obtain:

$$\begin{aligned}
\dot{\mu}^S(\lambda-1)\phi'\left(z^S\right) &= \left(\beta(1-\beta)^{-\frac{1}{\beta}} L - \left(z^S + \widehat{z}^S\right)\right)^{-\theta}(-\theta)\left(Q^S\right)^{-\theta-1}\dot{Q}^S(t) \\
&= -\left(\beta(1-\beta)^{-\frac{1}{\beta}} L - \left(z^S + \widehat{z}^S\right)\right)^{-\theta}\theta\left(Q^S\right)^{-\theta}. \\
&\quad \left((\lambda-1)\phi\left(z^S\right) + (\kappa-1)\widehat{z}^S(t)\eta\left(\widehat{z}^S\right)\right)
\end{aligned}$$

Therefore

$$\begin{aligned}
&\frac{\left(\beta(1-\beta)^{-\frac{1}{\beta}} L - \left(z^S + \widehat{z}^S\right)\right)^{-\theta}\theta\left((\lambda-1)\phi\left(z^S\right) + (\kappa-1)\widehat{z}^S(t)\eta\left(\widehat{z}^S\right)\right)}{(\lambda-1)\phi'\left(z^S\right)} \\
&+\rho\frac{\left(\beta(1-\beta)^{-\frac{1}{\beta}} L - \left(z^S + \widehat{z}^S\right)\right)^{-\theta}}{(\lambda-1)\phi'\left(z^S\right)} \\
= &\left(\beta(1-\beta)^{-\frac{1}{\beta}} L - \left(z^S + \widehat{z}^S\right)\right)\left(\beta(1-\beta)^{-\frac{1}{\beta}} L - \left(z^S + \widehat{z}^S\right)\right)^{-\theta} \\
&+\left((\lambda-1)\phi\left(z^S\right) + (\kappa-1)\widehat{z}^S\eta\left(\widehat{z}^S\right)\right)\frac{\left(\beta(1-\beta)^{-\frac{1}{\beta}} L - \left(z^S + \widehat{z}^S\right)\right)^{-\theta}}{(\lambda-1)\phi'\left(z^S\right)}.
\end{aligned}$$

Simplifying both sides by $\frac{\left(\beta(1-\beta)^{-\frac{1}{\beta}}L-\left(z^S+\widehat{z}^S\right)\right)^{-\theta}}{(\lambda-1)\phi'(z^S)}$ we have

$$\begin{aligned}
&\theta\left((\lambda-1)\phi\left(z^S\right)+(\kappa-1)\widehat{z}^S\eta\left(\widehat{z}^S\right)\right)\\
&+\rho\\
=\;&\left(\beta(1-\beta)^{-\frac{1}{\beta}}L-\left(z^S+\widehat{z}^S\right)\right)(\lambda-1)\phi'\left(z^S\right)\\
&+\left((\lambda-1)\phi\left(z^S\right)+(\kappa-1)\widehat{z}^S\eta\left(\widehat{z}^S\right)\right)
\end{aligned}$$

and

$$(\lambda-1)\phi'\left(z^S\right)=(\kappa-1)\left(\eta\left(\widehat{z}^S\right)+\widehat{z}^S\eta'\left(\widehat{z}^S\right)\right).$$

Therefore

$$g^S=\frac{\left(\beta(1-\beta)^{-\frac{1}{\beta}}L-\widehat{z}^S\right)(\lambda-1)\phi+(\kappa-1)\widehat{z}^S\eta\left(\widehat{z}^S\right)-\rho}{\theta}.$$

■

Derivation of the Functional Equation for the stationary distribution. Consider the evolution of the highest quality product in each sector. In the case of entry without imitation

$$q(t+\Delta t)=\begin{cases} q(t) & \text{with probability } 1-\phi\left(z\left(\widetilde{q}(t)\right)\right)\Delta t-\widehat{z}\left(\widetilde{q}(t)\right)\eta\left(\widehat{z}\left(\widetilde{q}(t)\right)\right)\Delta t+o(\Delta t)\\ \lambda q(t) & \text{with probability } \phi\left(z\left(\widetilde{q}(t)\right)\right)\Delta t+o(\Delta t)\\ \kappa q(t) & \text{with probability } \widehat{z}\left(\widetilde{q}(t)\right)\eta\left(\widehat{z}\left(\widetilde{q}(t)\right)\right)\Delta t+o(\Delta t)\end{cases}$$

in which $\widetilde{q}(t)=\frac{q(t)}{Q(t)}$, and the average product quality $Q(t)$ grows at a constant rate g. We also assume that when $q(t)\leq\epsilon Q(t)$, $q(t+)$ jumps to $\omega Q(t)$, $\omega>\epsilon$. Therefore the evolution of the normalize quality,$\widetilde{q}(t)=\frac{q(t)}{Q(t)}$,is

$$\widetilde{q}(t+\Delta t)=\begin{cases} q(t)(1-g\Delta t)+o(\Delta t)\\ \text{with probability } 1-\phi\left(z\left(\widetilde{q}(t)\right)\right)\Delta t-\widehat{z}\left(\widetilde{q}(t)\right)\eta\left(\widehat{z}\left(\widetilde{q}(t)\right)\right)\Delta t+o(\Delta t)\\ \lambda q(t)(1-g\Delta t)+o(\Delta t)\ \text{ with probability } \phi\left(z\left(\widetilde{q}(t)\right)\right)\Delta t+o(\Delta t)\\ \kappa q(t)(1-g\Delta t)+o(\Delta t)\ \text{ with probability } \widehat{z}\left(\widetilde{q}(t)\right)\eta\left(\widehat{z}\left(\widetilde{q}(t)\right)\right)\Delta t+o(\Delta t)\end{cases}$$

Moreover, when $\widetilde{q}(t)\leq\epsilon$, it jumps immediately to $\widetilde{q}(t+)=\omega>\epsilon$. By taking the limit $\Delta t\longrightarrow 0$

we can ignore the terms $o\left(\Delta t\right)$. Suppose we have a stationary distribution of relative product quality with the cdf $F\left(y\right)$.Then, for $y > \omega$

$$
\begin{aligned}
F\left(y\right) &= \Pr\left(\widetilde{q}\left(t+\Delta t\right) \leq y\right) \\
&= \mathbf{E}\left[\mathbf{1}_{\{\widetilde{q}(t+\Delta t)\leq y\}}\right] \\
&= \mathbf{E}\left[\mathbf{E}\left[\mathbf{1}_{\{\widetilde{q}(t+\Delta t)\leq y\}}|\widetilde{q}\left(t\right)\right]\right]
\end{aligned}
$$

We rewrite the iterated expectation as

$$
\begin{aligned}
&\mathbf{E}\left[\mathbf{E}\left[\begin{array}{c} \mathbf{1}_{\{\widetilde{q}(t)(1-g\Delta t)\leq y,1-\phi(z(\widetilde{q}(t)))\Delta t-\widehat{z}(\widetilde{q})\eta(\widehat{z}(\widetilde{q}))\Delta t,\widetilde{q}(t)(1-g\Delta t)>\varepsilon\}} + \mathbf{1}_{\{\widetilde{q}(t)(1-g\Delta t)\leq\varepsilon\}} \\ +\mathbf{1}_{\{\lambda\widetilde{q}(t)(1-g\Delta t)\leq y,\phi(z(\widetilde{q}(t)))\Delta t\}} + \mathbf{1}_{\{\kappa\widetilde{q}(t)(1-g\Delta t)\leq y,\widehat{z}(\widetilde{q})\eta(\widehat{z}(\widetilde{q}))\Delta t\}} \end{array}\middle|\,\widetilde{q}\left(t\right)\right]\right] \\
&= \mathbf{E}\left[\mathbf{E}\left[\begin{array}{c} \mathbf{1}_{\{\widetilde{q}(t)\leq y(1+g\Delta t),1-\phi(z(\widetilde{q}(t)))\Delta t-\widehat{z}(\widetilde{q})\eta(\widehat{z}(\widetilde{q}))\Delta t\}} + \mathbf{1}_{\{\widetilde{q}(t)(1-g\Delta t)\leq\varepsilon\}} \\ +\mathbf{1}_{\left\{\widetilde{q}(t)\leq\frac{y}{\lambda}(1+g\Delta t),\phi(z(\widetilde{q}(t)))\Delta t\right\}} + \mathbf{1}_{\left\{\widetilde{q}(t)\leq\frac{y}{\kappa}(1+g\Delta t),\widehat{z}(\widetilde{q})\eta(\widehat{z}(\widetilde{q}))\Delta t\right\}} \end{array}\middle|\,\widetilde{q}\left(t\right)\right]\right] \\
&= \mathbf{E}\left[\begin{array}{c} \left(1-\phi\left(z\left(\widetilde{q}\left(t\right)\right)\right)\Delta t-\widehat{z}\left(\widetilde{q}\right)\eta\left(\widehat{z}\left(\widetilde{q}\right)\right)\Delta t\right)\mathbf{1}_{\{q(t)\leq y(1+g\Delta t),\widetilde{q}(t)(1-g\Delta t)>\varepsilon\}} \\ +\mathbf{1}_{\{\widetilde{q}(t)(1-g\Delta t)\leq\varepsilon\}} + \phi\left(z\left(\widetilde{q}\left(t\right)\right)\right)\Delta t\mathbf{1}_{\left\{\widetilde{q}(t)\leq\frac{y}{\lambda}(1+g\Delta t)\right\}} \\ +\widehat{z}\left(\widetilde{q}\right)\eta\left(\widehat{z}\left(\widetilde{q}\right)\right)\Delta t\mathbf{1}_{\left\{\widetilde{q}(t)\leq\frac{y}{\kappa}(1+g\Delta t)\right\}} \end{array}\right]
\end{aligned} \tag{2.62}
$$

Replacing the last expectations by integrals, we obtain

$$
\begin{aligned}
F\left(y\right) &= \int_{\varepsilon(1+g\Delta t)}^{y(1+g\Delta t)} \left(1-\phi\left(z\left(\widetilde{q}\right)\right)\Delta t-\widehat{z}\left(\widetilde{q}\right)\eta\left(\widehat{z}\left(\widetilde{q}\right)\right)\Delta t\right)dF\left(\widetilde{q}\right) \\
&+ F\left(\varepsilon\left(1+g\Delta t\right)\right) + \int_{0}^{\frac{y}{\lambda}(1+g\Delta t)} \phi\left(z\left(\widetilde{q}\right)\right)\Delta t dF\left(\widetilde{q}\right) \\
&+ \int_{0}^{\frac{y}{\kappa}(1+g\Delta t)} \widehat{z}\left(\widetilde{q}\right)\eta\left(\widehat{z}\left(\widetilde{q}\right)\right)\Delta t dF\left(\widetilde{q}\right)
\end{aligned} \tag{2.63}
$$

Develop right hand side with respect to Δt, we have

$$
\begin{aligned}
F\left(y\right) &= F\left(y\right) + F_y\left(y\right)yg\Delta t - F_y\left(\varepsilon\right)\varepsilon g\Delta t + F\left(\varepsilon\left(1+g\Delta t\right)\right) \\
&- \int_0^y \phi\left(z\left(\widetilde{q}\right)\right)dF\left(\widetilde{q}\right)\Delta t - \int_0^y \widehat{z}\left(\widetilde{q}\right)\eta\left(\widehat{z}\left(\widetilde{q}\right)\right)dF\left(\widetilde{q}\right)\Delta t \\
&+ \int_0^{\frac{y}{\lambda}} \phi\left(z\left(\widetilde{q}\right)\right)dF\left(\widetilde{q}\right)\Delta t + \int_0^{\frac{y}{\kappa}} \widehat{z}\left(\widetilde{q}\right)\eta\left(\widehat{z}\left(\widetilde{q}\right)\right)dF\left(\widetilde{q}\right)\Delta t
\end{aligned}
$$

so, eliminate $F(y)$ from both side of the last equation and dividing Δt and send Δt to zero we arrive at

$$0 = F_y(y)\, yg - \int_{\frac{y}{\lambda}}^{y} \phi(z(\widetilde{q}))\, dF(\widetilde{q}) - \int_{\frac{y}{\kappa}}^{y} \widehat{z}(\widetilde{q})\, \eta(\widehat{z}(\widetilde{q}))\, dF(\widetilde{q})$$

as in (2.57). For $y < \omega$, we proceed exactly as above except now the terms $\mathbf{1}_{\{\widetilde{q}(t)(1-g\Delta t)\leq\varepsilon\}}$ and $F(\varepsilon(1+g\Delta t))$ do not appear in (2.62) and (2.63) due to the fact that any imitator entering into the economy will have relative quality ω exceeding y. We obtain the final expression

$$0 = F_y(y)\, yg - F_y(\varepsilon)\, \varepsilon g - \int_{\frac{y}{\lambda}}^{y} \phi(z(\widetilde{q}))\, dF(\widetilde{q}) - \int_{\frac{y}{\kappa}}^{y} \widehat{z}(\widetilde{q})\, \eta(\widehat{z}(\widetilde{q}))\, dF(\widetilde{q})$$

as in (2.58). For the special case in which $\epsilon = \omega = 0$, $g = g^*$, $z(\widetilde{q}) \equiv z^*$ and $\widehat{z}(\widetilde{q}) \equiv \widehat{z}^*$, we obtain the equation (2.52). ■

Derivation of the implied growth rate. The growth of the average product quality Q_t comes from three sources: Innovation of incumbent firms, of innovative entrants and imitators. Recall the definition of Q_t

$$Q_t = \int_0^1 q(v,t)\, dv$$

where $q(v,t)$ is the highest quality in sector v. We suppose that the investment of the incumbents in each sector are $z(\widetilde{q})$ and of the innovative entrants are $\widehat{z}(\widetilde{q})$, where $\widetilde{q}$ is the quality relative to the average quality that grows at the rate g then at time from time t to $t + \Delta t$:

$$\begin{aligned}
Q_{t+\Delta t} &= \int_0^1 q(v, t+\Delta t)\, dv \\
&= \int_{0,q(v,t)\geq\varepsilon Q(t)(1+g\Delta t)}^{1} \begin{pmatrix} \phi\left(z\left(\frac{q(v,t)}{Q_t}\right)\right)\Delta t\lambda q(v,t) + \widehat{z}\left(\frac{q(v,t)}{Q_t}\right)\eta\left(\widehat{z}\left(\frac{q(v,t)}{Q_t}\right)\right)\Delta t\kappa q(v,t) \\ +\left(1-\phi\left(z\left(\frac{q(v,t)}{Q_t}\right)\right)\Delta t - \widehat{z}\left(\frac{q(v,t)}{Q_t}\right)\eta\left(\widehat{z}\left(\frac{q(v,t)}{Q_t}\right)\right)\Delta t\right) q(v,t) \end{pmatrix} dv \\
&+ \int_{0,q(v,t)<\varepsilon Q(t)(1+g\Delta t)}^{1} \begin{pmatrix} \phi\left(z\left(\frac{q(v,t)}{Q_t}\right)\right)\Delta t\lambda q(v,t) + \widehat{z}\left(\frac{q(v,t)}{Q_t}\right)\eta\left(\widehat{z}\left(\frac{q(v,t)}{Q_t}\right)\right)\Delta t\kappa q(v,t) \\ +\left(1-\phi\left(z\left(\frac{q(v,t)}{Q_t}\right)\right)\Delta t - \widehat{z}\left(\frac{q(v,t)}{Q_t}\right)\eta\left(\widehat{z}\left(\frac{q(v,t)}{Q_t}\right)\right)\Delta t\right) \omega Q(t) \end{pmatrix} dv
\end{aligned}$$

Expand the right hand side around $\Delta t = 0$, we have

$$\begin{aligned}
Q_{t+\Delta t} = {} & \lambda Q_t \Delta t \int_0^1 \phi\left(z\left(\frac{q(v,t)}{Q_t}\right)\right)\frac{q(v,t)}{Q_t}dv + \kappa Q_t \Delta t \int_0^1 \widehat{z}\left(\frac{q(v,t)}{Q_t}\right)\eta\left(\widehat{z}\left(\frac{q(v,t)}{Q_t}\right)\right)\frac{q(v,t)}{Q_t}dv \\
& + \int_{0,q(v,t)\geq \varepsilon Q(t)(1+g\Delta t)}^1 \left(1-\phi\left(z\left(\frac{q(v,t)}{Q_t}\right)\right)\Delta t - \widehat{z}\left(\frac{q(v,t)}{Q_t}\right)\eta\left(\widehat{z}\left(\frac{q(v,t)}{Q_t}\right)\right)\Delta t\right) q(v,t)\,dv \\
& + \int_{0,q(v,t)< \varepsilon Q(t)(1+g\Delta t)}^1 \left(1-\phi\left(z\left(\frac{q(v,t)}{Q_t}\right)\right)\Delta t - \widehat{z}\left(\frac{q(v,t)}{Q_t}\right)\eta\left(\widehat{z}\left(\frac{q(v,t)}{Q_t}\right)\right)\Delta t\right) \omega Q(t)\,dv
\end{aligned}$$

We can rearrange to decompose the growth of average quality into three different component: innovation from incumbents, from entrants, and from imitators.

$$\begin{aligned}
Q_{t+\Delta t} = Q(t) + {} & \underbrace{(\lambda-1) Q_t \Delta t \int_0^1 \phi\left(z\left(\frac{q(v,t)}{Q_t}\right)\right)\frac{q(v,t)}{Q_t}dv}_{\text{Innovation from Incumbents}} \\
& \underbrace{+ (\kappa-1) Q_t \Delta t \int_0^1 \widehat{z}\left(\frac{q(v,t)}{Q_t}\right)\eta\left(\widehat{z}\left(\frac{q(v,t)}{Q_t}\right)\right)\frac{q(v,t)}{Q_t}dv}_{\text{Innovation from Entrants}} \\
& \underbrace{+ \int_{0,q(v,t)< \varepsilon Q(t)(1+g\Delta t)}^1 \left(\begin{array}{c} 1-\phi\left(z\left(\frac{q(v,t)}{Q_t}\right)\right)\Delta t \\ -\widehat{z}\left(\frac{q(v,t)}{Q_t}\right)\eta\left(\widehat{z}\left(\frac{q(v,t)}{Q_t}\right)\right)\Delta t \end{array}\right)(\omega Q(t) - q(v,t))\,dv}_{\text{Innovation from Imitators}}
\end{aligned}$$

We rewrite this growth accounting in term of stationary distribution with cumulative distribution function $F(.)$ over $\widehat{q} = \frac{q}{Q} > \varepsilon$ and probability density function $f(.)$

$$\begin{aligned}
Q(t+\Delta t) = Q(t) & + (\lambda-1) Q(t) \int \phi(z(\widehat{q}))\widehat{q}dF(\widehat{q})\Delta t \\
& + (\kappa-1) Q(t) \int \widehat{z}(\widehat{q})\eta(\widehat{z}(\widehat{q}))\widehat{q}dF(\widehat{q})\Delta t \\
& + F(\varepsilon(1+g\Delta t))(\omega-\varepsilon) Q(t).
\end{aligned}$$

So

$$g' = (\lambda-1)\mathbf{E}[\phi(z(\widehat{q}))\widehat{q}] + (\kappa-1)\mathbf{E}[\widehat{z}(\widehat{q})\eta(\widehat{z}(\widehat{q}))\widehat{q}] + \varepsilon g f(\varepsilon)(\omega-\varepsilon).$$

Equivalently

$$g' = \frac{(\lambda-1)\mathbf{E}[\phi(z(\widehat{q}))\widehat{q}] + (\kappa-1)\mathbf{E}[\widehat{z}(\widehat{q})\eta(\widehat{z}(\widehat{q}))\widehat{q}]}{1-\varepsilon f(\varepsilon)(\omega-\varepsilon)}$$

as in (2.59). When $\omega = 0$ we have

$$g' = (\lambda - 1)\,\mathbf{E}\left[\phi\left(z\left(\widetilde{q}\right)\right)\widetilde{q}\right] + (\kappa - 1)\,\mathbf{E}\left[\widehat{z}\left(\widetilde{q}\right)\eta\left(\widehat{z}\left(\widetilde{q}\right)\right)\widetilde{q}\right],$$

and when $z\left(\widetilde{q}\right) \equiv z^*$ and $\widehat{z}\left(\widetilde{q}\right) \equiv \widehat{z}^*$

$$g = (\lambda - 1)\,\phi\left(z^*\right) + (\kappa - 1)\,\widehat{z}^*\eta\left(\widehat{z}^*\right)$$

as in (2.24), given $E\left[\widetilde{q}\right] = 1$. ■

To prove Theorem 2.1, we need the following definition

Definition 2.3 *Let* $I_i(v) = \max_{z\geq 0}\phi(z)\,v - z$ *and* $I_e(u) = \frac{1}{\kappa u}\eta^{-1}\left(\frac{1}{\kappa u}\right)$.

Remark 2.4 $I_i(v)$ *is the value to incumbent firms from undertaking incremental innovation.* $I_e(u)$ *is the rate of entry by entrants with radical innovations. There are one-to-one mappings from the investment technology* ϕ *and* η *to the functions* I_i *and* I_e.

We prove Theorem 2.1 in three steps sketched in the body of the paper:

Step 1: We show the existence of $\widehat{V}_g\left(\widetilde{q}\right)$ under the form $\widetilde{q}U_g\left(\ln\left(\widetilde{q}\right) - \ln\epsilon_g\right)$ where U_g is shown to exist as follow.

Definition 2.4 $C^0\left(\left[g^* - \Delta, g^* + \Delta\right] \times \mathbf{R}^+, \mathbf{R}\right)$ *denotes the Banach space of continuous functions* $U : \left[g^* - \Delta, g^* + \Delta\right] \times \mathbf{R}^+ \longrightarrow \mathbf{R}$ *and* $U(g, 0) = 0\ \forall g \in \left[g^* - \Delta, g^* + \Delta\right]$ *with the norm*

$$\|U\| = \sup_{g^*-\Delta\leq g\leq g^*+\Delta}\ \sup_{0\leq p\leq\infty}\left|U(g,p)\right|.$$

Definition 2.5 *For each function* $u = U(g, .) \in C^0\left(\mathbf{R}^+, \mathbf{R}\right)$ *consider the operator* T_g

$$T_g u \in C^0\left(\mathbf{R}^+, \mathbf{R}\right)$$

satisfies the following ordinary differential equation[22]

$$\begin{aligned} & g\left(T_g u\right)'(p) + \left(r_g + I_e\left(u\left(p + \ln\kappa\right)\right)\right)\left(T_g u\right)(p) \\ & = \beta L + I_i\left(\lambda u\left(p + \ln\lambda\right) - T_g u\left(p\right)\right). \end{aligned} \tag{2.64}$$

with the initial condition $T_g u(0) = 0$. *Notice that*

$$r_g = \rho + \theta g. \tag{2.65}$$

Here $\widehat{z}(p)$ *is defined as:*$\eta\left(\widehat{z}(p)\right)\kappa u\left(p + \ln\kappa\right) = 1$. *The operator* T *is defined by*

$$TU(g,p) = T_g U(g,p).$$

To prove the existence of the value function of the incumbent, we need the following key lemma:

Lemma 2.1 *Suppose Assumption 2.1 is satisfied. Then there exists* $\Delta > 0$ *such that for each* $g \in [g^* - \Delta, g^* + \Delta]$, *there is a solution* $U_g \geq 0$ *to the functional equation*

$$\begin{aligned} & rU(p) + gU'(p) \\ & = \beta L + \max_z\left\{\phi(z)\left(\lambda U\left(p + \ln\lambda\right) - U(p)\right) - z\right\} - \widehat{z}(p)\,\eta\left(\widehat{z}(p)\right)U(p), \end{aligned} \tag{2.66}$$

where $r = \rho + \theta g$ *and*

$$\eta\left(\widehat{z}(p)\right)\kappa U\left(p + \ln\kappa\right) = 1.$$

U *should also satisfy the boundary conditions*

$$\begin{aligned} U(0) &= 0 \\ \lim_{p \longrightarrow \infty} U(p) &= v_g, \end{aligned} \tag{2.67}$$

[22] *Standard theorems in ODE theory imply the existence and uniqueness of* $T_g u(p)$ *if* $u \in F$ *defined below.*

where v_g is solution of the equation

$$v = \frac{\beta L + I_i\left((\lambda - 1)\, v\right)}{r + I_e\left(v\right)}. \tag{2.68}$$

Moreover, U_g is equicontinuous in g over any finite interval.

In order to prove this lemma, we again need the following lemmas

Lemma 2.2 *Suppose Assumption 2.1 is satisfied. Then there exists $\Delta > 0$ such that the set $\mathcal{F}$ of continuous function $U : [g^* - \Delta, g^* + \Delta] \times \mathbf{R}^+ \longrightarrow \mathbf{R}$, $U(g, 0) = 0$ and*

$$v_g - v_g e^{-\theta p} \leq U(g, p) \leq v_g + v_g e^{-\theta p} \;\forall p \geq 0 \tag{2.69}$$

satisfies $T(\mathcal{F}) \subset \mathcal{F}$.

Proof. Let $\overline{k}_g(p) = v_g + v_g e^{-\theta p}$ and $\underline{k}_g(p) = v_g + v_g e^{-\theta p}$. By definition, for each $U \in \mathcal{F}$, we have

$$\underline{k}_g(p) \leq U(g, p) \leq \overline{k}_g(p).$$

Let $k_g(p) = T_g U(g, p)$ then, also by definition (2.64) implies that

$$\begin{aligned} g k_g'(p) &= \beta L - I_i\left(\lambda u\left(p + \ln \lambda\right) - k_g(p)\right) - \left(r_g + I_e\left(u\left(p + \ln \kappa\right)\right)\right)\left(T_g u\right)(p) \\ &\leq \beta L - I_i\left(\lambda \overline{k}_g\left(p + \ln \lambda\right) - k_g(p)\right) - \left(r_g + I_e\left(\underline{k}_g\left(p + \ln \kappa\right)\right)\right)\overline{k}_g(p). \end{aligned}$$

So if we can show that

$$g\overline{k}_g'(p) > \beta L - I_i\left(\lambda \overline{k}_g\left(p + \ln \lambda\right) - \overline{k}_g(p)\right) - \left(r_g + I_e\left(\underline{k}_g\left(p + \ln \kappa\right)\right)\right)\overline{k}_g(p) \tag{2.70}$$

we will have $k_g(p) < \overline{k}_g(p)\; \forall p > 0$ given that $k_g(0) = 0 < \overline{k}_g(0)$. Similarly, if we can show that

$$g\underline{k}_g'(p) < \beta L - I_i\left(\lambda \underline{k}_g\left(p + \ln \lambda\right) - \underline{k}_g(p)\right) - \left(r_g + I_e\left(\overline{k}_g\left(p + \ln \kappa\right)\right)\right)\underline{k}_g(p) \tag{2.71}$$

we will have $k_g(p) > \underline{k}_g(p)\; \forall p > 0$ given that $k_g(0) = 0 = \underline{k}_g(0)$.

Below we will use Assumption 2.1 to show (2.70) and (2.71). Indeed, the two inequalities can be re-written as

$$\begin{aligned} -g\theta x > \beta L + I_i\left(\left(\lambda - 1\right)v_g + \left(\lambda^{1-\theta} - 1\right)x\right) \\ - \left(r_g + I_e\left(v_g - \kappa^{-\theta}x\right)\right)\left(v_g + x\right) \end{aligned} \tag{2.72}$$

and

$$\begin{aligned} g\theta x < \beta L + I_i\left(\left(\lambda - 1\right)v_g - \left(\lambda^{1-\theta} - 1\right)x\right) \\ - \left(r_g + I_e\left(v_g + \kappa^{-\theta}x\right)\right)\left(v_g - x\right) \end{aligned} \tag{2.73}$$

$\forall 0 < x \leq v_g$.

By definition of v_g in (2.68), we have equalities at $x = 0$. It is sufficient to show that the derivative of the left hand side of (2.72) is strictly greater than the derivative of its left hand side. Or equivalently,

$$\begin{aligned} -g\theta \;>\; & I_i'\left(\left(\lambda - 1\right)v_g + \left(\lambda^{1-\theta} - 1\right)x\right)\left(\lambda^{1-\theta} - 1\right) \\ & -r_g - I_e\left(v_g - \kappa^{-\theta}x\right) + I_e'\left(v_g - \kappa^{-\theta}x\right)\kappa^{-\theta}v_g. \end{aligned}$$

(2.65) yields $r_g > g\theta$ and $\theta \geq 1$ yields $I_i'\left(\left(\lambda - 1\right)v_g + \left(\lambda^{1-\theta} - 1\right)x\right)\left(\lambda^{1-\theta} - 1\right) \leq 0$. It remains to show that

$$I_e\left(v_g - \kappa^{-\theta}x\right)v_g \geq I_e'\left(v_g - \kappa^{-\theta}x\right)\kappa^{-\theta},$$

or

$$\left(v_g - \kappa^{-\theta}x\right) \geq \frac{1}{\min \epsilon_{I_e}}\kappa^{-\theta}v_g.$$

or

$$\frac{\min \epsilon_{I_e}}{\min \epsilon_{I_e} + 1} \geq \kappa^{-\theta} \tag{2.74}$$

Similarly, it is sufficient to show that the derivative of the left hand side of (2.73) is strictly greater than the derivative of its left hand side. Or equivalently,

$$\begin{aligned} g\theta &< I_i'\left((\lambda-1)\,v_g-\left(\lambda^{1-\theta}-1\right)x\right)\left(1-\lambda^{1-\theta}\right) \\ &\quad +r_g+I_e\left(v_g+\kappa^{-\theta}x\right)-I_e'\left(v_g+\kappa^{-\theta}x\right)\kappa^{-\theta}v_g. \end{aligned}$$

This is true if

$$\left(v_g+\kappa^{-\theta}x\right)\geq\frac{1}{\min\epsilon_{I_e}}\kappa^{-\theta}v_g$$

or

$$\min\epsilon_{I_e}\geq\kappa^{-\theta}. \tag{2.75}$$

Given that

$$\epsilon_{I_e}=\frac{1}{\epsilon_\eta}-1,$$

Assumption 2.1 implies both (2.74) and (2.75).

■

Lemma 2.3 *$T(\mathcal{F})$ is a relatively compact subset of $C^0([g^*-\Delta,g^*+\Delta]\times\mathbf{R}^+,\mathbf{R})$.*

Proof. Suppose $\{f_n\}_{n=1}^{\infty}\subset\mathcal{F}$, we will show that we can extract a Cauchy sequence from $\{Tf_n\}_{n=1}^{\infty}$. First, there exists a constant $K>0$ such that $\|U\|\leq K\ \forall U\in\mathcal{F}$. So

$$\left|\frac{\partial}{\partial p}Tf_n\right|=\left|\frac{d}{dp}T_gf_n(g,p)\right|\leq\frac{\beta L+I_i((\lambda+1)K)+(\rho+\theta(g^*+\Delta)+I_e(K))K}{g^*-\Delta}$$
$$\forall g,\forall p$$

Second, $D_g(p)=\frac{\partial}{\partial g}(Tf_n(p))$ is the solution of

$$\begin{aligned} & gD_g'(p)+Tf_n(p)+\left(\frac{dr_g}{dg}+I_e(f_n(p+\ln\kappa))\right)Tf_n(p) \\ = &\ \left(I_i'(\lambda f_n(p+\ln\lambda)-Tf_n(p))-(r_g+I_e(f_n(p+\ln\kappa)))\right)D_g(p) \end{aligned}$$

So $D_g(p)$ is uniformly bounded over $[g^*-\Delta,g^*+\Delta]\times[0,M]$ for any $M>0$. Therefore, for each $M=1,2,....$ we have $\{Tf_n(g,p)\}_{n=1}^{\infty}$ is equicontinuous over

$$C^0([g^*-\Delta,g^*+\Delta]\times[0,M],\mathbf{R}).$$

$M=1$ since $\{Tf_n\}_{n\geq 1}$ is equicontinuous over $[g^*-\Delta, g^*+\Delta]\times[0,M]$ there exists a subsequence $\{Tf_{1_k}\}_{k=1}^{\infty}$ converges uniformly to f_M^* over $[g^*-\Delta, g^*+\Delta]\times[0,M]$

$M \Longrightarrow M+1$: Since $\{Tf_{M_k}\}_{k=2}^{\infty}$ is equicontinuous over $[g^*-\Delta, g^*+\Delta]\times[0,M+1]$ there exists a subsequence $\left\{Tf_{(M+1)_k}\right\}_{k=1}^{\infty}$ converges uniformly to f_{M+1}^* over $[g^*-\Delta, g^*+\Delta]\times[0,M+1]$.

Because of the subsequence property: $f_{M+1}^*\big|_{[0,M]} = f_M^*$.

Let $f^* : [g^*-\Delta, g^*+\Delta]\times\mathbf{R}^+ \longrightarrow R$ be defined by $f^*|_{[g^*-\Delta,g^*+\Delta]\times[0,M]} = f_M^*\ \forall M\in\mathbf{N}^*$.

We show that

$$\lim_{M,N\longrightarrow\infty}\|Tf_{M_M}-f^*\|_{C^0([g^*-\Delta,g^*+\Delta]\times\mathbf{R}^+,\mathbf{R})}=0.$$

Indeed, $\forall\varepsilon>0$: Given (2.69) there exists $M_1\in\mathbf{N}^*$ such that $|Tf_{M_M}(g,p)-u_g|<\frac{\varepsilon}{2}\ \forall p\geq M_1$. Given $\lim_{M\longrightarrow\infty}Tf_{M_M}(g,p)=f^*(g,p)\ \forall p$, we have

$$|f^*(g,p)-u_g|\leq\frac{\varepsilon}{2}\forall p\geq M_1$$

and $g\in[g^*-\Delta, g^*+\Delta]$. Given M_1, there exists $M_2>M_1$ such that $|Tf_{M_M}(g,p)-f^*(g,p)|<\varepsilon\ \forall M_1\geq p\geq 0, g\in[g^*-\Delta, g^*+\Delta]$ and $M\geq M_2$. Therefore $\forall p\geq 0$, $g\in[g^*-\Delta, g^*+\Delta]$ and $M,N\geq M_2$

$$|Tf_{M_M}(p,g)-f^*(p,g)|<\varepsilon.$$

■

Lemma 2.4 *The mapping T is continuous over $\mathcal{F}$.*

Proof. Suppose $f_n\longrightarrow f$, by the Lebesgue dominated convergence theorem, we have Tf_n converges pointwise toward Tf. It remains to prove that

$$\lim_{n\longrightarrow\infty}\|Tf_n-Tf\|_{C^0([g^*-\Delta,g^*+\Delta]\times\mathbf{R}^+,\mathbf{R})}=0.$$

By the relative compactness of $T(\mathcal{F})$, from any subsequence of $\{Tf_n\}$ there is subsequence $\{h_M\}$ of $\{Tf_n\}$ that converges to h over $C^0([g^*-\Delta, g^*+\Delta]\times\mathbf{R}^+,\mathbf{R})$. Since $\{h_M\}$ also converges pointwise to Tf we have $h = Tf$. Therefore

$$\lim_{M \longrightarrow \infty} \|h_M - Tf\|_{C^0([g^*-\Delta, g^*+\Delta]\times \mathbf{R}^+, \mathbf{R})} = 0.$$

Thus

$$\lim_{n \longrightarrow \infty} \|Tf_n - Tf\|_{C^0([g^*-\Delta, g^*+\Delta]\times \mathbf{R}^+, \mathbf{R})} = 0.$$

■

Proof of Lemma 2.1. Given Lemma 2.2, 2.3, 2.4 we can apply the Schauder Fixed Point Theorem to show that T admits a fixed point U in $\mathcal{F}$: $TU = U$. Or equivalently for each $g \in [g^* - \Delta, g^* + \Delta]$, $u(.) = U(g, .)$ satisfies $u(0) = 0$ and (2.66). The limit at infinity in (2.67) follows directly from the definition of $\mathcal{F}$. Finally, equicontinuity is a consequence of the fact that $U(., .) \in C^0([g^* - \Delta, g^* + \Delta] \times \mathbf{R}^+, \mathbf{R})$. ■

Existence of the Value Function. Let $\underline{\mu} = U_{g^*}\left(\frac{1-\beta}{\beta}\log\left(\frac{1}{1-\beta}\right)\right)$ and $\overline{\mu} = U_{g^*}\left(\frac{1-\beta}{\beta}\log\left(\frac{1}{1-\beta}\right) + \delta\right)$, where $\delta > 0$ and $\mu_e \in (\underline{\mu}, \overline{\mu})$. Given $U(g,p)$ is equicontinuous in $g \in [g^*, g^* + \Delta]$, we can choose Δ sufficiently small such that $U_g\left(\frac{1-\beta}{\beta}\log\left(\frac{1}{1-\beta}\right)\right) < \mu_e < U_g\left(\frac{1-\beta}{\beta}\log\left(\frac{1}{1-\beta}\right) + \delta\right)$ $\forall g \in [g^*, g^* + \Delta]$. Therefore, there exists $\omega_g \in \left(\frac{1-\beta}{\beta}\log\left(\frac{1}{1-\beta}\right), \frac{1-\beta}{\beta}\log\left(\frac{1}{1-\beta}\right) + \delta\right)$ such that $\mu_e = U_g(\omega_g)$. For each ω, let

$$\begin{aligned} \epsilon_g &= \omega/\omega_g \\ &< \omega(1-\beta)^{\frac{1-\beta}{\beta}}. \end{aligned} \tag{2.76}$$

and let $\widehat{V}_g(\widetilde{q}) = \widetilde{q}U_g(\ln(\widetilde{q}) - \ln\epsilon_g)$. Then $\widehat{V}_g$ satisfies (2.54), (2.55) and (2.56). ■

Given the existence of $U(g,p)$, for each $g \in [g^*, g^* + \Delta]$, we define

$$z_g(p) = \arg\max_z (\lambda U_g(p+\lambda) - U_g(p))\,\phi(z) - z$$

and

$$\widehat{z}_g(p) = \eta^{-1}\left(\frac{U_g(p + \ln\kappa)}{\kappa}\right)$$

Given that

$$\lim_{p \longrightarrow \infty} U_g(p) = v_g$$

we have

$$\lim_{p \longrightarrow \infty} z_g(p) = z(v_g)$$
$$\lim_{p \longrightarrow \infty} \widehat{z}_g(p) = \widehat{z}(v_g).$$

Armed with the existence of the value function and the corresponding investment decisions, we are ready to prove the second step

Step 2: We show the existence of the stationary distribution under the form $f_g(y) = \frac{h_g(\ln y)}{y}$. Moreover, we show that f_g satisfies the asymptotic Pareto property in Theorem 2.2.

The implied stationary distribution $F_g(y)$ solves equations (2.57) and (2.58) with $z(\widetilde{q}) = z_g(\ln \widetilde{q})$ and $\widehat{z}(\widetilde{q}) = \widehat{z}_g(\ln \widetilde{q})$. Let $h_g(p) = e^p F_y(e^p)$. The equations (2.57) and (2.58) become:

If $p > \ln \omega_1$

$$0 = h_g(p) g - \int_{p-\ln \lambda}^{p} \phi(z_g(\widetilde{p})) h_g(\widetilde{p}) d\widetilde{p} - \int_{p-\ln \kappa}^{p} \widehat{z}_g(\widetilde{p}) \eta(\widehat{z}_g(\widetilde{p})) h_g(\widetilde{p}) d\widetilde{p}. \quad (2.77)$$

If $p \leq \ln \omega_1$

$$0 = h_g(p) g - h_g(0) g - \int_{p-\ln \lambda}^{p} \phi(z_g(\widetilde{p})) h_g(\widetilde{p}) d\widetilde{p} - \int_{p-\ln \kappa}^{p} \widehat{z}_g(\widetilde{p}) \eta(\widehat{z}_g(\widetilde{p})) h_g(\widetilde{p}) d\widetilde{p}. \quad (2.78)$$

We also have $h_g(p) = 0 \ \forall p \leq 0$. The conditions for h_g to be a well-defined distribution is

$$\int h_g(p) dp = 1.$$

Lemma 2.5 *Given the investment strategies $z_g(p), \widehat{z}_g(p)$, the stationary distribution $h_g(p)$ exists and unique.*

Proof. Differentiate both side of the integral equations on h_1, we have

$$g h_g'(p) = \phi(z_1(p)) h_g(p) - \phi(z_g(p - \ln \lambda)) h_g(p - \ln \lambda)$$
$$+ \widehat{z}_g(p) \eta(\widehat{z}_g(p)) h_g(p) - \widehat{z}_g(p - \ln \kappa) \eta(\widehat{z}_g(p - \ln \kappa)) h_g(p - \ln \kappa).$$

We rewrite this equation as

$$gh_g'(p) - (\phi(z_g(p)) + \widehat{z}_g(p)\eta(\widehat{z}_g(p)))h_g(p) = -\phi(z_g(p-\ln\lambda))h_g(p-\ln\lambda)$$
$$-\widehat{z}_g(p-\ln\kappa)\eta(\widehat{z}_g(p-\ln\kappa))h_g(p-\ln\kappa).$$

Using the variation of constant formula, this equation yields a unique equation for $0 \leq p < \omega_g$ given $h_g(0)$. For $p \geq \omega_g$ the equation also yields a unique solution, however the initial condition is now

$$h_g\left(\frac{1-\beta}{\beta}\ln\frac{1}{1-\beta}\right) = \frac{1}{g}\int_{\frac{1-\beta}{\beta}\ln\frac{1}{1-\beta}-\ln\lambda}^{\frac{1-\beta}{\beta}\ln\frac{1}{1-\beta}} \phi(z_g(\widetilde{p}))h_g(\widetilde{p})\,d\widetilde{p}$$
$$+\frac{1}{g_q}\int_{\frac{1-\beta}{\beta}\ln\frac{1}{1-\beta}-\ln\kappa}^{\frac{1-\beta}{\beta}\ln\frac{1}{1-\beta}} \widehat{z}_g(\widetilde{p})\eta(\widehat{z}_g(\widetilde{p}))h_g(\widetilde{p})\,d\widetilde{p}.$$

Since the system is linear in the initial condition $h_g(0)$, therefore, there exists a unique $h_g(0)$ such that $\int_0^\infty h_g(p)\,dp = 1$. Notice that given the lemma below $\int_0^\infty h_g(p)\,dp < \infty$. ■

Let $z_g^* = z(v_g)$ and $\widehat{z}_g^* = \widehat{z}_g(v_g)$. Then for each $g > g_q^*$, define the $\chi(g)$ is the unique number χ satisfying

$$g = \phi(z_g^*)\frac{\lambda^\chi - 1}{\chi} + \widehat{z}^*\eta(\widehat{z}_g^*)\frac{\kappa^\chi - 1}{\chi}.$$

Lemma 2.6 *$\chi(g^*) = 1$ and $\chi(g) > 1$ $\forall g > g^*$ and in the neighborhood of g^*.*

Proof. By definition of g^* we have $g^* = \phi(z_g^*)(\lambda - 1) + \widehat{z}^*\eta(\widehat{z}_g^*)(\kappa - 1)$, therefore $\chi(g^*) = 1$. For $g > g^*$

$$g > \phi(z_g^*)(\lambda - 1) + \widehat{z}_g^*\eta(\widehat{z}_g^*)(\kappa - 1)$$

Thus $\chi(g) > 1$. ■

Lemma 2.7 (Tail Index) *There exists $\forall\xi > 0$ there exist $\overline{B}$,$\underline{B}$ and p_0 such that*

$$h_g(p) < 2\overline{B}e^{-(\chi(g)-\xi)p}, \forall p \geq p_0$$

and

$$h_g(p) > \frac{1}{2}\underline{B}e^{-(\chi(g)+\xi)p}, \forall p \geq p_0.$$

In other words, $h_g(p) = e^{-\chi(g)p}\varphi_g(p)$*, where* $\varphi_g(p)$ *is a slow-varying function.*

In order to prove this lemma we need the following lemma

Lemma 2.8 *For each* $\xi > 0$*, there exists a* $\delta > 0$ *such that*

$$\left(\frac{1}{g}\phi\left(z_g^*\right) + \delta\right)\frac{\lambda^{\chi-\xi} - 1}{\chi - \xi} + \left(\frac{1}{g}\widehat{z}_g^*\eta\left(\widehat{z}_g^*\right) + \delta\right)\frac{\kappa^{\chi-\xi} - 1}{\chi - \xi} < 1$$

and

$$\left(\frac{1}{g}\phi\left(z_g^*\right) - \delta\right)\frac{\lambda^{\chi+\xi} - 1}{\chi + \xi} + \left(\frac{1}{g}\widehat{z}_g^*\eta\left(\widehat{z}_g^*\right) - \delta\right)\frac{\kappa^{\chi+\xi} - 1}{\chi + \xi} > 1$$

Proof. This is true given $\frac{1}{g}\phi\left(z_g^*\right)\frac{\lambda^{\chi}-1}{\chi} + \frac{1}{g}\widehat{z}_g^*\eta\left(\widehat{z}_g^*\right)\frac{\kappa^{\chi}-1}{\chi} = 1$ and the functions

$$\frac{\lambda^{\chi-\xi} - 1}{\chi - \xi}, \frac{\kappa^{\chi-\xi} - 1}{\chi - \xi}$$

are strictly increasing in ξ. ■

For each $\xi > 0$ let $\delta > 0$ be such a δ. Given limit result in the last section, there exists a $p_0 = p_0(\delta) \geq \omega_g$ such that, for all $p \geq p_0$

$$\left|\left(\frac{1}{g}\phi\left(z_g(p)\right) + \frac{1}{g}\widehat{z}_g(p)\,\eta\left(\widehat{z}_g(p)\right)\right) - \left(\frac{1}{g}\phi\left(z_g^*\right) + \frac{1}{g}\widehat{z}_g^*\eta\left(\widehat{z}_g^*\right)\right)\right| < \delta$$

and

$$\left|\frac{1}{g}\phi\left(z_g\left(p - \ln\lambda\right)\right) - \frac{1}{g}\phi\left(z_g^*\right)\right| < \delta$$
$$\left|\frac{1}{g}\widehat{z}_g\left(p - \ln\kappa\right)\eta\left(\widehat{z}_g\left(p - \ln\kappa\right)\right) - \frac{1}{g}\widehat{z}_1^*\eta\left(\widehat{z}_g^*\right)\right| < \delta.$$

Let define $\overline{B}(\delta) = \max_{p_0 \leq p \leq p_0 + \ln\kappa} h_g(p)\, e^{(\chi-\xi)p}$ and $\underline{B}(\delta) = \min_{p_0 \leq p \leq p_0 + \ln\kappa} h_g(p)\, e^{(\chi+\xi)p}$.

Proof of the Tail Index Lemma. We show that

$$h_g(p) < 2\overline{B}(\delta)\, e^{-(\chi-\xi)p}, \forall p \geq p_0$$

and

$$h_g(p) > \frac{1}{2}\underline{B}(\delta)\, e^{-(\chi+\xi)p}, \forall p \geq p_0.$$

These inequalities hold for $p_0 \le p \le p_0 + \ln\kappa$ by definition. We will show that they also hold for all $p \ge p_0$ using contradiction. Suppose that there is $p > p_0 + \ln\kappa$ such that

$$h_g(p) \ge 2\overline{B}(\delta)\, e^{-(\chi-\xi)p}.$$

Consider the infimum of those p, then

$$h_g(p) = 2\overline{B}(\delta)\, e^{-(\chi-\xi)p}.$$

In the other hand, the equation determining h_g implies

$$\begin{aligned}
h_g(p) &= \frac{1}{g}\int_{p-\ln\lambda}^{p} \phi\left(z_g(\widetilde{p})\right) h_g(\widetilde{p})\, d\widetilde{p} + \frac{1}{g}\int_{p-\ln\kappa}^{p} \widehat{z}_g(\widetilde{p})\,\eta\left(\widehat{z}_g(\widetilde{p})\right) h_g(\widetilde{p})\, d\widetilde{p} \\
&< \int_{p-\ln\lambda}^{p} \left(\frac{1}{g}\phi\left(z_g^*\right) + \delta\right) 2\overline{B}(\delta)\, e^{-(\chi-\xi)p} d\widetilde{p} + \int_{p-\ln\kappa}^{p} \left(\frac{1}{g}\widehat{z}_g^*\eta\left(\widehat{z}_g^*\right) + \delta\right) 2\overline{B}(\delta)\, e^{-(\chi-\xi)p} d\widetilde{p} \\
&= 2\overline{B}(\delta)\left(\frac{1}{g}\phi\left(z_g^*\right) + \delta\right)\frac{\lambda^{\chi-\xi}-1}{\chi-\xi} e^{-(\chi-\xi)p} + 2\overline{B}(\delta)\left(\frac{1}{g}\widehat{z}_g^*\eta\left(\widehat{z}_1^*\right) + \delta\right)\frac{\kappa^{\chi-\xi}-1}{\chi-\xi} e^{-(\chi-\xi)p} \\
&< 2\overline{B}(\delta)\, e^{-(\chi-\xi)p}.
\end{aligned}$$

This yields a contraction. Therefore

$$h_g(p) < 2\overline{B}(\delta)\, e^{-(\chi-\xi)p}, \forall p \ge p_0.$$

Similarly, we can show that:

$$h_g(p) > \frac{1}{2}\underline{B}(\delta)\, e^{-(\chi+\xi)p}, \forall p \ge p_0.$$

■

As a consequence, if $g > g^*$, then $\chi(g) > 1$, Lemma 2.7 for $\xi = \frac{\chi(g)-1}{2}$ implies $\int h_g(p)\,dp < C\int e^{-\frac{1+\chi}{2}p}dp < \infty$.

When $g = g^*$ we have a better lower bound of h_g in the following lemma

Lemma 2.9 (Tail Index at the Limit) *There exists* $\underline{B}$ *and* p_0 *such that*

$$h_{g^*}(p) > \frac{1}{2}\underline{B}\frac{e^{-p}}{p}, \forall p \geq p_0.$$

Proof. Let choose $\underline{B} > 0$ such that the inequality hold for $p_0 \leq p \leq p_0 + \ln\kappa$. We will show that they also hold for all $p \geq p_0$ using contradiction. Suppose that there is $p > p_0 + \ln\kappa$ such that

$$h_{g^*}(p) < \frac{1}{2}\underline{B}\frac{e^{-p}}{p}.$$

Consider the infimum of those p, then

$$h_{g^*}(p) = \frac{1}{2}\underline{B}\frac{e^{-p}}{p}.$$

In the other hand, the equation determining h_{g^*} implies

$$\begin{aligned}
h_{g^*}(p) &= \frac{1}{g^*}\int_{p-\ln\lambda}^{p}\phi\left(z_{g^*}(\widetilde{p})\right)h_{g^*}(\widetilde{p})\,d\widetilde{p} + \frac{1}{g^*}\int_{p-\ln\kappa}^{p}\widehat{z}_{g^*}(\widetilde{p})\,\eta\left(\widehat{z}_{g^*}(\widetilde{p})\right)h_{g^*}(\widetilde{p})\,d\widetilde{p} \\
&> \int_{p-\ln\lambda}^{p}\left(\frac{1}{g^*}\phi\left(z_{g^*}^*\right) - Ce^{-\psi\widetilde{p}}\right)\frac{1}{2}\underline{B}\frac{e^{-\widetilde{p}}}{\widetilde{p}}d\widetilde{p} + \int_{p-\ln\kappa}^{p}\left(\frac{1}{g^*}\widehat{z}_{g^*}^*\eta\left(\widehat{z}_{g^*}^*\right) - Ce^{-\psi\widetilde{p}}\right)\frac{1}{2}\underline{B}\frac{e^{-\widetilde{p}}}{\widetilde{p}}d\widetilde{p} \\
&= \frac{1}{2}\underline{B}\frac{1}{g^*}\phi\left(z_{g^*}^*\right)(\lambda-1)\frac{e^{-p}}{p} + \frac{1}{2}\underline{B}\frac{1}{g^*}\widehat{z}_{g^*}^*\eta\left(\widehat{z}_{g^*}^*\right)(\kappa-1)\frac{e^{-p}}{p} \\
&+ \frac{1}{2}\underline{B}C'\frac{e^{-p}}{p^2} - \frac{1}{2}\underline{B}C''e^{-\left(1+\frac{\psi}{2}\right)p} \\
&> \frac{1}{2}\underline{B}\frac{e^{-p}}{p}.
\end{aligned}$$

(we choose p_0 such that $C'\frac{e^{-p}}{p^2} - \frac{1}{2}C''e^{-\left(1+\frac{\psi}{2}\right)p} > 0\ \forall p \geq p_0$) This yields a contraction. Therefore

$$h_1(p) > \frac{1}{2}\underline{B}\frac{e^{-p}}{p}, \forall p \geq p_0.$$

So

$$\int_0^\infty h_g(p)\,e^p dp > \int_{p_0}^\infty \frac{1}{2}\underline{B}\frac{1}{p}dp = \infty.$$

■

Lemma 2.10 *h_g is uniformly continuous in g. And for $g > g_q^*$*

$$\Phi(g) = \int h_g(p)\, e^p dp < \infty$$

is continuous in g. Moreover $\lim_{g \downarrow g_q^*} \Phi(g) = +\infty$.

Proof. The fact that h_g is uniformly continuous in g is a result of uniform continuity of $\{U_g\}$, thus of $z_g(.)$ and $\widehat{z}_g(.)$ as well. $\Phi(g)$ is finite given Theorem 2.2. $\Phi(g)$ is continuous by the Lebesgue dominated convergence theorem. Finally, as we show above $\Phi(g^*) = +\infty$ and by the uniform continuity of h_g:$\lim_{g \downarrow g_q^*} \Phi(g) = +\infty$. ■

For any $\omega > 0$, ε is defined as in (2.76). The corresponding stationary distribution is $h_\varepsilon(p) = h_g(p - \ln \varepsilon)\ \forall p \geq \ln \varepsilon$.Given g, g' is defined in (2.59) can be written using $h_g = F_y$ as

$$\begin{aligned} g' &= \frac{(\lambda - 1) \int_{\ln \varepsilon}^{\infty} e^p \phi(z_\varepsilon(p))\, h_\varepsilon(p)\, dp + (\kappa - 1) \int_{\ln \varepsilon}^{\infty} e^p \widehat{z}_\varepsilon(p)\, \eta(\widehat{z}_\varepsilon(p))\, h_\varepsilon(p)\, dp}{1 - h_\varepsilon(\ln \varepsilon)(\omega - \varepsilon)} \\ &= \varepsilon \frac{(\lambda - 1) \int_0^{\infty} e^p \phi(z_g(p))\, h_g(p)\, dp + (\kappa - 1) \int_0^{\infty} e^p \widehat{z}_g(p)\, \eta(\widehat{z}_g(p))\, h_g(p)\, dp}{1 - h_g(0)(\omega - \varepsilon)} \end{aligned}$$

We obtain an BGP if $g' = g$.

Step 3: There exists $g^*(\omega)$ such that $g' = g$.

Theorem 2.1 (Existence of the Epsilon Economy) Suppose

$$\mu_e \in \left(U_{g^*}\left(\frac{1-\beta}{\beta} \log\left(\frac{1}{1-\beta} \right) \right), U_{g^*}\left(\frac{1-\beta}{\beta} \log\left(\frac{1}{1-\beta} \right) + \delta \right) \right).$$

Consider $\Delta > 0$ such that

$$\mu_e \in \left(U_g\left(\frac{1-\beta}{\beta} \log\left(\frac{1}{1-\beta} \right) \right), U_g\left(\frac{1-\beta}{\beta} \log\left(\frac{1}{1-\beta} \right) + \delta \right) \right)$$
$$\forall g \in [g^*, g^* + \Delta]$$

and $\varpi > 0$ sufficiently small such that $\forall 0 < \omega < \varpi : \varepsilon = \frac{\omega}{\omega_{g^*+\Delta}}$ satisfies

$$\varepsilon \frac{(\lambda - 1) \int_0^\infty e^p \phi\left(z_{g^*+\Delta}(p)\right) h_{g^*+\Delta}(p)\, dp + (\kappa - 1) \int_0^\infty e^p \widehat{z}_{g^*+\Delta}(p)\, \eta\left(\widehat{z}_{g^*+\Delta}(p)\right) h_{g^*+\Delta}(p)\, dp}{1 - h_{g^*+\Delta}(0)(\omega - \varepsilon)}$$

$$< g^* + \Delta.$$

Notice also that Δ should also be small enough to apply Lemma 2.1.For each ω such that $0 < \omega < \varpi$ There exists a $g = g(\omega) \in (g^*, g^* + \Delta)$ such that $g' = g$. Moreover

$$\lim_{\omega \longrightarrow 0} g(\omega) = g^*.$$

Proof. Since

$$\mu_e \in \left(U_g\left(\frac{1-\beta}{\beta} \log\left(\frac{1}{1-\beta}\right)\right), U_g\left(\frac{1-\beta}{\beta} \log\left(\frac{1}{1-\beta}\right) + \delta\right)\right)$$

there exists $\omega_g : \log(\omega_g) \in \left(\frac{1-\beta}{\beta} \log\left(\frac{1}{1-\beta}\right), \frac{1-\beta}{\beta} \log\left(\frac{1}{1-\beta}\right) + \delta\right)$ such that $U_g(\omega_g) = \mu_e$. Let define $\varepsilon = \frac{\omega}{\omega_g}$ and

$$D(g) = \varepsilon \frac{(\lambda - 1) \int_0^\infty e^p \phi\left(z_g(p)\right) h_g(p)\, dp + (\kappa - 1) \int_0^\infty e^p \widehat{z}_g(p)\, \eta\left(\widehat{z}_g(p)\right) h_g(p)\, dp}{1 - h_g(0)(\omega - \varepsilon)}$$

$$- g.$$

Using Lemma 2.10, we can show that $D(g)$ is continuous in g. Moreover

$$D(g^* + \Delta) < 0$$

and

$$\lim_{g \longrightarrow g^*} D(g) = +\infty$$

therefore there exists $g(\omega)$ such that $D(g) = 0$. Moreover if $g(\omega) > g^* + \varpi$ when $\omega \longrightarrow 0$ then the first part of $D(g(\omega))$

$$\varepsilon \frac{(\lambda - 1) \int_0^\infty e^p \phi\left(z_g(p)\right) h_g(p)\, dp + (\kappa - 1) \int_0^\infty e^p \widehat{z}_g(p)\, \eta\left(\widehat{z}_g(p)\right) h_g(p)\, dp}{1 - h_g(0)(\omega - \varepsilon)} \longrightarrow 0$$

given that $\varepsilon \longrightarrow 0$. So $D(g(\omega)) < -(g^* + \varpi) < 0$. This yields a contradiction with the fact that $D(g(\omega)) = 0$. Therefore

$$\lim_{\omega \longrightarrow 0} g(\omega) = g^*.$$

■

Chapter 3

Racing: When Should We Handicap the Advantaged Competitor?

3.1 Introduction

A race is a contest among two or more competitors who exert effort to win a prize. Sport contests, such as bicycle races, golf tournaments and basketball championships, are the most popular forms of races. Races studied in economic theory include patent races and contests for job promotion. An important question in the study of races is how a race designer who wants to promote effort from all competitors should treat competitors with different capabilities differently? Should he give more incentives to the advantaged competitors or the disadvantaged ones?

We observe in reality both types of races, in which advantaged or disadvantaged competitors are given higher incentives. In golf's stroke play competition, the competitor's handicap is subtracted from the total "gross" score at the end of each round, to calculate a "net" score against which standings are calculated. Therefore, golfers with higher handicaps are given advantage. In labor markets, Lazear and Rosen (1981) find that reverse discrimination, where less able workers are given a head start or rewarded more lucratively if they happen to accomplish the unlikely and win job promotion contests, can be consistent with efficient incentive mechanisms. However, in NSF research grants allocation processes, researchers with higher past performance are more likely to win the grants.

This paper introduces a continuous time, continuous state-space model based on Harris and Vickers (1987)'s tug-of-war to address the question above. In this benchmark model, two players compete for a final reward. The reward is won by the first player who achieves a given distance over his rival. At any moment when the race is ongoing, each player puts in effort which influences the distance between him and his rival: a Brownian motion with a drift that depends on the players' effort. The cost of effort functions are strictly convex. The model is particularly suitable to describe job promotion contests. An employee is promoted when he demonstrates significantly better qualities than his co-worker do.

The criteria I choose to select among different designs of races are the expected completion time and the final rewards. Beside final rewards, completion time is especially important in patent races. For example, the Human Genome Project, officially founded in 1990, was expected to be completed before 2005.

I consider the set of Markovian Perfect Equilibria (MPEs) in which equilibrium strategies of the players are conditioned only on the current distance between them. As in the discrete state-space model in Harris and Vickers (1987), the Markov Perfect Equilibrium (MPE) strategies exhibit a *discouragement effect*: The players exert high effort only when they are close to each other. When a player is left further behind by his rival, he reduces his effort given his slim chance of winning. The rival who gets further ahead therefore faces less competition and can safely reduces his effort. A larger distance between the two players thus discourages both players.

My answer to the initially asked question is that when the discouragement effect is strong we should reduce the promised reward to the advantaged player or handicap him. This is because reducing the promised reward to him or handicapping him mitigates the effect, therefore increases both players' effort and reduces the expected completion time. However, when the discouragement effect is weak, the direct incentive effect of reducing the promised reward to the advantaged player or handicapping him decreases his effort. Consequently, the expected completion time increases instead of decreases.

The intuition behind this result is the following. If we handicap the advantaged player, or if we promise him a lower reward, and hold the strategy of the disadvantaged player constant, the advantaged player exerts lower effort because of direct incentives. However, the disadvantaged player increases his effort because of his higher chance of winning the race. This increase in effort of the disadvantaged player in turn induces a higher effort from the advantaged player, due to their strategic interaction. In total, lower incentives to the advantaged player has an ambiguous effect on his effort. By reducing incentives to the advantaged player, we can, therefore, induce higher effort from him, as well as higher effort from the disadvantaged player.

Moreover, I show that the discouragement effect is stronger under the following conditions: higher final rewards, lower cost of effort, less uncertainty, more patient players or less convex cost-of-effort functions. Therefore, under these conditions, reducing incentives to the advantaged players will raise the players' effort and shorten the expected completion time more effectively.

In addition to the tug-of-war model above, I consider another continuous time, continuous state-space model of races which is more suitable to describe patent races. The departure from the benchmark model is that a player wins when he reaches a given finish line, independent of his distance to his rival. The MPE strategies now depend on the two distances of the two players to the finish line. Numerical exercises suggest that all the results for the benchmark model carry over to this model.

The same question about competition design has been asked by Acemoglu and Akcigit (2006) in the context of optimal intellectual property right policy. They find that the welfare-maximizing policy involves state-dependent intellectual property right protection: providing greater protection to the technological leaders that are further ahead than to those that are close to their followers. This is a result of the "trickle-down effect": greater protection to the leaders with higher technological gaps and lower protection to the leaders with lower technological gaps does not only encourage the R&D investment of the former but also of the latter because of the prospect to the latter of reaching a higher technological gap to benefit from higher protection. In my model, the final reward can be considered as a degree of patent protection to the winner of a R&D race. Thus, offering a high final reward and handicapping the advantaged player simultaneously encourage his effort.

Moscarini and Smith (2007) is the first paper to address the optimal design of the race in a similar continuous time, continuous state-space model. The authors study the optimal score function as a function of the distance between the two players. However, they only ask the question how to treat ex-post different players that are otherwise ex-ante identical. The optimal score function "taxes" the a leader and "subsidizes" the follower at any moment of the race.

Moscarini and Smith take a different approach to solving the model, relying on the symmetry of Markov Perfect Equilibria. Besides restricting their attention to symmetric equilibria, the authors mostly work with quadratic cost functions and consider only the case of no discounting. In contrast, I use the theory of boundary value problems for systems of second-order differential equations developed in Hartman (1964); this theory allows me to consider the model in full generality without restricting attention to symmetric equilibria and with discounting. Especially, I can consider the case in which players have different abilities to address the question in the title. My results are related to Moscarini and Smith's: They find that the total expected effort is decreasing in the size of the final reward at high levels due to the same discouragement effect mentioned earlier.

My paper also contributes to the relatively sparse literature of modelling dynamic competition.Harris and Vickers (1987) is a pioneering paper with a model in discrete state-space. In their model, they prove that at least one equilibrium exists and characterize some of its properties. However, they only prove uniqueness of the symmetric equilibria and they do not allow players to discount future. In the continuous time and continuous state-space version of their paper that I consider, I prove the existence and uniqueness of equilibrium MPE strategies under some weak conditions on the cost functions and final rewards. Also, the general model I consider does not rule out discounting. In some special cases, the continuous time framework delivers a closed-form solution of MPEs which facilitates the characterization of equilibrium strategies[1]. The equilibrium strategies share basic properties with the equilibrium strategies in the discrete time model.

Budd, Harris, and Vickers (1993) also solve a similar model using boundary value representations. Their method only applies when the discount rate r goes to infinity. Another continuous time continuous state-space version of the Harris and Vickers model is developed in Horner (1999). He restricts the action space to be finite, allowing for only two levels of effort.

[1] In the case with identical players, Moscarini and Smith (2007) find a closed-form identical to mine up to an affine transformation. Their method relies on symmetric equilibria. My method covers asymmetric equilibria also.

Hence, the MPE strategies are such that players switch their actions only infrequently based one some threshold rule.

In the next section, I present the model. I prove the existence and uniqueness of MPEs with general cost functions under some weak restrictions, both with and without discounting. I also illustrate these theorems for the case of quadratic cost functions. In Section 3, I study the properties of equilibrium strategies including the discouragement effect. For that purpose, in the case of quadratic costs and no discounting, I use closed-form solutions for MPE and in the case of general cost functions with discounting, I use numerical analysis. In Section 4, I show that if the discouragement effect is strong we should lower the promised reward to the advantaged player to encourage effort provision from both players, and to reduce the expected completion time. Section 6 concludes.

3.2 The Model

Two players, A and B, engage in a contest for a final reward in continuous time. At each moment, each player chooses an effort, x_A for player A which costs him $c_A(x_A)$ and x_B for player B which costs him $c_B(x_B)$. In Harris and Vickers (1987), A and B are two research firms competing for an exclusive patent. The effort can then be interpreted as money spent on laboratories, equipment, researchers, etc. Let $z \in \mathbf{R}$ denote the distance between the two players. The race starts at $z = 0$, and a player wins the race if he attains a certain lead over the other player: Player A wins the race with reward P_A when he reaches his lead over B, $z = K_A$, and B wins the race with reward P_B when he reaches his lead over A, $z = -K_B$. Therefore, z is also the only payoff relevant state of the race. Without loss of generality I assume $K_A, K_B > 0$.

The uncertainty is incorporated in the temporal evolution of the state z_t :

$$dz_t = (x_{At} - x_{Bt})\,dt + \sigma dW_t, \tag{3.1}$$

where W_t is a standard Brownian motion and $W_0 = 0$.

I assume that the cost functions are twice continuously differentiable, strictly increasing and strictly convex: $c_i'(.) > 0$ and $c_i''(.) > 0$, $i \in \{A, B\}$. I also assume that players do not bear any cost if they do not exert effort, $c_i(0) = 0$ and the Inada conditions at 0 and ∞ are satisfied:

$$c_i'(0) = 0 \text{ and } \lim_{x \longrightarrow +\infty} c_i'(x) = +\infty.$$

These conditions ensure that in each moment, the effort choice of each player is a well-defined maximization problem. In addition, the last assumption implies that players always have an incentive to exert effort. Lastly, I restrict players' efforts to the compact set $[0, \overline{x}]$, where the upper bound $\overline{x}$ is chosen large enough[2]. This condition ensures that the evolution of the distance z_t as well as the expected payoff function of each player are well-defined.

The expected payoff to player i is

$$E\left[e^{-r\tau} P_i \mathbf{1}_{\{i \text{ wins}\}} - \int_0^\tau e^{-rt} c_i(x_{it})\, dt\right], \tag{3.2}$$

where τ is the finish time of the race, it is the first time where either z_t reaches K_A, player A wins the race, or z_t reaches $-K_B$, player B wins the race. The indicator function indicates who wins the race. Notice that τ is a random variable depending on the uncertain evolution of the race; or more precisely, it is a stopping time. The race starts at $z_0 \in (-K_B, K_A)$. There are two components of the payoff functions. The first part is the discounted reward $e^{-r\tau} P_i$ if player i wins the race, and the second part is the discounted cost of effort, $e^{-rt}\ c_i(x_{it})$, that player i continuously makes during the race. Each player chooses a strategy maximizing his expected payoff given his rival's strategy.

Under the restrictions on the cost functions and the effort choice of the players, the payoff functions are well-defined. The problem now is to find the strategy functions, $X_A = \{x_{At}\}$ and $X_B = \{x_{Bt}\}$, such that each player maximizes his expected payoff given his rival's strategy:

$$J_i(z) = \max_{X_i \text{ given } X_{-i}} E_0\left[e^{-r\tau} P_i \mathbf{1}_{\{i \text{ wins}\}} - \int_0^\tau e^{-rt} c_i(x_{it})\, dt \,\middle|\, z_0 = z\right],$$

in which i is again either A or B and $-i$ is the other player, i.e., B or A respectively.

[2] It will be shown later that $\overline{x}$ can be $\max\left\{c_A^{-1}(M), c_B^{-1}(M)\right\}$ in which M is determined as $\int_{\frac{P_A+P_B}{K_A+K_B}}^{M} \frac{s ds}{\frac{2r}{\sigma^2}\max(P_A, P_B) + \frac{2}{\sigma^2}\left(\left(c_A'\right)^{-1}(s) + \left(c_B'\right)^{-1}(s)\right)s} = P_A + P_B$. Harris and Vickers (1987, p7) also assumes this. However, they do not show conditions under which the bound is not binding.

It is well-known that any MPE is a subgame perfect equilibrium[3]. An analogy for continuous time games is that, if strategy x_{Bt} is Markovian, i.e. function of z_t only, then x_{At} can be chosen from the class of Markovian strategies, and vice versa; therefore, I can restrict myself to cases where both strategies are Markovian.

I further restrict myself to the set of equilibria with twice differentiable value functions in order to write the second derivatives. We can then obtain the Hamilton-Jacobi-Bellman equations using the dynamic programming principle:

$$\max_{0\leq x\leq \overline{x}} \left(-c_i(x) - rJ_i(z_t) + (x - x_{-i}(z_t)) J_i'(z_t) + \frac{\sigma^2}{2} J_i''(z_t)\right) = 0. \tag{3.3}$$

At each moment, the effort choice of each player involves the trade-off between the current convex cost of effort, $-c_i(x)$ with higher chance of winning, taken the other player's strategy as given,$(x - x_{-i}(z_t)) J_i'(z_t)$. Each player also discounts the future payoff,$-rJ_i(z_t)$, and takes into account the uncertainty evolution of the state z, $\frac{\sigma^2}{2} J_i''(z_t)$. The first order conditions from (3.3) determine the effort levels of players as functions of the derivatives of their value functions:

$$(x_A(z), x_B(z)) = \left(f_A\left(J_A'(z)\right), f_B\left(J_B'(z)\right)\right) \tag{3.4}$$

where

$$f_i(k) = \begin{cases} 0 & \text{if } k \leq 0 \\ \min\left\{\left(c_i'\right)^{-1}(k), \overline{x}\right\} & \text{otherwise} \end{cases} \tag{3.5}$$

These equations determine x_A as a function of $J_A'(z)$ and x_B as a function of $J_B'(z)$. Finally, the boundary conditions for J_A and J_B are

$$\begin{aligned} J_A(K_A) &= P_A, J_B(K_A) = 0 \\ J_A(-K_B) &= 0, J_B(-K_B) = P_B. \end{aligned} \tag{3.6}$$

These boundary conditions are intuitive: when A is K_A ahead of B, he wins the reward P_A, B receives nothing. The interpretation for the case in which B is K_B ahead of A is similar.

Definition 3.1 *A Markov Perfect Equilibrium (MPE) is a pair of equilibrium payoff func-*

[3]See for example Acemoglu (2008).

tions $(J_A(z), J_B(z))$ *satisfying the Hamilton-Jacobi-Bellman equations* (3.3) *and the boundary conditions* (3.6) *and a pair of equilibrium strategies* $(x_A(z), x_B(z))$ *given by* (3.4).

In this model, the only payoff relevant state is the distance between the two players because the outcome of the race only depends on the distance. While this model might be a suitable description of some type of races such a race for job promotion or tie-breaks in tennis, it is not a good model for patent races in which a player wins if he achieves a certain discovery, not his progress relative to the other player. In a separate paper, I develop a model for this situation. The payoff relevant state is a vector of two numbers, distance of each player to a finish line. However, that model is less tractable, and I can only solve it numerically. On the other hand, in discrete state spaces, Harris and Vickers (1987) argue that the tug-of-war race is a close approximation of that model.

The task of finding MPE strategies becomes solving a second-order boundary value problem on $(J_A(z), J_B(z))$. We first solve the effort choice given the incentive $J_i'(z)$ as in (3.4).Then, plug the effort choice into(3.3), we can re-write the Hamilton-Jacobi-Bellman equations as an explicit second-order boundary value problem:

$$\begin{pmatrix} J_A''(z) \\ J_B''(z) \end{pmatrix} = \frac{2}{\sigma^2} \begin{pmatrix} rJ_A(z) + F_A(J_A'(z), J_B'(z)) \\ rJ_B(z) + F_B(J_A'(z), J_B'(z)) \end{pmatrix} \tag{3.7}$$

with the boundary conditions

$$(J_A(K_A), J_B(K_A)) = (P_A, 0) \text{ and } (J_A(-K_B), J_B(-K_B)) = (0, P_B)$$

where

$$\begin{aligned} F_A(J_A', J_B') &= (f_A(J_A') - f_B(J_B'))\, J_A' - c_A(f_A(J_A')) \\ F_B(J_A', J_B') &= (f_A(J_A') - f_B(J_B'))\, J_B' - c_B(f_B(J_B')) \end{aligned} \tag{3.8}$$

Rewriting the Hamilton-Jacobi-Bellman equations as a boundary value problem allows me to use the theory of boundary value for system of second order differential equation developed in Hartman (1964). Using this system, some preliminary properties of the payoff functions can be shown. First, the payoff functions are strictly positive except at the two boundaries. This is because, whenever the race is not yet concluded, a player can choose to stay in the race and to exert no effort, but he still has a positive probability of winning due to the uncertain evolution of the state z. Second, the closer a player is to his goal, the higher his expected payoff is because he has more chance of winning. Hence, the slope of the payoff function, which is the incentive determining the effort level of each player, is strictly positive; so each player will exert a strictly positive effort at any moment of the race.

Lemma 3.1 *Suppose the players' discount rate is positive, $r \geq 0$, a solution of the payoff functions (J_A, J_B) to the system (3.7) satisfies*

1. ***Strict positivity of the payoff functions:*** *$J_A(z), J_B(z) > 0$ for all $z \in (-K_B, K_A)$. Given the option to exert no effort, and the Brownian evolution of the distance between the two players, each player has a strictly positive probability of winning the race without incurring any cost of effort, their payoff functions are strictly positive whenever the race is not yet concluded.*

2. ***Strict positivity of incentives:*** *$J'_A(z) > 0$ and $J'_B(z) < 0$ for all $z \in (-K_B, K_A)$. As each player moves closer to his goal, he has higher probability to win the race, therefore, his payoff function is higher. Since the incentives are strictly positive, the players always exert a non-zero level of effort.*

Proof. In the Appendix, using the Gronwall's Inequality. ■

Example 3.1 (Quadratic Cost) *In this example I consider the case with quadratic cost which will be used extensively to explore the main economic results of this paper. This example is also studied in Moscarini and Smith (2007). However the general formulation here allows for discounting and asymmetric cost function. I also establish the equivalence between lower cost of effort, lower uncertainty and higher final reward, which will be useful for studying the case with heterogeneity in cost functions.*

Let the cost of effort functions be quadratic

$$c_A(x) = \frac{x^2}{2\alpha}, c_B(x) = \frac{x^2}{2\beta}. \tag{3.9}$$

The higher α is (β), the less effort costs to player A (B). Again by choosing a large the upper bound on the effort of player, we can suppose this bound is not binding. Therefore, (3.4) *implies a linear relationship between efforts and slopes of the payoff functions:*

$$\begin{aligned} x_A(z) &= \alpha J_A'(z) \\ x_B(z) &= -\beta J_B'(z). \end{aligned} \tag{3.10}$$

Thus (3.8) *simplifies to* $F_A(u,v) = -\alpha u^2 - 2\beta uv, F_B(u,v) = -2\alpha uv - \beta v^2$. *Let* $\widetilde{J}_A(z) = \frac{\alpha}{\sigma^2} J_A(\sigma z)$ *and* $\widetilde{J}_B(z) = \frac{\beta}{\sigma^2} J_B(\sigma z)$, *the boundary conditions become*

$$\begin{aligned} \widetilde{J}_A\left(-\frac{K_B}{\sigma}\right) &= 0, \widetilde{J}_A\left(\frac{K_A}{\sigma}\right) = \frac{\alpha P_A}{\sigma^2} = \widetilde{P}_A \\ \widetilde{J}_B\left(-\frac{K_B}{\sigma}\right) &= \frac{\beta P_B}{\sigma^2} = \widetilde{P}_B, \widetilde{J}_B\left(\frac{K_A}{\sigma}\right) = 0 \end{aligned}$$

with the differential equations on $J_i(z)$, $i \in \{A, B\}$

$$\begin{aligned} -2r\widetilde{J}_i(z) + \left(\widetilde{J}_i'(z)\right)^2 + 2\widetilde{J}_i'(z)\widetilde{J}_{-i}'(z) + \widetilde{J}_i''(z) &= 0 \\ \forall z &\in (-K_B, K_A) \end{aligned} \tag{3.11}$$

Therefore, holding everything else constant, a player would be indifferent between seeing its cost decreases from $\frac{x^2}{\alpha}$ *to* $\frac{x^2}{\alpha'}$ *and seeing the final reward augmented by* $\frac{\alpha'}{\alpha}$. *It is then enough to consider the system where* $\alpha = \beta = \sigma = 1$ *and then, interpret the result using the system* (3.11). *In this special case,* (3.10) *becomes*

$$\begin{aligned} x_A(z) &= J_A'(z) \\ x_B(z) &= -J_B'(z). \end{aligned} \tag{3.12}$$

This intuition is that the higher the slopes of the payoff functions, the higher the incentives to the players to exert more effort. Substituting these effort functions into (3.3), *we have finally* *the system of differential equation*

$$\begin{pmatrix} J''_A(z) \\ J''_B(z) \end{pmatrix} = \begin{pmatrix} 2rJ_A(z) - J'_A(z)\,J'_B(z) - (J'_A(z))^2 \\ 2rJ_B(z) - J'_A(z)\,J'_B(z) - (J'_B(z))^2 \end{pmatrix}, \tag{3.13}$$

with the boundary conditions as in (3.7). *Lemma 3.1 shows that* $J'_A(z) > 0$ *and* $J'_B(z) < 0$, *so* $0 < J_A(z) < P_A$ *and* $0 < J_B(z) < P_B$ $\forall z \in (-K_B, K_A)$.

3.3 Existence and Uniqueness of Markov Perfect Equilibrium

Before analyzing the equilibrium strategies and outcomes of the race, it is important to prove the existence of Markov Perfect Equilibria and their uniqueness, or equivalently the existence and uniqueness of the solution to the boundary value problem (3.7).The steps of the existence and uniqueness proof are in the Appendix.

Theorem 3.1 *Suppose that*

$$\int^{+\infty} \frac{ds}{(c'_A)^{-1}(s) + (c'_A)^{-1}(s)} = +\infty \tag{3.14}$$

then (3.7) *has at least one solution.*

Remark 3.1 *This condition is satisfied if* $c''_i(x)$ *are bounded below from* 0 *at infinity; i.e., there exists an* δ *and an* $x^* > 0$ *such that* $c''_i(x) > \delta$ $\forall x > x^*$. *Geometric cost functions* $c_i(x) = c_i x^{k_i}$ *with* $k_i \geq 2$ *satisfy this condition, in particular quadratic cost functions satisfy this condition since they have constant second derivatives.*

This condition means that $c_A(.)$ and $c_B(.)$ are "sufficiently" convex at least at infinity. If $c_A(.)$ and $c_B(.)$ are not too convex, for example, in the extreme, when they are both linear, players will exert high effort and might reach any upper bound on the efforts. I rule out this situation to avoid imposing any ad-hoc bound on effort of the players.Horner (1999) is an example of races where $c_A(.)$ and $c_B(.)$ are linear. In equilibrium, the players only choose between two levels of effort which can be interpreted as the bounds that he imposes on the efforts of the players given the linearity of the cost functions. The MPE strategies are such that players switch their actions only infrequently based on some threshold rule. This structure of MPEs is thus too different from equilibria in my model.

As in other mathematical and economic models, it is more difficult to ensure the uniqueness of equilibria. As a result, the condition to ensure the uniqueness of the MPE is more stringent. It requires conditions on the cost of effort functions and that the final rewards are sufficiently small.

Theorem 3.2 *Suppose that* (3.14) *is satisfied and that* $c_i''(x)$ *are bounded below away from* 0 *when* x *goes to* 0, *i.e., there exists* ϵ *and* δ *such that*

$$c_i''(x) > \delta \; \forall 0 < x < \epsilon.$$

Then there exists a $\overline{P}$ *finite such that* (3.7) *has a unique solution when* $0 < P_A, P_B < \overline{P}$.

The conditions on the cost functions guarantee that the players' equilibrium efforts go to 0 as P goes to 0. When the equilibrium efforts or equivalently the first derivative of the value functions are sufficiently small, the boundary value problem (3.7) admits a unique solution following Hartman (1964). Going back to Example 3.1, we show in the Appendix that the equilibrium effort of players are bounded by

$$M = \sqrt{\exp\left(2\left(P_A + P_B\right)\right)\left(2rP + \left(\frac{P_A + P_B}{K_A + K_B}\right)^2\right) - 2r\max\left(P_A, P_B\right)} \tag{3.15}$$

and the MPE is unique if

$$M < \sqrt{\frac{1}{3}\left(4r + \frac{\pi^2}{\left(K_A + K_B\right)^2}\right)}. \tag{3.16}$$

When there is no discounting, the closed-form derivation of the MPE in the Appendix also shows the uniqueness without any of these restrictions.

3.4 The Discouragement Effect

The previous section establishes the general existence and uniqueness of the MPE. In this section, I investigate some properties of the MPE strategies. A striking property is that higher distance between the leader and the follower discourages both from exerting effort, which is often mentioned as the *discouragement effect.* This effect leads to an ambiguous effect of incentives, such as higher final reward to the winner of the race on the total expected effort of the players.Moscarini and Smith (2007) show that a higher final reward does not necessarily increase the total expected effort of the players. This discouragement effect is the key determinant of why handicapping the advantaged player will reduce the expected completion time of the race. The factors that affect the intensity of this effect are the final rewards, the amount of uncertainty, the level of the cost of effort to the two players, their discount rates and the degree of convexity of their cost of effort functions. In the case of quadratic costs and no discounting, I show the first three factors analytically using the closed-form solutions, and in the case of general cost functions and with discounting, I show the last two factors numerically.

3.4.1 Quadratic Cost and No Discounting

Consider the case of quadratic cost and no discounting. Following Example 3.1, I can restrict myself to the case where the cost functions are $c_A(x) = c_B(x) = \frac{x^2}{2}$ and $r = 0$, $\sigma = 1$. Under these restrictions, the model has a closed-form solution for the Markov Perfect Equilibria. The closed-form solution also allows me to obtain an analytical expression for the expected completion time of the race as function of model parameters.

I study the case in which players do not discount future ($r = 0$) to disentangle strategic interaction effects from the discounting effect on strategies of the players. I find two properties of the MPE strategies which are similar to the discrete time MPE strategies in Harris and Vickers (1987). First, the leader in the race puts in higher efforts than the follower does. Second, efforts increase as the gap between players decreases. Other R&D competition models share the second property of MPE strategies. For instance, Aghion, Harris, Howitt, and Vickers (2001) and Acemoglu and Akcigit (2006), both find that effort is highest when firms are technologically close to each other. The first property, however, does not hold in all models. For instance, the models of Acemoglu and Akcigit (2006) and Reinganum (1983) have the opposite property. In their model, there is an Arrow's replacement effect, i.e., the leading firm receives flow profits before successful new innovations, so it has relatively weaker incentive than the follower to stochastically shorten the random time to the next innovation. In contrast, in my model, players only receive reward at the end; the Arrow's replacement effect is thus not present.

The model admits closed-form solutions of players' strategies as functions of the state z, the distance between the two players. The pair of the strategy functions is a solution to a vector-valued first-order boundary problem. The closed-form solution is derived in the Appendix with a special parameterization $g = g(z)$ where

$$g - \frac{1}{g} + 2\ln(g) = C_1 z + C_2 \tag{3.17}$$

In the Appendix, I show that g is the ratio of player A's effort to player B's effort. Since $f(g) = g - \frac{1}{g} + 2\ln(g)$ is strictly increasing over the interval $(0, +\infty)$ and

$$\begin{aligned} \lim_{g \longrightarrow 0} f(g) &= -\infty \\ \lim_{g \longrightarrow \infty} f(g) &= +\infty, \end{aligned}$$

for each z there exists a unique $g(z)$ satisfies (3.17).We have C_1 greater than 0, thus $g(z)$ is increasing in z, i.e., a player exerts relatively higher effort than his rival does when the former is closer to his goal. I define the "pivot" of the race as the state z^*, where joint expected payoffs are minimized:

$$z^* = \arg\min_z (J_A(z) + J_B(z))$$

Since

$$J_A(z) + J_B(z) = \ln\left(\frac{(1 + g(z))^2}{g(z)}\right) + const,$$

the joint expected payoff is minimized at $g = 1$, or, equivalently, at $z^* = g^{-1}(1)$. The definition of being leader or follower here is with respect to the pivotal point z^*. Since we do not rule out that players have different cost functions or are offered different rewards, being a leader does not necessarily imply being closer to one's goal under this definition of leadership.

Proposition 3.1 *Suppose that $z > z^*$, i.e., player A is relatively closer to his goal, $z = K_A$, than player B is to his goal, $z = -K_B$ then*

1. *Player A exerts higher effort than player B does: $x_A(z) > x_B(z)$.*

2. *Player B reduces his effort as he gets further behind: $x_B(z)$ is decreasing in z.*

3. *Once further enough ahead, player A will start slowing down: There exists a* $z_A^* > z^*$ *such that* $x_A(z)$ *is decreasing over* (z_A^*, K_A).

Proof. *Using the closed forms of the equilibrium strategies in the Appendix.* ■

Literally, the leader works harder than the follower does, the follower slows down as he gets further behind, and once a leader has started slowing down, he will continue to do so. The second and third properties are respectively the discouragement effect on the follower and the leader.

The higher the final reward, the stronger the discouragement effect. When P_A and P_B are large, the two players will both exert high effort only when they are close to each other, however when one player gets further ahead of his rival, he wants to reduce his effort because the cost of effort is too high to him. He can safely reduce his effort since the continuous time, continuous state-space and perfect information features of the race allow him to commit to engage in a war phase with high effort when his rival gets closer to him. Given this strategy, his rival also reduces effort because of the smaller chance to win the race. As $P_A = P_B = P$ goes to infinity, both players only exert infinitely high effort over a infinitely small distance to each other. As one of them takes the lead, the other reduces his effort to almost 0, and the leader exerts an infinitesimal effort level.

Moreover, the equivalence result in Example 3.1 shows that lower cost of effort delivers the same equilibrium strategies as if the cost of effort unchanged but the final rewards are higher. Thus, the lower the cost of effort to the players, the stronger the discouragement effect. Lower cost of effort allows the players to sustain their strategy more cheaply.

Finally, also by the equivalence result, a lower uncertainty on the evolution of the state of the race, i.e. lower σ, corresponds to higher P_A and P_B, and thus a stronger discouragement effect. Indeed, the equivalent strategies are the same as in the case of unit uncertainty $\sigma = 1$ and the final rewards are respectively $\widetilde{P}_A = \frac{P_A}{\sigma^2}$ and $\widetilde{P}_B = \frac{P_B}{\sigma^2}$. In the limiting case when, there is no uncertainty, i.e., $\sigma = 0$, the disadvantaged player knows that the advantaged player will rationally outdo any effort he makes. This credible threat discourages the weaker player from making any effort.Fudenberg et al. (1983) and Harris and Vickers (1985) stress the same point.

3.4.2 General Case

Other factors that affect the players' strategic behaviors are the discount rate and the degree of convexity of the cost of effort functions. Higher discount rate and higher degree of convexity weaken the discouragement effect. I do not have closed-form solutions for MPE in this general case; however the Hamilton-Jacobi-Bellman equations can be solved numerically using discretization to obtain the corresponding MPE strategies and the expected completion time.

I first use the numerical solution to study the interaction between the players' impatience and strategic motives in their effort choice. When a player is behind, the discouragement effect and discounting both serve to lower effort provision. However, when a player is sufficiently ahead, the strategic motivation, as analyzed in the previous sections, reduces his incentive to provide greater effort, whereas discounting operates in the opposite direction. When the discount rates are high, the impatience is strong enough to cancel the slowing down interval in which the leader of the race reduces his effort after getting further ahead from the follower.

Second, the numerical solutions also shed light on the interaction between the convexity of the cost of effort functions and strategic motives. When the cost functions are sufficiently convex, the players tend to smooth their effort. Thus, they hold on a more constant level of effort even if they get further ahead or behind of their rival.

3.5 Handicapping the Advantaged Player

In this section I consider the design of the race to answer the question in title. Consider a principal engaging two players A, B in a contest. His objective is to minimize the expected completion time taking into account the reward he pays to the winner. In the case of quadratic cost functions and no discounting, I use the closed-forms of the equilibrium strategies, x_A, x_B, and the expected completion time, $E[\tau]$, to investigate the question whether the principal should choose to encumber the more able player with a handicap. The principal can handicap the advantaged player by either offering him a lower final reward or by weakening his ability.

If the discouragement effect introduced in the previous section is sufficiently strong, handicapping the advantaged player will mitigate the effect. First, the disadvantaged player raises his equilibrium effort because of his higher chance of winning the race. Second, anticipating this behavior of the disadvantaged player, the advantaged player increases his equilibrium effort. This strategic increasing in effort of the advantaged player can dominate the weaker incentive effect coming from his lower reward. Therefore, handicapping the advantaged player can increase his effort and stochastically shortens the completion time of the race. At the same time, the principal pays less to the winner of the race because of the following. If the principal lowers the final reward promised to the advantaged player, he will pay strictly less in expectation to the winner of the race than in the original race. If the principal chooses to reduce the ability of the advantaged player, the expected payment to the winner remains the same as in the original race. In both cases, the race finishes earlier in expectation while the principal has to pay weakly less than in the original race.

In the general model with general cost of effort functions and with discounting, the discouragement effect is weaker if the players have a higher discount rate or if the cost of effort functions are more convex. Under these circumstances, handicapping the advantaged player is less effective in shortening the completion time of the race.

3.5.1 Quadratic Cost Function and No-discounting

In this special case, MPE strategies admit a closed-form and so does the expected completion time. I use these closed-form solutions to show that when the final rewards are sufficiently high, the discouragement effect is strong. In this case, mitigating this effect, by either reducing the final reward promised to the advantaged player or by increasing his cost, will reduce the expected completion time.

Consider the case where the two players are different in their cost of effort. Without loss of generality, suppose effort costs $\frac{1}{1+s}$, $s > 0$, less to player A then than to player B:

$$\begin{aligned} c_A(x) &= \frac{1}{1+s}\frac{x^2}{2} \\ c_B(x) &= \frac{x^2}{2}. \end{aligned} \tag{3.18}$$

The two players have the same distance requirement to win their final reward:

$$\begin{aligned} K_A &= K_B = 1 \\ \sigma &= 1 \\ z_0 &= 0, \end{aligned}$$

and the principal starts with the same final reward promised to each player:

$$P_A = P_B = P.$$

From Example 3.1, the MPE strategies and the expected completion time, $E[\tau]$, are analytically the same as in the race in which the two players have the same cost of effort function:

$$\widetilde{c}_A(x) = \widetilde{c}_B(x) = \frac{x^2}{2},$$

but the player A has a proportionally higher promised final reward:

$$\begin{aligned} \widetilde{P}_A &= (1+s) P_A \\ \widetilde{P}_B &= P_B. \end{aligned}$$

This equivalence implies that reducing the promised reward to player A or increasing his cost of effort have the same impact on the MPE strategies and the stochastic completion time.

Proposition 3.2 *Given s, there exists a reward level $\overline{P}(s)$ such that if $P_A = P_B = P > \overline{P}(s)$, a decrease in P_A will also decrease the expected completion time. In addition, there exists a reward level $\underline{P}(s)$ such that if $P_A = P_B = P < \underline{P}(s)$, a decrease in P_A will increase the expected completion time instead.*

Notice that s is the degree at which player A is more competitive than player B. For a given s, the higher P the stronger the discouragement effect, the player will exert high effort when they are sufficiently close to each other, but when one players get further ahead, both of them reduce significantly their effort level. When this effect is strong enough, lowering the promised reward to the advantaged player A, or equivalently increase his cost of effort function, will encourage effort of both player and stochastically reduce the expected completion time. Example 3.1 also shows that the discouragement effect is stronger at lower level of the cost of effort and lower uncertainty. Thus, given P and s, when α and β in (3.9) is sufficiently large and σ in(3.1) is sufficiently small, handicapping player A will also reduces the expected completion time.

In contrast, when P is small, the discouragement effect is weak. Lowering the promised reward to player A reduces his incentive. This reduction of his incentive dominates the strategic effect on the player's effort. Overall, A exerts lower effort. The expected completion time thus increases.

The detail of the proof of Proposition 3.2 using exponential and Taylor expansions is in the Appendix

3.5.2 General Model

In this subsection, I study numerically how Proposition 3.2 changes if we depart from the case of quadratic cost of effort functions and no discounting.

Other factors that affect the players' strategic behaviors are the discount rate and the degree of convexity of the cost of effort functions. In Subsection 3.4.2, I argue that higher discount rate and higher degree of convexity alleviate the discouragement effect. Therefore, handicapping the advantaged player will be less effective in reducing the expected completion time. I do not have closed-form solutions for MPEs in this general case. However, the Hamilton-Jacobi-Bellman equations can be solved numerically using a discretization procedure. This procedure also allows me compute the corresponding MPE strategies and the expected completion time. I present the details of this procedure in the Appendix. Let k denote the degree of convexity of the cost of effort functions:

$$\begin{aligned} c_A(x) &= \frac{1}{1+s}\frac{x^{1+k}}{1+k} \\ c_B(x) &= \frac{x^{1+k}}{1+k}. \end{aligned} \tag{3.19}$$

I will show numerically that: Given the reward level $P > \overline{P}(s)$ in Proposition 3.2, there exists a critical discount rate r^* and a critical degree of convexity k^* such that if the discount rate of the players is higher than r^* , or the degree of convexity of the cost of effort functions is higher than k^*, handicapping the advantaged player will increase the expected completion time instead of decreasing it.

This claim is complementary to Proposition 3.2 in which reducing incentives to the more advantaged player, will reduce the expected completion time only if the discouragement effect is strong and increase the expected completion time otherwise. Here, the effect is weaker as the players are more impatient, or the cost of effort function is highly convex. Consequently, reducing incentives to the more advantaged player will increase the expected completion time in these cases.

For the numerically exercise, I fix the finish lines K_A ,K_B at $K = 1$ and the two players start at $z_0 = 0$. The advantaged player, player A is twice as productive as player B, i.e., $s = 1$. The benchmark case is $r = 0$ and $k = 0$ in which I have the closed-form solution. Then I calculate the equilibrium strategies and the expected completion time for the cases $r > 0, k = 0$ and $r = 0, k > 0$ to study the effect of discounting and cost convexity on MPEs. Regarding the final rewards, I start each calculation with $P_A = P_B = 1.5$. Then, I vary P_A locally holding P_B constant to study how the expected completion time changes with respect to P_A.

The strategies of the advantaged player under different discount rates, r, is shown in the right panel of Figure 3-1 and the expected completion time is shown in left panel of the same figure.

In the right panel, the horizontal axis shows the distance between the two players on which the MPE strategies are conditioned. If the advantaged player leads his disadvantaged rival by a distance $K_A = 1$, he wins the race. If instead, his rival leads him by a distance $K_B = 1$ then he loses the race. The vertical axis shows the effort levels of the advantaged player. The solid and dashed lines are respectively his equilibrium effort profiles as functions of the distance to his rival for different discount rates: $r = 0$ and 1. In all cases, he exerts a high level of effort when he is close to his rival ($z \simeq 0$). He reduces his effort once he is left further behind or he gets further ahead of his rival. However, when he is less patient, (r is high enough), he also wants to finish the race early. Therefore, even if he gets further ahead of his rival (z is close to $K_A = 1$), he maintains a high effort level in order to win the race in a shorter time ($r =$ 1 in the right panel). The discouragement effect is diminished when r is high enough.

When the discouragement effect is weak, handicapping the advantaged player has a direct incentive effect of reducing his effort level. Handicapping the advantaged player thus increases the expected completion time instead of decreasing it. This result is shown in the left panel of Figure 3-1. The horizontal axis shows the promised final reward to the advantaged player P_A , keeping constant the promised final reward to his rival $P_B = 1.5$. The vertical axis shows the expected completion times of the races in which players start at the same position $z_0 = 0$. The solid and dashed lines are respectively the expected completion time as functions of the promised reward to the advantaged player for different discount rates: $r = 0$ and 1. First, when $r = 0$, a lower promised reward to the advantaged player indeed reduces the expected completion time. Second, in contrast to the former case, when r is higher, $r = 1$, a lower promised reward to the advantaged player increases the expected completion time of the race.

The strategies of the advantaged player under different values of cost convexity, k, is shown in Figure 3-2. When the cost function is highly convex, he also wants to smooth his cost of effort. Therefore, he maintains a almost constant level of effort as k is sufficiently high. This result is shown in the right panel. Handicapping the advantaged player increases the expected completion time. This result is shown in the left panel.

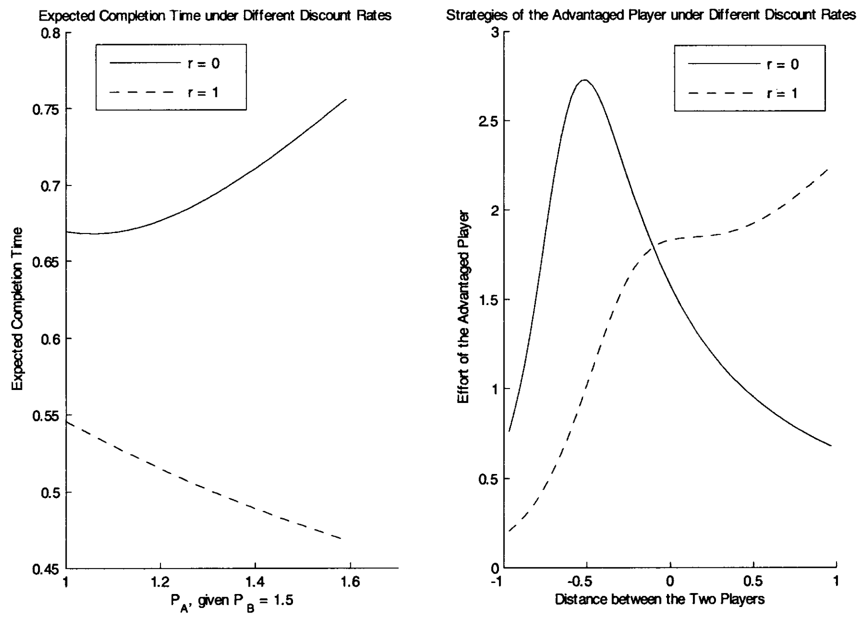


Figure 3-1: **Expected Completion Time and Equilibrium Strategies of the Advantaged Player under Different r's.**
The solid line, $r = 0$, in the right panel shows that the advantaged player reduces significantly his effort when he gets further ahead (strong discouragement effect). The dashed line, $r = 1$, shows that he does not reduce his effort when he is impatient (weak discouragement effect).
In the left panel, the solid line shows that when the discouragement effect is strong, lower promised reward to the advantaged player decreases the expected completion time. The dashed line shows the opposite result when the discouragement effect is weak.

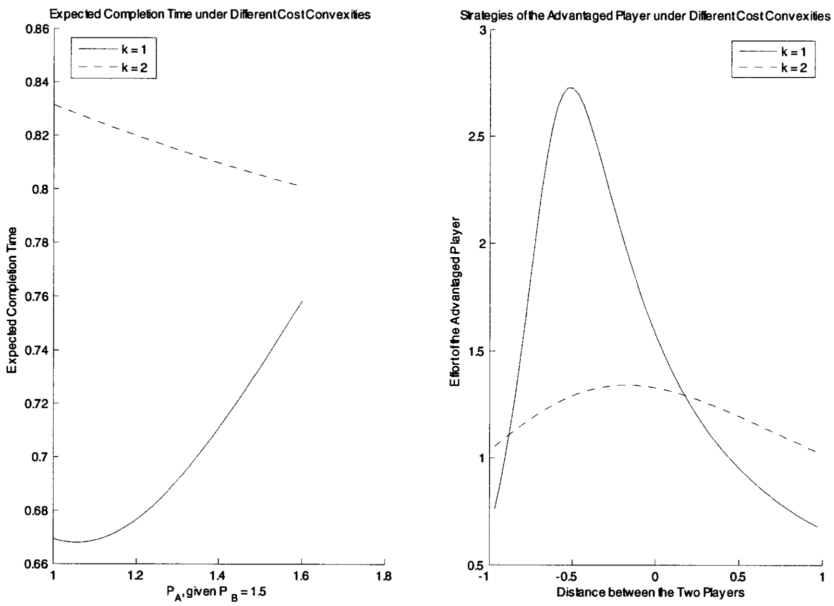


Figure 3-2: **Expected Completion Time and Equilibrium Strategies of the Advantaged Player under Different k's.**
The solid line, $k = 1$, in the right panel shows that the advantaged player reduces significantly his effort when he gets further ahead (strong discouragement effect). The dashed line, $k = 2$, shows that he also reduces his effort but to lesser degree (weak discouragement effect).
In the left panel,the solid line shows that when the discouragment effect is strong, lower promised reward to the advantaged player decreases the expected completion time. The dashed line shows the opposite result when the discouragement effect is weak.

3.6 Conclusion

In this paper, I develop a simple continuous time model of racing under uncertainty to analyze the question initially asked in the abstract. I prove the existence of Markov Perfect Equilibria and, in some cases, also their uniqueness. The equilibria have similar properties to those in the original discrete time model. In addition, for some special cases, I can derive the closed-form of these MPE strategies, which facilitates the study of the comparative statics, and also allows me to show that handicapping the advantaged player in a race might be useful. A future research direction is to develop a model with more realistic features of certain races. This paper has made some progress along these lines. For example, I have allowed for more general cost functions, discounting, and for a finish line instead of distance between players. Even though these models do not have closed-form MPEs, it is still possible to numerically compute the equilibria, and examine their properties. Interestingly, the properties of the MPEs and the answer to the initial question in these models are consistent with the results from the less general model. Another potentially fruitful avenue for future research is to incorporate asymmetric information into the model, which would allow for the study of the interaction between asymmetric information and dynamic features studied here and how the answer to the initial question is affected.

Appendix

Derivation of Hamilton-Jacobi-Bellman equations. For example, for firm A at time t, assume that it optimizes from $t + \Delta t$ forward and solves

$$\begin{aligned}
J_A(z_t) &= \max_x E_t \left[- \int_0^{\Delta t} e^{-rs} c_A(x_{t+s})\, ds + e^{-r\Delta t} J_A(z_{t+\Delta t}) \right] \\
&= \max_x E_t \left[- \int_0^{\Delta t} e^{-rs} c_A(x_{t+s})\, ds + e^{-r\Delta t} J_A(z_{t+\Delta t}) \right] \\
&= \max_x E_t \left(-\Delta t c_A(x) + e^{-r\Delta t} \left(J_A(z_t) + \Delta J_A(z_t) \right) + o(\Delta t) \right) \qquad (3.20)
\end{aligned}$$

The first part of this expression is the flow of the cost of R&D effort during a time interval of length Δt. The second part is the discounted continuation value after this time interval. The continuation value is discounted by the factor $e^{-r\Delta t} = 1 - r\Delta t + o(\Delta t)$,where, from now on, $o(\Delta t)$ denotes second-order terms. This continuation value depends on the evolution of z_t to $z_{t+\Delta t}$. By Ito's Lemma, we have:

$$\begin{aligned}
\Delta J_A(z_t) &= J_A(z_{t+\Delta t}) - J_A(z_t) \\
&= J_A'(z_t)\, \Delta z_t + \frac{\sigma^2}{2} J_A''(z_t)\, \Delta t + o(\Delta t) \\
&= J_A'(z_t)\, (x_{At} - x_{Bt})\, \Delta t + J_A'(z_t)\, \sigma \Delta W_t \\
&\quad + \frac{\sigma^2}{2} J_A''(z_t)\, \Delta t + o(\Delta t)\,.
\end{aligned}$$

Taking expectation of both side, and using the normal independent increments property of Brownian noise, we have $E_t [J_A'(z_t)\, \sigma \Delta W_t] = 0$. Thus,

$$\begin{aligned}
E_t [\Delta J_A(z_t)] &= J_A'(z_t)\, (x_{At} - x_{Bt})\, \Delta t \\
&\quad + \frac{\sigma^2}{2} J_A''(z_t)\, \Delta t + o(\Delta t)\,.
\end{aligned}$$

Now, substitute these results into (3.20) and subtract $J_A(z_t)$ from both sides. Dividing all terms by Δt, and taking the limit as $\Delta t \longrightarrow 0$, we obtain the Hamilton-Jacobi-Bellman equation (3.3) for the value function of firm i. ∎

We will make use of the following theorem (Gronwall's Inequality)[4] in (Hartman 1964, pg. 24)

Proof of Lema 3.1. 1) Let z^* be a minimum of $J_A(z)$ over the interval $[-K_B, K_A]$ (since $J_A(.)$ is continuous, that minimum exists). If z^* is an interior point then we have $J_A'(z^*) = 0$. From (3.3) we have $J_A''(z^*) = 2\,rJ_A(z^*)$. In addition since z^* is a minimum, we have $J_A''(z^*) \geq 0$ so $J_A(z^*) \geq 0$. Furthermore, at the two boundaries, $J_A \geq 0$ therefore $J_A(z) \geq 0$ for all $z \in [-K_B, K_A]$.

Now if there exists an interior point z^* such that $J_A(z^*) = 0$, let z^{**} be a maximum of $J_A(.)$ over $[-K_B, z^*]$; then $J_A'(z^*) = 0$ and $J_A''(z^{**}) = 2rJ_A(z^{**})$. Since z^{**} is a maximum and $J_A''(z^{**}) \leq 0$, we have $J_A(z^{**}) \leq 0$, thus $J_A(z) = 0$ for all $z \in [-K_B, z^*]$. And from the fact that z^* is strictly interior,

$$J_A(z^*) = J_A'(z^*) = 0.$$

We can show that this yields a contradiction because J_A would be identically 0 over $[z^*, K_A]$. First of all, we have the following inequality:

$$\begin{aligned} \left|J_A''(z)\right| &= \frac{2}{\sigma^2}\left|rJ_A(z) + c_A\left(f_A\left(J_A'(z)\right)\right) - \left(f_A\left(J_A'(z)\right) - x_B(z)\right)J_A'(z)\right| \\ &\leq \frac{2r}{\sigma^2}\left|J_A(z)\right| + 3\overline{x}\left|J_A'(z)\right|, \end{aligned} \tag{3.21}$$

[4]**Grownwall's Inequality, Hartman (1964) II-1.1** Let $u(t)$ and $v(t)$ be non-negative, continuous functions on $[a,b]$; $C \geq 0$ a constant; and

$$v(t) \leq C + \int_a^t v(s)\,u(s)\,ds \text{ for } a \leq t \leq b.$$

Then

$$v(t) \leq C \exp\left(\int_a^t u(s)\,ds\right) \text{ for } a \leq t \leq b,$$

in particular, if $C = 0$, then $v \equiv 0$.

where the inequalities is obtained from the three inequalities

$$\begin{aligned}
0 &\leq x_B(z) \leq \overline{x} \\
0 &\leq f_A\left(J_A'(z)\right) \leq \overline{x} \\
0 &\leq c_A\left(f_A\left(J_A'(z)\right)\right) \\
&\leq c_A'\left(f_A\left(J_A'(z)\right)\right) f_A\left(J_A'(z)\right) \\
&\leq \left|J_A'(z)\right| \overline{x}.
\end{aligned}$$

Apply the Gronwall's inequality for $|J_A(z)|^2 + |J_A'(z)|^2$, we have $J_A(z) = J_A'(z) = 0\ \forall z \in [z^*, K_A]$.This is a contradiction with the fact that $J_A(K_A) = P_A > 0$.

So we have $J_A(z) > 0$ for all $z \in (-K_B, K_A)$. The proof for $J_B(z)$ is analogous.

2) By the mean value theorem, there exists a $z^0 \in (-K_B, K_A)$ such that

$$J_A'\left(z^0\right) = \frac{J_A(K_A) - J_A(-K_B)}{K_A - (-K_B)} = \frac{P_A}{K_A + K_B} > 0.$$

If there exists some $z^1 \in (-K_A, K_B)$ such that $J_A'\left(z^1\right) < 0$, then, by the intermediate value theorem, there exists an interior point z^* between z^0 and z^1 such that $J_A'(z^*) = 0$. Hence, from the first part,

$$J_A''(z^*) = 2rJ_A(z^*) > 0.$$

Consider the interval $[-K_B, z^*]$,at $z = -K_B$, $J_A(-K_B) = 0$.The extreme $-K_B$ cannot be a maximum of J_A over this interval. And in a neighborhood $z = z^* - \varepsilon$ of z^*,

$$J_A(z) = J_A(z^*) + \frac{1}{2}J_A''(z^*)\varepsilon^2 + o\left(\varepsilon^2\right) > J_A(z^*),$$

so this extreme z^* cannot be a maximum over the interval, either. Thus, J_A has an interior maximum in the interval. Denote this maximum z^{**}. We have $J_A'(z^{**}) = 0$. This yields a contradiction because it implies $J_A''(z^{**}) > 0$, or z^{**} is a local minimum.

We have established that $J_A'(z) > 0\ \forall z \in (-K_B, K_A)$. The argument for $J_B'(z) < 0\ \forall z \in (-K_B, K_A)$ is analogous. ■

The proof for the case $r = 0$ is easier. For example, if there exists $z^* \in [-K_B, K_A]$ such that $J_A'(z^*) = 0$, then, as derived in (3.21)

$$\left|J_A''(z)\right| \leq 3\overline{x}\left|J_A'(z)\right|.$$

Again, by applying the Gronwall's inequality, we have $J_A'(z) = 0\ \forall z \in [-K_B, K_A]$. But we know that,

$$J_A(-K_B) = 0 < P_A = J_A(K_A)$$

and hence we have a contradiction. It follows that $J_A'(z) > 0\ \forall z \in [-K_B, K_A]$. The argument for $J_B'(z) < 0\ \forall z \in (-K_B, K_A)$ is analogous.

Proof of Theorem 3.2. The steps of the existence proof are the following. I will show that there exist constants P, M and a globally bounded vector-valued function g satisfying

1)$\forall |J_i| \leq P, |J_i'| \leq M, i = A, B$

$$g\left(\begin{pmatrix} J_A \\ J_B \end{pmatrix}, \begin{pmatrix} J_A' \\ J_B' \end{pmatrix}\right) = \frac{2}{\sigma^2}\begin{pmatrix} rJ_A + F_A(J_A', J_B') \\ rJ_B + F_B(J_A', J_B') \end{pmatrix}.$$

However, g can be different from the right hand side outside this region

2)Any solution to the boundary value problem

$$\begin{aligned} \begin{pmatrix} J_A''(z) \\ J_B''(z) \end{pmatrix} &= g\left(\begin{pmatrix} J_A \\ J_B \end{pmatrix}, \begin{pmatrix} J_A' \\ J_B' \end{pmatrix}\right) \\ \begin{pmatrix} J_A(-K_B) \\ J_B(-K_B) \end{pmatrix} &= \begin{pmatrix} 0 \\ P_B \end{pmatrix}, \begin{pmatrix} J_A(K_A) \\ J_B(K_A) \end{pmatrix} = \begin{pmatrix} P_A \\ 0 \end{pmatrix}, \end{aligned} \tag{3.22}$$

will satisfy $|J_i(z)| \leq P, |J_i'(z)| \leq M, i = A, B\ \forall z \in [-K_B, K_A]$

Therefore, any solution to the boundary value problem (3.22) is also a solution to the original problem (3.7). ■

In order to prove the existence and the uniqueness of the solution to the boundary problem (3.7), we first provide a bound on the effort intensity of each firm.

Lemma 3.2 *There exists some M depending only on P_A, P_B, K_A, K_B and $c(.)$ such that* $0 < J_A'(z), -J_B'(z) < M\ \forall z \in (-K_A, K_B)$

Proof. Let $D(z) = J_A(z) - J_B(z)$ then $D'(z) = J'_A(z) - J'_B(z)$ and $0 < J'_A(z), -J'_B(z) < D'(z)$. Substituting the effort functions (3.5) into the Hamilton-Jacobi-Bellman equation (3.7), we have:

$$\begin{aligned}
J''_A(z) &= \frac{2r}{\sigma^2} J_A(z) + \frac{2}{\sigma^2}\left(\begin{array}{c} -c_A\left(\left(f_A\left(J'_A(z)\right)\right)\right) + \left(f_A\left(J'_A(z)\right)\right) J'_A(z) \\ -\left(f_B\left(-J'_B(z)\right)\right) J'_A(z) \end{array}\right) \\
J''_B(z) &= \frac{2r}{\sigma^2} J_B(z) + \frac{2}{\sigma^2}\left(\begin{array}{c} -c_B\left(\left(f_B\left(-J'_B(z)\right)\right)\right) - \left(f_B\left(-J'_B(z)\right)\right) J'_B(z) \\ +\left(f_A\left(J'_A(z)\right)\right) J'_B(z) \end{array}\right).
\end{aligned}$$

By subtracting these two equalities, we obtain

$$\left|D''(z)\right| \leq \frac{2r}{\sigma^2} P + \frac{2}{\sigma^2}\left(\left(\left(c'_A\right)^{-1}\left(D'(z)\right)\right) D'(z) + \left(\left(c'_B\right)^{-1}\left(D'(z)\right)\right) D'(z)\right),$$

where $P = \max(P_A, P_B)$ .This is due to

$$\begin{aligned}
0 &\leq -c_A\left(\left(f_A\left(J'_A(z)\right)\right)\right) + \left(f_A\left(J'_A(z)\right)\right) J'_A(z) \leq \left(c'_A\right)^{-1}\left(J'_A(z)\right) J'_A(z) \\
0 &\leq \left(f_B\left(-J'_B(z)\right)\right) J'_A(z) \leq \left(c'_B\right)^{-1}\left(-J'_B(z)\right) J'_A(z) \\
0 &\leq -c_B\left(\left(f_B\left(-J'_B(z)\right)\right)\right) - \left(f_B\left(-J'_B(z)\right)\right) J'_B(z) \leq \left(c'_B\right)^{-1}\left(-J'_B(z)\right)\left(-J'_B(z)\right) \\
0 &\leq -\left(f_A\left(J'_A(z)\right)\right) J'_B(z) \leq \left(c'_A\right)^{-1}\left(J'_A(z)\right)\left(-J'_B(z)\right)
\end{aligned}$$

and

$$\begin{aligned}
\left(c'_A\right)^{-1}\left(J'_A(z)\right) &\leq \left(c'_A\right)^{-1}\left(D'(z)\right) \\
\left(c'_B\right)^{-1}\left(-J'_B(z)\right) &\leq \left(c'_B\right)^{-1}\left(D'(z)\right).
\end{aligned}$$

By the mean value theorem, there exists a $z^* \in (-K_B, K_A)$ such that $D'(z^*) = \frac{P_A+P_B}{K_A+K_B}$.

$\forall z \in [-K_B, K_A]$. It then follows that

$$
\begin{aligned}
& P_A + P_B \\
> & \left| \int_{z^*}^{z} D'(t)\, dt \right| \\
> & \left| \int_{z^*}^{z} D'(t) \frac{D''(t)}{\frac{2r}{\sigma^2} P + \frac{2}{\sigma^2} \left(\left((c'_A)^{-1} (D'(z)) \right) D'(z) + \left((c'_B)^{-1} (D'(z)) \right) D'(z) \right)} dt \right| \\
\overset{D'(t)=s}{=} & \left| \int_{\frac{P_A+P_B}{K_A+K_B}}^{D'(z)} \frac{sds}{\frac{2r}{\sigma^2} P + \frac{2}{\sigma^2} \left(\left((c'_A)^{-1} (s) \right) + \left((c'_B)^{-1} (s) \right) \right) s} \right| .
\end{aligned}
$$

The last equality is a result of a change of integration variables from t to $s = D'(t)$: $D'(t)\, D''(t)\, dt = D'(t)\, dD'(t)$. Condition (3.14) implies that

$$\int_{\frac{P_A+P_B}{K_A+K_B}}^{\infty} \frac{sds}{\frac{2r}{\sigma^2} P + \frac{2}{\sigma^2} \left(\left((c'_A)^{-1} (s) \right) + \left((c'_B)^{-1} (s) \right) \right) s} = +\infty.$$

Thus, there exists an M such that

$$\int_{\frac{P_A+P_B}{K_A+K_B}}^{M} \frac{sds}{\frac{2r}{\sigma^2} P + \frac{2}{\sigma^2} \left(\left((c'_A)^{-1} (s) \right) + \left((c'_B)^{-1} (s) \right) \right) s} = P_A + P_B.$$

We conclude that

$$D'(z) < M \ \forall z \in (-K_B, K_A).$$

■

Using these bounds on $J'_A(z)$ and $J'_B(z)$, we now can prove the existence of a solution for any value of r, P_A, P_B, K_A, K_B[5].To this end, the following classical lemma from (G.Scorza-Dragoni 1935) will be useful:

Lemma *Let $g(t, x, x')$ be a continuous and bounded (vector-valued) function for $0 \le t \le T$ and arbitrary (x, x'). Then, for arbitrary x_0 and x_T the system of differential equations*

$$x'' = g(t, x, x')$$

[5]This proof follows closely Hartman (1960)

has at least one solution $x = x(t)$ satisfying

$$x(0) = x_0, x(T) = x_T.$$

It is been pointed out by Bass (1958) that this lemma is easily derived from the Schauder's fixed point theorem. In order to use this lemma, we need to transform the system (3.7) into a bounded system over $\begin{pmatrix} J_A \\ J_B \end{pmatrix}$ and $\begin{pmatrix} J'_A \\ J'_B \end{pmatrix}$.

Proof of the Theorem 3.1. First, we can easily find two bounded, strictly increasing and infinitely differentiable functions φ, χ such that

$$\begin{aligned} \varphi(x) &= x \text{ if } |x| \le P \text{ and } |\varphi'| \le 1 \\ \chi(x) &= x \text{ if } |x| \le M \text{ and } |\chi'| \le 1 \end{aligned}$$

Consider the function

$$\begin{aligned} & g\left(\begin{pmatrix} J_A \\ J_B \end{pmatrix}, \begin{pmatrix} J'_A \\ J'_B \end{pmatrix}\right) \\ = & \frac{1}{\sigma^2}\begin{pmatrix} 2r\varphi(J_A) + F_A(\chi(J'_A), \chi(J'_B)) \\ 2r\varphi(J_B) + F_B(\chi(J'_A), \chi(J'_B)) \end{pmatrix}. \end{aligned}$$

Since φ, χ are bounded g is bounded, then by the Lemma from (G.Scorza-Dragoni 1935), the boundary value problem

$$\begin{aligned} \begin{pmatrix} J''_A \\ J''_B \end{pmatrix} &= g\left(\begin{pmatrix} J_A \\ J_B \end{pmatrix}, \begin{pmatrix} J'_A \\ J'_B \end{pmatrix}\right) \\ \begin{pmatrix} J_A(-K_B) \\ J_B(-K_B) \end{pmatrix} &= \begin{pmatrix} 0 \\ P_B \end{pmatrix}, \begin{pmatrix} J_A(K_A) \\ J_B(K_A) \end{pmatrix} = \begin{pmatrix} P_A \\ 0 \end{pmatrix} \end{aligned}$$

has at least one solution $\begin{pmatrix} J_A(z) \\ J_B(z) \end{pmatrix}$ $-K_B \le z \le K_A$. We can proceed exactly the same way as in the proof of Lemma 3.1 and 3.2 to show that $0 < J_A(z), J_B(z) < P$ and $0 < J'_A(z), -J'_B(z) <$

M, so $\begin{pmatrix} J_A(z) \\ J_B(z) \end{pmatrix}$ $-K_B \leq z \leq K_A$ is also the solution to the original system. ∎

In order to prove uniqueness, we use theorem XII-4.3[6] from (Hartman 1964, pg. 425).

Proof of Theorem 3.2. First from the proof of the Lemma 3.2, we show that the bound M can be taken such that M goes to 0 as P goes to 0. Indeed, since

$$\int_{\frac{P_A+P_B}{K_A+K_B}}^{M} \frac{sds}{\frac{2r}{\sigma^2}P + \frac{2}{\sigma^2}\left(\left((c'_A)^{-1}(s)\right) + \left((c'_B)^{-1}(s)\right)\right)s} = P_A + P_B$$

and the right hand side goes to 0 as P goes to 0. If M does not go to zero we can extract a subsequence such that M_n converges to $M^* > 0$ and P_n goes to zero. Then taking the limit as n goes to infinity, we obtain

$$\int_0^{M^*} \frac{sds}{\frac{2}{\sigma^2}\left(\left((c'_A)^{-1}(s)\right) + \left((c'_B)^{-1}(s)\right)\right)s} = 0.$$

This yields a contradiction. Therefore M must go to 0 as P goes to 0.

In order to apply the Theorem XII - 4.3 (Hartman 1964, pg. 425), we need to verify the condition

$$2\left(B - \frac{1}{4}FF^*\right)s.s > -\frac{\pi^2}{(K_A+K_B)^2}\left\|s\right\|^2, \tag{3.23}$$

[6]**Theorem XII-4.3 (Hartman 1964, pg 425)**Let $f(t,x,x')$ be continuous for $0 \leq t \leq p$ and for (x,x') on some 2d-dimensional convex set. Let $f(t,x,x')$ have continuous partial derivatives with respect to the components of x and x'. Let the Jacobian matrices of f with respect to x,x'

$$\begin{aligned} B(t,x,x') &= \partial_x f(t,x,x') \\ F(t,x,x') &= \partial_{x'} f(t,x,x') \end{aligned}$$

satisfy

$$2\left(B - \frac{1}{4}FF^*\right)\mathbf{z}.\mathbf{z} > -\frac{\pi^2}{p^2}\left\|\mathbf{z}\right\|^2$$

for all constant vectors $\mathbf{z} \neq \mathbf{0}$. Then the boundary value problem

$$\begin{aligned} x'' &= f(t,x,x') \\ x(0) &= x_0 \\ x(p) &= x_p \end{aligned}$$

has at most one solution.

where

$$B\left(J,J'\right)=\begin{pmatrix}\frac{2r}{\sigma^2} & 0\\ 0 & \frac{2r}{\sigma^2}\end{pmatrix}$$

and

$$F\left(J,J'\right)=\frac{2}{\sigma^2}\begin{pmatrix}\left(c_A'\right)^{-1}\left(J_A'\right)-\left(c_B'\right)^{-1}\left(-J_B'\right) & -\frac{J_A'}{c_B''\left(\left(c_B'\right)^{-1}\left(-J_B'\right)\right)}\\ -\frac{J_B'}{c_A''\left(\left(c_A'\right)^{-1}\left(J_A'\right)\right)} & \left(c_A'\right)^{-1}\left(J_A'\right)-\left(c_B'\right)^{-1}\left(-J_B'\right)\end{pmatrix}.$$

Since c_i'' are bounded below around 0, the norm of F is bounded by $|J_A'|$ and $|J_B'|$, thus by M, which goes to 0 as P goes to 0. We then have(3.23) as P goes to 0. ■

Proof of (3.15). Using the proof method in Theorem 3.2, we show that in the case of quadratic cost M can be chosen as

$$\sqrt{\exp\left(2\left(P_A+P_B\right)\right)\left(2rP+\left(\frac{P_A+P_B}{K_A+K_B}\right)^2\right)-2rP}$$

where $P=\max\left(P_A,P_B\right)$.Subtracting the first equation in (3.13) from the second, we have a new equation in terms of $D\left(z\right)=J_A\left(z\right)-J_B\left(z\right)$

$$-rD\left(z\right)+\frac{1}{2}D'\left(z\right)\left(J_A'\left(z\right)+J_B'\left(z\right)\right)+\frac{1}{2}D''\left(z\right)=0.$$

Since $J_A'\left(z\right)>0$ and $J_B'\left(z\right)<0$, it is implied that $D'\left(z\right)>\left|J_A'\left(z\right)+J_B'\left(z\right)\right|$.In addition, because $0<J_A\left(z\right),J_B\left(z\right)<P,$ we also have $\left|D\left(z\right)\right|<P$. So

$$\left|D''\left(z\right)\right|<2rP+\left(D'\left(z\right)\right)^2$$

By the mean value theorem, there exists $z^*\in\left(-K_B,K_A\right)$ such that $D'\left(z^*\right)=\frac{P_A+P_B}{K_A+K_B}$ $\forall z\in\left[-K_B,K_A\right]:$

$$\begin{aligned}P_A+P_B &> \left|\int_{z^*}^{z}D'\left(t\right)dt\right|\\ &> \left|\int_{z^*}^{z}D'\left(t\right)\frac{D''\left(t\right)}{2rP+\left(D'\left(t\right)\right)^2}dt\right|\\ &= \left|\int_{D'\left(z^*\right)}^{D'\left(z\right)}\frac{sds}{2rP+s^2}\right|.\end{aligned}$$

So

$$
\begin{aligned}
D'(z) &< \sqrt{\exp\left(2\left(P_A+P_B\right)\right)\left(2rP+\left(\frac{P_A+P_B}{K_A+K_B}\right)^2\right)-2rP} \\
&= M.
\end{aligned}
$$

Since $J_A'(z) > 0$ and $J_B'(z) < 0$, they are then both smaller than $D'(z)$ in absolute value. ■

Proof (3.16). Again, in order to apply the Theorem XII - 4.3 in (Hartman 1964, pg. 425), we need to verify the condition

$$
2\left(B-\frac{1}{4}FF^*\right)s.s > -\frac{\pi^2}{\left(K_A+K_B\right)^2}\left\|s\right\|^2
$$

where $B(J,J') = \begin{pmatrix} 2r & 0 \\ 0 & 2r \end{pmatrix}$ and $F(J,J') = \begin{pmatrix} 2\left(J_A'+J_B'\right) & 2J_A' \\ 2J_B' & 2\left(J_A'+J_B'\right) \end{pmatrix}$. Substituting these expressions for $B(J,J')$ and $F(J,J')$, we have

$$
\begin{aligned}
& 2\left(B-\frac{1}{4}FF^*\right)s.s \\
= {} & 4r\left(s_1^2+s_2^2\right) \\
& -2\begin{pmatrix} \left(\left(J_A'+J_B'\right)^2+\left(J_A'\right)^2\right)s_1^2 \\ +\left(\left(J_A'+J_B'\right)^2+\left(J_B'\right)^2\right)s_2^2+2\left(J_A'+J_B'\right)^2 s_1 s_2 \end{pmatrix} \\
> {} & \left(4r-3M^2\right)\left(s_1^2+s_2^2\right)
\end{aligned}
$$

Therefore, if $4r-3M^2 > -\frac{\pi^2}{(K_A+K_B)^2}$ or $M < \sqrt{\frac{1}{3}\left(4r+\frac{\pi^2}{(K_A+K_B)^2}\right)}$, then, the system has a unique solution. As we notice above that M goes to 0 as P goes to 0; so if P is sufficiently small, we have the required inequality. ■

Closed form Derivation. Substituting the functional forms of the cost functions into (3.4), we have

$$
\begin{aligned}
x_A(z) &= J_A'(z) \\
x_B(z) &= -J_B'(z).
\end{aligned} \tag{3.24}
$$

Substituting these equations into (3.3) , we finally obtain the second order ordinary differential equations on (J_A, J_B)

$$
\begin{aligned}
\frac{1}{2}\left(J_A'(z)\right)^2 + J_A'(z)\, J_B'(z) + \frac{1}{2}J_A''(z) &= 0 \\
\frac{1}{2}\left(J_B'(z)\right)^2 + J_A'(z)\, J_B'(z) + \frac{1}{2}J_B''(z) &= 0.
\end{aligned}
\tag{3.25}
$$

Denote $x = x_A$ and $y = x_B$. Differentiating both sides of(3.24) gives $J_A''(z) = x'(z)$ and $J_B''(z) = -y'(z)$. Thus, we can rewrite (3.25) as a system of first-order differential equations with unknowns are strategy functions x and y :

$$
\begin{aligned}
\frac{1}{2}x^2 - xy + \frac{1}{2}x' &= 0 \\
\frac{1}{2}y^2 - xy - \frac{1}{2}y' &= 0.
\end{aligned}
\tag{3.26}
$$

We derive the boundary conditions for (3.26) using Lebnitz's rule:

$$
\begin{aligned}
P_A &= J_A(K_A) - J_A(-K_B) \\
&= \int_{-K_B}^{K_A} J_A'(z)\, dz
\end{aligned}
$$

and

$$
\begin{aligned}
P_B &= J_B(-K_B) - J_B(K_A) \\
&= -\int_{-K_B}^{K_A} J_B'(z)\,(z)\, dz.
\end{aligned}
$$

To recapitulate, the two boundary conditions are:

$$
\begin{aligned}
\int_{-K_B}^{K_A} x(z)\, dz &= P_A \\
\int_{-K_B}^{K_A} y(z)\, dz &= P_B.
\end{aligned}
\tag{3.27}
$$

Let $g(z) = \frac{x(z)}{y(z)}$. g is well-defined since $y > 0$. We will solve g as a function of y

$$\begin{aligned}
x'(z) &= g'(y)\, y'(z)\, y(z) + g(y)\, y'(z) \\
&\Longrightarrow g'(y)\, y = g(y)\frac{2 - g(y)}{1 - 2g(y)} - g(y) \\
&= g(y)\frac{1 + g(y)}{1 - 2g(y)}.
\end{aligned}$$

Rewrite this in a differential form:

$$\begin{aligned}
\frac{dg\,(1 - 2g)}{g\,(g + 1)} &= \frac{dy}{y} \\
&\Longrightarrow dg\left(\frac{1}{g\,(g+1)} - \frac{2}{g+1}\right) = \frac{dy}{y} \\
&\Longrightarrow dg\left(\frac{1}{g} - \frac{3}{g+1}\right) = \frac{dy}{y} \\
&\Longrightarrow \ln g - 3\ln(g+1) = \ln y - \ln C_1 \\
&\Longrightarrow \frac{C_1 g(y)}{(g(y) + 1)^3} = y.
\end{aligned}$$

where $C_1 > 0$ is a constant pinned down by the boundary conditions. So

$$\begin{aligned}
y &= \frac{C_1 g}{(g+1)^3} \\
x &= \frac{C_1 g^2}{(g+1)^3}.
\end{aligned} \tag{3.28}$$

Now with these expressions, we determine $g(.)$ as a function of z

$$\begin{aligned}
x' &= g'C_1\frac{2g(g+1)-3g^2}{(g+1)^4} \\
&= \frac{C_1g^2}{(g+1)^3}\left(2\frac{C_1g}{(g+1)^3}-\frac{C_1g^2}{(g+1)^3}\right) \\
&\Longrightarrow g'(2(g+1)-3g) = C_1g^2\left(2\frac{1}{(g+1)^2}-\frac{g}{(g+1)^2}\right) \\
&\Longrightarrow g'(z) = \frac{C_1g(z)^2}{(g(z)+1)^2} \\
\text{or equivalently } \frac{dg(g+1)^2}{g^2} &= C_1dz \\
&\Longleftrightarrow \frac{dg}{g^2}+2\frac{dg}{g}+dg = C_1dz \\
&\Longleftrightarrow -\frac{1}{g(z)}+2\ln(g(z))+g(z) = C_1z+C_2 \\
&\Longleftrightarrow g(z)-\frac{1}{g(z)}+2\ln g(z) = C_1z+C_2. \qquad (3.29)
\end{aligned}$$

Again, the constant C_2 is pinned down by the boundary conditions. We come back to write equations determining C_1 and C_2. Since $x = J_A'(z)$ and $y = -J_B'(z)$, the Lebnitz's rule implies

$$\begin{aligned}
P_A &= J_A(K_A) - J_A(-K_B) \\
&= \int_{-K_B}^{K_A} x(z)\,dz \\
P_B &= -J_B(K_A) + J_B(-K_B) \\
&= \int_{-K_B}^{K_A} y(z)\,dz
\end{aligned}$$

Using the previous closed-form yields

$$\int_{-K_2}^{K_1} x(z)\,dz = \int_{-K_2}^{K_1} \frac{C_1g(z)^2}{(g(z)+1)^3}dz$$

Differentiate (3.29) with respect to z we have

$$\frac{dg\,(g+1)^2}{g^2} = C_1 dz$$

The integral becomes

$$\begin{aligned}
\int\limits_{-K_B}^{K_A} x\,(z)\,dz &= \int\limits_{-K_B}^{K_A} \frac{C_1 g\,(z)^2}{(g\,(z)+1)^3} dz \\
&= \int\limits_{g(-K_B)}^{g(K_A)} \frac{g^2}{(g+1)^3} \frac{(g+1)^2}{g^2} dg \\
&= \int\limits_{g(-K_B)}^{g(K_A)} \frac{1}{(g+1)} dg \\
&= \ln\left(\frac{1+g\,(K_A)}{1+g\,(-K_B)}\right) \\
&= P_A
\end{aligned}$$

Similarly

$$\begin{aligned}
\int\limits_{-K_B}^{K_A} y\,(z)\,dz &= \int\limits_{-K_B}^{K_A} \frac{C_1 g\,(z)}{(g\,(z)+1)^3} dz \\
&= \int\limits_{g(-K_B)}^{g(K_A)} \frac{g}{(g+1)^3} \frac{(g+1)^2}{g^2} dg \\
&= \int\limits_{g(-K_B)}^{g(K_A)} \frac{1}{(g+1)\,g} dg \\
&= \ln\left(\frac{g\,(K_A)}{g\,(-K_B)}\right) - \ln\left(\frac{1+g\,(K_A)}{1+g\,(-K_B)}\right) \\
&= P_B
\end{aligned}$$

We can compute then $g(K_A)$ and $g(-K_B)$ explicitly in functions of K_A, K_B, P_A, P_B

$$\begin{aligned} \frac{1+g(K_A)}{1+g(-K_B)} &= \exp(P_A) \\ \text{and } \frac{g(K_A)}{g(-K_B)} &= \exp(P_A+P_B) \end{aligned}$$

i.e.

$$\begin{aligned} 1+g(K_A) &= \exp(P_A)(1+g(-K_B)) \\ &\Longleftrightarrow g(K_A) = \exp(P_A+P_B)\, g(-K_B) \\ &\Longleftrightarrow 1+g(K_A) = \exp(P_A) + \exp(-P_B)\, g(K_A) \end{aligned}$$

so

$$\begin{aligned} g(K_A) &= \frac{\exp(P_A)-1}{1-\exp(-P_B)} \\ &= \frac{\exp(P_A+P_B)-\exp(P_B)}{\exp(P_B)-1}, \end{aligned} \tag{3.30}$$

and

$$\begin{aligned} g(-K_B) &= \frac{1-\exp(-P_A)}{\exp(P_B)-1} \\ &= \frac{\exp(P_A)-1}{\exp(P_A+P_B)-\exp(P_A)}. \end{aligned} \tag{3.31}$$

Together with the two equations (3.29) on g at $z=-K_B$ and K_A, C_1 and C_2 are then

$$C_1 = \frac{1}{K_A+K_B}\begin{pmatrix} \frac{\exp(P_A+P_B)-\exp(P_B)}{\exp(P_B)-1} \\ -\frac{\exp(P_B)-1}{\exp(P_A+P_B)-\exp(P_B)} \\ -\frac{\exp(P_A)-1}{\exp(P_A+P_B)-\exp(P_A)} \\ +\frac{\exp(P_A+P_B)-\exp(P_A)}{\exp(P_A)-1} \\ +2P_A+2P_B \end{pmatrix}, \tag{3.32}$$

and

$$C_2 = \frac{1}{K_A + K_B} \left(\begin{array}{c} K_B \left(\begin{array}{c} \frac{\exp(P_A+P_B)-\exp(P_B)}{\exp(P_B)-1} \\ -\frac{\exp(P_B)-1}{\exp(P_A+P_B)-\exp(P_B)} \\ +2\ln\left(\frac{\exp(P_A+P_B)-\exp(P_B)}{\exp(P_B)-1}\right) \end{array} \right) \\ +K_A \left(\begin{array}{c} \frac{\exp(P_A)-1}{\exp(P_A+P_B)-\exp(P_A)} \\ -\frac{\exp(P_A+P_B)-\exp(P_A)}{\exp(P_A)-1} \\ +2\ln\left(\frac{\exp(P_A)-1}{\exp(P_A+P_B)-\exp(P_A)}\right) \end{array} \right) \end{array} \right). \tag{3.33}$$

To conclude we have: The strategy functions are

$$\begin{aligned} x(z) &= C_1 \frac{g(z)^2}{(1+g(z))^3} \\ y(z) &= C_1 \frac{g(z)}{(1+g(z))^2}. \end{aligned} \tag{3.34}$$

and the payoff function for each player is:

$$\begin{aligned} J_A(z) &= \ln\left(\frac{1+g(z)}{1+g(-K_B)}\right) \\ J_B(z) &= \ln\left(\frac{g(K_A)}{g(z)}\right) - \ln\left(\frac{1+g(K_A)}{1+g(z)}\right). \end{aligned} \tag{3.35}$$

■

Proof of Proposition 3.1. Since $g(.)$ is increasing, $g(z) > g(z^*) = 1$ so $x(z) = g(z)\,y(z) > y(z)$.Second, since $g(z) > 1$ and $g'(z) > 0$ we have $y'(z) = C_1 \frac{1-2g(z)}{(1+g(z))^4} g'(z) < 0$. Finally, $x'(z) = C_1 \frac{(2-g(z))g(z)}{(1+g(z))^4} g'(z)$, so for $z > z_A = g^{-1}(2) > g^{-1}(1) = z^*$ we have $2 - g(z) < 0$. Thus $x'(z) > 0$. ■

Derivation of the Expected Completion Time. To compute $E_0[\tau]$,we first make use of the change of variable from z to g using (3.17).By Ito's Lemma, we have

$$\begin{aligned} dg(Z_t) &= g'(Z_t)\,dZ_t + \frac{1}{2} g''(Z_t)\,(dZ_t)^2 \\ &= \left(g'(Z_t)(x_t - y_t) + \frac{1}{2} g''(Z_t)\right) dt + g'(Z_t)\,dW_t. \end{aligned}$$

From (3.17), we have

$$\begin{aligned} g' &= C_1 \frac{g^2}{(g+1)^2} \\ g'' &= 2C_1^2 \frac{g^3}{(g+1)^5}. \end{aligned}$$

So

$$\begin{aligned} dg_t &= \left(C_1 \frac{g^2}{(g+1)^2} C_1 \frac{g(g-1)}{(g+1)^3} + C_1^2 \frac{g^3}{(g+1)^5} \right) dt \\ &\quad + C_1 \frac{g^2}{(g+1)^2} dW_t \\ &= C_1^2 \frac{g_t^4}{(g_t+1)^5} dt + C_1 \frac{g_t^2}{(g_t+1)^2} dW_t. \end{aligned}$$

Thus, in terms of g, we can define the completion time τ as $\tau := \inf_t g_t \notin (g(-K_2), g(K_1)) = (g_2, g_1)$. ■

Lemma 3.3 $E_0[\tau | g(t=0) = g_0(z_0)] = v(g_0(z_0))$, *where v is the unique solution of the boundary problem*

$$\frac{1}{2} C_1^2 \frac{g^4}{(g+1)^4} v''(g) + C_1^2 \frac{g^4}{(g+1)^5} v'(g) + 1 = 0 \tag{3.36}$$

with

$$v(g_2) = v(g_1) = 0.$$

The solution to (3.36) *is*

$$v(g) = -\frac{CC1}{g+1} - \frac{g^2}{3C_1^2} - \frac{14g}{3C_1^2} - \frac{1}{3C_1^2 g^2} - \frac{14}{3C_1^2 g} + \frac{20\log(g)}{C_1^2} - \frac{40\log(g)\, g}{C_1^2 (g+1)} + CC2 \tag{3.37}$$

where $CC1$ and $CC2$ are constant and depend only on K_1, K_2, P_1, P_2 such that $v(g_2) = v(g_1) = 0$.

Proof. Using Dynkin's formula for the process g_t and the function v, we have $E[v(g_\tau)] = v(g_0) + E\left[\int_0^\tau \frac{1}{2} C_1^2 \frac{g_s^4}{(g_s+1)^4} v''(g_s) + C_1^2 \frac{g_s^4}{(g_s+1)^5} v'(g_s)\, ds\right] = v(g_0) - E[\tau]$. Since $v(g_\tau) = v(g_1)$ or $v(g_2), = 0$ then $E[\tau] = v(g_0)$. ■

Proof of Proposition 3.2. I expand the partial derivative $\frac{\partial E[\tau]}{\partial P_A}$ with $P_A = P_B = P \longrightarrow$

∞ using the following approximation

$$\begin{aligned} g_A &\simeq \exp\left(\left(1+s\right)P_A\right)+\exp\left(\left(1+s\right)P_A-P_B\right) \\ g_B &\simeq \exp\left(-P\right). \end{aligned}$$

Then

$$\begin{aligned} C_1 &\simeq \frac{1}{2}\left(g_A+\exp\left(P\right)\right) \\ g_0 &\simeq \frac{1}{2}\left(g_A-\exp\left(P\right)\right). \end{aligned}$$

Using (3.37), I obtain

$$\begin{aligned} \frac{\partial E\left[\tau\right]}{\partial P_A} &\simeq \frac{1}{C_1^3}\frac{1}{2}g_A e^P \\ &\simeq \frac{1}{C_1^3}\frac{1}{2}e^{(2+s)P} > 0. \end{aligned}$$

I also expand the partial derivative $\frac{\partial E[\tau]}{\partial P_A}$ with $P_A = P_B = P \longrightarrow 0$. Plugging in the expression of g_A, g_B, C_1 and g_0 into Mathematica using the Series[] built-in function, I obtain

$$\frac{\partial E\left[\tau\right]}{\partial P_A} = -\frac{1}{3}P + O\left(P^2\right).$$

Therefore $\frac{\partial E[\tau]}{\partial P_A} < 0$ as P goes to 0. ■

Discretization. I discretize the interval $(-K_B, K_A)$ into

$$(-K_B = z_0 < z_1 < \ldots < z_{N+1} = K_A)$$

with equal steps Δ. For all functions $f\left(z\right)$ over $\left(-K_B, K_A\right)$, we have

$$\begin{aligned} f''\left(z_i\right) &\approx \frac{f\left(z_{i-1}\right)+f\left(z_{i+1}\right)-2f\left(z_i\right)}{\Delta^2} \\ f'\left(z_i\right) &\approx \frac{f\left(z_{i+1}\right)-f\left(z_{i-1}\right)}{2\Delta}. \end{aligned}$$

The discretized version of (3.7) is

$$\begin{pmatrix} \frac{J_A(z_{i-1})+J_A(z_{i+1})-2J_A(z_i)}{\Delta^2} \\ \frac{J_B(z_{i-1})+J_B(z_{i+1})-2J_B(z_i)}{\Delta^2} \end{pmatrix} = \begin{pmatrix} 2rJ_A(z_i) + 2F_A\left(\frac{J_A(z_{i+1})-J_A(z_{i-1})}{2\Delta}, \frac{J_B(z_{i+1})-J_B(z_{i-1})}{2\Delta}\right) \\ 2rJ_B(z_i) + 2F_B\left(\frac{J_A(z_{i+1})-J_A(z_{i-1})}{2\Delta}, \frac{J_B(z_{i+1})-J_B(z_{i-1})}{2\Delta}\right) \end{pmatrix}$$

which can be rewritten iteratively as

$$\begin{pmatrix} J_A(z_i) \\ J_B(z_i) \end{pmatrix} = \frac{1}{2(1+r\Delta^2)} \begin{pmatrix} J_A(z_{i-1}) + J_A(z_{i+1}) - \Delta^2 F_A\left(\frac{J_A(z_{i+1})-J_A(z_{i-1})}{2\Delta}, \frac{J_B(z_{i+1})-J_B(z_{i-1})}{2\Delta}\right) \\ J_B(z_{i-1}) + J_B(z_{i+1}) - \Delta^2 F_B\left(\frac{J_A(z_{i+1})-J_A(z_{i-1})}{2\Delta}, \frac{J_B(z_{i+1})-J_B(z_{i-1})}{2\Delta}\right) \end{pmatrix}. \tag{3.38}$$

In the case where the cost functions are given by (3.19)

$$\begin{aligned} x_A(z_i) &= \left((1+s)\frac{J_A(z_{i+1}) - J_A(z_{i-1})}{2\Delta}\right)^{\frac{1}{k}} \\ x_B(z_i) &= \left(\frac{J_B(z_{i+1}) - J_B(z_{i-1})}{2\Delta}\right)^{\frac{1}{k}}. \end{aligned}$$

By Dynkin's formula the expected completion time is $v(0)$ where v is solution of the following boundary value problem$\frac{1}{2}v''(z)$

$$\begin{gathered} \frac{1}{2}v''(z) + (x_A(z) - x_B(z))\, v'(z) + 1 = 0 \\ v(-K_B) = v(+K_A) = 0. \end{gathered} \tag{3.39}$$

Again, I can solve this boundary value problem by discretization

$$\begin{gathered} \frac{1}{2}\frac{v(z_{i+1}) - 2v(z_i) + v(z_{i-1})}{\Delta^2} + (x_A(z_i) - x_B(z_i))\frac{v(z_{i+1}) - v(z_{i-1})}{2\Delta} + 1 = 0 \\ v(-K_B) = v(+K_A) = 0. \end{gathered}$$

■

Bibliography

Abernathy, W. (1980). *The Productivity Dilemma.* Johns Hopkins University Press, Baltimore.

Acemoglu, D. (2008). *Political Economy Lecture Notes.* TBA.

Acemoglu, D. and U. Akcigit (2006). State dependent intellectual property right policy. NBER Working Paper, No. 12775.

Aghion, P., N. Bloom, R. Blundell, R. Griffith, and P. Howitt (2005). Competition and innovation: An inverted-u relationship. *Quarterly Journal of Economics 120*(2), 701–728.

Aghion, P., R. Burgess, S. Redding, and F. Zilibotti (2005). Entry liberalization and inqequality in industrial performance. *Journal of the European Economic Association Papers and Proceed 3*, 291–302.

Aghion, P., C. Harris, P. Howitt, and J. Vickers (2001). Competition, imitation and growth with step-by-step innovation. *Reviews of Economic Studies 68*, 467–492.

Aghion, P. and P. Howitt (1992). A model of growth through creative destruction. *Econometrica 60*(2), 323–351.

Alvarez, F. and U. J. Jermann (2000, July). Efficiency, equilibrium, and asset pricing with risk of default. *Econometrica 68*(4), 775–798.

Alvarez, F. and U. J. Jermann (2001). Quantitative asset pricing implications of endogenous solvency constraints. *The Review of Financial Studies 14*(4), 1117–1151.

Araujo, A., F. Kubler, and S. Schommer (2009). Regulating collateral requirements when markets are incomplete. Working Paper.

Arrow, K. (1974). *The Limits of Organization.* Norton, New York.

Arrow, K. J. (1962). The economic implications of learning by doing. *Review of Economic Studies 29*, 155–173.

Atkeson, A. and A. Burstein (2007). Innovation, firm dynamics and international trade. UCLA mimeo.

Axtell, R. (2001). Zipf distribution of us firm sizes. *Science 293*, 1818–1820.

Bartelsman, E. J. and M. Doms (2000). Understanding productivity: Lessons from longitudinal microdata. *Journal of Economic Literature 38*, 569–594.

Bass (1958). On non-linear repulsive forces. *Contributions to the theory of non-linear oscillations 4*, 201–211.

Beker, P. and S. Chattopadhyay (2009). Consumption dynamics in general equilibrium: A characterisation when markets are incomplete. Working Paper, University of Warwick.

Beker, P. and E. Espino (2010). A general equilibrium explanation for financial market anomalies: Belief heterogeoneity under limited enforceability. University of Warwick Working Paper.

Blume, L. and D. Easley (2006). If you're so smart, why aren't you rich? belief selection in complete and incomplete markets. *Econometrica 74*(4), 929–966.

Broda, C. and D. E. Weinstein (1996). Globalization and the gains from variety. *Quarterly Journal of Economics CXXI*(2), 541–585.

Budd, C., C. Harris, and J. Vickers (1993). A model of the evolution of duopoly: Does the asymmetry between firms tend to increase or decrease? *Reviews of Economic Studies 60*, 543–573.

Chien, Y. and H. Lustig (2009). The market price of aggregate risk and the wealth distribution. Review of Financial Studies, forthcoming.

Cogley, T. and T. Sargent (2008). Diverse beliefs, survival, and the market price of risk. Working Paper, NYU.

Coury, T. and E. Sciubba (2005). Belief heterogeneity and survival in incomplete markets. Working Paper, University of Oxford.

Debreu, G. (1959). *Theory of Value.* John Wiley.

Dobkins, L. and Y. Ioannides (1998). Dynamic evolution of the u. s. city size distribution. In J. M. Huriot and J. F. Thisse (Eds.), *The Economics of Cities.* Cambridge University Press, New York.

Duffie, D., J. Geanakoplos, A. Mas-Colell, and A. McLennan (1994). Stationary markov equilibria. *Econometrica 62*(4), 745–781.

Dunne, Timothy, M. J. R. and L. Samuelson (1988). Patterns of firm entry and exit in us manufacturing industries. *Rand Journal of Economics 19*, 495–515.

Dunne, Timothy, M. J. R. and L. Samuelson (1989). The growth and failure of us manufacturing plants. *Quarterly Journal of Economics 104*, 671–698.

Eeckhout, J. and B. Jovanovic (2002). Knowledge spillovers and inequality. *American Economic Review 92*, 1290–1307.

Ericson, R. and A. Pakes (1995). Markov perfect industry dynamics: A framework for empirical work. *Review of Economic Studies 62*, 53–82.

Fostel, A. and J. Geanakoplos (2008). Leverage cycles and the anxious economy. *American Economic Reviews 98*(4), 1211–1244.

Foster, Lucia, J. H. and C. Krizan (2000). Aggregate procuctivity growth: Lessons from microeconomic evidence. NBER Working Paper No.6803.

Freeman, C. (1982). *The Economics of Industrial Innovation.* MIT Press Cambridge, Massachusetts.

Friedman, M. (1953). *Essays in Positive Economics.* University of Chicago Press.

Fudenberg et al. (1983). Pre-emption, leapfrogging and competition in patent races. *European Economic Review 22*, 3–31.

Gabaix, X. (1999). Zipf's law for cities: An explanation. *Quarterly Journal of Economics 114*, 739–767.

Geanakoplos, J. (2009). The leverage cycle. Cowels Foundation Discussion Paper.

Geanakoplos, J. and H. Polemarchakis (1986). Existence, regularity, and constrained suboptimality of competitive allocations when the asset market is incomplete. In W. Heller,

R. Starr, and D. Starrett (Eds.), *Essays in Honor of Kenneth Arrow*, Volume 3, pp. 65–95. Cambridge University Press.

Geanakoplos, J. and W. Zame (2002). Collateral and the enforcement of intertemporal contracts. Discussion Paper, Yale University.

Geanakoplos, J. and W. Zame (2007). Collateralized asset market. Discussion Paper, Yale University.

Grossman, G. and E. Helpman (1991a). *Innovation and Growth in the Global Economy.* Cambridge, MA, MIT Press.

Grossman, G. and E. Helpman (1991b). Quality ladders in the theory of growth. *Review of Economic Studies 68*, 43–61.

G.Scorza-Dragoni (1935). Sul problema dei valori ai limiti per i systemi di equazioni differenziali del secondo ordine. *Boll.Un.Mat.Ital 14*, 225–230.

Harris and Vickers (1985). Perfect equilibrium in a model of a race. *Reviews of Economic Studies 53*, 193–209.

Harris and Vickers (1987, January). Racing with uncertainty. *Reviews of Economic Studies 54*(1), 1–21.

Harrison, M. and D. Kreps (1978). Speculative investor behavior in a stock market with heterogeneous expectations. *Quarterly Journal of Economics 92*, 323–336.

Hartman (1960, September). On boundary value problems for systems of ordinary, nonlinear, second order differential equation. *Trans.Am.Math.*

Hartman, P. (1964). *Ordinary Differential Equation.* John Wiley.

Hopenhayn, H. A. (1992). Entry, exit, and firm dynamics in long run equilibrium. *Econometrica 65*(5), 1127–1150.

Horner (1999). A perpetual race to stay ahead: the continous-time analogue. mimeo UPenn.

Jankowski, T. (2005). Advanced differential equations with nonlinear boundary condition. *Journal of Mathematical Analysis and Applications 304*, 490–503.

Jones, C. I. (1995). R and d-based models of economic growth. *Journal of Political Economy 103*, 759–784.

Jovanovic, B. (1982). Selection and evolution of industry. *Econometrica 50*, 649–670.

Kilenthong, W. T. (2009). Collateral premia and risk sharing under limited commitment. Working Paper, University of California, Santa barbara.

Kilenthong, W. T. and R. Townsend (2009). Market based, segregated exchanges in securities with default risk. Working Paper, MIT.

Kiyotaki, N. and J. Moore (1997). Credit cycles. *Journal of Political Economy 105*(2), 211–248.

Klepper, S. (1996). Entry, exit, growth, and innovation over the product life cycle. *The American Economic Review 86*(3), 562–583.

Klette, T. J. and S. Kortum (2004). Innovating firms and aggregate innovation. *Journal of Political Economy 112*(5), p986 – 1018.

Kogan, L., S. Ross, J. Wang, and M. Westerfield (2006). The price impact and survival of irrational traders. *Journal of Finance 61*, 195–229.

Krusell, P. and J. Smith, Anthony A. (1998). Income and wealth heterogeneity in the macroeconomy. *The Journal of Political Economy 106*(5), 867–896.

Kubler, F. and K. Schmedders (2003). Stationary equilibria in asset-pricing models with incomplete markets and collateral. *Econometrica 71*(6), 1767–1793.

Lagos, R. (2001). A model of tfp. NYU Working Paper.

Lazear and Rosen (1981). Rank-order tournaments as optimum labor contracts. *Journal of Political Economy 89*, 841–864.

Lentz, R. and D. T. Mortensen (2008). On the size distribution of business firms. *Econometrica 76*(6), 1317–1373.

Lieberman, M. B. (1984). The learning curve and pricing curve and chemical processing industries. *Rand Journal of Economics 15*, 213–228.

Long, J. B. D., A. Shleifer, L. H. Summers, and R. J. Waldmann (1990). Noise trader risk in financial markets. *The Journal of Political Economy 98*(4), 703–738.

Long, J. B. D., A. Shleifer, L. H. Summers, and R. J. Waldmann (1991). The survival of noise traders in financial markets. *The Journal of Business 64*(1), 1–19.

Lorenzoni, G. and K. Walentin (2009). Financial frictions, investment and tobin's q. MIT Working Paper.

Lucas, Robert E., J. (1978a). Asset prices in an exchange economy. *Econometrica 46*(6), 1429–1445.

Lucas, Robert E., J. and E. C. Prescott (1971). Investment under uncertainty. *Econometrica 39*(5), 659–681.

Lucas, R. E. J. (1978b). On the size distribution of business firms. *Bell Journal of Economics 9*(2), 508–523.

Luttmer, E. (2004). The size distribution of firms in an economy with fixed entry costs. Federal Reserve Bank of Minneapolis, working paper.

Luttmer, E. (2007). Selection, growth and the size distribution of firms. *Quarterly Journal of Economics 122*, 1103–1144.

Magill, M. and M. Quinzii (1994). Infinite horizon incomplete markets. *Econometrica 62*(4), 853–880.

Melitz, M. (2003). The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica 71*(6), 1695–1725.

Miao, J. (2006). Competitive equilibria of economies with a continuum of consumers and aggregate shocks. *Journal of Economic Theory 128*, 274–298.

Moscarini and Smith (2007, November). Optimal dynamic contests. Cowles Foundation Working Paper, Yale University.

Pennings, J. M. and A. Buitendam (1987). *New Technology As Organizational Innovation: The Development and Diffusion of Microelectronics.* Bollinger, Cambridge, MA.

Radner, R. (1972). Existence of equilibrium of plans, prices, and price expectations in a sequence of markets. *Econometrica 40*(2), 289–303.

Reinganum (1983). Uncertain innovation and the persistence of monopoly. *American Economic Review 73*, 741–748.

Romer, P. (1990). Endogenous technical change. *Journal of Political Economy 98*(5), 71–102.

Rossi-Hansberg, E. and M. L. Wright (2007a). Establishment size dynamics in the aggregate economy. *American Economic Reviews 97*(5), 1639Ű1666.

Rossi-Hansberg, E. and M. L. Wright (2007b). Urban structure and growth. *Review of Economic Studies 74*, 597Ű624.

Sandroni, A. (2000, November). Do markets favor agents able to make accurate predicitions? *Econometrica 68*(6), 1303–1342.

Scheinkman, J. A. and W. Xiong (2003, December). Overconfidence and speculative bubbles. *Journal of Political Economy 111*(6), 1183–1219.

Scherer, F. (1984). *Innovation and Growth: Schumpeterian Perspectives.* MIT Press Cambridge, Massachusetts.

Schumpeter, J. A. (1934). *The Theory of Economic Development.* Harvard University Press, Cambridge, MA.

Schumpeter, J. A. (1942). *Capitalism, Socialism and Democracy.* Harper and Brothers, New York.

Segerstrom, Paul S., T. A. A. and E. Dinopoloulos (1990). A schumpterian model of the product life cycle. *American Economic Reviews 80*, 1077–1091.

Simon, H. and C. P. Bonini (1958). The size distribution of business firms. *American Economic Reviews 48*, 607–617.

Simsek, A. (2009a). Speculation and risk sharing with new financial assets. MIT Working Paper.

Simsek, A. (2009b). When optimism needs credit: Asymmetric filtering of optimism and implications for asset prices. MIT Working Paper.

Staikos, V. and Tsamatos (1981). On the terminal value problem for differential equations with deviating arguments. *Archivum Mathematicum 21*(2), 43–50.

Sutton, J. (1997). Gibrat's legacy. *Journal of Economic Literature 35*, 40–59.

Sutton, J. (1998). *Technology and Market Structure:Theory and History.* MIT Press, Cambridge, MA.

Tushman, M. L. and P. Anderson (1986). Technological discontinuities and organizational environments. *Administrative Science Quarterly 31*, 439–465.

Xiong, W. and H. Yan (2009). Heterogeneous expectations and bond markets. Cowels Foundation Discussion Paper.