Essays on Business Cycles, Unemployment, and Investment

Jiayi Li

Submitted in Partial Fulfillment of the Requirements for the Degree

Doctor of Philosophy

in

Economics

at

Tepper School of Business

Carnegie Mellon University

Dissertation Committee:

Professor Stephen Spear (co-chair)

Professor Bryan Routledge (co-chair)

Professor Laurence Ales

Professor David Childers

Professor Pierre Liang

Abstract

Economic agents adapt to expected and unexpected shocks in their decision-making. This thesis develops three theoretical chapters about the business cycle, and studies three different sets of economic activities through numerical analysis.

Chapter 1 studies the consumption and saving behavior that can lead to endogenous fluctuations in interest rates. This chapter is based on a joint paper with Eungsik Kim. We analyze the implications of quasi-hyperbolic discounting preference for two types of endogenous economic fluctuations, endogenous deterministic cycles and local sunspot equilibria, in a three-period overlapping generations (OLG) economy with pure exchange. We provide a sufficient condition for the existence of two-period endogenous cycles and a necessary and sufficient condition for the existence of a local sunspot equilibrium characterized by local indeterminacy. We show that introducing the present bias into preferences shrinks the set of two-period cycles but enlarges the set of locally indeterminate equilibria. Moreover, our model suggests that the locally indeterminate equilibria exist under a reasonable value of time discount factor.

Chapter 2 studies the employment decisions of firms in terms of wage offering and new job creation under an environment with volatile labor productivities. This chapter is based on a joint paper with Eungsik Kim and Stephen Spear. We introduce a novel model that incorporates both search friction and imperfect competition in the labor market through a two-stage game. We find that the level of competition increases wages, unemployment, and labor market volatility. Moreover, by varying how much labor assignment depends on wage bidding versus vacancy posting, we find that the labor market becomes more volatile as the weight of wages on labor assignment increases. The effect of competition level among firms is also more significant when labor assignment is decided by wages.

Chapter 3 studies the innovation and investment decisions of large technology firms and venture capitalists facing a newly emerged technology. This chapter is based on my job market paper. We develop a theory which connects corporate innovation with VC investment through corporate takeover activities of startups. We explore the mechanism where corporate decision makers use the level of VC investment to predict the acquisition opportunities in the near future, and make in-house R&D decisions accordingly. We show that increase in VC investment deters corporate internal R&D, and the deterrent effect is stronger for low-profit technologies. A strategic venture capitalist has more incentives to invest if corporate R&D can be more easily deterred, since it increases the demand to acquire their startups. The theory thus predicts high VC investment in technologies with lower profit than those firms invest in.

Acknowledgement

I am deeply indebted to my advisors in developing this thesis. Professor Stephen Spear has been very supportive since the beginning of my PhD career. He encouraged each and every one of my research idea when I was developing my first and second year summer papers, and provided insightful opinions. He was also extremely open-minded when I expressed interest in Machine Learning. Professor Bryan Routledge always enlightened me with his expertise in Machine Learning and Finance. I was constantly inspired by his enthusiasm towards research. I also benefited from Professor Laurence Ales, Professor David Childers, Professor Shu Lin Wee, and Professor Pierre Liang for their invaluable feedback and support. I am also grateful to all other colleagues who have provided feedback to my research.

My gratitude also goes to my peers and friends. My co-author Eungsik Kim contributed greatly to our joint project, and motivated me with his self-discipline. I benefited from the discussions with my fellow PhD students Tianyi Yang, Siyu Lu, and Yue Zhang on research, coursework, job market, and other aspects. I have enjoyed the past six years with their companion. My remote friend Xijia Zheng provided tremendous mental support throughout my PhD life. I also thank my partner Wenjie Xue for his unreserved support and constant encouragement.

I thank Lawrence Rapp and Laila Lee for their help in facilitating everything in the program. I also thank all other professional staff at Carnegie Mellon University, and everyone I met in Pittsburgh who have helped me. I am lucky to have spent six years in such a beautiful city.

Finally, I would like to thank my parents Sen Li and Chunqi Zu for supporting all my decisions and believing in me. I would never have achieved this far without the values and principles I learned from them. I dedicate this thesis to them.

Contents

Chapter 1

Endogenous Fluctuations in Behavioral Overlapping Generations Model

We analyze the implications of quasi-hyperbolic discounting preference for two types of endogenous economic fluctuations, endogenous deterministic cycles and local sunspot equilibria, in a three-period overlapping generations (OLG) economy with pure exchange. We provide a sufficient condition for the existence of two-period endogenous cycles and a necessary and sufficient condition for the existence of a local sunspot equilibrium characterized by local indeterminacy. Through numerical characterization, we show that introducing the present bias into preferences shrinks the set of two-period cycles but enlarges the set of locally indeterminate equilibria. Moreover, unlike in a standard two-period OLG model with exponential discounting preferences, our model suggests that the locally indeterminate equilibria exist under a reasonable value of time discount factor, supporting sunspot as a source of economic fluctuation. We also find that two-period cycles arise under a skewed endowment profile toward the young, whereas a local sunspot equilibrium exists under a hump-shaped endowment profile. This result breaks down the conventional finding in a two-period OLG economy with a single commodity where the set of economies for the two types of endogenous economic fluctuations coincide.

1.1 Introduction

Understanding the origin of economic fluctuations is crucial to shaping stabilizing macroeconomic policies. While the real business cycle (RBC) literature argues that exogenous productivity shocks generate fluctuations in economies (Kydland & Prescott, 1982), the endogenous business cycle (EBC) and sunspot equilibrium (SE) theories provide frameworks where economic fluctuations are internally driven by changes in the beliefs of agents under overlapping generations (OLG) (See Benhabib and Day (1982); Grandmont (1985) for EBC, and Azariadis (1981); Cass and Shell (1983) for SE). Due to the different perspectives on the source of economic fluctuations, the two paradigms deliver different policy conclusions. The RBC school holds that economic fluctuations arise as the optimal allocations and there is little room for government interventions. On the other hand, the EBC and SE schools argue that economic fluctuations are suboptimal and support the adoption of countercyclical fiscal and monetary policies.

Despite providing endogenous mechanisms behind economic fluctuations, EBC and SE theories have been criticized as unrealistic for the following reasons. Firstly, these studies have restricted their attention to two-period-lived OLG models with standard exponential discounting preferences where business cycles have the same length as the lifetime of the households. Secondly, the existence of EBC and SE in these models requires extremely low discount factors, an endowment structure concentrated to the young generation and high curvature of the utility function (Bhattacharya & Russell, 2003; Orrego, 2014). To better evaluate the plausibility of endogenous fluctuations, more realistic extensions from the over-simplified standard two-period model are needed.

With the above motivation, this paper studies both EBC and SE in a three-period OLG model with quasi-hyperbolic discounting (QHD) preferences. The QHD preferences discount the future differently than standard exponential discounting preferences (discussed below). A three-period OLG model introduces a new middle-aged generation. Thus, these extensions can change the requirement of the existence of EBC and SE on the parameter space. In addition to overcoming the criticisms above, we also directly compare and differentiate the conditions for EBC and SE. Here, we focus on local sunspot equilibria which arise around locally stable steady states with a continuum of convergent equilibrium sequences (Woodford, 1986).[1] The two types of endogenous fluctuations have equivalent conditions in the standard two-period model with one good. We check whether the result still holds in our model. (We show that the result does not hold in our model.) We study these questions based on a combination of analytical and computational approaches.

Quasi-hyperbolic discounting refers to the type of discounting scheme where agents have a bias towards the current period in evaluating their lifetime consumption allocations, rather than simply being impatient. Such a present bias can justify some prevailing yet puzzling saving patterns such as extremely low saving rates and aggresive borrowing (Angeletos, Laibson, Repetto, Tobacman, & Weinberg, 2001; Laibson, 1997; Laibson, Repetto, & Tobacman, 2007). Changes in these intertemporal decisions are crucial to the existence of endogenous economic fluctuations. In fact, for a two-period EBC to exist in our model, the income effect of agents

[1] The literature also examines other types of SE. For example, (Azariadis & Guesnerie, 1986) show the existence of two-state sunspot equilibria implied by the existence of deterministic monetary cycles and vice versa.

facing an increase in interest rate must dominate the substitution effect.[2] In the model, the parameter value of the quasi-hyperbolic discount factor controls the strength of the present bias, which then affect the magnitude of agents' income and substitution effects and the relative dominance between the two effects. Therefore, incorporating QHD preference into the model changes the set of economies where EBC exist. The existence of local SE depends on whether steady states are locally stable. The stability of steady states is determined by the assets/goods market clearing conditions. Household saving and consumption decisions affect the market clearing conditions. Thus, incorporating QHD preference into the model also changes the set of economies where local SE exist.

The extension to a three-period lifetime itself also affects the existence of endogenous fluctuation. With a middle-aged generation behaving similarly to the young generation facing changes in the interest rate, the requirement that aggregate income effect dominates substitution effect is more easily satisfied. The existence of a middle-aged generation also affects the stability of steady states since its saving influences aggregate savings and consumption and thus assets/goods market clearing conditions. The three-period setting allows us to have a two-period cycle length which is shorter than the lifetime and model a more realistic hump-shaped endowment structure. Moreover, a three-period lifetime is the minimum length for QHD preferences to reveal. We do not further extend the lifetime more than three periods so that we can maintain the analytical tractability of the model.

We provide an analytical sufficient condition for the existence of EBC, which is similar to that of the two-period model, i.e., the intertemporal elasticity of subtitution should be less than $-\frac{1}{2}$. As is discussed above, this condition essentially requires that the aggregate income effect dominates the substitution effect when interest rates rise. We stress that we have not found any examples numerically that cycles arise when the sufficient conditions is violated. We also derive the condition of local indeterminacy where local SE exists. Based on the analysis of Blanchard and Kahn (1980), this requires the dimension of stable manifolds around the steady-state is larger than the number of predetermined variables. In our three-period

[2] We use the following analysis to provide an intuition. Let real interest rates alternate between R_H and R_L over two-period cycles where $R_H > 1$ and $R_L < 1$, and let (c_{1H}, c_{2H}, c_{3H}) and (c_{1L}, c_{2L}, c_{3L}) be consumption streams for agents who face R_H and R_L in their first-period of life, respectively. The values of the consumption streams depend on the relative strengths of income and substitution effects. If the substitution effect dominates the income effect, then $c_{1H} < c_1$, $c_{2H} > c_2$, $c_{3H} < c_3$ and $c_{1L} > c_1$, $c_{2L} < c_2$, $c_{3L} > c_3$ where (c_1, c_2, c_3) is a consumption stream at a monetary steady state with $R = 1$. Note that these consumption profiles violate resource constraints because all (c_{1H}, c_{2L}, c_{3H}) is lower than, and all (c_{1L}, c_{2H}, c_{3L}) is higher than their corresponding values at the monetary steady state. Therefore, a necessary condition for two-period cycles is that the income effect should dominate the substitution effect to satisfy the resource constraint. (Then, $c_{1H} > c_1$, $c_{2H} > c_2$, $c_{3H} > c_3$ and $c_{1L} < c_1$, $c_{2L} < c_2$, $c_{3L} < c_3$. Hence, the goods market clearing conditions can be satisfied.)

OLG model, there is one predetermined price variable whereas the price dynamics system is three-dimensional. Moreover, in the special case where the QHD discount factor is one, the model reduces to the case with standard exponential discounting, which allows us to separately study the effect of extending the lifetime to three periods alone.

We find that endogenous economic fluctuations can exist under more reasonable parameter sets in our model. Firstly, EBC is more likely to emerge when the endowment structure is more concentrated on the young generation, the discount factor is lower and the risk aversion is higher, as in the standard two-period model. However, the quantitative requirement on the time discount factor is relaxed in our model given the same level of risk-aversion and endowment profile concentrated on the young. This more realistic range of parameters is facilitated by the additional middle-aged generation that strengthens the aggregate income effect. On the contrary, the QHD preference shrinks the set of EBC economies. Secondly, a local SE exists when either the endowment is extremely concentrated on the young or moderately centered on the middle-aged, unlike the two-period model. The former case requires a very low discount factor comparable to the two-period model given the same value of risk aversion. However, the latter case allows a moderately low value of discount factor that is acceptable for calibration task. This result is enhanced by adding QHD preferences because a stronger present bias enlarges the set of SE economies. We also find that unlike in the two-period standard model, the set of economies where EBC is possible does not completely overlap with that of local SE. In fact, the relative dominance between income and substitution effects does not play a role in determining the existence of local SE. This result reveals the importance of studying the two types of endogenous fluctuations simultaneously.

Our study contributes to the literature in two main ways. Firstly, we are the first to connect QHD preferences with endogenous fluctuations. We extend from previous work that study the conditions of EBC and local SE separately under standard exponential discounting (see Bhattacharya and Russell (2003) for EBC, and Kehoe and Levine (1990) for local SE). Our results suggest that the QHD preference plays an important role in generating endogenous fluctuations. However, we are not the first to incorporate behavioral preferences into the model. Bunzel (2006); Lahiri and Puhakka (1998); Orrego (2014) provide insights about both EBC and SE under habit formation preferences. Secondly, we are the first to directly compare the conditions of EBC and local SE. All of the above work only focus on one type of endogenous fluctuations, either EBC or local SE. Guesnerie (1986) is the only work to our knowledge that studies the conditions for both in a two-period model with multiple goods, but it does not clarify the relationship between them. We provide a comprehensive comparison between the two different types of endogenous fluctuations.

The rest of the paper is organized as follows. Section 1.2 describes our general overlapping generations model with QHD and its equilibrium. In Section 1.3, we study a sufficient condition for two-period cycles and characterize its set in the parameter space. Section 1.4 examines a necessary and sufficient condition for the local indeterminacy and identify where it exists in the parameter space. We relate endogenous cycles and local indeterminate equilibria in Section 1.5. The final section concludes this paper. Proofs and numerical algorithms are in appendices.

1.2 Model

In this section, we develop a pure exchange overlapping generation model with outside money. Time is discrete and indexed by t from 0 to infinity. In each period, a representative agent is born and live three-periods labeled as young, middle-aged and old. We assume no population growth. In the initial period, there are two households: an old household who lives period 0 only and a middle-aged household who lives periods 0 and 1.

The representative agent is endowed with a single perishable consumption good defined by a deterministic nonnegative vector $\boldsymbol{\omega} = (\omega_1, \omega_2, \omega_3)$ where ω_1 is the endowment when young, ω_2 is the endowment when middle-aged and ω_3 is the endowment when old. We assume $\omega_i > 0$ for $\forall i$. The consumptions of the representative household born in time t are denoted by $(c_{1,t}, c_{2,t+1}, c_{3,t+2})$ where $c_{1,t}$ is the consumption when young, $c_{2,t+1}$ is the consumption when middle-aged and $c_{3,t+2}$ is the consumption when old.

Agents trade the single commodity with the outside money to transfer income over time. In period 0, only the initial middle-aged and old obtain fixed money endowments $\bar{m}_{1,-1}$ and $\bar{m}_{2,-1}$ where $\bar{m}_{1,-1} + \bar{m}_{2,-1} = \bar{m}$. The young born in time t demands $m_{1,t}$ amount of money today and $m_{2,t+1}$ tomorrow. The aggregate money supply is assumed to be fixed at $\bar{m}$ for all the times.

Following Laibson (1997), we assume consumers are quasi-hyperbolic and their lifetime preferences are time-additively separable described by a utility function $U : \mathbb{R}^3_+ \to \mathbb{R}$:

$$U(c_{1,t}, c_{2,t+1}, c_{3,t+2}) = u(c_{1,t}) + \beta\delta u(c_{2,t+1}) + \beta\delta^2 u(c_{3,t+2}) \tag{1.1}$$

where $u(\cdot)$ is strictly increasing, strictly concave, and satisfies the Inada condition. β is the degree of present bias in intertemporal preferences. δ is the time discount factor. We assume $\beta \in (0, 1]$ and $\delta \in (0, 1]$. Note that if the present bias parameter sets at 1, i.e. $\beta = 1$, then the quasi-hyperbolic specification degenerates to the standard exponential discounting preference.

In (1.1), quasi-hyperbolic consumers employ time-variant discount factors by discounting immediate rewards at a high rate and distant results at a low rate. These time-varying discount functions give rise to dynamically inconsistent preferences under which the principle of optimality does not hold. Thus, agents might not commit to their current optimal plans in the future and would re-optimize plans in every period. We model an individual as a set of distinct 'temporal selves' who is in control for one period, respectively.

We can classify hyperbolic consumers into two types relying on how to expect the behaviors of future selves. A 'sophisticated' consumer rationally expects that their future selves will continue being quasi-hyperbolic and thus she solves her problem via backward induction. On the other hand, a 'naive' consumer forms the wrong belief that their future selves will implement the current optimal plans. We explain later how actual consumptions are different for the two types in more detail.

Let p_t be the price of the consumption good measured by the outside money. The representative agent born in time t chooses $(c_{1,t}, c_{2,t+1}, c_{3,t+2})$ and $(m_{1,t}, m_{2,t+1})$ to maximize (1.1), subject to the following sequential budget constraints:

$$\begin{aligned} c_{1,t} &\leq \omega_1 - m_{1,t}/p_t \\ c_{2,t+1} &\leq \omega_2 + \left(m_{1,t} - m_{2,t+1}\right)/p_{t+1} \\ c_{3,t+2} &\leq \omega_3 + m_{2,t+1}/p_{t+2} \end{aligned} \tag{1.2}$$

and $(c_{1,t}, c_{2,t+1}, c_{3,t+2}) \geq 0$. The agent's problem has a solution by the Weierstrass theorem since U is continuous and the constrained choice set is compact. Our strict concavity assumption for the utility function guarantees a unique solution.

Following Orrego (2014), we replace the nominal money holdings in period t with real ones measured by the consumption good as $a_{1,t} = m_{1,t}/p_t$ and $a_{2,t+1} = m_{2,t+1}/p_{t+1}$. We let A_t be the sum of the assets held by the young and middle-aged agents in time t, i.e. $A_t = a_{1,t} + a_{2,t}$. The gross rate of return is denoted by $R_t = p_t/p_{t+1}$. With these notations, the budget constraints in (1.2) can be re-written as:

$$\begin{aligned} c_{1,t} &\leq \omega_1 - a_{1,t} \\ c_{2,t+1} &\leq \omega_2 + R_t a_{1,t} - a_{2,t+1} \\ c_{3,t+2} &\leq \omega_3 + R_{t+1} a_{2,t+1} \end{aligned} \tag{1.3}$$

By eliminating money holdings and using the strict monotonicity of the preference, we can combine the three sequential budget constraints in (1.3) into a lifetime budget constraint:

$$c_{1,t} + \frac{c_{2,t+1}}{R_t} + \frac{c_{3,t+2}}{R_t R_{t+1}} = \omega_1 + \frac{\omega_2}{R_t} + \frac{\omega_3}{R_t R_{t+1}} \tag{1.4}$$

1.2.1 Sophisticated consumers

The sophisticated consumers rationally expect the behaviors of their future selves via backward induction. They know their future selves will keep the present bias toward the immediate gratification, and re-optimize the future consumptions higher than what they optimally choose today. The intertemporal Euler equation of the sophisticated consumer in age i at date t is given by:

$$u'(c_{i,t}) = R_t \underbrace{\left\{\beta\delta c'_{i+1}(x_{i+1,t+1}) + \delta\left(1 - c'_{i+1}(x_{i+1,t+1})\right)\right\}}_{\text{effective discount factor}} u'(c_{i+1,t+1}) \tag{1.5}$$

where $x_{i+1,t+1}$ is the wealth of an age-$(i+1)$ agent in time $t+1$. Thus, $x_{1,t+1} = \omega_1$, $x_{2,t+1} = \omega_2 + R_t a_{1,t} = \omega_2 + R_t(x_{1,t} - c_{1,t})$ and $x_{3,t+1} = \omega_3 + R_t a_{2,t} = \omega_2 + R_t(x_{2,t} - c_{2,t})$. See Appendix 1.A.1 for the derivation of (1.5).

For an agent of age i, the effective discount factor can be interpreted as a weighted sum of the short-run discount factor $\beta\delta$ and the long-run discount factor δ using $c'_{i+1}(x_{t+1})$ and $\left[1 - c'_{i+1}(x_{t+1})\right]$ as weights, where $c'_{i+1}(x_{t+1})$ is the marginal propensity to consumption (MPC) in time $t+1$. The effective discount factor is diminishing in MPC.

The naive consumer does not know that tomorrow she will choose a higher consumption than what she selects today. Unlike the naive consumer, the sophisticated consumer expects her future self will save less tomorrow which decreases the consumption two-periods later. Such present bias in the next period raises the marginal utility of consumption two-periods later from the current point of view. Thus, the sophisticated consumer does save more today than the naive consumer to carry on more assets to the next period which will increase the consumption after two-periods. This mechanism is reflected as the term $\delta(1 - \text{MPC})$ exists in the effective discount factor.

The optimal consumption plan of the three-period-lived sophisticated consumer born in date t is defined by the lifetime budget constraint (1.4) and the following two Euler equations:

$$u'(c_{1,t}) = R_t\left\{\beta\delta c'_2(x_{2,t+1}) + \delta(1 - c'_2(x_{2,t+1}))\right\} u'(c_{2,t+1}) \tag{1.6}$$

and

$$u'(c_{2,t+1}) = R_{t+1}\left\{\beta\delta c'_3(x_{3,t+2}) + \delta(1 - c'_3(x_{3,t+2}))\right\} u'(c_{3,t+2}) \tag{1.7}$$

where $c_3(x_{3,t+2}) = x_{3,t+2}$ since the agent consumes all her saving in the last period and thus, $c'_3(x_{3,t+2}) = 1$. (1.7) can be re-written as

$$u'(c_{2,t+1}) = \beta\delta R_{t+1} u'(c_{3,t+2}) \tag{1.8}$$

Since $c_{3,t+2} = \omega_3 + R_{t+1}(x_{2,t+1} - c_{2,t+1})$, we can change the equation above again into

$$u'(c_{2,t+1}) = \beta\delta R_{t+1} u'(\omega_3 + R_{t+1}(x_{2,t+1} - c_{2,t+1})) \tag{1.9}$$

By solving this equation under certain utility specifications or using the implicit function theorem, we can obtain $c_{2,t+1} = c_2(x_{2,t+1})$ and thus $c_2'(x_{2,t+1})$ as well. Lastly, we can solve for $c_{1,t}$ in $x_{1,t}$ by plugging the two terms ahead and $x_{2,t+1} = \omega_2 + R_t(x_{1,t} - c_{1,t})$ into (1.6) and then resorting to the implicit function theorem.

1.2.2 Naive consumers

The naive consumers conceive a wrong belief that their future selves will commit to the optimal plans of the current selves. The intertemporal Euler equation of the naive consumer in age i at date t is determined by:

$$u'\left(c_{i,t}^n\right) = \beta\delta R_t u'\left(c_{i+1,t+1}^e\right) \tag{1.10}$$

where $c_{i,t}^n$ is the actual consumption of a naive agent in age i at time t and $c_{i+1,t+1}^e$ is the planned consumption when being age $i+1$ at date $t+1$ of the current self in period t expecting that her future selves will be patient and not show present bias. The intertemporal Euler equation determining the planned consumption in time $t+1$ onward of the naive consumer in age i at date t is given by:

$$u'\left(c_{i+1,t+1}^e\right) = \delta R_{t+1} u'\left(c_{i+2,t+2}^e\right) \tag{1.11}$$

Note that the time discount factor is δ instead of $\beta\delta$ in (1.11) since naive consumers believe their future selves will be exponential discounting agents. See Appendix 1.A.2 for the derivations of (1.10) and (1.11).

The planned consumptions of the three-period-lived naive consumer born in date t are given by the lifetime budget constraint (1.4) and the following two Euler equations:

$$u'\left(c_{1,t}^n\right) = \beta\delta R_t u'\left(c_{2,t+1}^e\right) \tag{1.12}$$

and

$$u'\left(c_{2,t+1}^e\right) = \delta R_{t+1} u'\left(c_{3,t+2}^e\right) \tag{1.13}$$

However, her actual consumptions are defined by the lifetime budget constraint (1.4) and the following second period Euler equation instead of (1.13), after $c_{1,t}^n$ is determined by (1.12) and (1.13):

$$u'\left(c_{2,t+1}^n\right) = \beta\delta R_{t+1} u'\left(c_{3,t+2}^n\right) \tag{1.14}$$

where we replace the second period Euler equation with the one derived under the QHD assumption instead of the standard exponential discounting assumption. Note that we also substitute the acutal consumptions in the second and third periods, $c^n_{2,t+1}$ and $c^n_{3,t+2}$, for the planned consumptions under the quasi hyperbolic discounting, $c^e_{2,t+1}$ and $c^e_{3,t+2}$, in equation (1.13).

1.2.3 Perfect foresight equilibrium

We assume the entire population is either sophisticated or naive. This assumption does not significantly affect the endogenous business cycle and the local indeterminacy of equilibria below. In our deterministic model, agents have the perfect foresight on the equilibrium prices. The competitive equilibria with outside money for the sophisticated and naive cases are described in Definition (1.1) and Definition (1.2), respectively. We include the optimal choices by the initial generations in period 0 as well. The problems of the initial generations are the same across sophisticated and naive consumers because they can live either one or two periods. The initial old's problem is given by maximizing $u\left(c_{3,0}\right)$ subject to $c_{3,0} \leq \omega_3 + \alpha_{2,-1}$ where $\alpha_{2,-1}$ is the real wealth carried by the initial old from the previous period evaluated at today's price. The initial middle's problem is given by maximizing $u\left(c_{2,0}\right) + \beta\delta u\left(c_{3,1}\right)$ subject to $c_{2,0} \leq \omega_2 + \alpha_{1,-1} - a_{2,0}$ $c_{3,1} \leq \omega_3 + R_0 a_{2,0}$ where $\alpha_{1,-1}$ is the real wealth carried by the initial middle-aged from the previous period evaluated at today's price.[3]

Definition 1.1. *Given preferences, endowment structures and initial real wealth, a monetary competitive equilibrium in the three-period overlapping generations model populated by sophisticated consumers is a sequence of consumption bundles, money holdings and gross rates of returns* $\{c_{3,0}, c_{2,0}, c_{3,1}, a_{2,0}, c_{1,t}, c_{2,t+1}, c_{3,t+2}, a_{1,t}, a_{2,t+1}, R_t\}_{t=0}^{\infty}$ *such that*

1. Given the rates of return, the young sophisticated consumers choose consumptions and money holdings to fulfill the lifetime budget constraint (1.4) and two optimality conditions (1.6) and (1.7) plus $(c_{1,t}, c_{2,t+1}, c_{3,t+2}) \geq 0$ for $\forall t$.

2. Given $\alpha_{1,-1}$ and R_0, $c_{2,0}, c_{3,1}$ and $a_{2,0}$ are the optimal choices for the problem of the initial middle-aged. Given $\alpha_{2,-1}$, $c_{3,0}$ is the optimal choice for the problem of the initial old.

3. For $\forall t$, $A_t = \bar{m}/p_t$, i.e. the asset market clears at all times.

Definition 1.2. *Given preferences, endowment structures and initial money allocations, a monetary competitive equilibrium in the three-period overlapping generations models populated by*

[3] These initial generations' problems also apply to naive consumers.

naive consumers is a sequence of actual consumption bundles, money holdings and gross rates of returns $\{c^n_{3,0}, c^n_{2,0}, c^n_{3,1}, a^n_{2,0}, c^n_{1,t}, c^n_{2,t+1}, c^n_{3,t+2}, a^n_{1,t}, a^n_{2,t+1}, R_t\}_{t=0}^{\infty}$ *such that*

1. Given the rates of return, the young naive consumers choose actual consumptions and money holdings to satisfy the lifetime budget constraint (1.4) and two optimality conditions (1.14) and (1.14) plus $\left(c^n_{1,t}, c^n_{2,t+1}, c^n_{3,t+2}\right) \geq 0$ for $\forall t$.

2. Given $\alpha_{1,-1}$ and R_0, $c^n_{2,0}, c^n_{3,1}$ and $a^n_{2,0}$ are the optimal choices for the problem of the initial middle-aged. Given $\alpha_{2,-1}$, $c^n_{3,0}$ is the optimal choice for the problem of the initial old.

3. For $\forall t$, $A_t = \bar{m}/p_t$, i.e. the asset market clears at all times.

For both types of consumers, the objective function is strictly concave and the lifetime budget constraint is a convex and compact set given interest rates. The Lagrange theorem can be applied to the convex problem since the nonnegativity constraints on consumptions do not bind due to the Inada condition. Thus, the first order conditions are necessary and sufficient conditions for the unique interior optimum. By applying the implicit function theorem to the optimality conditions for actual consumptions, we can derive the following relationships:

$$a_{1,t} = a_{1,t}\left[R_{t+1}, R_t\right], \; a_{2,t+1} = a_{2,t+1}\left[R_{t+1}, R_t\right] \tag{1.15}$$

Note that we use notations for the sophisticated consumers but the entire analysis hereafter is also applicable to the naive consumers as well. With (1.15), we can express the asset market clearing condition as:

$$\begin{gathered} A_{t+1} = R_t A_t \\ \Longleftrightarrow a_{1,t+1}\left[R_{t+2}, R_{t+1}\right] + a_{2,t+1}\left[R_{t+1}, R_t\right] = R_t\left(a_{1,t}\left[R_{t+1}, R_t\right] + a_{2,t}\left[R_t, R_{t-1}\right]\right) \end{gathered} \tag{1.16}$$

By rearranging the second equation in (1.16), we derive a nonlinear function:

$$G\left[R_{t+2}, R_{t+1}, R_t, R_{t-1}\right] = 0 \tag{1.17}$$

The implicit function theorem is applicable to (1.17) for R_{t+2} to describe the interest rate as a function of other rates. We summarize this finding in the following proposition.

Proposition 1.1. *Assuming* $\frac{\partial c_{1,t+1}[R_{t+2}, R_{t+1}]}{\partial R_{t+2}} \neq 0$ *in the models inhabited by either sophisticated or naive consumers, the perfect foresight dynamics of gross return rates can be expressed by a third-order nonlinear difference equation given by:*

$$R_{t+2} = F\left[R_{t+1}, R_t, R_{t-1}\right] \tag{1.18}$$

Proof. See Appendix 1.A.3. □

In our models with outside money, there exists a unique stationary equilibrium where actual consumptions, real money holdings, and return rates stay constant at all times. The unique steady-state gross interest rate is 1.

Proposition 1.2. *There exists a unique steady-state gross rate of return $R^* = 1$ where the equilibrium is a monetary steady state, i.e. when the steady-state aggregate saving, A^*, is positive.*

Proof. See Appendix 1.A.4. □

According to the efficiency criterion in Balasko and Shell (1980), the monetary steady-state with $R^* = 1$ is Pareto optimal.

1.2.4 Closed forms of optimal consumptions

We introduce a utility specification to compare the aggregate savings patterns of the sophisticated and the naive type explicitly.

Assumption 1.1. *We assume the instantaneous utility function u is an iso-elastic function with constant relative risk aversion (CRRA):*

$$u(c) = \frac{c^{1-\gamma} - 1}{1-\gamma}$$

where $\gamma > 0$ is the risk aversion parameter. When $\gamma = 1$,

$$u(c) = \mathit{ln}(c)$$

Now, we can derive the closed-form expression for the optimal consumptions of the sophisticated under the iso-elastic utility function with a constant risk aversion as below.

Proposition 1.3. *Under Assumption 1.1, the following closed form expression describes the optimal consumptions of sophisticated consumers born in time t:*

$$\begin{aligned} c_{1,t} &= \frac{\omega_1 + \omega_2/R_t + \omega_3/(R_t R_{t+1})}{1 + \eta_t^{\frac{1}{\gamma}} R_t^{\frac{1}{\gamma}-1} + (\beta\delta\eta_t)^{\frac{1}{\gamma}} (R_t R_{t+1})^{\frac{1}{\gamma}-1}} \\ c_{2,t+1} &= \eta_t^{\frac{1}{\gamma}} R_t^{\frac{1}{\gamma}} c_{1,t} \\ c_{3,t+2} &= (\beta\delta)^{\frac{1}{\gamma}} R_{t+1}^{\frac{1}{\gamma}} c_{2,t+1} \end{aligned} \tag{1.19}$$

where $\eta_t = \frac{\delta\left\{\beta + (\beta\delta)^{\frac{1}{\gamma}} R_{t+1}^{\frac{1}{\gamma}-1}\right\}}{1+(\beta\delta)^{\frac{1}{\gamma}} R_{t+1}^{\frac{1}{\gamma}-1}}$.

Proof. See Appendix 1.A.5. □

Next, we obtain the closed-form expression for the planned and actual consumptions of the naive under the same utility configuration.

Proposition 1.4. *Under Assumption 1.1, the following closed form expression describes the planned and actual consumptions of naive consumers born in time t respectively:*

** Planned consumptions:*

$$\begin{aligned} c^n_{1,t} &= \frac{\omega_1 + \omega_2/R_t + \omega_3/\left(R_t R_{t+1}\right)}{1 + \left(\beta\delta\right)^{\frac{1}{\gamma}} R_t^{\frac{1}{\gamma}-1} + \left(\beta\delta^2\right)^{\frac{1}{\gamma}} \left(R_t R_{t+1}\right)^{\frac{1}{\gamma}-1}} \\ c^e_{2,t+1} &= \left(\beta\delta\right)^{\frac{1}{\gamma}} R_t^{\frac{1}{\gamma}} c^n_{1,t} \\ c^e_{3,t+2} &= \delta^{\frac{1}{\gamma}} R_{t+1}^{\frac{1}{\gamma}} c^e_{2,t+1} \end{aligned} \tag{1.20}$$

** Actual consumptions:*

$$\begin{aligned} c^n_{1,t} &= \frac{\omega_1 + \omega_2/R_t + \omega_3/\left(R_t R_{t+1}\right)}{1 + \left(\beta\delta\right)^{\frac{1}{\gamma}} R_t^{\frac{1}{\gamma}-1} + \left(\beta\delta^2\right)^{\frac{1}{\gamma}} \left(R_t R_{t+1}\right)^{\frac{1}{\gamma}-1}} \\ c^n_{2,t+1} &= \left(\beta\delta\right)^{\frac{1}{\gamma}} R_t^{\frac{1}{\gamma}} \frac{\left(1 + \delta^{\frac{1}{\gamma}} R_{t+1}^{\frac{1}{\gamma}-1}\right)}{1 + \left(\beta\delta\right)^{\frac{1}{\gamma}} R_{t+1}^{\frac{1}{\gamma}-1}} c^n_{1,t} \\ c^n_{3,t+2} &= \left(\beta\delta\right)^{\frac{1}{\gamma}} R_{t+1}^{\frac{1}{\gamma}} c^n_{2,t+1} \end{aligned} \tag{1.21}$$

Proof. See Appendix 1.A.6. □

From Propositions (1.3) and (1.4), we find that the actual choices of the two types are equivalent under the logarithmic preference, as is stated in Corollary 1.1.

Corollary 1.1. *When $\gamma = 1$, i.e.$u\left(c\right) = ln\left(c\right)$, the actual consumptions of naive consumers are equivalent to the optimal consumptions of sophisticated consumers born in the same period:*

$$\begin{aligned} c_{1,t} = c^n_{1,t} &= \frac{\omega_1 + \omega_2/R_t + \omega_3/\left(R_t R_{t+1}\right)}{1 + \beta\delta + \beta\delta^2} \\ c_{2,t+1} = c^n_{2,t+1} &= \frac{\beta\delta\left(1+\delta\right)}{1+\beta\delta} R_t c_{1,t} \\ c_{3,t+2} = c^n_{3,t+2} &= \beta\delta R_{t+1} c_{2,t+1} \end{aligned} \tag{1.22}$$

Proof. See Appendix 1.A.7. □

This corollary implies that there are identical aggregate consumptions/savings between sophisticated and naive consumers under a low-risk aversion. In the next subsection, we numerically show that aggregate savings are indistinguishable even with a larger risk-aversion when discount factor and present-bias set at a calibration value widely used in the literature. Note that aggregate savings mainly determine both EBC and local SE. Thus, we focus on the sophisticated consumer case through the remainder of the paper since there are similar results for naive consumers.

1.2.5 Aggregate saving behaviors

We provide a numerical comparison of the optimal choices between the two types of consumers in the steady-state with $R = 1$ using the closed-form expressions above. For this, we focus on the young-age consumption since there are similar patterns in consumption in other ages and savings. We normalize the sum of endowments to be 1. Note that when $R = 1$, the endowment profile does not matter for optimal consumption and saving.

The left panel in Figure 1.1 shows how the steady-state consumptions move with the hyperbolic discounting parameter fixing $\delta = 0.4$ and $\gamma = 4$.[4] The solid line represents the sophisticated whereas the dashed line corresponds to the naive. The consumptions of the two types coincide at two extreme points $\beta = 0$ and $\beta = 1$. When $\beta = 0$, the preference reduces to extreme myopia, and thus the two types consume all their endowment in the young period. When $\beta = 1$, the preference reduces to the standard exponential discounting in which the two types are identical. Anywhere in $0 < \beta < 1$, the naive consumers consume more than the sophisticated when young under $\gamma > 1$ because the sophisticated consumers, unlike the naive consumers, anticipate that their future selves will consume more than their current plan in the next period, which makes them increase saving to indirectly maintain adequate consumption level when old from a low saving when middle-aged. We can see that c_1^* is decreasing in β for both consumer types which implies the young's saving is reduced under the QHD relative to the regular exponential discounting. This pattern also holds for the middle-aged. Therefore, the QHD drops the aggregate saving.

The right panel in Figure 1.1 illustrates how the steady-state consumptions in the young period change with the risk-aversion parameter fixing $\delta = 0.4$ and $\beta = 0.6$.[5] Given these parameter values, the difference in consumption of the sophisticated and naive is negligible for the range of risk aversion of interest in the analysis below. These results justify we con-

[4] To compute δ, we use a calibration value for one-year time discount factor and consider one-period in our model corresponding to 20 years.

[5] We take $\beta = 0.6$ following Laibson (1997).

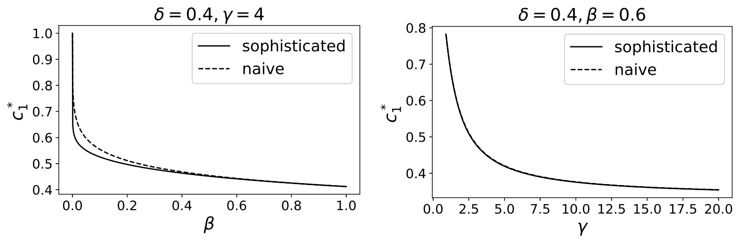


Figure 1.1: Young consumption at monetary steady state

centrate on the sophisticated consumer case in the analysis below although we provide the result for the naive consumer case in the appendix.

As seen in Proposition 1.2, the perfect foresight stationary equilibria do not guarantee positive aggregate savings under which the price of money is positive and a true monetary equilibrium exists. When the endowment structure is decreasing in age, both the young and the middle-aged must save in equilibrium to transfer wealth for consumption smoothing. However, young households borrow from the middle-aged to increase the current consumption under a hump-shaped endowment profile where the endowment is concentrated on the middle-age period. Thus, aggregate saving can be negative. In order to have positive aggregate savings, we introduce the following assumptions on $\boldsymbol{\omega}$.

Assumption 1.2. *If the endowment vector* $\boldsymbol{\omega}$ *satisfies* $\omega_3 = 0$ *and* $(2-\rho_i)\,\omega_1 + (1-\rho_i)\,\omega_2 > 0$*, the steady state with* $R^* = 1$ *has stationary aggregate savings equal to* $A^* > 0$*, i.e., it qualifies as a monetary steady-state. More specifically, when consumers are sophisticated,*

$$\rho_s = \frac{2+\eta^{\frac{1}{\gamma}}}{1+\eta^{\frac{1}{\gamma}}+(\beta\delta\eta)^{\frac{1}{\gamma}}}, \tag{1.23}$$

where $\eta = \frac{\delta\left[\beta+(\beta\delta)^{\frac{1}{\gamma}}\right]}{1+(\beta\delta)^{\frac{1}{\gamma}}}$.

When consumers are naive,

$$\rho_n = \left(2+\frac{(\beta\delta)^{\frac{1}{\gamma}}\left(1+\delta^{\frac{1}{\gamma}}\right)}{1+(\beta\delta)^{\frac{1}{\gamma}}}\right)\frac{1}{1+(\beta\delta)^{\frac{1}{\gamma}}+(\beta\delta^2)^{\frac{1}{\gamma}}}. \tag{1.24}$$

Note that when $\gamma = 1$*,* $\rho_s = \rho_n$*, which is consistent with previous results.*

If Assumption 1.2 is satisfied, the aggregate savings are positive and the stationary interest rate is at the golden rule level, $R^* = 1$, we call this case the Samuelson case in Gale (1973). Under QHD, we need high ω_1 and δ to fulfill the conditions for the positive money price since patient young households with a larger endowment will save enough to offset the reduced saving from the hyperbolic preferences. When ω_1 is not high enough, decreasing β is likely to violate the conditions above, and thus the aggregate saving will be negative. This implies introducing the QHD can turn the Samuelson case into the Classical case. Therefore, government currency plays no role in transferring wealth to the next period.

1.3 Two-period Cycle

In this section, we study the effect of the QHD preferences on the existence of a two-period endogenous business cycle in a three-period OLG model. We restrict our attention to the two-period cycle following the justification of Bhattacharya and Russell (2003) that the length of cycles should be shorter than the lives of households in models to be consistent with observed business cycles.

The two-period cycle implies the equilibrium allocations alternate over two states. For example, $R_t = R_{t+2}$ and $A_t = A_{t+2}$. From the equilibrium conditions, we have:

$$A_{t+1} = R_t A_t,\ A_{t+2} = R_{t+1} A_{t+1} \tag{1.25}$$

Combining the two equations above with $A_t = A_{t+2}$ yields:

$$R_t R_{t+1} = 1 \tag{1.26}$$

We define two interest rates $R_H > 1$ and $R_L < 1$ where $R_H R_L = 1$. Without loss of generality, denote $R > 1$ as R_H, and $R_L = \frac{1}{R}$. Denote $\{c_{1H}, c_{2H}, c_{3H}\}$ as the consumption plan of the generations born in high interest rate periods, and $\{c_{1L}, c_{2L}, c_{3L}\}$ as the consumption plan of generations born in low interest rate periods. By replacing interest rates in (1.19) with $\left\{R, \frac{1}{R}\right\}$ and $\left\{\frac{1}{R}, R\right\}$, we can find closed form expressions for the optimal consumptions of sophisticated consumers born in high and low states respectively:

$$\begin{aligned}
c_{1H}(R) &= \frac{\omega_1 + \frac{\omega_2}{R} + \omega_3}{1 + \eta\left(\frac{1}{R}\right)^{\frac{1}{\gamma}} R^{\frac{1}{\gamma}-1} + \left(\beta\delta\eta\left(\frac{1}{R}\right)\right)^{\frac{1}{\gamma}}},\ c_{1L}(R) = c_{1H}(1/R) \\
c_{2H}(R) &= \left(\eta\left(\frac{1}{R}\right)R\right)^{\frac{1}{\gamma}} c_{1H}(R),\ c_{2L}(R) = c_{2H}(1/R) \\
c_{3H}(R) &= \left(\beta\delta\eta\left(\frac{1}{R}\right)\right)^{\frac{1}{\gamma}} c_{1H}(R),\ c_{3L}(R) = c_{3H}(1/R)
\end{aligned} \tag{1.27}$$

where $\eta(R) = \frac{\delta\left[\beta + (\beta\delta)^{\frac{1}{\gamma}} R^{\frac{1}{\gamma}-1}\right]}{1 + (\beta\delta)^{\frac{1}{\gamma}} R^{\frac{1}{\gamma}-1}}$.

Likewise, we can derive closed-form expressions for both planned and actual consumptions of naive consumers born in high and low states respectively (See Appendix 1.A.8). Henceforth, we proceed with an analysis of cycles in the economy with only sophisticated consumers in the manuscript. It is straightforward to derive a similar result for naive consumers by substituting their actual consumptions for the sophisticated's ones.

The two-period cycle exists if and only if the market clearing condition $A(1/R) = RA(R)$ has a solution $R > 1$ at which $A(R) > 0$. We reexpress $A(R)$ and $A(1/R)$ with consumptions as:

$$A(R) = a_{1H}(R) + a_{2L}(R) = (\omega_1 - c_{1H}(R)) + \left(\frac{1}{R}(\omega_1 - c_{1L}(R)) + \omega_2 - c_{2L}(R)\right) \quad (1.28)$$

and

$$A(1/R) = a_{1L}(R) + a_{2H}(R) = (\omega_1 - c_{1L}(R)) + (R(\omega_1 - c_{1H}(R)) + \omega_2 - c_{2H}(R)) \quad (1.29)$$

Then, by plugging these expressions into the market clearing condition and rearranging it, we can derive a new market clearing condition as

$$(R-1)\omega_2 = Rc_{2L}(R) - c_{2H}(R) \quad (1.30)$$

Note that (1.27) implies the consumption plan depends only on the sum of ω_1 and ω_3, not each individual term. In the following lemma, we identify this property of consumption plans.

Lemma 1.1. *If R^* is a solution to (1.30) for an endowment profile with $\{\bar{\omega}_1, \bar{\omega}_2, \bar{\omega}_3\}$, then it is also a solution to (1.30) for any alternative profiles with $\{\hat{\omega}_1, \hat{\omega}_2, \hat{\omega}_3\}$ in which $\hat{\omega}_1 + \hat{\omega}_3 = \bar{\omega}_1 + \bar{\omega}_3$.*

Proof. Suppose we consider a class of economies that are identical to each other except for the values of ω_1 and ω_3, which are required to sum to a constant value: $\omega_1 + \omega_3 = w_{13} > 0$. It follows from the equation (1.27) that

$$c_{1H} = \frac{w_{13} + \frac{\omega_2}{R}}{1 + \eta\left(\frac{1}{R}\right)^{\frac{1}{\gamma}} R^{\frac{1}{\gamma}-1} + \left(\beta\delta\eta\left(\frac{1}{R}\right)\right)^{\frac{1}{\gamma}}} \quad (1.31)$$

The value of c_{1H} depends on only the sum of ω_1 and ω_3, not on their individual values. Since c_{1L}, c_{2H}, and c_{2L} are expressed with c_{1H} and other variables independent of ω_1 or ω_3, it is clear that the equation (1.30) also depends on only w_{13}. Hence, a solution $R^* > 1$ to the equation (1.30) should be the same in the class of economies with $\omega_1 + \omega_3 = w_{13}$ if it exists. □

Without loss of generality, we assume $\omega_3 = 0$ hereafter from Lemma 1.1. The following lemma summarizes a sufficient condition for the existence of a solution to equation (1.30).

Proposition 1.5. *Under Assumption 1.2, a sufficient condition for the existence of solution $R^* > 1$ to equation (1.30) is given by:*

$$\frac{A'(1)}{A(1)} < -\frac{1}{2} \tag{1.32}$$

where $A(1)$ is the aggregate saving as defined in equation (1.28) evaluated at $R = 1$, and $A'(1)$ denotes its derivative which is $A'(R) = a'_{1H}(R) + a'_{2L}(R)$ evaluated at $R = 1$.

Proof. See Appendix 1.A.9. □

When $\omega_1 = 1$, the aggregate saving is positive for all $R > 1$ as long as β and δ are positive since both the young and middle-aged households save. Then, the condition driven from Proposition 1.5 is also a sufficient condition for the existence of a two-period cycle in this model. This is the main content of the following corollary.

Corollary 1.2. *When $\omega_1 = 1$ and $\omega_2 = \omega_3 = 0$, a sufficient condition for the existence of a two-period cycle in a three-period OLG model with QHD preferences is*

$$\frac{A'(1)}{A(1)} < -\frac{1}{2} \tag{1.33}$$

Proof. The proof is straightforward following the arguement above. □

The condition (1.33) is analogous to the necessary and sufficient condition of cycles in two-period models with the same CRRA preferences under exponential discounting: the intertemporal elasticity of substitution, $\frac{a'(1)}{a(1)}$, is lower than $-\frac{1}{2}$,where $a(1)$ is the young's saving. The necessary and sufficient condition requires a strong income effect and a weak substitution effect when real interest rates rise. Indeed, the sufficient condition for cycles in a three-period model also permits similar intuition. The strong negativity of $A'(1)$ is required in (1.33), which needs both $a'_{1H}(1)$ and $a'_{2L}(1)$ to be strongly negative as well since $A'(1) = a'_{1H}(1) + a'_{2L}(1)$. The negativity of $a'_{1H}(1)$ and $a'_{2L}(1)$ implies that the young born in high interest rate periods and the middle-aged born in low interest rate periods decrease saving when real interest rates increase. A rise in real interest rates generates both income and substitution effects on current consumption although their effects are opposite: the income effect increases current consumption, but the substitution effect decreases current consumption. Thus, if the income effect dominates the substitution effect, an increase in the interest rate leads to a decrease in aggregate saving.

. In detail, from the lifetime budget constraint for the generation born in high interest rate periods, $R(c_{1H} + c_{3H}) + c_{2H} = R\omega_1 + \omega_2$. An increase in R creates an income effect due to a net wealth change by $\Delta R\left(\omega_1 - c_{1H} - c_{3H}\right)$ and a substitution effect because of a relative price change bewteen $\{c_{1H}, c_{3H}\}$ and c_{2H}. If the net endowment, $\left(\omega_1 - c_{1H} - c_{3H}\right)$, is positive, the income effect increases all lifetime consumptions whereas the substitution effect raises c_{2H} but drops c_{1H} and c_{3H}. Note that $a'_{1H}\left(1\right) = -c'_{1H}\left(1\right)$, $a'_{2H}\left(1\right) = c'_{3H}\left(1\right) + c_{3H}\left(1\right)$ and $a'_{2H}\left(1\right) = -a'_{2L}\left(1\right)$.[6] Thus, if the net endowment, $\left(\omega_1 - c_{1H} - c_{3H}\right)$, is positive and large enough, then an income effect dominating a substitution effect will decrease both $a'_{1H}\left(1\right)$ and $a'_{2L}\left(1\right)$ so that $A'\left(1\right)$ can be negative.

For the income effect to dominate the substitution effect, we need a high ω_1 to make the net endowment $\left(\omega_1 - c_{1H} - c_{3H}\right)$ increase and generate a stronger income effect while the substitution effect is not affected. A high risk-aversion is also required because it weakens the substitution effect by making goods over periods less substitutable because the indifference curve becomes closer to the one under a Leontief utility function. Under a high risk-aversion, the income effect is strengthened. A stronger consumption smoothing need decreases c_{1H} and increases c_{3H} closer to an equal consumption profile over periods. Since $c_{1H} > c_{3H}$ when discounting future consumption, c_{1H} declines faster than the rise of c_{3H} from an increase in the risk-aversion parameter value. Thus, the net endowment is larger for a higher γ, which leads to a stronger income effect. However, there is not a clear prediction about whether reducing the discounting factor or introducing a present bias helps the income effect to dominate the substitution effect. Both a decrease in the discounting factor and having a present bias make goods over periods less substitutable, but they also reduce the income effect because c_{1H} rises faster than the rate that c_{3H} falls.

In addition, a strongly negative $A'\left(1\right)$ is not enough to generate cycles because if $A\left(1\right)$ is largely positive, then the sufficient condition can be violated. Thus, $A\left(1\right)$ should be a small postive number for the sufficient condition to be satisfied. One interesting fact is that the absolute values of $A'\left(1\right)$ and $A\left(1\right)$ can be positively correlated. In other word, if $A\left(1\right)$ further diverges from zero to the right , then $A'\left(1\right)$ can also diverge further from zero to the left and vice versa. The intuition is that a large income effect arises when the net endowment at $R = 1$, $\left(\omega_1 - c_1\left(1\right) - c_3\left(1\right)\right)$, is largely positive. The net endowment can be written as $a_1\left(1\right) - a_2\left(1\right)$. Thus, the net endowment can be positively corrleated with $A\left(1\right)$ through $a_1\left(1\right)$. Therefore, a large income effects leads to both a strongly negative $A'\left(1\right)$ and a strongly positive $A\left(1\right)$.

[6] From the budget constraints, $a_{1H} = \omega_1 - c_{1H}$ and $c_{3H} = \frac{1}{R}a_{2H}$. Taking their derivatives with respect to R and then equating $R = 1$ yields that $a'_{1H}\left(1\right) = -c'_{1H}\left(1\right)$ and $a'_{2H}\left(1\right) = c'_{3H}\left(1\right) + a_{2H}\left(1\right) = c'_{3H}\left(1\right) + c_{3H}\left(1\right)$. Since $c_{1H}\left(R\right) = c_{1L}\left(\frac{1}{R}\right)$, $c'_{1H}\left(1\right) = -c'_{1L}\left(1\right)$ and $a'_{1H}\left(1\right) = -a'_{1L}\left(1\right)$. Likewise, $c'_{2H}\left(1\right) = -c'_{2L}\left(1\right)$. From $a_{2H} = \omega_2 - c_{2H} + Ra_{1H}$ and $a_{2L} = \omega_2 - c_{2L} + \frac{1}{R}a_{1L}$, $a'_{2H}\left(1\right) = -c'_{2H}\left(1\right) + a_{1H}\left(1\right) + a'_{1H}\left(1\right)$ and $a'_{2L}\left(1\right) = -c'_{2L}\left(1\right) - a_{1L}\left(1\right) + a'_{1L}\left(1\right)$. Therefore, $a'_{2H}\left(1\right) = -a'_{2L}\left(1\right)$ because $a_{1H}\left(1\right) = a_{1L}\left(1\right)$ and $a'_{1H}\left(1\right) = -a'_{1L}\left(1\right)$.

Due to the ex-ante unclear prediction and the possible co-movement of $A'(1)$ and $A(1)$ in absolute value, examining the effect of changes in parameters to the sufficient condition requires a quantitative analysis. Figure 1.2 represents the set of two-period endogenous business cycles in (δ, β) Caretesian plane for two pairs of (γ, ω_1). Thus, the parameter space of interest is $(\delta, \beta) \in [0, 1] \times [0, 1]$.

We numerically find out $R^* > 1$ satisfying (1.30) and $A(R^*) > 0$ for every pair of (δ, β) in the grid. The dark grey area in Figure 1.2 represents the set of parameter values in which there exists $R^* > 1$ satisfying the two conditions for EBC.

There are several interesting things to note from the numerical analysis. First, although it is very hard to analytically check the sufficient condition (1.32) is also necessary under QHD preferences, the numerical analysis supports that (1.32) is both necessary and sufficient for the existence of solution $R^* > 1$ satisfying (1.30). We check the set characterized by (1.32) is identical with the set of (δ, β) where $R^* > 1$ satisfying (1.30) exists from the numerical analysis for a myriad of pairs of (γ, ω_1). From the numerical analysis, we also check that the aggregate saving is positive for $R^* > 1$ satisfying (1.30). Therefore, the sufficient condition well characterizes the set of two-period endogenous business cycles as identical to the set from the numerical analysis.

The numerical analysis confirms that the effects of changes in endowment and risk-aversion parameters are consistent with intuition: cycles are more likely to emerge as the young's endowment share and risk-aversion increase. This result accords with two-period OLG models (See Lahiri and Puhakka (1998)). The effect of a time discount factor is mixed. For a relatively low risk-aversion – say $\gamma = 4$, a lower time discount factor allows cycles to exist as seen in the right first panel. On the other hand, a higher time discount factor is required for cycles to appear for a relatively large risk-aversion such as $\gamma = 12$. In detail, a drop in δ makes commodities over periods less substitutable and decreases aggregate saving. Such changes are more significant under a low γ than under a high γ. When consumers are very risk-averse, a fall in δ does not reduce the aggregate saving much because of a strong consumption smoothing effect. Since the indifference curve is already closer to the one from a Leontief utility function, a drop in δ does not weaken the substitution effect significantly. Thus, there are opposite effects of changes in the time discount factor to the existence of cycles depending on the risk aversion.

Unlike the time discount factor, a higher value for the QHD parameter is essential for EBC in both high and low risk-aversions. This result implies that introducing QHD preferences shrinks the set of EBC in the parameter space. Compared to δ, a fall in β decreases the aggregate saving less because the young agents reflect a change in δ for saving decisions both this

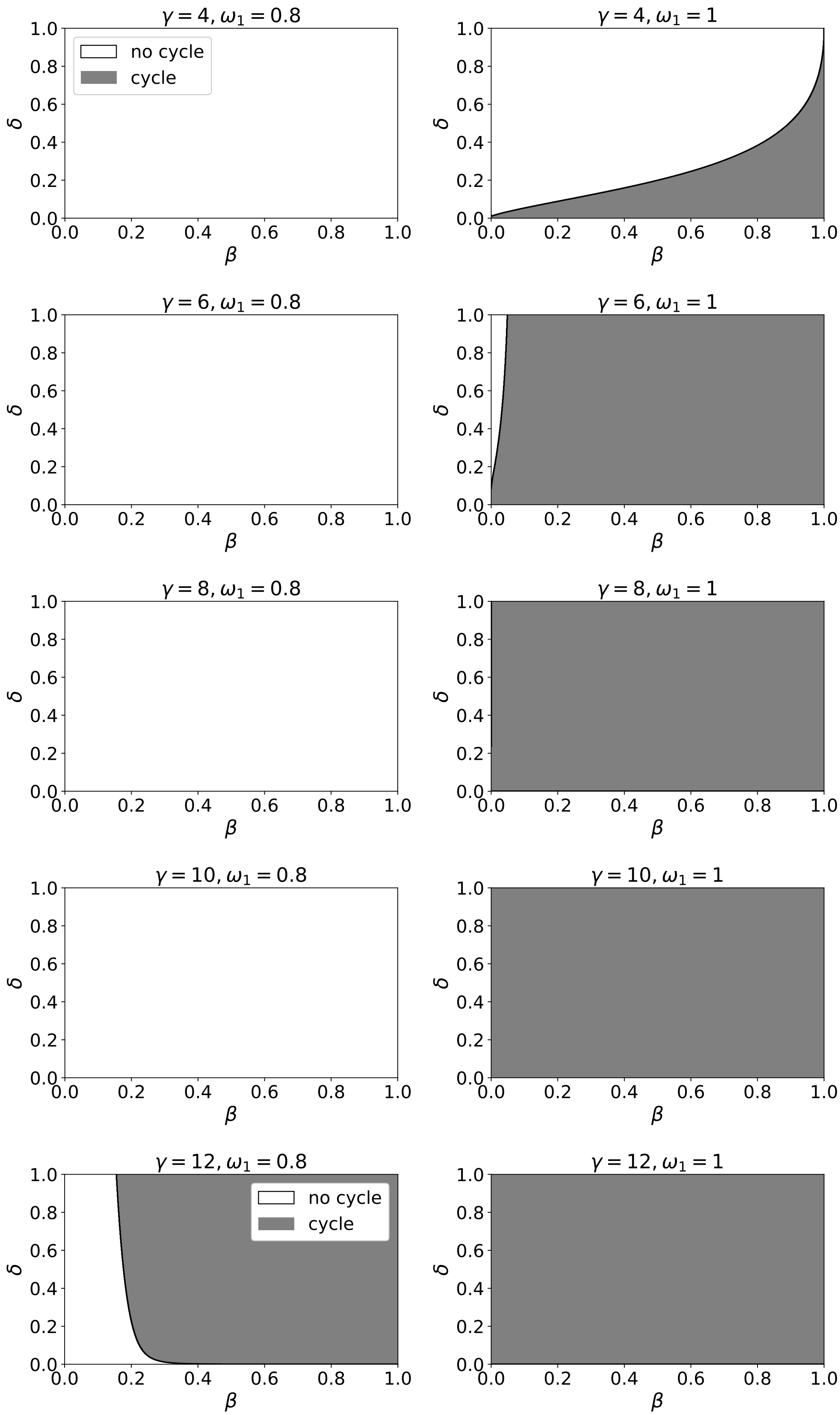


Figure 1.2: Endogenous business cycles for different parameter configurations, when consumers are sophisticated.

and next periods. However, they partially consider a variation in β when saving in the next period as sophisticated types or ignore it as naive ones due to the wrong belief of no present bias in the next period. Then, $\frac{A'(1)}{A(1)}$ can rather increase when β diminishes unlike the case that δ decreases. However, as we show below, QHD preferences contribute to the existence of indeterminate equilibria where sunspot equilibrium can be constructed.

Now, we compare our model with the two-period model in terms of the parameter values for cycles to emerge. For this comparison, we consider the case with the endowment profile fully concentrated on the young, i.e. $\omega_1 = 1$ and $\gamma = 4$. The two-period model requires the time discount factor to be less than 0.0625 for a cycle.[7] However, in our model with $\beta = 0.6$, a cycle emerges if the time discount factor is less than 0.247.[8] One year time discount factor is $0.247^{1/20} \approx 0.93$ in our model assuming one-period is 20 years. Likewise, one year time discount factor is $0.0625^{1/30} \approx 0.91$ in two-period model assuming one-period is 30 years. Moreover, as β increases, the upper limit for the time discount factor increases. This result implies that the set of cycles in our model is at least comparable with the one in a two-period model with standard preferences, although introducing a present bias reduces the set of EBC, compared to the exponential discounting case.[9]

1.4 Locally Indeterminate Equilibria

This section examines the implication of the QHD preferences to the existence of a local sunspot equilibrium characterized by a locally indeterminate equilibrium in a three-period OLG model. Following Woodford (1986), a monetary steady state has a local sunspot equilibrium if one can find a stationary sunspot equilibrium for any small neighborhoods of a monetary steady state.[10] An equivalent definition provided by Dávila, Gottardi, and Kajii (2007) is that a local sunspot equilibrium exists if there is a converging sequence of a stationary sunspot equilibrium to the monetary steady state. One well-known result in the literature is that a locally indeterminate monetary steady state has a local sunspot equilibrium (See Woodford (1986) and Peck (1988)). Thus, we characterize a condition for the existence of the local indeterminacy and its set in the parameter space.

[7] The necessary and sufficient condition for a two-period cycle is $\frac{\omega_2}{\omega_1} < \frac{\left(1-\frac{2}{\gamma}\right)-\beta^{\frac{1}{\gamma}}}{\beta^{\frac{-1}{\gamma}}-\left(1-\frac{2}{\gamma}\right)}$ under CRRA preferences.

[8] Here, we consider $\beta = 0.6$ following Laibson (1997).

[9] With $\beta = 1$, a cycle exists for δ between 0 and 1.

[10] A stationary sunspot equilibrium is a stationary rational expectations equilibrium with a Markov transition with at least one non-absorbing state.

Blanchard and Kahn (1980) provide conditions for local determinacy and indeterminacy and stability of dynamic systems in general, which regards the dimension of convergent sequences starting in a neighborhood of the unique monetary steady state. Orrego (2014) explicitly expresses the Blanchard and Kahn conditions in a three-period OLG model where there exists one predetermined variable in time t, R_{t-1}. We follow Orrego (2014) to examine the equilibrium manifolds around the stationary point in our model.

To analyze the local dynamics around the monetary steady state with $R^* = 1$, we convert the third-order nonlinear difference equation, (1.18), into a first-order vector difference system:

$$\begin{pmatrix} R_{t+2} \\ R_{t+1} \\ R_t \end{pmatrix} = \begin{pmatrix} F\left[R_{t+1}, R_t, R_{t-1}\right] \\ R_{t+1} \\ R_t \end{pmatrix}. \tag{1.34}$$

At $R^* = 1$, the Jacobian matrix of this system can be written as

$$J\left[1\right] = \begin{bmatrix} \partial F\left[1,1,1\right]/\partial R_{t+1} & \partial F\left[1,1,1\right]/\partial R_t & \partial F\left[1,1,1\right]/\partial R_{t-1} \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}. \tag{1.35}$$

The derivation of $J\left[1\right]$ for sophisticated consumers and naive consumers can be found in Appendix 1.A.10 and 1.A.11 respectively. The position of the three eigenvalues of the above Jacobian matrix determines the dimension of the stable equilibrium manifolds around the unique monetary steady state. This is the content of the following proposition.

Proposition 1.6. *(Blanchard-Kahn conditions) Let $J\left[1\right]$ be invertible and $R^* = 1$ be a hyperbolic steady state. If the number of eigenvalues of $J\left[1\right]$ whose moduli lie inside the unit circle is:*

1. *more than one, then the steady state is **locally stable** and the equilibrium is **indeterminate** (an infinity of competitive equilibria).*

2. *exactly one, then the steady state is **locally saddle-path stable** and the equilibrium is **determinate** (a unique competitive equilibrium).*

3. *zero, then there is **no nonexplosive solution** satisfying the return dynamics in (1.18).*

Note that the invertible Jacobian matrix assumption is generic since one can perturb the second-order derivative of utility functions to have a non-singular matrix. As is shown below, the Lebesgue measure of the non-hyperbolic steady-state set is zero. Thus, it justifies our restriction to the hyperbolic steady state. Under these assumptions, the Jacobian matrix cannot have both 0 and 1 as its eigenvalues. The Hartman-Grobman theorem (local stable manifold theorem) in dynamical system theory states that there is an open neighborhood $\Re$ of $R^* = 1$ where a linearized system with the Jacobian matrix J is topologically equivalent to the original nonlinear system (1.34) if $R^* = 1$ is a hyperbolic steady-state of (1.34) and the Jacobian matrix J is invertible. This theorem justifies us to study equilibrium manifolds around a steady-state in (1.34) with their counterparts in an affine approximation represented by J.

In the indeterminate equilibrium case, there is an infinite number of solutions for the sequence of return rates. This implies that given an initial wealth distribution, there exists a continuum of convergent sequences consistent with the equilibrium conditions. The dimension of indeterminate equilibrium manifolds is determined by the number of eigenvalues inside the unit circle net of the number of predetermined variables. To pin down an equilibrium sequence, the solution should depend on the past interest rates before time 0 directly instead of the predetermined wealth condition. Or the solution requires any coordination devices such as sunspot to select an equilibrium path. Thus, sunspot equilibria can arise when the steady-state is locally indeterminate. If three eigenvalues are inside the unit circle, any equilibrium sequences generated by the model dynamics should converge to the steady state since all eigenvector manifolds are stable and they can span any three-dimensional vectors as long as they are linearly independent.

When the equilibrium is determinate, there exists a unique convergent sequence of rates of returns corresponding to an initial condition for the wealth distribution in period 0. In the explosive case, one cannot find a sequence of return rates converging to the steady state unless the economy starts at the steady-state.

Now, we numerically characterize the set of economies based on the relationship between its monetary steady-states and the Blanchard-Kahn eigenvalue conditions under CRRA preference and some normalizations. We first examine an economy entirely populated by sophisticated consumers. The results for the naive case are quite similar to the sophisticated case, and thus we leave the analysis for naive consumers in Appendix 1.C. When the two types of households coexist, the results will be in between the two cases. We expect the results in the mixing case should not be much different because the results in both cases resemble each other.

Figure 1.3 depicts four types of equilibria over parameter spaces. The figure examines the stability properties of hyperbolic equilibria in (β, δ) Caretesian plane for six pairs of (γ, ω_1). We make a normalization that $\omega_1 + \omega_2 = 1$ and $\omega_3 = 0$ as we did in the endogeneous business cycle analysis and thus the parameter space of interest is $(\beta, \delta) \in [0, 1] \times [0, 1]$. We calculated

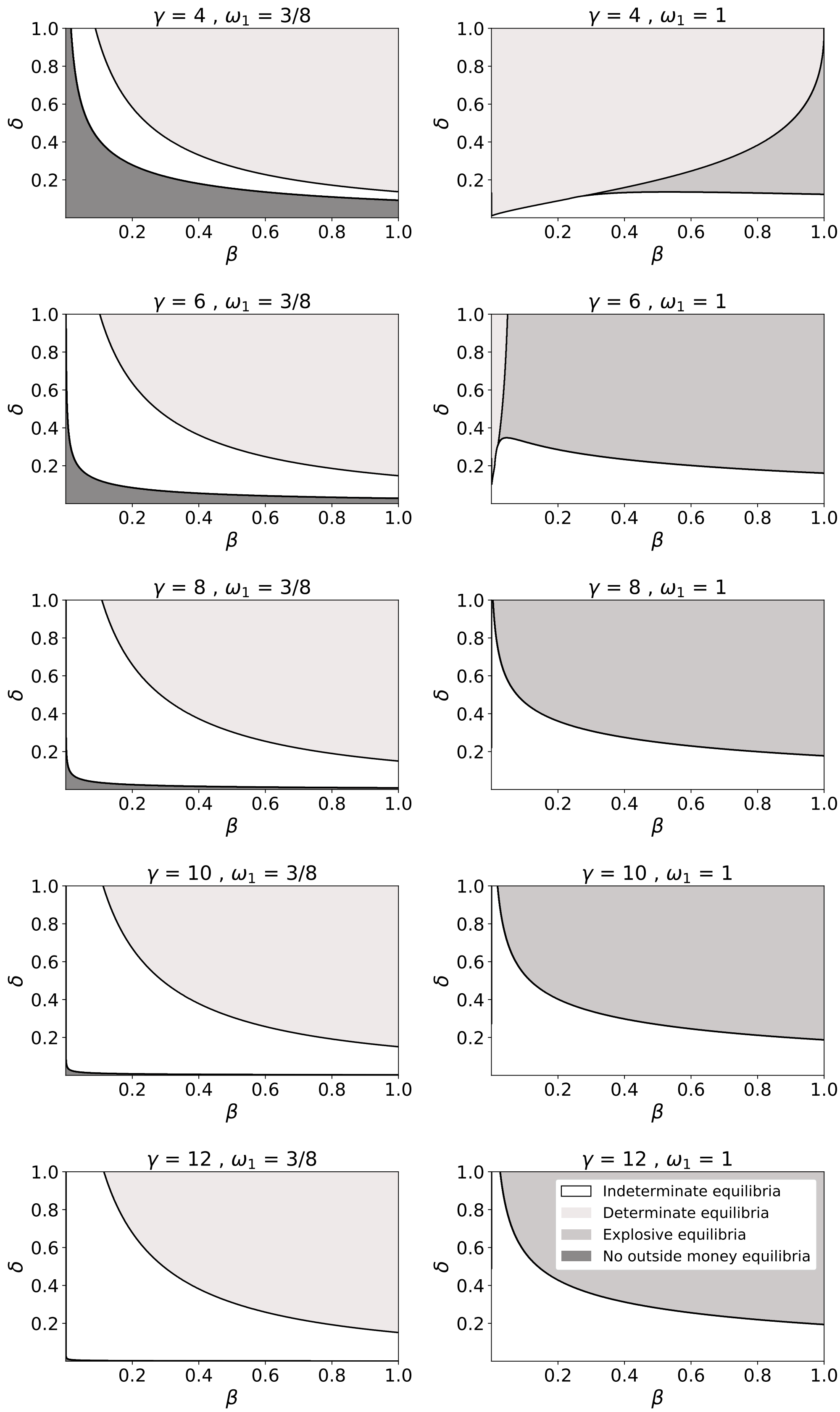


Figure 1.3: Stability properties of the monetary steady state for different parameter configurations, when consumers are sophisticated.

the number of eigenvalues inside the unit circle for every pair of (β, δ) in the grid under the Cartesian product of three values of $\gamma \in \{4, 6, 8, 10, 12\}$ and two values of $\omega_1 \in \{3/8, 1\}$.

First, to compare the two-period model and our model, we consider the case with the endowment profile fully concentrated on the young because a locally indeterminate equilibrium exists in the two-period model when the young's endowment share is very high. When $\gamma = 4$, the two-period model requires the time discount factor to be less than $1/16$ for local indeterminacy as noted in the cycle section.[11] However, in our model with $\beta = 0.6$, a locally indeterminate equilibrium emerges if the time discount factor is less than 0.136.[12] One year time discount factor is $0.136^{1/20} \approx 0.91$ in our model assuming one-period is 20 years. Likewise, one year time discount factor is $0.0625^{1/30} \approx 0.91$ in a two-period model assuming one-period is 30 years. This result implies that the range of the time discount factor is comparable with the two-period model for local indeterminacy.

Another important finding is that the local SE arises under a low share of the young's endowment set at $\omega_1 = 3/8$ (hump-shaped endowment profile as observed in data) as seen in the left panels in Figure 1.3. In this endowment profile, EBC is not observed.

Moreover, compared to the exponential discounting case of $\beta = 1$, the area for the local SE increases as the present bias is introduced in both high and low young endowments. Thus, the QHD contributes to the local SE.

Figure 1.4 depicts four types of equilibria over parameter spaces. The figure examines the stability properties of hyperbolic equilibria in (ω_1, β) Caretesian plane for six pairs of (γ, δ). We make a normalization that $\omega_1 + \omega_2 = 1$ and $\omega_3 = 0$ as we did in the endogeneous business cycle analysis and thus the parameter space of interest is $(\omega_1, \beta) \in [0, 1] \times [0, 1]$. We calculated the number of eigenvalues inside the unit circle for every pair of (ω_1, β) in the grid under the Cartesian product of three values of $\gamma \in \{4, 5, 6\}$ and two values of $\delta \in \{0.3, 0.7\}$.

We denote the dark gray area on the lower left of each panel as "no outside money equilibria", where the conditions in Assumption 1.2 are not satisfied and thus, the aggregate saving is negative and so is the price of money. Overall, the area of this region shrinks for high γ and δ. As γ increases, households are willing to smooth over lifetime consumption, especially due to the zero endowment when old. This leads to a positive aggregate saving by making the middle-aged save more. As δ increases, both the young and the middle-aged are willing to save more. Thus, even with a low ω_1, the aggregate saving can be positive. In each panel, no outside money equilibrium disappears at high β and large ω_1. Here, β works as an additional multiplicative time discount factor. Thus, as β goes up, both the young and middle-aged save

[11] The necessary and sufficient condition for a two-period cycle under CRRA preferences is the same as that for a locally indeterminate equilibrium.

[12] Here, we also consider $\beta = 0.6$ following Laibson (1997).

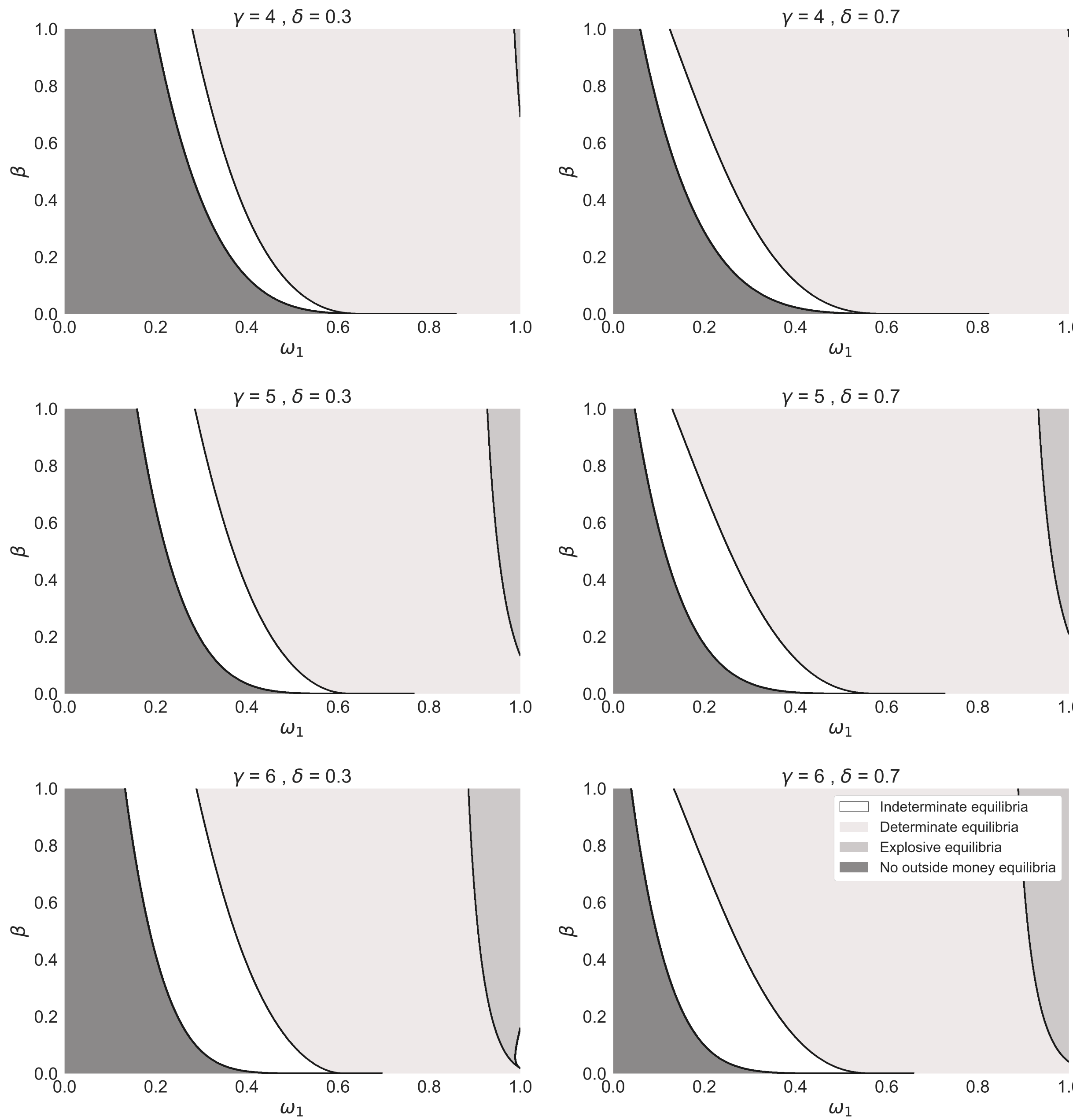


Figure 1.4: Stability properties of the monetary steady state for different parameter configurations, when consumers are sophisticated.

more and thus, the no outside money equilibria diminishes. As ω_1 decreases, the endowment structure is hump-shaped. Thus, the young want to borrow which leads the aggregate saving to be negative. In contrast, as ω_1 increases, both the young and the middle-aged save and thus the aggregate saving is positive.

The indeterminate equilibria stay at the white region where the number of eigenvalues inside the unit circle is more than one. A lower β, i.e. more hyperbolic discounting, can sometimes turn an otherwise determinate (light gray) equilibrium into an indeterminate one given ω_1. Indeed, the indeterminate equilibria can arise in an economy under the QHD preferences with a wide range of moderate ω_1 unlike in the exponential discounting case which requires a very low ω_1. Thus, a sunspot equilibrium can exist in the model with hump-shaped income profiles consistent with data.[13]

Figure 1.5 describes dynamic equilibrium paths starting at different initial interest rates for both models inhabited by exponential and hyperbolic discounting consumers under a certain set of parameters: $\{\omega_1, \delta, \gamma\} = \{0.2, 0.7, 4\}$ assuming $\omega_3 = 0$. The purpose of this figure is to show that introducing QHD converts a determinate equilibrium into an indeterminate one as shown in Figure 1.4. Thus, any sequences of real returns starting at arbitrary initial interest rates converge to the monetary steady state whereas it is not in the exponential discounting case.[14] We explain this result in more detail later after talking about the existence of non-hyperbolic steady-state on the boundary between no outside money and indeterminate equilibria areas.

In the light gray area, the stationary equilibria are locally saddle-path stable or determinate. In this case, there is one stable eigenvalue, and thus the competitive equilibrium is unique given an initial condition for the wealth distribution. Finally, the gray area indicates the explosive case where there are no eigenvalues inside the unit circle. This case is not in our interest since it requires a high ω_1 while we often consider the hump-shaped endowment structure to be consistent with data.

[13] Following Laibson (1997), we take $\beta = 0.6$ as a reasonable hyperbolic discounting factor for one-period in our three-period model regarded as being equivalent to about 20 years. Although Laibson (1997) uses $\beta = 0.6$ for one year, it is reasonable to accept it as a QHD factor for one period as well since agents keep hyperbolic-discounting every year.

[14] We have three sequences of equilibrium interest rates as follows in the exponential discounting case. Given an initial distribution of money holdings, equilibrium path P_1 has initial three interest rates, R_0, R_1 and R_2, to satisfy the equilibrium condition in the first period and be on the unique stable manifold around the monetary steady state. On the other hand, paths P_2 and P_3 have different values of the initial interest rates to meet the first-period equilibrium condition but not to be on the unique stable manifold. We generate the rest of real returns in all paths using a linearized system with the Jacobian matrix in (1.35). For the hyperbolic discounting case, all three paths take R_0, R_1 and $\bar{m}_{1,-1}$ from the corresponding ones in the exponential discounting case as initial starting points. Then,R_2 is determined so that it satisfies the equilibrium condition in the first period with given R_0, R_1 and $\bar{m}_{1,-1}$. The rest of the real returns are produced using a linearized system as above.

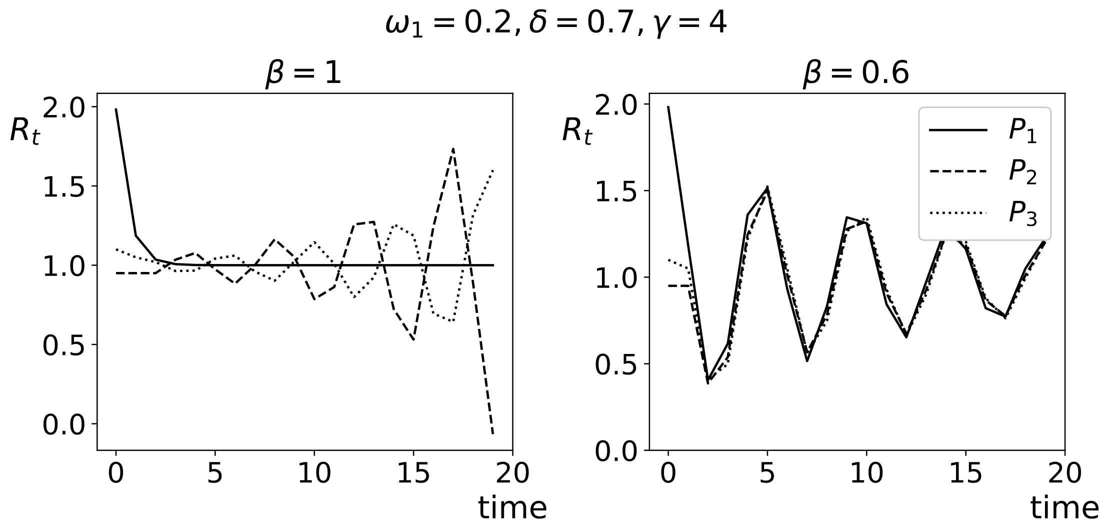


Figure 1.5: Dynamic paths with different initial conditions

Now, we analyze the characteristics of eigenvalues when an equilibrium is locally indeterminate. We first show the monetary steady-state is non-hyperbolic at the boundary between the no outside money area and the indeterminate equilibria area. In other words, one of three eigenvalues of $J[1]$ is 1 on the boundary where the aggregate money demand is zero at $R^* = 1$. This result is the content of the following proposition.

Proposition 1.7. *The steady-state is non-hyperbolic with an eigenvalue of the Jacobian matrix* $J[1]$ *at* 1 *if and only if* $A^* = 0$.

Proof. See Appendix 1.A.12.

This result implies that the set of the non-hyperbolic steady states is a Lebesgue measure zero set as observed in Figures 1.4 and 1.10. In our model, the characteristic function defining three eigenvalues has a negative coefficient for the third power as seen in the following lemma.

Lemma 1.2. *The coefficient for the third power in the cubic characteristic function of the (1.35) is negative.*

From Proposition 1.7, the cubic characteristic function has 1 as a real solution when $A^* = 0$. The characteristic function shifts downward when the aggregate asset demand is positive whereas it shifts upward when the demand is negative as seen in (1.A.84) and (1.A.85). Therefore, if the steady-state has complex eigenvalues at the boundary, the indeterminate equilibria area has a real eigenvalue less than 1 because the characteristic function with a negative coefficient for the third-order term shifts downward under a positive aggregate money demand. Indeed, both the no outside money and indeterminate equilibria areas have two complex eigenvalues inside the unit circle which are conjugates to each other from the numerical analysis above.[15] Then, the real eigenvalue is less than 1 in the indeterminate equilibria area. Therefore, the indeterminate area has all three eigenvalues inside the unit circle which implies all invariant manifolds around the monetary steady-state are stable.

Since it is very hard to analytically identify the effects of changes in parameter values to eigenvalues, we resort to numerical analysis to understand the relationship. The numerical exercise implies that a rise in ω_1 reduces the value of a real eigenvalue and the module of complex eigenvalues. Therefore, as ω_1 increases, the module of complex eigenvalues falls, which flips the indeterminate area into the determinate area. Likewise, a lower QHD parameter also decreases the module of complex eigenvalues. Thus, a determinate equilibrium converts into an indeterminate one as β declines.

1.5 Comparisons of Two-Period Cycles and Locally Indeterminate Equilibria

To compare the sets for two-period cycles and locally indeterminate equilibria, we restrict our attention to the pure exponential discounting case by setting $\beta = 1$. In this case, we know the sufficient condition for a cycle is also necessary by Bhattacharya and Russell (2003). We show the two sets in a (ω_1, δ) Caretesian plane for six different values of $\gamma \in \{4, 6, 8, 10, 12, 20\}$. Again, we make a normalization that $\omega_1 + \omega_2 = 1$ and $\omega_3 = 0$ as we did in the analysis above and thus the parameter space of interest is $(\omega_1, \delta) \in [0, 1] \times [0, 1]$.

First, both low time discount factor and high-risk aversion yield two-period cycles and local indeterminate equilibria. Their set expands as the time discount factor falls and risk aversion rises. However, the ratio of endowments between the young and middle-aged contributes to the existence of EBC and local SE in a reverse way. When the endowment is concentrated on the young, local indeterminate equilibria exist, but two-period cycles not. On the other hand, when the endowment is concentrated on the middle-aged, two-period cycles exist, but not local indeterminate equilibria. This result contrasts with the finding in a two-period OLG economy with a single commodity that the set of economies for two endogenous economic fluctuations coincide when the intertemporal elasticity of saving with respect to the

[15] As the risk aversion converges to infinity, the determinant of the characteristic equation has a global maximum at $\omega_1 = 0.5$, which is negative. Thus, complex eigenvalues exist in all the parameter space under high-risk aversion.

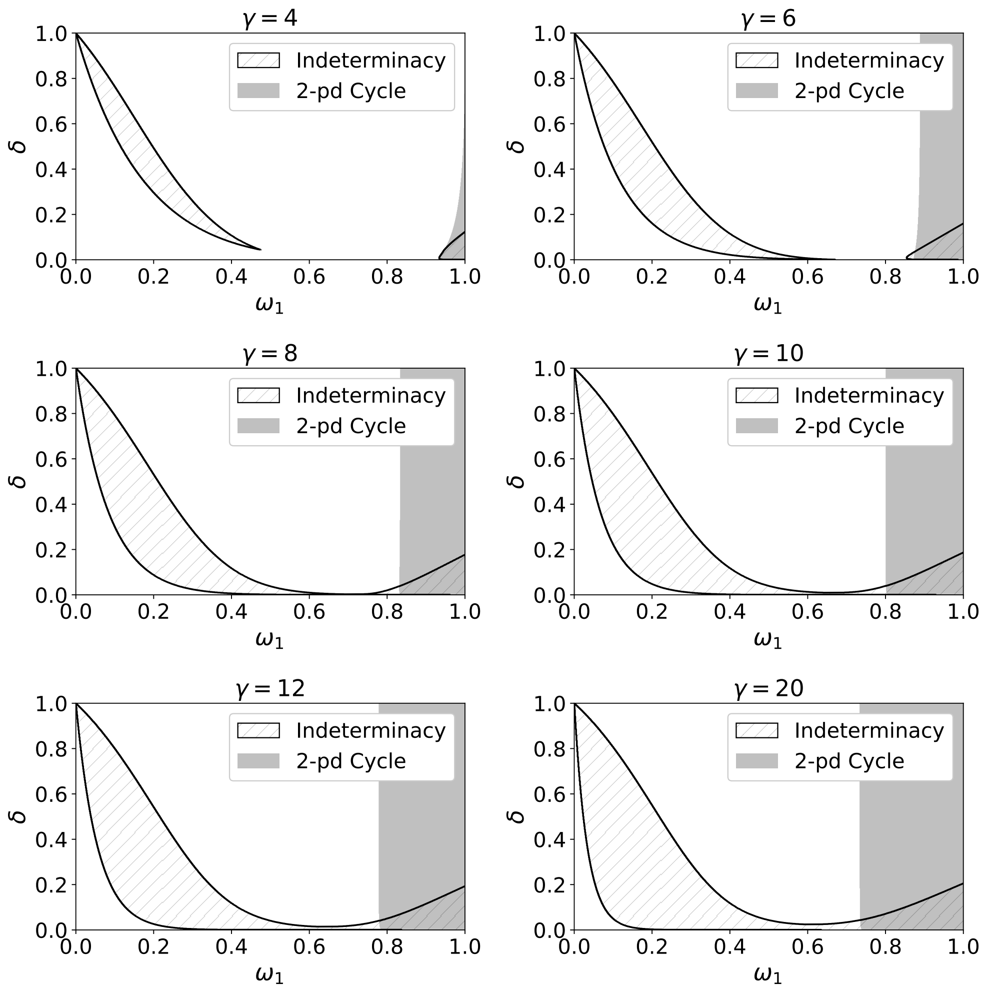


Figure 1.6: Local Indeterminacy vs Two-period Cycle

interest rate is less than $-\frac{1}{2}$. Another implication for the non-coincidence in a three-period OLG model is that the set of economies for endogenous fluctuations appears in the parameter space more extensively than in a two-period model since they do not significantly overlap.

To understand the non-coincidence result, we analyze the relationship between the eigenvalues of a linearized equilibrium dynamics and the necessary and sufficient condition for a cycle. This is the content of the following Proposition.

Lemma 1.3. *When the necessary and sufficient condition for the existence of the cycle is satisfied under the exponential discounting, there must be a real eigenvalue λ_r of (1.35) less than -1. The reverse is also true if there are complex conjugate eigenvalues.*

Proof. See Appendix 1.A.14. □

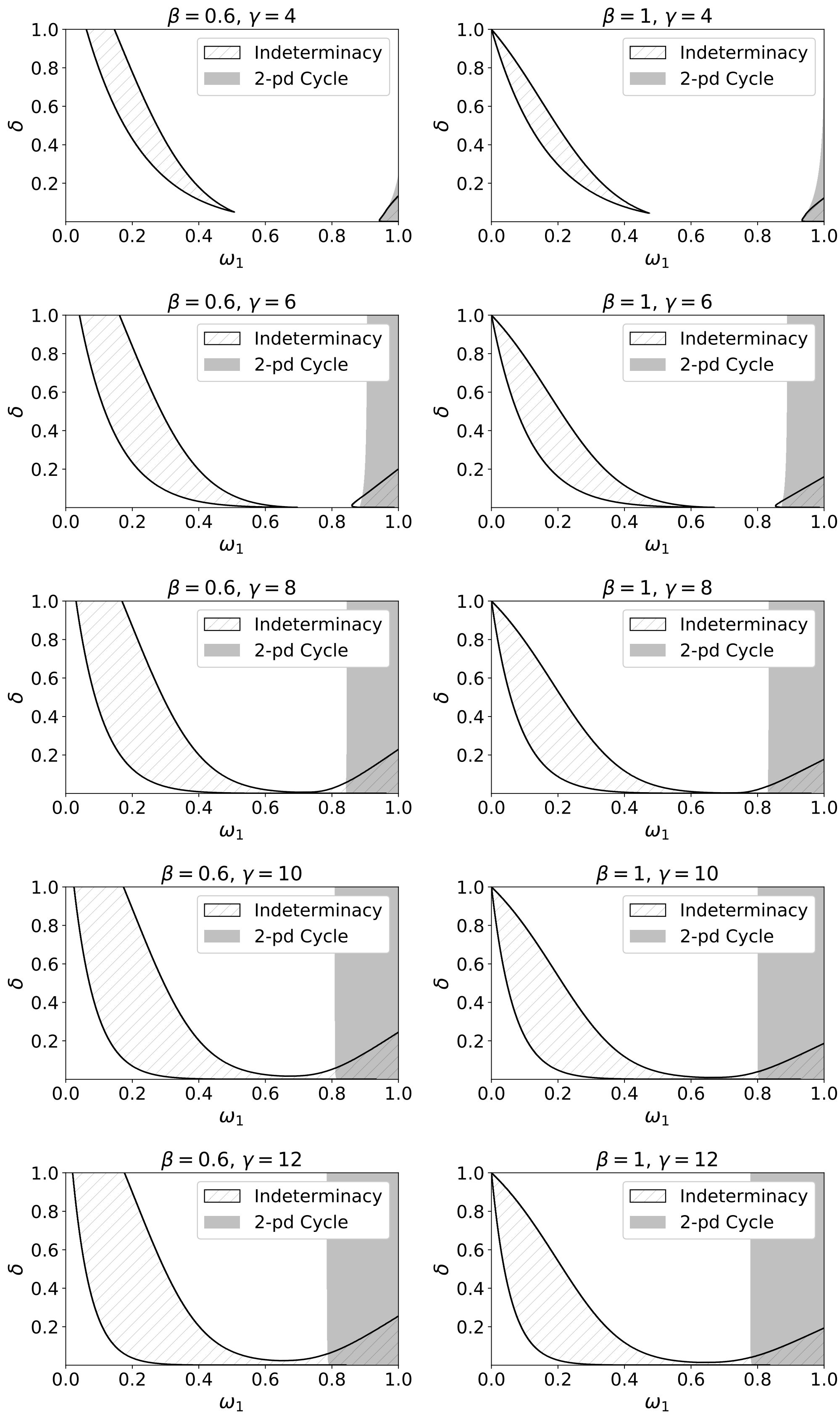


Figure 1.7: Local Indeterminacy vs Two-period Cycle

Lemma 1.3 provides a partial explanation for the non-coincidence result. If a two-period cycle exists, then there should be one real eigenvalue less than -1. Thus, there is already one eigenvalue outside the unit circle, which leads to the non-overlapping local indeterminate equilibria and cycles. In a two-period OLG model, the unique real eigenvalue should be between -1 and 0 for cycle existence under the exponential discounting. However, assuming there are complex eigenvalues, the unique real eigenvalue should be less than -1 in a three-period model. This discrepancy explains why the conventional finding of perfectly overlapped local indeterminate equilibria and cyclic equilibria breaks down in a three-period OLG economy.

1.6 Conclusion

We analyze the implications of QHD preference for endogenous economic fluctuations such as endogenous deterministic cycles and local sunspot equilibria in a three-period OLG model with pure exchange. We provide a sufficient condition for the existence of two-period endogenous cycles and a necessary and sufficient condition for the existence of a local sunspot equilibrium characterized by local indeterminacy. Through numerical characterization, we show that introducing the present bias into preferences shrinks the set of two-period cycles and/but enlarges the set of locally indeterminate equilibria. For the existence of cycles, there should be strong income effects and weak substitution effects when interest rates rise. Introducing QHD preferences reduces both income and substitution effects by decreasing savings and making goods across periods less substitutable. However, income effects are dampened more than substitution effects, which leads to the contraction of the EBC set. A local sunspot equilibrium needs the complex eigenvalues of the locally linearized equilibrium system to be inside the unit circle. The present bias contributes to a decrease in the module of the eigenvalues. Thus, the locally indeterminate equilibria can exist with a reasonable value of time discount factor, unlike in a standard two-period OLG model with exponential discounting preferences.

We also find that two-period cycles arise under a skewed endowment profile toward the young for a strong income effect. However, a local sunspot equilibrium exists under a hump-shaped endowment profile because an increase in the endowment share of the young increases the module of complex eigenvalues and decreases a real eigenvalue to be outside the unit circle, which can violate the requirement for the local indeterminacy. This result breaks down the conventional finding in a two-period OLG economy with a single commodity where the set of economies for two endogenous economic fluctuations coincide. Hence, endogenous fluctuations exist in more extensive parameter space, and thus they can be observed under calibration parameter values consistent with empirical findings when lengthening the lifetime of households.

This paper has several limitations and possible extensions. First, this paper does not show whether the sufficient condition for two-period cycles is also necessary, although it is checked numerically. Also, future research can extend the results of this paper in a more realistic OLG model with longer lifespans and other preferences such as habit formation. This extension can provide robustness for the existence of endogenous fluctuations.

1.A Appendix: Proofs

1.A.1 The intertemporal Euler equation of sophisticated consumers in (1.5):

The current value of the sophisticated agent is:

$$W_{i,t}\left(x_t\right) = \left\{u\left(c_i\left(x_t\right)\right) + \beta\delta V_{i+1,t+1}\left(R_t\left(x_t - c_i\left(x_t\right)\right) + y_{t+1}\right)\right\} \tag{1.A.1}$$

where

$$c_i\left(x_t\right) \quad = \quad \underset{c_{i,t}\in[0,x_t]}{\text{argmax}}\left\{u\left(c_{i,t}\right) + \beta\delta V_{i+1,t+1}\left(R_t\left(x_t - c_{i,t}\right) + y_{t+1}\right)\right\} \tag{1.A.2}$$

The continuation-value function of the sophisticated agent is:

$$V_{i+1,t+1}\left(x_{t+1}\right) = \left\{u\left(c_{i+1}\left(x_{t+1}\right)\right) + \delta V_{i+2,t+2}\left(R_{t+1}\left(x_{t+1} - c_{i+1}\left(x_{t+1}\right)\right) + y_{t+2}\right)\right\} \tag{1.A.3}$$

where

$$c_{i+1}\left(x_{t+1}\right) = \underset{c_{i+1,t+1}\in[0,x_{t+1}]}{\text{argmax}}\left\{u\left(c_{i+1,t+1}\right) + \beta\delta V_{i+2,t+2}\left(R_{t+1}\left(x_{t+1} - c_{i+1,t+1}\right) + y_{t+2}\right)\right\} \tag{1.A.4}$$

Note that in $W_t\left(x_t\right)$, the continuation value function $V_{t+1}\left(x_{t+1}\right)$ is evaluated at the policy function $c\left(x_{t+1}\right)$ under the rational expectations that the future selves continue to be quasi-hyperbolic. Therefore, in (1.A.2), the current self-choose c_t with the future behavior of the future selves obtained via backward induction. The continuation value function is discounted at δ to respect the utility function of the current self in date t. In every period, the choice problem under the hyperbolic discounting is stationary. Thus, the policy function is time-invariant in the infinite horizon problem.

To derive the intertemporal Euler equation, take the first order condition of (1.A.2)

$$\partial c_{i,t}:\ u'\left(c_{i,t}\right) = \beta\delta R_t\left[V'_{i+1,t+1}\left(x_{t+1}\right)\right] \tag{1.A.5}$$

From the total differentiation of the identity equation (1.A.3) with respect to x_{t+1},

$$V'_{i+1,t+1}\left(x_{t+1}\right) = u'\left(c_{i+1}\left(x_{t+1}\right)\right)\frac{\partial c_{i+1}\left(x_{t+1}\right)}{\partial x_{t+1}} + \delta R_{t+1}\left(1 - \frac{\partial c_{i+1}\left(x_{t+1}\right)}{\partial x_{t+1}}\right)V'_{i+2,t+2}\left(x_{t+2}\right) \tag{1.A.6}$$

The first order condition of (1.A.4) is:

$$\partial c_{i+1,t+1}:\ u'\left(c_{i+1,t+1}\right) = \beta\delta R_{t+1}V'_{i+2,t+2}\left(x_{t+2}\right) \tag{1.A.7}$$

By combining (1.A.6) and (1.A.7),

$$V'_{i+1,t+1}\left(x_{t+1}\right)=u'\left(c_{i+1}\left(x_{t+1}\right)\right)\frac{\partial c_{i+1}\left(x_{t+1}\right)}{\partial x_{t+1}}+\frac{1}{\beta}\left(1-\frac{\partial c_{i+1}\left(x_{t+1}\right)}{\partial x_{t+1}}\right)u'\left(c_{i+1}\left(x_{t+1}\right)\right) \tag{1.A.8}$$

By combining (1.A.5) and (1.A.8),

$$\partial c_{i,t}:\ u'\left(c_{i,t}\right)=R_t\underbrace{\left\{\beta\delta c'_{i+1}\left(x_{t+1}\right)+\delta\left(1-c'_{i+1}\left(x_{t+1}\right)\right)\right\}}_{\text{effective discount factor}}u'\left(c_{i+1,t+1}\right) \tag{1.A.9}$$

1.A.2 The intertemporal Euler equation of naive consumers in (1.10) and (1.11):

A naive consumer forecasts that her future selves commit its optimal plan today and thus, her future selves follow exponential discounting. We construct the standard exponential discounting Bellman equation as follows:

$$V^e_{i+1,t+1}\left(x_{t+1}\right)=\max_{c^e_{i+1,t+1}\in[0,x_{t+1}]}\left\{u\left(c^e_{i+1,t+1}\right)+\delta V^e_{i+2,t+2}\left(R_{t+1}\left(x_{t+1}-c^e_{i+1,t+1}\right)+y_{t+2}\right)\right\} \tag{1.A.10}$$

$$c^e_{i+1}\left(x_{t+1}\right)=\underset{c^e_{i+1,t+1}\in[0,x_{t+1}]}{\text{argmax}}\left\{u\left(c^e_{i+1,t+1}\right)+\delta V^e_{i+2,t+2}\left(R_{t+1}\left(x_{t+1}-c^e_{i+1,t+1}\right)+y_{t+2}\right)\right\} \tag{1.A.11}$$

where V is a continuation-value function.

The current value of the naive agent is:

$$W^n_{i,t}\left(x_t\right)=\max_{c^n_{i,t}\in[0,x_t]}\left\{u\left(c^n_{i,t}\right)+\beta\delta V^e_{i+1,t+1}\left(R_t\left(x_t-c^n_{i,t}\right)+y_{t+1}\right)\right\} \tag{1.A.12}$$

$$c^n_i\left(x_t\right)=\underset{c^n_{i,t}\in[0,x_t]}{\text{argmax}}\left\{u\left(c^n_{i,t}\right)+\beta\delta V^e_{i+1,t+1}\left(R_t\left(x_t-c^n_{i,t}\right)+y_{t+1}\right)\right\} \tag{1.A.13}$$

where W is a current-value function.

The naive agents believe that from date $t+1$ onwards, he will choose according to the exponential discounter's policy function.

$$\partial c^n_{i,t}:\ u'\left(c^n_{i,t}\right)=\beta\delta R_t V^{e'}_{i+1,t+1}\left(x_{t+1}\right) \tag{1.A.14}$$

and

$$\partial c^e_{i+1,t+1}:\ u'\left(c^e_{i+1,t+1}\right)=\delta R_{t+1}V^{e'}_{i+2,t+2}\left(x_{t+2}\right) \tag{1.A.15}$$

By the envelop theorem,

$$V^{e'}_{i+1,t+1}\left(x_{t+1}\right)=\delta R_{t+1}V^{e'}_{i+2,t+2}\left(x_{t+2}\right) \tag{1.A.16}$$

Thus, combining (1.A.15) and (1.A.16), we obtain:

$$u'\left(c^{e}_{i+1,t+1}\right) = V^{e'}_{i+1,t+1}\left(x_{t+1}\right) \tag{1.A.17}$$

By plugging (1.A.17) at time $t+2$ into (1.A.15), we get the Euler equation which defines the optimal future plan of the time-t self

$$u'\left(c^{e}_{i+1,t+1}\right) = \delta R_{t+1} u'\left(c^{e}_{i+2,t+2}\right) \tag{1.A.18}$$

By inserting (1.A.17) into (1.A.14), we get the Euler equation which defines the actual plan of the naive consumers onward

$$u'\left(c^{n}_{i,t}\right) = \beta\delta R_t u'\left(c^{e}_{i+1,t+1}\right) \tag{1.A.19}$$

In every period, agents actually decide consumptions following (1.A.19).

1.A.3 Proof of Proposition 1.1:

Note that $a_{1,t}$ and $a_{2,t}$ can be written as:

$$a_{1,t}\left[R_{t+1}, R_t\right] = \omega_1 - c_{1,t},\ a_{2,t}\left[R_t, R_{t-1}\right] = R_{t-1}\omega_1 + \omega_2 - \left(R_{t-1}c_{1,t-1} + c_{2,t}\right) \tag{1.A.20}$$

Using the market-clearing condition,

$$A_{t+1}\left[R_{t+2}, R_{t+1}, R_t\right] = R_t A_t\left[R_{t+1}, R_t, R_{t-1}\right] \tag{1.A.21}$$

we can denote the equilibrium condition as:

$$\begin{aligned} G\left[R_{t+2}, R_{t+1}, R_t, R_{t-1}\right] =& \omega_1 + \omega_2 - c_{1,t+1}\left[R_{t+2}, R_{t+1}\right] - c_{2,t+1}\left[R_{t+1}, R_t\right] \\ & - R_t\left\{R_{t-1}\omega_1 + \omega_2 - R_{t-1}c_{1,t-1}\left[R_t, R_{t-1}\right] - c_{2,t}\left[R_t, R_{t-1}\right]\right\} \\ =& 0 \end{aligned} \tag{1.A.22}$$

Since

$$\frac{\partial G\left[R_{t+2}, R_{t+1}, R_t, R_{t-1}\right]}{\partial R_{t+2}} = -\frac{\partial c_{1,t+1}\left[R_{t+2}, R_{t+1}\right]}{\partial R_{t+2}} \tag{1.A.23}$$

by implicit function theorem, we can write R_{t+2} as:

$$R_{t+2} = F\left[R_{t+1}, R_t, R_{t-1}\right] \tag{1.A.24}$$

whenever $\frac{\partial c_{1,t+1}\left[R_{t+2}, R_{t+1}\right]}{\partial R_{t+2}} \neq 0$.

Note that we use notations for the sophisticated consumers. However, the entire analysis here is also applicable to the naive consumers as well by just replacing the notations with ones for the native consumers.

1.A.4 Proof of Proposition 1.2:

At any stationary equilibrium, $R_t = R^*$. By the market clearing condition $A^* = R^* A^*$, we have $R^* = 1$ as a unique monetary steady-state if $A^* = a_1^* + a_2^* > 0$.

1.A.5 Proof of Proposition 1.3:

Under the utility function with a constant relative risk aversion, the optimal consumptions of the three-period-lived sophisticated consumer born in date t are given by the following three equations:

$$(c_{1,t})^{-\gamma} = \{\beta\delta c_2'(x_{t+1}) + \delta(1 - c_2'(x_{t+1}))\} R_t (c_{2,t+1})^{-\gamma} \tag{1.A.25}$$

$$(c_{2,t+1})^{-\gamma} = \beta\delta R_{t+1} (c_{3,t+2})^{-\gamma} \tag{1.A.26}$$

and

$$c_{1,t} + \frac{c_{2,t+1}}{R_t} + \frac{c_{3,t+2}}{R_t R_{t+1}} = \omega_1 + \frac{\omega_2}{R_t} + \frac{\omega_3}{R_t R_{t+1}} \tag{1.A.27}$$

By solving (1.A.26),

$$c_{3,t+2} = (\beta\delta)^{\frac{1}{\gamma}} R_{t+1}^{\frac{1}{\gamma}} c_{2,t+1} \tag{1.A.28}$$

From the budget constraint over the rest of life at time $t+1$,

$$c_{2,t+1} + \frac{c_{3,t+2}}{R_{t+1}} = x_{t+1} + \frac{\omega_3}{R_{t+1}} \tag{1.A.29}$$

By combining (1.A.27) and (1.A.29),

$$c_{2,t+1} = \frac{x_{t+1} + \omega_3 / R_{t+1}}{1 + (\beta\delta)^{\frac{1}{\gamma}} R_{t+1}^{\frac{1}{\gamma}-1}} \tag{1.A.30}$$

Thus,

$$c_2'(x_{t+1}) = \frac{1}{1 + (\beta\delta)^{\frac{1}{\gamma}} R_{t+1}^{\frac{1}{\gamma}-1}} \tag{1.A.31}$$

By plugging this result into (1.A.25), we obtain:

$$(c_{1,t})^{-\gamma} = \left\{ \frac{\delta\left(\beta + (\beta\delta)^{\frac{1}{\gamma}} R_{t+1}^{\frac{1}{\gamma}-1}\right)}{1 + (\beta\delta)^{\frac{1}{\gamma}} R_{t+1}^{\frac{1}{\gamma}-1}} \right\} R_t (c_{2,t+1})^{-\gamma} = \eta_t R_t (c_{2,t+1})^{-\gamma} \tag{1.A.32}$$

Therefore,

$$c_{2,t+1} = \eta_t^{\frac{1}{\gamma}} R_t^{\frac{1}{\gamma}} c_{1,t} \tag{1.A.33}$$

By inserting these results into the lifetime budget constraint (1.A.27), we get:

$$c_{1,t} + \frac{\eta_t^{\frac{1}{\gamma}} R_t^{\frac{1}{\gamma}} c_{1,t}}{R_t} + \frac{(\beta\delta\eta_t)^{\frac{1}{\gamma}} R_t^{\frac{1}{\gamma}} R_{t+1}^{\frac{1}{\gamma}}}{R_t R_{t+1}} = \omega_1 + \frac{\omega_2}{R_t} + \frac{\omega_3}{R_t R_{t+1}} \tag{1.A.34}$$

Hence,

$$c_{1,t} = \frac{\omega_1 + \omega_2/R_t + \omega_3/\left(R_t R_{t+1}\right)}{1 + \eta_t^{\frac{1}{\gamma}} R_t^{\frac{1}{\gamma}-1} + (\beta\delta\eta_t)^{\frac{1}{\gamma}} \left(R_t R_{t+1}\right)^{\frac{1}{\gamma}-1}} \tag{1.A.35}$$

1.A.6 Proof of Proposition 1.4:

Under the utility function with a constant relative risk aversion, the consumption plan of the three-period-lived naive consumer born in date t is given by the following three equations:

$$\left(c_{1,t}^n\right)^{-\gamma} = \beta\delta R_t \left(c_{2,t+1}^e\right)^{-\gamma} \tag{1.A.36}$$

$$\left(c_{2,t+1}^e\right)^{-\gamma} = \delta R_t \left(c_{3,t+2}^e\right)^{-\gamma} \tag{1.A.37}$$

and

$$c_{1,t}^n + \frac{c_{2,t+1}^e}{R_t} + \frac{c_{3,t+2}^e}{R_t R_{t+1}} = \omega_1 + \frac{\omega_2}{R_t} + \frac{\omega_3}{R_t R_{t+1}} \tag{1.A.38}$$

By solving the first two equations,

$$c_{2,t+1}^e = (\beta\delta)^{\frac{1}{\gamma}} R_t^{\frac{1}{\gamma}} c_{1,t}^n \tag{1.A.39}$$

$$c_{3,t+2}^e = \delta^{\frac{1}{\gamma}} R_{t+1}^{\frac{1}{\gamma}} c_{2,t+1}^e \tag{1.A.40}$$

and

$$c_{1,t}^n + \frac{(\beta\delta)^{\frac{1}{\gamma}} R_t^{\frac{1}{\gamma}} c_{1,t}^n}{R_t} + \frac{(\beta\delta^2)^{\frac{1}{\gamma}} \left(R_t R_{t+1}\right)^{\frac{1}{\gamma}} c_{1,t}^n}{R_t R_{t+1}} = \omega_1 + \frac{\omega_2}{R_t} + \frac{\omega_3}{R_t R_{t+1}} \tag{1.A.41}$$

Therefore,

$$c_{1,t}^n = \frac{\omega_1 + \omega_2/R_t + \omega_3/\left(R_t R_{t+1}\right)}{1 + (\beta\delta)^{\frac{1}{\gamma}} R_t^{\frac{1}{\gamma}-1} + (\beta\delta^2)^{\frac{1}{\gamma}} \left(R_t R_{t+1}\right)^{\frac{1}{\gamma}-1}} \tag{1.A.42}$$

However, the actual consumptions in the second and third period of life are defined by the following equation instead of (1.A.37):

$$\left(c_{2,t+1}^n\right)^{-\gamma} = \beta\delta R_{t+1} \left(c_{3,t+2}^n\right)^{-\gamma} \tag{1.A.43}$$

Thus,

$$c_{3,t+2}^n = (\beta\delta)^{\frac{1}{\gamma}} R_{t+1}^{\frac{1}{\gamma}} c_{2,t+1}^n \tag{1.A.44}$$

From the lifetime budget constraint,

$$c_{1,t}^n + \frac{c_{2,t+1}^n}{R_t} + \frac{(\beta\delta)^{\frac{1}{\gamma}} R_{t+1}^{\frac{1}{\gamma}} c_{2,t+1}^n}{R_t R_{t+1}} = \omega_1 + \frac{\omega_2}{R_t} + \frac{\omega_3}{R_t R_{t+1}} \tag{1.A.45}$$

Considering the closed form expression of $c_{1,t}^n$, this equation can be re-expressed as:

$$c_{1,t}^n + \frac{c_{2,t+1}^n}{R_t} + \frac{(\beta\delta)^{\frac{1}{\gamma}} R_{t+1}^{\frac{1}{\gamma}} c_{2,t+1}^n}{R_t R_{t+1}} = \left\{1 + (\beta\delta)^{\frac{1}{\gamma}} R_t^{\frac{1}{\gamma}-1} + \left(\beta\delta^2\right)^{\frac{1}{\gamma}} (R_t R_{t+1})^{\frac{1}{\gamma}-1}\right\} c_{1,t}^n \tag{1.A.46}$$

Therefore,

$$c_{2,t+1}^n = (\beta\delta)^{\frac{1}{\gamma}} R_t^{\frac{1}{\gamma}} \frac{\left(1 + \delta^{\frac{1}{\gamma}} R_{t+1}^{\frac{1}{\gamma}-1}\right)}{1 + (\beta\delta)^{\frac{1}{\gamma}} R_{t+1}^{\frac{1}{\gamma}-1}} c_{1,t}^n \tag{1.A.47}$$

1.A.7 Proof of Corollary 1.1:

When $\gamma = 1$, the optimal consumptions of the sophisticated consumer degenerate as follows.

$$c_{1,t} = \frac{\omega_1 + \omega_2/R_t + \omega_3/\left(R_t R_{t+1}\right)}{1 + \eta_t + \beta\delta\eta_t} \text{ and } \eta_t = \frac{\delta\left(\beta + \beta\delta\right)}{1 + \beta\delta} \tag{1.A.48}$$

By inserting η_t into $c_{1,t}$ and rearranging the expression,

$$c_{1,t} = \frac{\omega_1 + \omega_2/R_t + \omega_3/\left(R_t R_{t+1}\right)}{1 + \beta\delta + \beta\delta^2} \tag{1.A.49}$$

$c_{2,t+1}$ and $c_{3,t+2}$ reduce to:

$$c_{2,t+1} = \frac{\delta\left(\beta + \beta\delta\right)}{1 + \beta\delta} R_t c_{1,t} \text{ and } c_{3,t+2} = \beta\delta R_{t+1} c_{2,t+1} \tag{1.A.50}$$

Likewise, the actual consumptions of the native consumer reduce as follows:

$$c_{1,t}^n = \frac{\omega_1 + \omega_2/R_t + \omega_3/\left(R_t R_{t+1}\right)}{1 + \beta\delta + \beta\delta^2} \tag{1.A.51}$$

$$c_{2,t+1}^n = \frac{\delta\left(\beta + \beta\delta\right)}{1 + \beta\delta} R_t c_{1,t}^n \tag{1.A.52}$$

and

$$c_{3,t+2}^n = \beta\delta R_{t+1} c_{2,t+1}^n \tag{1.A.53}$$

Hence,

$$c_{1,t} = c_{1,t}^n,\ c_{2,t+1} = c_{2,t+1}^n \text{ and } c_{3,t+2} = c_{3,t+2}^n \tag{1.A.54}$$

1.A.8 Closed form expressions for both planned and actual consumptions of naive consumers in two-period business cycles

By replacing interest rates in (1.20) and (1.21) with $\left\{R, \frac{1}{R}\right\}$ and $\left\{\frac{1}{R}, R\right\}$, we can derive closed form expressions for both planned and actual consumptions of naive consumers born in high and low states respectively:

Planned consumptions:

$$\begin{aligned} c_{1H}^{n}(R) &= \frac{\omega_1 + \frac{\omega_2}{R} + \omega_3}{1 + (\beta\delta)^{\frac{1}{\gamma}} R^{\frac{1}{\gamma}-1} + (\beta\delta^2)^{\frac{1}{\gamma}}},\ c_{1L}^{n}(R) = c_{1H}^{n}(1/R) \\ c_{2H}^{e}(R) &= (\beta\delta R)^{\frac{1}{\gamma}} c_{1H}^{n}(R),\ c_{2L}^{e}(R) = c_{2H}^{e}(1/R) \\ c_{3H}^{e}(R) &= (\delta/R)^{\frac{1}{\gamma}} c_{2H}^{e},\ c_{3L}^{e}(R) = c_{3H}^{e}(1/R) \end{aligned} \tag{1.A.55}$$

Actual consumptions:

$$\begin{aligned} c_{1H}^{n}(R) &= \frac{\omega_1 + \frac{\omega_2}{R} + \omega_3}{1 + (\beta\delta)^{\frac{1}{\gamma}} R^{\frac{1}{\gamma}-1} + (\beta\delta^2)^{\frac{1}{\gamma}}},\ c_{1L}^{n}(R) = c_{1H}^{n}(1/R) \\ c_{2H}^{n}(R) &= (\beta\delta R)^{\frac{1}{\gamma}} \frac{\left(1 + \delta^{\frac{1}{\gamma}} R^{1-\frac{1}{\gamma}}\right)}{1 + (\beta\delta)^{\frac{1}{\gamma}} R^{1-\frac{1}{\gamma}}} c_{1H}^{n}(R),\ c_{2L}^{n}(R) = c_{2H}^{n}(1/R) \\ c_{3H}^{n}(R) &= (\beta\delta/R)^{\frac{1}{\gamma}} c_{2H}^{n}(R),\ c_{3L}^{n}(R) = c_{3H}^{n}(1/R) \end{aligned} \tag{1.A.56}$$

With consumption allocation for naive consumers, households savings can be written as $a_{1H}^{n} = \omega_1 - c_{1H}^{n}$, $a_{1L}^{n} = \omega_1 - c_{1L}^{n}$, $a_{2H}^{n} = \omega_2 - c_{2H}^{n} + R a_{1H}^{n}$ and $a_{2L}^{n} = \omega_2 - c_{2L}^{n} + \frac{1}{R} a_{1L}^{n}$.

1.A.9 Proof of Lemma 1.5:

In (1.30), LHS $= (R-1)\,\omega_2$, and RHS $= R c_{2L}(R) - c_{2H}(R)$. For the sophisticated consumers, LHS $=$ RHS $= 0$ when $R = 1$. When $R \to \infty$ and assuming $\gamma > 1$, based on the closed form expression of consumptions in (1.19), we have:

$$\lim_{R\to\infty} \eta(R) = \lim_{R\to\infty} \frac{\delta\left[\beta + (\beta\delta)^{\frac{1}{\gamma}} R^{\frac{1}{\gamma}-1}\right]}{1 + (\beta\delta)^{\frac{1}{\gamma}} R^{\frac{1}{\gamma}-1}} = \beta\delta \tag{1.A.57}$$

and

$$\lim_{R\to\infty} \eta\left(\frac{1}{R}\right) = \lim_{R\to\infty} \frac{\delta\left[\beta + (\beta\delta)^{\frac{1}{\gamma}} R^{1-\frac{1}{\gamma}}\right]}{1 + (\beta\delta)^{\frac{1}{\gamma}} R^{1-\frac{1}{\gamma}}} = \delta \tag{1.A.58}$$

Therefore, we have:

$$\lim_{R\to\infty} c_{1H} = \lim_{R\to\infty} \frac{\omega_1 + \omega_2/R}{1 + \eta\left(\frac{1}{R}\right)^{\frac{1}{\gamma}} R^{\frac{1}{\gamma}-1} + \left(\beta\delta\eta\left(\frac{1}{R}\right)\right)^{\frac{1}{\gamma}}} = \frac{\omega_1}{1 + (\beta\delta^2)^{\frac{1}{\gamma}}}$$

$$\lim_{R\to\infty} c_{2H} = \lim_{R\to\infty} \left(\eta\left(\frac{1}{R}\right)R\right)^{\frac{1}{\gamma}} c_{1H} = \infty \tag{1.A.59}$$
$$\lim_{R\to\infty} c_{3H} = \lim_{R\to\infty} \left(\beta\delta\eta\left(\frac{1}{R}\right)\right)^{\frac{1}{\gamma}} c_{1H} = \frac{\left(\beta\delta^2\right)^{\frac{1}{\gamma}}\omega_1}{1+\left(\beta\delta^2\right)^{\frac{1}{\gamma}}}$$

and

$$\lim_{R\to\infty} c_{1L} = \lim_{R\to\infty} \frac{\omega_1+\omega_2 R}{1+\eta\left(R\right)^{\frac{1}{\gamma}}R^{1-\frac{1}{\gamma}}+\left(\beta\delta\eta\left(R\right)\right)^{\frac{1}{\gamma}}} = \lim_{R\to\infty}\frac{\omega_2}{\left(\beta\delta\right)^{\frac{1}{\gamma}}}R^{\frac{1}{\gamma}} = \infty$$
$$\lim_{R\to\infty} c_{2L} = \lim_{R\to\infty} \left(\eta\left(R\right)\frac{1}{R}\right)^{\frac{1}{\gamma}} c_{1L} = \lim_{R\to\infty}\left(\beta\delta\right)^{\frac{1}{\gamma}}R^{-\frac{1}{\gamma}}\frac{\omega_2}{\left(\beta\delta\right)^{\frac{1}{\gamma}}}R^{\frac{1}{\gamma}} = \omega_2 \tag{1.A.60}$$
$$\lim_{R\to\infty} c_{3L} = \lim_{R\to\infty} \left(\beta\delta\eta\left(R\right)\right)^{\frac{1}{\gamma}} c_{1L} = \infty$$

Then, we have:

$$\begin{aligned}\lim_{R\to\infty} \text{LHS}-\text{RHS} &= \lim_{R\to\infty}\left(R-1\right)\omega_2-\left(Rc_{2L}-c_{2H}\right)\\ &= \lim_{R\to\infty} R\left(\omega_2-c_{2L}\right)-\omega_2+c_{2H}\\ &= \infty\end{aligned} \tag{1.A.61}$$

because

$$\lim_{R\to\infty} R\left(\omega_2-c_{2L}\right) = \lim_{R\to\infty}\frac{\omega_2\left(1+\left(\beta\delta\eta\left(R\right)\right)^{\frac{1}{\gamma}}\right)R-\omega_1\eta\left(R\right)^{\frac{1}{\gamma}}R^{1-\frac{1}{\gamma}}}{1+\eta\left(R\right)^{\frac{1}{\gamma}}R^{1-\frac{1}{\gamma}}+\left(\beta\delta\eta\left(R\right)\right)^{\frac{1}{\gamma}}} = \infty \tag{1.A.62}$$

Therefore, we can say LHS $>$ RHS when R is large enough. Combining with LHS $=$ RHS when $R=1$, a sufficient condition for a solution $R^{*}>1$ to (1.30) to exist is:

$$\frac{\partial RHS}{\partial R}|_{R=1} > \frac{\partial LHS}{\partial R}|_{R=1} \tag{1.A.63}$$

This condition can be expressed as:

$$c'_{2H}\left(1\right) < c_2\left(1\right)+c'_{2L}\left(1\right)-\omega_2 = c'_{2L}\left(1\right)+a_1\left(1\right)-a_2\left(1\right) \tag{1.A.64}$$

where $c_2\left(1\right)=c_{2H}\left(1\right)=c_{2L}\left(1\right)$, $a_1\left(1\right)=a_{1H}\left(1\right)=a_{1L}\left(1\right)$ and $a_2\left(1\right)=a_{2H}\left(1\right)=a_{2L}\left(1\right)$.

Since $c'_{2H}\left(1\right)=a_1\left(1\right)+a'_{1H}\left(1\right)-a'_{2H}\left(1\right)$ and $c'_{2L}\left(1\right)=-a_1\left(1\right)+a'_{1L}\left(1\right)-a'_{2L}\left(1\right)$, (1.A.64) reduces to:

$$a_1\left(1\right)+a'_{1H}\left(1\right)-a'_{2H}\left(1\right) < a'_{1L}\left(1\right)-a'_{2L}\left(1\right)-a_2\left(1\right) \tag{1.A.65}$$

By the fact that $c'_{2H}(1) = -c'_{2L}(1)$, $c'_{1H}(1) = -c'_{1L}(1)$, $a'_{2H}(1) = -a'_{2L}(1)$, $a'_{1H}(1) = -a'_{1L}(1)$, and $a_1(1) + a_2(1) = A(1)$, the equation can be transformed into:[16]

$$A(1) < -2A'(1). \tag{1.A.66}$$

Under Assumption 1.2, $A(1) > 0$ because both the young and middle-aged save. Then, we can derive the following condition:

$$\frac{A'(1)}{A(1)} < -\frac{1}{2}. \tag{1.A.67}$$

One can obtain the same sufficient condition even when $0 < \gamma \leq 1$. In addition, it is straightforward to find a similar sufficient condition for naive consumers following the same approach above with (1.21).

1.A.10 Derivation of Jacobian matrix for sophisticated consumers:

We can derive the total differentiation of (1.17) as:

$$\frac{\partial G}{\partial R_{t+2}} dR_{t+2} + \frac{\partial G}{\partial R_{t+1}} dR_{t+1} + \frac{\partial G}{\partial R_t} dR_t + \frac{\partial G}{\partial R_{t-1}} dR_{t-1} = 0. \tag{1.A.68}$$

Therefore, for the function $R_{t+2} = F[R_{t+1}, R_t, R_{t-1}]$, we have:

$$\begin{aligned} \frac{\partial F}{\partial R_{t+1}} &= -\frac{\partial G}{\partial R_{t+1}} \Big/ \frac{\partial G}{\partial R_{t+2}}, \\ \frac{\partial F}{\partial R_t} &= -\frac{\partial G}{\partial R_t} \Big/ \frac{\partial G}{\partial R_{t+2}}, \\ \frac{\partial F}{\partial R_{t-1}} &= -\frac{\partial G}{\partial R_{t-1}} \Big/ \frac{\partial G}{\partial R_{t+2}}. \end{aligned} \tag{1.A.69}$$

Recall that consumptions of sophisticated consumers satisfy (1.19), we further have:

$$\begin{aligned} \frac{\partial G}{\partial R_{t+2}} &= -\frac{\partial c_{1,t+1}}{\partial R_{t+2}}, \\ \frac{\partial G}{\partial R_{t+1}} &= -\frac{\partial c_{1,t+1}}{\partial R_{t+1}} - \frac{1}{\gamma} \eta_t^{\frac{1}{\gamma}-1} \frac{\partial \eta_t}{\partial R_{t+1}} R_t^{\frac{1}{\gamma}} c_{1,t} - \eta_t^{\frac{1}{\gamma}} R_t^{\frac{1}{\gamma}} \frac{\partial c_{1,t}}{\partial R_{t+1}}, \\ \frac{\partial G}{\partial R_t} &= -\frac{1}{\gamma} \eta_t^{\frac{1}{\gamma}} R_t^{\frac{1}{\gamma}-1} c_{1,t} - \eta_t^{\frac{1}{\gamma}} R_t^{\frac{1}{\gamma}} \frac{\partial c_{1,t}}{\partial R_t} - \left\{ R_{t-1}\omega_1 + \omega_2 - R_{t-1} c_{1,t-1} - \eta_{t-1}^{\frac{1}{\gamma}} R_{t-1}^{\frac{1}{\gamma}} c_{1,t-1} \right\} \\ &\quad + R_t \left\{ R_{t-1} \frac{\partial c_{1,t-1}}{\partial R_t} + \frac{1}{\gamma} \eta_{t-1}^{\frac{1}{\gamma}-1} \frac{\partial \eta_{t-1}}{\partial R_t} R_{t-1}^{\frac{1}{\gamma}} c_{1,t-1} + \eta_{t-1}^{\frac{1}{\gamma}} R_{t-1}^{\frac{1}{\gamma}} \frac{\partial c_{1,t-1}}{\partial R_t} \right\}, \end{aligned} \tag{1.A.70}$$

[16] Notice that $c_{iH}(R) = c_{iL}\left(\frac{1}{R}\right)$ for all i. Thus, $c'_{iH}(1) = -c'_{iL}(1)$ for all i. Since $a_{1s} = \omega_1 - c_{1s}$ for all s, $a'_{1H}(1) = -a'_{1L}(1)$. Likewise, it is straightforward to derive that $a'_{2H}(1) = -a'_{2L}(1)$ from $a_{2H} = \omega_2 - c_{2H} + Ra_{1H}$ and $a_{2L} = \omega_2 - c_{2L} + \frac{1}{R} a_{1L}$ with $a'_{1H}(1) = -a'_{1L}(1)$.

$$\frac{\partial G}{\partial R_{t-1}} = - R_t\omega_1 + R_t c_{1,t-1} + R_t R_{t-1}\frac{\partial c_{1,t-1}}{\partial R_{t-1}} + \frac{1}{\gamma} R_t \eta_{t-1}^{\frac{1}{\gamma}} R_{t-1}^{\frac{1}{\gamma}-1} c_{1,t-1} + R_t \eta_{t-1}^{\frac{1}{\gamma}} R_{t-1}^{\frac{1}{\gamma}} \frac{\partial c_{1,t-1}}{\partial R_{t-1}}.$$

in which

$$\begin{aligned}\frac{\partial c_{1,t}}{\partial R_{t+1}} = &- \frac{\omega_3/\left(R_t R_{t+1}^2\right)}{1+\eta_t^{\frac{1}{\gamma}} R_t^{\frac{1}{\gamma}-1} + (\beta\delta\eta_t)^{\frac{1}{\gamma}} (R_t R_{t+1})^{\frac{1}{\gamma}-1}} \\ &- \frac{\omega_1 + \omega_2/R_t + \omega_3/(R_t R_{t+1})}{\left\{1+\eta_t^{\frac{1}{\gamma}} R_t^{\frac{1}{\gamma}-1} + (\beta\delta\eta_t)^{\frac{1}{\gamma}} (R_t R_{t+1})^{\frac{1}{\gamma}-1}\right\}^2} \\ &\cdot \left\{\left(\frac{1}{\gamma}-1\right)(\beta\delta\eta_t)^{\frac{1}{\gamma}} R_t^{\frac{1}{\gamma}-1} R_{t+1}^{\frac{1}{\gamma}-2} + \frac{1}{\gamma}\eta_t^{\frac{1}{\gamma}-1}\frac{\partial \eta_t}{\partial R_{t+1}} R_t^{\frac{1}{\gamma}-1}\left(1+(\beta\delta)^{\frac{1}{\gamma}}(R_{t+1})^{\frac{1}{\gamma}-1}\right)\right\},\end{aligned} \tag{1.A.71}$$

$$\begin{aligned}\frac{\partial c_{1,t}}{\partial R_t} = &- \frac{\omega_2/R_t^2 + \omega_3/(R_t^2 R_{t+1})}{1+\eta_t^{\frac{1}{\gamma}} R_t^{\frac{1}{\gamma}-1} + (\beta\delta\eta_t)^{\frac{1}{\gamma}} (R_t R_{t+1})^{\frac{1}{\gamma}-1}} \\ &- \frac{\omega_1 + \omega_2/R_t + \omega_3/(R_t R_{t+1})}{\left\{1+\eta_t^{\frac{1}{\gamma}} R_t^{\frac{1}{\gamma}-1} + (\beta\delta\eta_t)^{\frac{1}{\gamma}} (R_t R_{t+1})^{\frac{1}{\gamma}-1}\right\}^2}\left\{\left(\frac{1}{\gamma}-1\right)\eta_t^{\frac{1}{\gamma}} R_t^{\frac{1}{\gamma}-2}\left(1+(\beta\delta)^{\frac{1}{\gamma}} R_{t+1}^{\frac{1}{\gamma}-1}\right)\right\},\end{aligned} \tag{1.A.72}$$

and

$$\frac{\partial \eta_t}{\partial R_{t+1}} = \frac{\delta(\beta\delta)^{\frac{1}{\gamma}}(1-\beta)\left(\frac{1}{\gamma}-1\right)R_{t+1}^{\frac{1}{\gamma}-2}}{\left(1+(\beta\delta)^{\frac{1}{\gamma}} R_{t+1}^{\frac{1}{\gamma}-1}\right)^2}. \tag{1.A.73}$$

Evaluated at $R_{t+1} = R_t = R_{t-1} = R^* = 1$, we can simplify the above derivatives by dropping the time subscripts and get:

$$\begin{aligned}\eta' =& \frac{(1-\beta)\left(\frac{1}{\gamma}-1\right)\beta^{\frac{1}{\gamma}}\delta^{1+\frac{1}{\gamma}}}{\left(1+(\beta\delta)^{\frac{1}{\gamma}}\right)^2}, \\ \eta =& \frac{\delta\left(\beta+(\beta\delta)^{\frac{1}{\gamma}}\right)}{1+(\beta\delta)^{\frac{1}{\gamma}}}, \\ c_1' = \frac{\partial c_{1,t}}{\partial R_{t+1}} =& - \frac{\omega_3}{\left(1+\eta^{\frac{1}{\gamma}}+(\beta\delta\eta)^{\frac{1}{\gamma}}\right)} \\ &- \frac{(\omega_1+\omega_2+\omega_3)}{\left(1+\eta^{\frac{1}{\gamma}}+(\beta\delta\eta)^{\frac{1}{\gamma}}\right)^2}\left\{\frac{1}{\gamma}\eta^{\frac{1}{\gamma}-1}\eta'\left(1+(\beta\delta)^{\frac{1}{\gamma}}\right)+\left(\frac{1}{\gamma}-1\right)(\beta\delta\eta)^{\frac{1}{\gamma}}\right\}, \\ c_0' = \frac{\partial c_{1,t}}{\partial R_t} =& - \frac{\omega_2+\omega_3}{\left(1+\eta^{\frac{1}{\gamma}}+(\beta\delta\eta)^{\frac{1}{\gamma}}\right)}\end{aligned} \tag{1.A.74}$$

$$
\begin{aligned}
&- \frac{(\omega_1 + \omega_2 + \omega_3)}{\left(1 + \eta^{\frac{1}{\gamma}} + (\beta\delta\eta)^{\frac{1}{\gamma}}\right)^2} \left\{ \left(\frac{1}{\gamma} - 1\right) \eta^{\frac{1}{\gamma}} \left(1 + (\beta\delta)^{\frac{1}{\gamma}}\right) \right\}, \\
c_1 = c_{1,t} =& \frac{\omega_1 + \omega_2 + \omega_3}{1 + \eta^{\frac{1}{\gamma}} + (\beta\delta\eta)^{\frac{1}{\gamma}}}.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\frac{\partial F}{\partial R_{t+1}}|_{R^*=1} =& - \left\{ \frac{c_0'}{c_1'} + \frac{\eta^{\frac{1}{\gamma}-1}\eta' c_1}{\gamma c_1'} + \eta^{\frac{1}{\gamma}} \right\}, \qquad (1.A.75)\\
\frac{\partial F}{\partial R_t}|_{R^*=1} =& - \left\{ \frac{\eta^{\frac{1}{\gamma}} c_1}{\gamma c_1'} + \frac{\eta^{\frac{1}{\gamma}} c_0'}{c_1'} + \frac{\omega_1 + \omega_2 - c_1 - \eta^{\frac{1}{\gamma}} c_1}{c_1'} - 1 - \frac{\eta^{\frac{1}{\gamma}-1}\eta' c_1}{\gamma c_1'} - \eta^{\frac{1}{\gamma}} \right\} \\
\frac{\partial F}{\partial R_{t-1}}|_{R^*=1} =& - \frac{\omega_1}{c_1'} + \frac{c_1}{c_1'} + \frac{c_0'}{c_1'} + \frac{\eta^{\frac{1}{\gamma}} c_1}{\gamma c_1'} + \frac{\eta^{\frac{1}{\gamma}} c_0'}{c_1'}.
\end{aligned}
$$

Note that the Jacobian matrix evaluated at $R^* = 1$ is:

$$
J[1] = \begin{bmatrix} \frac{\partial F}{\partial R_{t+1}} & \frac{\partial F}{\partial R_t} & \frac{\partial F}{\partial R_{t-1}} \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}. \qquad (1.A.76)
$$

Thus the eigenvalues of the Jacobian matrix must satisfy:

$$
\lambda^3 - \frac{\partial F}{\partial R_{t+1}}\lambda^2 - \frac{\partial F}{\partial R_t}\lambda - \frac{\partial F}{\partial R_{t-1}}|_{R^*=1} = 0. \qquad (1.A.77)
$$

1.A.11 Derivation of Jacobian matrix for naive consumers:

Note that (1.A.68) and (1.A.69) still apply here. However, since the consumptions of naive consumers satisfy (1.21), we now have:

$$
\begin{aligned}
\frac{\partial G}{\partial R_{t+2}} =& - \frac{\partial c_{1,t+1}}{\partial R_{t+2}}, \qquad (1.A.78)\\
\frac{\partial G}{\partial R_{t+1}} =& - \frac{\partial c_{1,t+1}}{\partial R_{t+1}} - \frac{(\beta\delta)^{\frac{1}{\gamma}} R_t^{\frac{1}{\gamma}}}{1 + (\beta\delta)^{\frac{1}{\gamma}} R_{t+1}^{\frac{1}{\gamma}-1}} \left\{ \frac{\delta^{\frac{1}{\gamma}} R_{t+1}^{\frac{1}{\gamma}-2} \left(\frac{1}{\gamma} - 1\right)\left(1 - \beta^{\frac{1}{\gamma}}\right) c_{1,t}}{\left(1 + (\beta\delta)^{\frac{1}{\gamma}} R_{t+1}^{\frac{1}{\gamma}-1}\right)} + \left(1 + \delta^{\frac{1}{\gamma}} R_{t+1}^{\frac{1}{\gamma}-1}\right) \frac{\partial c_{1,t}}{\partial R_{t+1}} \right\}, \\
\frac{\partial G}{\partial R_t} =& - (\beta\delta)^{\frac{1}{\gamma}} R_t^{\frac{1}{\gamma}-1} \frac{1 + \delta^{\frac{1}{\gamma}} R_{t+1}^{\frac{1}{\gamma}-1}}{1 + (\beta\delta)^{\frac{1}{\gamma}} R_{t+1}^{\frac{1}{\gamma}-1}} \left(\frac{1}{\gamma} c_{1,t} + R_t \frac{\partial c_{1,t}}{\partial R_t}\right) \\
&- \left\{ R_{t-1}\omega_1 + \omega_2 - R_{t-1}c_{1,t-1} - (\beta\delta)^{\frac{1}{\gamma}} R_{t-1}^{\frac{1}{\gamma}} \frac{1 + \delta^{\frac{1}{\gamma}} R_t^{\frac{1}{\gamma}-1}}{1 + (\beta\delta)^{\frac{1}{\gamma}} R_t^{\frac{1}{\gamma}-1}} c_{1,t-1} \right\} + R_{t-1}R_t \frac{\partial c_{1,t-1}}{\partial R_t}
\end{aligned}
$$

$$+\frac{(\beta\delta)^{\frac{1}{\gamma}}R_{t-1}^{\frac{1}{\gamma}}R_t}{1+(\beta\delta)^{\frac{1}{\gamma}}R_t^{\frac{1}{\gamma}-1}}\left\{\frac{\delta^{\frac{1}{\gamma}}R_t^{\frac{1}{\gamma}-2}\left(\frac{1}{\gamma}-1\right)\left(1-\beta^{\frac{1}{\gamma}}\right)c_{1,t-1}}{1+(\beta\delta)^{\frac{1}{\gamma}}R_t^{\frac{1}{\gamma}-1}}+\left(1+\delta^{\frac{1}{\gamma}}R_t^{\frac{1}{\gamma}-1}\right)\frac{\partial c_{1,t-1}}{\partial R_t}\right\},$$

$$\frac{\partial G}{\partial R_{t-1}}=-R_t\omega_1+R_tc_{1,t-1}+R_tR_{t-1}\frac{\partial c_{1,t-1}}{\partial R_{t-1}}+R_tR_{t-1}^{\frac{1}{\gamma}-1}(\beta\delta)^{\frac{1}{\gamma}}\frac{1+\delta^{\frac{1}{\gamma}}R_t^{\frac{1}{\gamma}-1}}{1+(\beta\delta)^{\frac{1}{\gamma}}R_t^{\frac{1}{\gamma}-1}}\left(\frac{1}{\gamma}c_{1,t-1}+R_{t-1}\frac{\partial c_{1,t-1}}{\partial R_{t-1}}\right).$$

in which

$$\begin{aligned}\frac{\partial c_{1,t}}{\partial R_{t+1}}=&-\frac{\omega_3/\left(R_tR_{t+1}^2\right)}{1+(\beta\delta)^{\frac{1}{\gamma}}R_t^{\frac{1}{\gamma}-1}+(\beta\delta^2)^{\frac{1}{\gamma}}(R_tR_{t+1})^{\frac{1}{\gamma}-1}}\\&-\frac{\left(\frac{1}{\gamma}-1\right)(\beta\delta^2)^{\frac{1}{\gamma}}\left(\omega_1+\omega_2/R_t+\omega_3/\left(R_tR_{t+1}\right)\right)R_t^{\frac{1}{\gamma}-1}R_{t+1}^{\frac{1}{\gamma}-2}}{\left\{1+(\beta\delta)^{\frac{1}{\gamma}}R_t^{\frac{1}{\gamma}-1}+(\beta\delta^2)^{\frac{1}{\gamma}}(R_tR_{t+1})^{\frac{1}{\gamma}-1}\right\}^2},\end{aligned}\tag{1.A.79}$$

and

$$\begin{aligned}\frac{\partial c_{1,t}}{\partial R_t}=&-\frac{\omega_2/R_t^2+\omega_3/\left(R_t^2R_{t+1}\right)}{1+(\beta\delta)^{\frac{1}{\gamma}}R_t^{\frac{1}{\gamma}-1}+(\beta\delta^2)^{\frac{1}{\gamma}}(R_tR_{t+1})^{\frac{1}{\gamma}-1}}\\&-\frac{\left(\frac{1}{\gamma}-1\right)(\beta\delta)^{\frac{1}{\gamma}}\left(\omega_1+\omega_2/R_t+\omega_3/\left(R_tR_{t+1}\right)\right)R_t^{\frac{1}{\gamma}-2}\left(1+\delta^{\frac{1}{\gamma}}R_{t+1}^{\frac{1}{\gamma}-1}\right)}{\left\{1+(\beta\delta)^{\frac{1}{\gamma}}R_t^{\frac{1}{\gamma}-1}+(\beta\delta^2)^{\frac{1}{\gamma}}(R_tR_{t+1})^{\frac{1}{\gamma}-1}\right\}^2}.\end{aligned}\tag{1.A.80}$$

Evaluated at $R_{t+1}=R_t=R_{t-1}=R^*=1$, we can simplify the above derivatives by dropping the time subscipts and get:

$$\begin{aligned}c_1'=&-\frac{\omega_3}{\left\{1+(\beta\delta)^{\frac{1}{\gamma}}+(\beta\delta^2)^{\frac{1}{\gamma}}\right\}}\\&-\frac{\left(\frac{1}{\gamma}-1\right)(\beta\delta^2)^{\frac{1}{\gamma}}(\omega_1+\omega_2+\omega_3)}{\left\{1+(\beta\delta)^{\frac{1}{\gamma}}+(\beta\delta^2)^{\frac{1}{\gamma}}\right\}^2},\\c_0'=&-\frac{\omega_2+\omega_3}{\left\{1+(\beta\delta)^{\frac{1}{\gamma}}+(\beta\delta^2)^{\frac{1}{\gamma}}\right\}}\\&-\frac{\left(\frac{1}{\gamma}-1\right)(\beta\delta)^{\frac{1}{\gamma}}(\omega_1+\omega_2+\omega_3)\left(1+\delta^{\frac{1}{\gamma}}\right)}{\left\{1+(\beta\delta)^{\frac{1}{\gamma}}+(\beta\delta^2)^{\frac{1}{\gamma}}\right\}^2},\\c_1=&\frac{\omega_1+\omega_2+\omega_3}{1+(\beta\delta)^{\frac{1}{\gamma}}+(\beta\delta^2)^{\frac{1}{\gamma}}}.\end{aligned}\tag{1.A.81}$$

Therefore,

$$\frac{\partial G}{\partial R_{t+2}}|_{R^*=1} = -c_1' \tag{1.A.82}$$

$$\frac{\partial G}{\partial R_{t+1}}|_{R^*=1} = -c_0' - \frac{(\beta\delta)^{\frac{1}{\gamma}}}{1+(\beta\delta)^{\frac{1}{\gamma}}}\left\{\frac{\delta^{\frac{1}{\gamma}}\left(\frac{1}{\gamma}-1\right)\left(1-\beta^{\frac{1}{\gamma}}\right)c_1}{1+(\beta\delta)^{\frac{1}{\gamma}}}+\left(1+\delta^{\frac{1}{\gamma}}\right)c_1'\right\},$$

$$\frac{\partial G}{\partial R_t}|_{R^*=1} = c_1 + \frac{(\beta\delta^2)^{\frac{1}{\gamma}}\left(\frac{1}{\gamma}-1\right)\left(1-\beta^{\frac{1}{\gamma}}\right)}{\left(1+(\beta\delta)^{\frac{1}{\gamma}}\right)^2}c_1 + \frac{(\beta\delta)^{\frac{1}{\gamma}}}{1+(\beta\delta)^{\frac{1}{\gamma}}}\left(1+\delta^{\frac{1}{\gamma}}\right)\left(1-\frac{1}{\gamma}\right)c_1$$

$$-\omega_1-\omega_2+c_1'+\frac{(\beta\delta)^{\frac{1}{\gamma}}}{1+(\beta\delta)^{\frac{1}{\gamma}}}\left(1+\delta^{\frac{1}{\gamma}}\right)(c_1'-c_0'),$$

$$\frac{\partial G}{\partial R_{t-1}}|_{R^*=1} = -\omega_1 + c_1 + c_0' + \frac{(\beta\delta)^{\frac{1}{\gamma}}}{1+(\beta\delta)^{\frac{1}{\gamma}}}\left(1+\delta^{\frac{1}{\gamma}}\right)\left(\frac{1}{\gamma}c_1+c_0'\right),$$

Thus the eigenvalues of the Jacobian matrix must satisfy

$$\lambda^3\frac{\partial G}{\partial R_{t+2}}+\lambda^2\frac{\partial G}{\partial R_{t+1}}+\lambda\frac{\partial G}{\partial R_t}+\frac{\partial G}{\partial R_{t-1}}=0. \tag{1.A.83}$$

1.A.12 Proof of Proposition 1.7:

The eigenvalues of $J[1]$ are solutions to the equation (1.A.77), which can be rearranged as:

$$\frac{\partial G}{\partial R_{t+2}}\lambda^3+\frac{\partial G}{\partial R_{t+1}}\lambda^2+\frac{\partial G}{\partial R_t}\lambda+\frac{\partial G}{\partial R_{t-1}}=0, \tag{1.A.84}$$

where the function G$(\cdot)$ is defined as in (1.17). Therefore, we have:

$$\begin{aligned}\frac{\partial G}{\partial R_{t+2}} &= \frac{\partial A_{t+1}}{\partial R_{t+2}},\\ \frac{\partial G}{\partial R_{t+1}} &= \frac{\partial A_{t+1}}{\partial R_{t+1}} - R_t\frac{\partial A_t}{\partial R_{t+1}},\\ \frac{\partial G}{\partial R_t} &= \frac{\partial A_{t+1}}{\partial R_t} - A_t - R_t\frac{\partial A_t}{\partial R_t},\\ \frac{\partial G}{\partial R_{t-1}} &= -R_t\frac{\partial A_t}{\partial R_{t-1}}.\end{aligned} \tag{1.A.85}$$

Substitute the above expressions and $\lambda = 1$ into the left hand side of the equation, we have:

$$\frac{\partial G}{\partial R_{t+2}}+\frac{\partial G}{\partial R_{t+1}}+\frac{\partial G}{\partial R_t}+\frac{\partial G}{\partial R_{t-1}} \tag{1.A.86}$$

$$
\begin{aligned}
=&\frac{\partial A_{t+1}}{\partial R_{t+2}}+\frac{\partial A_{t+1}}{\partial R_{t+1}}-R_t\frac{\partial A_t}{\partial R_{t+1}}+\frac{\partial A_{t+1}}{\partial R_t}-A_t-R_t\frac{\partial A_t}{\partial R_t}-R_t\frac{\partial A_t}{\partial R_{t-1}}\\
=&\left(\frac{\partial A_{t+1}}{\partial R_{t+2}}+\frac{\partial A_{t+1}}{\partial R_{t+1}}+\frac{\partial A_{t+1}}{\partial R_t}\right)-R_t\left(\frac{\partial A_t}{\partial R_{t+1}}+\frac{\partial A_t}{\partial R_t}+\frac{\partial A_t}{\partial R_{t-1}}\right)-A_t.
\end{aligned}
$$

At the hyperbolic monetary steady state where $R^*=1$ and $A^*=0$, we have:

$$
\begin{aligned}
&\frac{\partial G}{\partial R_{t+2}}+\frac{\partial G}{\partial R_{t+1}}+\frac{\partial G}{\partial R_t}+\frac{\partial G}{\partial R_{t-1}} \qquad (1.A.87)\\
=&\left(\frac{\partial A_{t+1}}{\partial R_{t+2}}+\frac{\partial A_{t+1}}{\partial R_{t+1}}+\frac{\partial A_{t+1}}{\partial R_t}\right)-R_t\left(\frac{\partial A_t}{\partial R_{t+1}}+\frac{\partial A_t}{\partial R_t}+\frac{\partial A_t}{\partial R_{t-1}}\right)-A_t\\
=&\left(\frac{\partial A}{\partial R_{+1}}+\frac{\partial A}{\partial R_0}+\frac{\partial A}{\partial R_{-1}}\right)|_{A=A^*}-1\times\left(\frac{\partial A}{\partial R_{+1}}+\frac{\partial A}{\partial R_0}+\frac{\partial A}{\partial R_{-1}}\right)|_{A=A^*}-0\\
=&0.
\end{aligned}
$$

Therefore, 1 is always an eigenvalue of $J[1]$ when $A^*=0$. From (1.A.87), it is straightforward to check the aggregate saving is zero if a linearized system in this model has 1 as its eigenvalue.

1.A.13 The Proof of Lemma 1.2

We have

$$
\begin{aligned}
\frac{\partial G}{\partial R_{t+2}}|_{R^*=1}&=\frac{\partial A_{t+1}}{\partial R_{t+2}}|_{R^*=1} \qquad (1.A.88)\\
&=\frac{\partial\left(a_1\left(R_{t+1},R_{t+2}\right)+a_2\left(R_t,R_{t+1}\right)\right)}{\partial R_{t+2}}|_{R^*=1}\\
&=\frac{\partial a_1\left(R_{t+1},R_{t+2}\right)}{\partial R_{t+2}}|_{R^*=1}\\
&=-\frac{\partial c_1\left(R_{t+1},R_{t+2}\right)}{\partial R_{t+2}}|_{R^*=1}
\end{aligned}
$$

Since

$$
c_1\left(R_{t+1},R_{t+2}\right)=\frac{\omega_1+\omega_2/R_{t+1}}{1+\eta\left(R_{t+2}\right)^{\frac{1}{\gamma}}R_{t+1}^{\frac{1}{\gamma}-1}+\left(\beta\delta\eta\left(R_{t+2}\right)\right)^{\frac{1}{\gamma}}\left(R_{t+1}R_{t+2}\right)^{\frac{1}{\gamma}-1}}, \qquad (1.A.89)
$$

assuming $\omega_3=0$ where the function

$$
\eta\left(R\right)=\frac{\delta\left[\beta+\left(\beta\delta\right)^{\frac{1}{\gamma}}R^{\frac{1}{\gamma}-1}\right]}{1+\left(\beta\delta\right)^{\frac{1}{\gamma}}R^{\frac{1}{\gamma}-1}} \qquad (1.A.90)
$$

is decreasing in R.

Therefore, c_1 is increasing in R_{t+2} and a_1 is decreasing in R_{t+2}, which leads to that $\frac{\partial G}{\partial R_{t+2}}|_{R^*=1}$ is negative.

1.A.14 The Proof of Lemma 1.3

From the previous proof in Appendix (1.A.9), we know the necessary and sufficient condition for cycles under the exponential discounting is equivalent to the equation (1.A.66). We can further express

$$a'_{1H}(R) = \frac{\partial a_1\left(R, \frac{1}{R}\right)}{R} = a'_{1,1}\left(R, \frac{1}{R}\right) + a'_{1,2}\left(R, \frac{1}{R}\right)\left(-\frac{1}{R^2}\right) \tag{1.A.91}$$
$$a'_{2H}(R) = \frac{\partial a_2\left(R, \frac{1}{R}\right)}{R} = a'_{2,1}\left(R, \frac{1}{R}\right) + a'_{2,2}\left(R, \frac{1}{R}\right)\left(-\frac{1}{R^2}\right)$$

By evaluating at the steady state $R = 1$, we can rewrite the necessary and sufficient condition as

$$-a'_{1,2}(1) + \left(a'_{1,1}(1) + a'_{2,2}(1) - a'_{1,2}(1)\right) - \left(a'_{2,1}(1) - a'_{1,1}(1) - a'_{2,2}(1) - A(1)\right) - a'_{2,1}(1) < 0 \tag{1.A.92}$$

Recall the characteristic equation (1.A.84) of $J[1]$ that we can define as

$$ch(\lambda) = \frac{\partial G}{\partial R_{t+2}}\lambda^3 + \frac{\partial G}{\partial R_{t+1}}\lambda^2 + \frac{\partial G}{\partial R_t}\lambda + \frac{\partial G}{\partial R_{t-1}}|_{R=1}.$$

From (1.A.85), we can also plug in the steady state $R = 1$ and get

$$\frac{\partial G}{\partial R_{t+2}} = \frac{\partial A_{t+1}}{\partial R_{t+2}}|_{R^*=1} = a'_{1,2}(1), \tag{1.A.93}$$
$$\frac{\partial G}{\partial R_{t+1}} = \frac{\partial A_{t+1}}{\partial R_{t+1}} - R_t\frac{\partial A_t}{\partial R_{t+1}}|_{R^*=1} = a'_{1,1}(1) + a'_{2,2}(1) - a'_{1,2}(1),$$
$$\frac{\partial G}{\partial R_t} = \frac{\partial A_{t+1}}{\partial R_t} - A_t - R_t\frac{\partial A_t}{\partial R_t}|_{R^*=1} = a'_{2,1}(1) - a'_{1,1}(1) - a'_{2,2}(1) - A(1),$$
$$\frac{\partial G}{\partial R_{t-1}} = -R_t\frac{\partial A_t}{\partial R_{t-1}}|_{R^*=1} = -a'_{2,1}(1).$$

Therefore, the necessary and sufficient condition in the form of (1.A.92) is equivalent to

$$ch(-1) < 0.$$

Since the cubic coefficient $a'_{1,2}(1)$ is negative as shown in Lemma 1.2, there must be a real eigenvalue λ_r satisfying $ch(\lambda_r) = 0$ such that $\lambda_r < -1$. If there are complex conjugate eigenvalues, then the existence of a real eigenvalue $\lambda_r < -1$ implies that the necessary and sufficient condition for cycles holds.

1.B Appendix: Numerical approach for finding the set of economies with two-period cycles

We proceed with the numerical exercise to characterize the set of two-period endogenous business cycles seen in 1.2 as follows. The basic idea is to identify combinations of (β, δ) with $R^* > 1$ that satisfies both the equilibrium equation (1.30) and the positive aggregate saving using a line search method that creates a vector of fine grids for $R^* > 1$.

- We first create a vector of line space from 10^{-5} to 1 with 100000 equally spaced grid points, and we take the pointwise inverse so that they are all larger than 1. Note that the $R^* > 1$ generated in this way are not equally spaced.
- We then substitute them into the equilibrium equation (1.30). If two consecutive interest rates make the excess demand flip sign, we know a solution exists in between.
- Then we verify the aggregate saving under such interest rate is positive. If both conditions are satisfied, such a point in the parameter space qualifies for a cycle.
- We do this for all combinations of (β, δ) under 0.01 increments for selected (γ, ω_1) pairs, and compare the boundaries between existence and non-existence with those suggested by the sufficient condition.

This method does not guarantee that there is no miss because there can be counter-examples that two consecutive grid points have the same sign while there is a solution in between. One way to overcome this issue is to increase the number of grid points by decreasing the grid width. We find that the set of two-period endogenous business cycles does not change when decreasing the grid width. This result implies that our numerical approach with the line size 10^{-5} is acceptable to identify the set with cycles. We also stress that this method helps prevent missing potential cycles that violate the sufficient condition. The comparison of the boundaries between existence and non-existence featured by the sufficient condition and found by the numerical procedure indicates that the errors are all within the magnitude of the grid widths in the (β, δ) plane. Therefore, we can say the necessary and sufficient condition for cycle existence coincides with the sufficient condition (1.32). Also, the sufficient condition for cycles implies positive aggregate savings.

1.C Appendix: Equilibrium characterization for naive consumers case

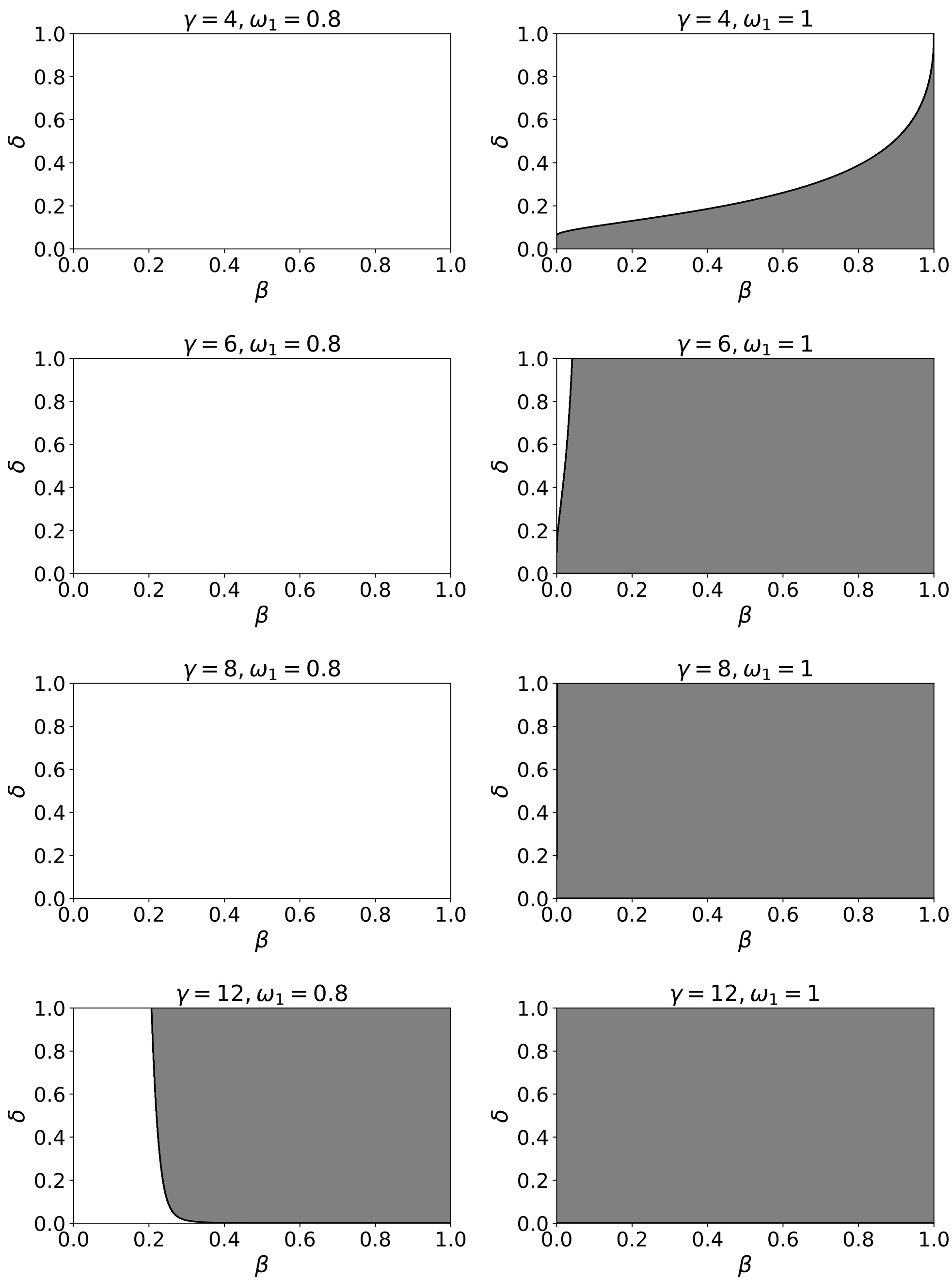


Figure 1.8: Endogenous business cycles for different parameter configurations, when consumers are naive.

Naive consumers save less when young than sophisticated consumers as seen in Figure 1.1. As a result, naive consumers are less wealthy when middle aged than their counterparts, resulting in less saving in that age as well. Therefore, the no outside money equilibrium area is relatively larger in an economy entirely populated by the naive consumers.

1.4 References

Angeletos, G.-M., Laibson, D., Repetto, A., Tobacman, J., & Weinberg, S. (2001). The hyperbolic consumption model: Calibration, simulation, and empirical evaluation. *The Journal of Economic Perspectives*, *15*(3), 47–68.

Azariadis, C. (1981). Self-fulfilling prophecies. *Journal of Economic theory*, *25*(3), 380–396.

Azariadis, C., & Guesnerie, R. (1986). Sunspots and cycles. *The review of economic studies*, *53*(5), 725–737.

Balasko, Y., & Shell, K. (1980). The overlapping-generations model, i: The case of pure exchange without money. *Journal of Economic Theory*, *23*(3), 281–306.

Benhabib, J., & Day, R. H. (1982). A characterization of erratic dynamics in, the overlapping generations model. *Journal of Economic Dynamics and Control*, *4*, 37–55.

Bhattacharya, J., & Russell, S. (2003). Two-period cycles in a three-period overlapping generations model. *Journal of Economic Theory*, *109*(2), 378–401.

Blanchard, O. J., & Kahn, C. M. (1980). The solution of linear difference models under rational expectations. *Econometrica: Journal of the Econometric Society*, 1305–1311.

Bunzel, H. (2006). Habit persistence, money, and overlapping generations. *Journal of Economic Dynamics and Control*, *30*(12), 2425–2445.

Cass, D., & Shell, K. (1983). Do sunspots matter? *Journal of political economy*, *91*(2), 193–227.

Dávila, J., Gottardi, P., & Kajii, A. (2007). Local sunspot equilibria reconsidered. *Economic Theory*, *31*(3), 401–425.

Gale, D. (1973). Pure exchange equilibrium of dynamic economic models. *Journal of Economic Theory*, *6*(1), 12–36.

Grandmont, J.-M. (1985). On endogenous competitive business cycles. *Econometrica: Journal of the Econometric Society*, 995–1045.

Guesnerie, R. (1986). Stationary sunspot equilibria in an n commodity world. *Journal of Economic Theory*, *40*(1), 103–127.

Henriksen, E., & Spear, S. (2012). Endogenous market incompleteness without market frictions: Dynamic suboptimality of competitive equilibrium in multiperiod overlapping generations economies. *Journal of Economic Theory*, *147*(2), 426–449. doi: 10.1016/j.jet.2010.09.006

Kehoe, T. J., & Levine, D. K. (1990). The economics of indeterminacy in overlapping generations models. *Journal of Public Economics*, *42*(2), 219–243.

Kydland, F. E., & Prescott, E. C. (1982). Time to build and aggregate fluctuations. *Econometrica: Journal of the Econometric Society*, 1345–1370.

Lahiri, A., & Puhakka, M. (1998). Habit persistence in overlapping generations economies under pure exchange. *Journal of Economic Theory*, *78*(1), 176–186.

Laibson, D. (1997). Golden eggs and hyperbolic discounting. *The Quarterly Journal of Economics*, *112*(2), 443–478.

Laibson, D., Repetto, A., & Tobacman, J. (2007). *Estimating discount functions with consumption choices over the lifecycle* (Tech. Rep.). National Bureau of Economic Research.

Orrego, F. (2014). Habit formation and indeterminacy in overlapping generations models. *Economic Theory*, *55*(1), 225–241.

Peck, J. (1988). On the existence of sunspot equilibria in an overlapping generations model. *Journal of Economic Theory*, *44*(1), 19–42.

Spear, S. E., Srivastava, S., & Woodford, M. (1990). Indeterminacy of stationary equilibrium in stochastic overlapping generations models. *Journal of Economic Theory*, *50*(2), 265–284.

Woodford, M. (1986). Stationary sunspot equilibria-the case of small fluctuations around a deterministic steady state.

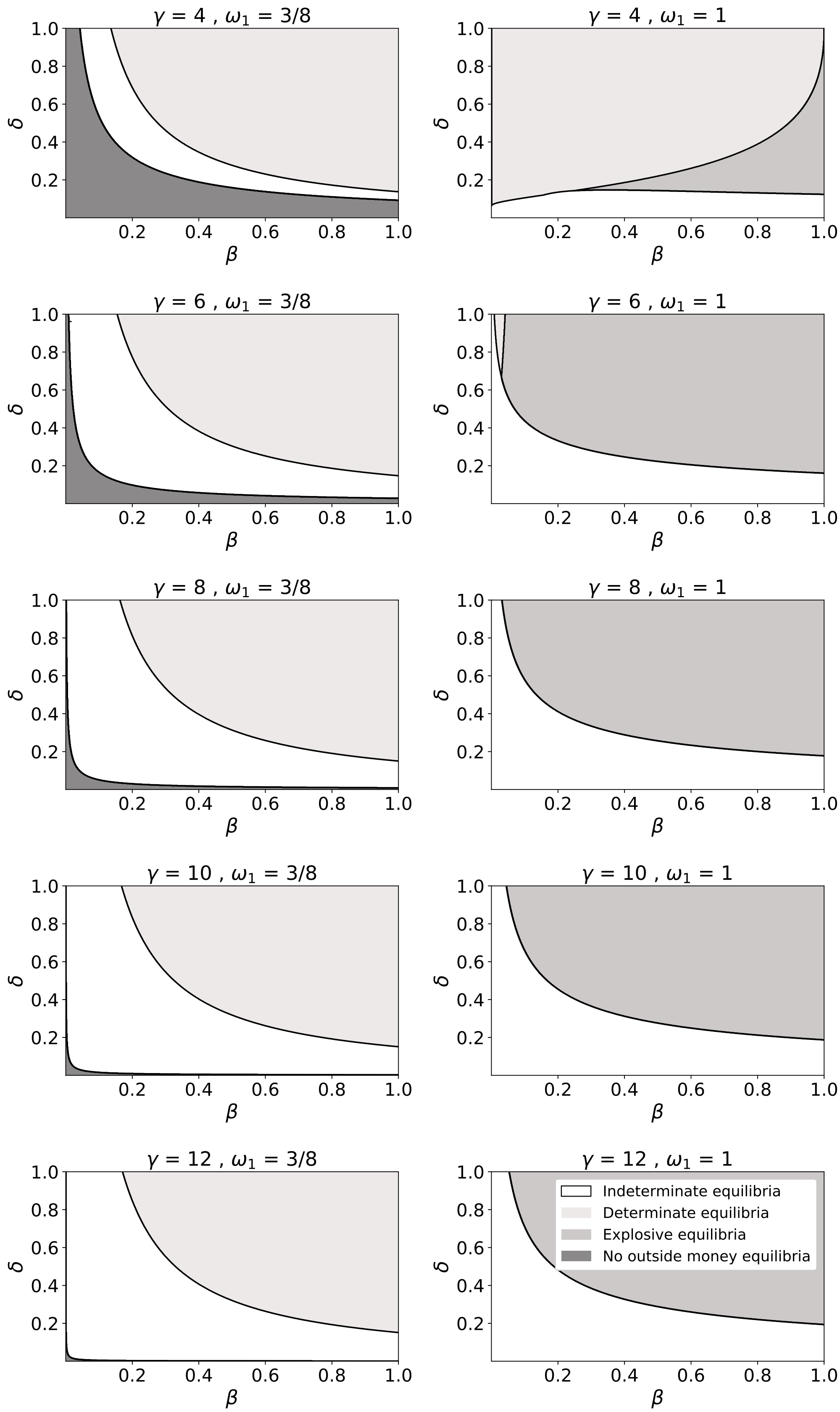


Figure 1.9: Stability properties of the monetary steady state when consumers are naive.

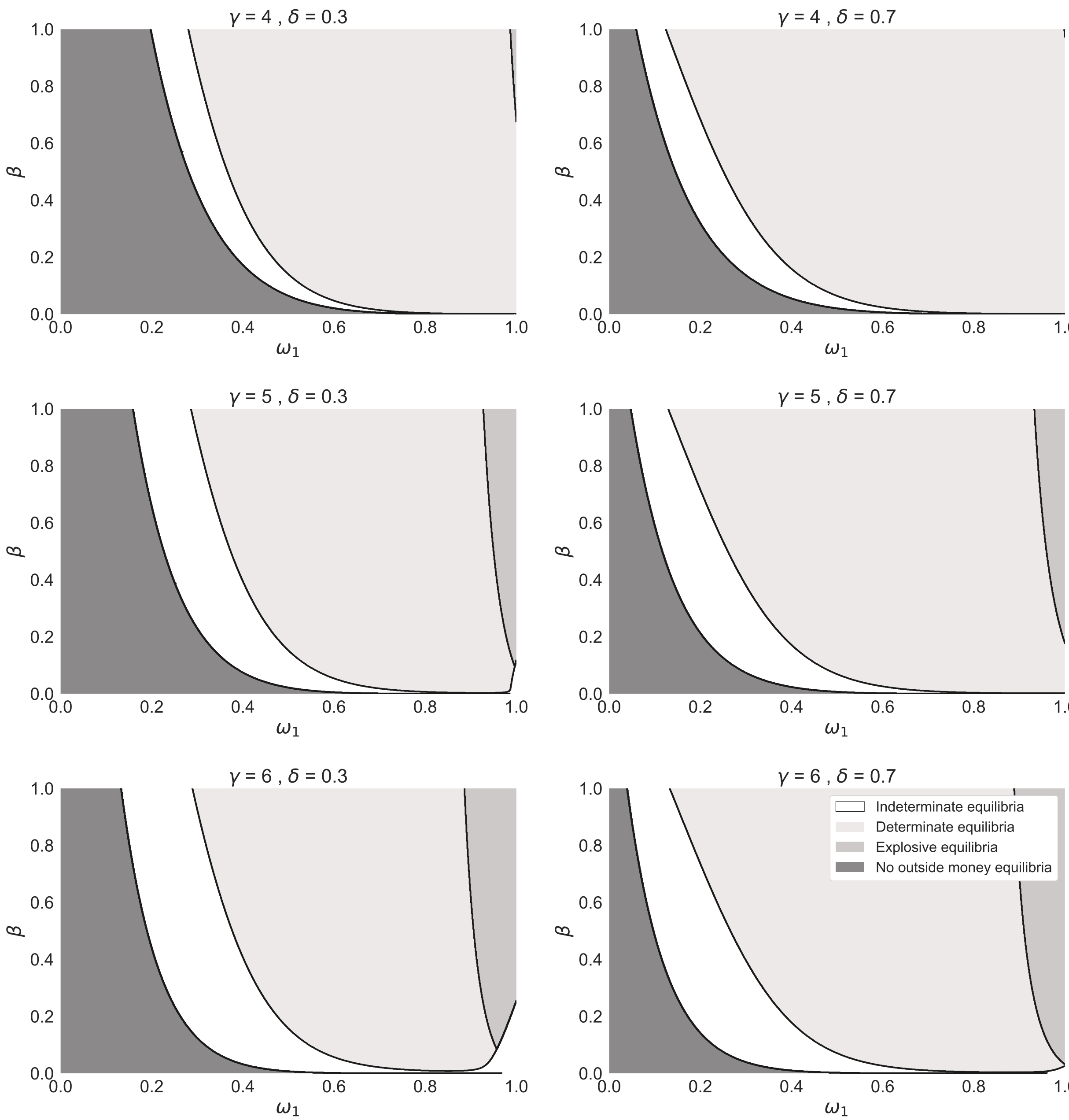


Figure 1.10: Stability properties of the monetary steady state when consumers are naive.

Chapter 2

Search Externalities, Imperfect Competition, and Labor Market Volatility in General Equilibrium

Wage and employment level do not always exhibit standard price and quantity properties as predicted in supply and demand models due to the existence of frictions in the search and matching process. While the random search model in Diamond (1982); Mortensen and Pissarides (1994); Pissarides (2000) provides a nice framework to incorporate these frictions, it abstracts away the effect of labor demand shocks on wages. Moreover, with this abstraction, no implication can be made with regard to the level of competition in the labor market. In this paper, we introduce a novel model that incorporates both search friction and imperfect competition in the labor market through a two-stage game. We find that the level of competition increases wages, unemployment, and labor market volatility. Moreover, by varying how much labor assignment depends on wage bidding versus vacancy posting, we find that the labor market becomes more volatile as the weight of wages on labor assignment increases. The effect of competition level among firms is also more significant when labor assignment is decided by wages.

2.1 Introduction

Search and matching models of the labor market (see Diamond (1982); Mortensen and Pissarides (1994); Pissarides (2000), hereafter DMP) provide a productive alternative to standard Walrasian models for explaining why workers may be involuntarily unemployed, and why short-run wage adjustments might fail to eliminate that unemployment. These models provide alternative mechanisms for wage setting that can differ markedly from the standard role of wages as the price of labor in the Walrasian model. The search friction embodied in these models reflect the costly process of a potential employer's recruiting process, which determines the equilibrium hiring strategy under productivity shocks. These models can thus generate meaningful predictions on unemployment itself, as well as the duration and volatility of unemployment.

Despite these successes, one of the major short-comings of the DMP-based labor market models has been their failure to preserve the fight for labor between firms on the intensive margin, which is done through wage bidding. In the DMP setup, most models rely on the simple Nash bargaining mechanism to determine the wage, and while this mechanism can capture aspects of differential worker versus firm market power, it has failed spectacularly when put to empirical tests (see the survey of Steinbaum (2014)). A second short-coming with these simple search and matching models stems from the mechanical way in which search proceeds, where the exogenous randomness in each firm's searching process is completely independent, without regard for the possible spillover externalities these mechanisms generate. In other words, the DMP model does not take into account the effect of one firm's recruiting activities on other firms search process, as well as on the aggregate labor market.

With the development of the personnel management technology, firms (and even workers) nowadays are delegating their search to professionals: human resource departments, headhunters, placement agencies, and career fairs. Theses services that help actively manage and overcome labor frictions have become a sector on its own. This indicates that rather than searching individually and randomly in the darkness (as in DMP), firms and workers are now searching in a more centralized pattern.

In this paper, we propose a novel labor market mechanism featuring a professional employment agency. Our model seperates the match generating technology from the wage determination process in the form of a two-stage game. In the first stage, employers post vacancies at a cost through the agency, and the agency then attract workers to its own platform using the resources from firms to overcome search frictions. In the second stage, employers meet with the workers who have been discovered by the agency and smoothly negotiate wage and offers. In such a two-stage game, a firm can attract workers on both the extensive and intensive margin. It can either invest more search effort, i.e. post more vacancies through the employment agency, which increases the visibility of the firm among unemployed workers. Alternatively, it can also offer a higher wage, which increases the willingness of workers to accept its offer as opposed to offers from other employers, conditional on being matched. We facilitate the wage determination in the second stage through a Shapley-Shubik market game model (Shapley & Shubik, 1977), which is distinctly different from the conventional Nash bargaining process in standard DMP models. We then combine it with the vacancy posting activities in the first stage, so that the new workers each firm can hire is dependent on both its vacancies and its wage bid. As a result, externalities are generated when the labor allocation among firms is not completely based on their vacancy posting activities. The effect of these externalities are affected by the competition level as well as other factors of the labor market environment. Our use of the Shapley-Shubik model also allows us to explicitly consider the effects of imperfect competition in the labor market.

We use this framework to ask and answer the following questions about the labor market. Firstly, how do firms allocate their hiring expenses in vacancy posting versus wage bidding? Secondly, how is this decision affected by the competition level within the labor market? How is the decision affected by the weight of wage versus vacancy in the labor allocaiton rule? Lastly, how is the labor market volatility in terms of wage, unemployment and job finding rate affected by these factors?

We find that under the baseline model setting, where firms attract new workers completely based on wage bidding, and vacancy posting create positive spillovers for other firms, imperfect competition leads to a reduction in wage, an increase in employment level, and a decrease in labor market volatility with respect to productivity shocks. Since the number of competing firms represent the amount of outside options of workers, wage increases with the competition level as well as aggregate productivity. Furthermore, since wages are determined in the second stage where the search and matching results are already revealed, wages are not affected by search friction parameters in the model. On the other hand, vacancy posting creates a positive externality, as it attracts workers to the aggregate recruiting agency, instead of the posting firm itself. Therefore, lower competiton allows firms to better internalize this benefit, and increases the equilibrium employment level. Finally, due to the above two effects, firms make more profit per worker in the low competition case. The higher profit margin allows them to stabilize wage and unemployment level under volatile aggregate productivity shocks by absorbing the corresponding ups and downs. On the other hand, as competition goes up, profit becomes thinner and firms have to adjust their strategies on both wage and vacancies as productivity varies. Therefore, labor market volatility is higher when there is more competition.

To directly compare our model with DMP, we study in the extension how our results change as the labor allocation rule gradually changes. As the weight on vacancy shares increases, the labor allocation rule converges to the standard DMP setting. In the extreme case, firms internalize all the benefits of their vacancy postings, and have no incentives to bid wages higher than the reservation wage. As a result, there is zero wage volatility in this case, which is actually similar to DMP since its wage volatility is completely driven by exogenously given positive bargaining power of workers. On the other hand, not only do firms internalize all the benefits of vacancy posting, they actually create a negative externality to other firms due to the diminishing return of the aggregate matching function. In equilibrium, firms post a lot of vacancies which leads to a high employment level. Since firms are making more profit by saving from the wage premiums, they are more tolerant on productivity shocks, which makes the labor market less volatile. The high employment level also narrows down the difference caused by the level of competition. Therefore, labor market volatility does not vary a lot with respect to the number of firms in the DMP setup.

Our model implies that a change in labor market volatility relative to the volatility in aggregate volatility might be caused by a change in the labor market paradigm (that determines the allocation rule) and/or a change in the level of competition. This finding supplements to the related literature that mainly focuses on other factors of the labor market. The rest of the paper is organized as follows. Section 2.2 describes the baseline model. Section 2.3 calibrates the model parameters based on empirical moments. Section 2.4 extends the baseline model. Section 2.5 concludes.

2.2 Model

2.2.1 Deterministic Case

Time is discrete. There are N ex-ante identical firms in the economy that rely on labor to produce a single good. The labor force is a continuum of workers with mass 1. The relationship between units of output y and units of input (labor) n follows the linear production function

$$y = f(n) = z \cdot n, \tag{2.2.1}$$

where $z > 0$ is the labor productivity. We denote the unit of labor of firm i at the end of period t with $n_{i,t}$, and the total employment level at the end of time t as n_t. Naturally, we have

$$n_t = \sum_{i=1}^{N} n_{i,t}. \tag{2.2.2}$$

The labor market is frictional. At the beginning of each period, s proportion of employed workers from the previous period will be separated from their jobs by nature. In other words, the mass of unemployed workers at the beginning of time t is

$$u_t = 1 - (1 - s)n_{t-1}. \tag{2.2.3}$$

Another friction of the labor market is in the hiring process. The matching of unemployed workers and hiring firms happen at a trading post, which can be interpreted as a centralized labor market. Firms can post vacancies at the labor market, which increases the visibility of the labor market, and helps attracting unemployed workers to the trading post. The unit cost of posting a vacancy is c. Denote the number of vacancies posted by firm i at time t as $v_{i,t}$, and the aggregate visibility of the labor market as v_t. Specifically, v_t is determined by the following aggregation process

$$v_t = \left(\sum_{i=1}^{N} v_{i,t}^{r} \right)^{\frac{1}{r}}, \tag{2.2.4}$$

where $r > 0$ controls the elasticity of labor substitution between firms. The amount of workers that have successfully landed in the labor market at time t is determined by

$$h_t = h\left(u_t, v_t\right). \tag{2.2.5}$$

The function $h\left(\cdot,\cdot\right)$ has the following properties.

1. $h\left(u, v\right) \leq u, \forall u, v$.
2. $h_u\left(u, v\right) \geq 0$ and $h_v\left(u, v\right) \geq 0, \forall u, v$.
3. $h_{vv}\left(u, t\right) \leq 0, \forall u, v$.

Once a certain amount of workers are attracted to the trading post, firms can attract more workers by offering a higher wage. A worker needs at least $\underline{w}$ to be willing to work, where $0 \leq \underline{w} \leq z$ is the reservation wage, or value of leisure. Denote the wage bid from firm i at time t as $w_{i,t}$, which is the total wage premium it is willing to pay above the reservation wage. The aggregate wage bid as

$$W_t = \sum_{i=1}^{N} w_{i,t}. \tag{2.2.6}$$

The number of new hires of firm i, denoted as $h_{i,t}$, is determined by

$$h_{i,t} = \frac{w_{i,t}}{W_t} h_t. \tag{2.2.7}$$

Therefore the amount of workers firm i has at the end of period t is

$$n_{i,t} = \left(1 - s\right) n_{i,t-1} + h_{i,t}. \tag{2.2.8}$$

Finally, the profit of firm i at time t can be denoted as

$$\begin{aligned} \pi_{i,t} &= \left(z - \underline{w}\right) \cdot n_{i,t} - w_{i,t} - c \cdot v_{i,t} \\ &= \left(z - \underline{w}\right)\left(1 - s\right) n_{i,t-1} + z \cdot h_{i,t} - w_{i,t} - c \cdot v_{i,t} \end{aligned} \tag{2.2.9}$$

The timeline of the model is as shown in Figure 2.1.

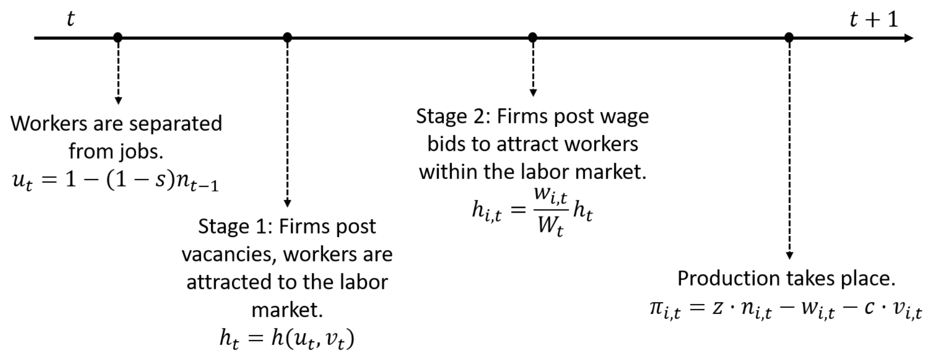


Figure 2.1: Timeline

Given the state variables at the beginning of each period $\{n_{i,t-1}\}_{i=1}^N$, denote the continuation value of firm i as

$$V_{i,t}\left(\{n_{i,t-1}\}_{i=1}^N\right) = \pi_{i,t} + \beta V_{i,t+1}\left(\{n_{i,t}\}_{i=1}^n\right). \tag{2.2.10}$$

Stage 2 Firms take h_t and $\{v_{i,t}\}_{i=1}^N$ which are determined at state 1, and other state variables $\{n_{i,t-1}\}_{i=1}^N$ as given. The stage game is resolved in a Nash equilibrium where given other firms' actions $\{w_{j,t}\}_{j\neq i}^N$, $w_{i,t}$ is the solution to the problem:

$$\max_{w_{i,t}} \pi_{i,t} + \beta V_{i,t+1}\left(\{n_{i,t}\}_{i=1}^n\right) \tag{2.2.11}$$

subject to equations (2.2.7), (2.2.8) and (2.2.9). Equilibrium is such that all wage bids are best responses of each other. Denote such strategy as $w_{i,t}\left(h_t, \{n_{i,t-1}\}_{i=1}^N\right)$.

Stage 1 Firms rationally infer the equilibrium wage bids in stage two as in $w_{i,t}\left(h_t, \{n_{i,t-1}\}_{i=1}^N\right)$. Given other firms' actions $\{v_{j,t}\}_{j\neq i}^N$, firm i chooses $v_{i,t}$ such that

$$\begin{aligned}&\max_{v_{i,t}} \pi_{i,t} + \beta V_{i,t+1}\left(\{n_{i,t}\}_{i=1}^n\right)\\ &\quad = (z-\underline{w})(1-s)\, n_{i,t-1} + (z-\underline{w}) \cdot h_{i,t} - w_{i,t}\left(h_t, \{n_{i,t-1}\}_{i=1}^N\right) - c \cdot v_{i,t} + \beta V_{i,t+1}\left(\{n_{i,t}\}_{i=1}^n\right)\end{aligned} \tag{2.2.12}$$

subject to

$$h_{i,t} = \frac{w_{i,t}\left(h_t, \{n_{i,t-1}\}_{i=1}^N\right)}{\sum_{i'} w_{i',t}\left(h_t, \{n_{i,t-1}\}_{i=1}^N\right)} h_t \tag{2.2.13}$$

$$h_t = h\left(u_t, v_t\right) \tag{2.2.14}$$

$$v_t = \left(\sum_{i'=1}^{N} v_{i',t}^r\right)^{\frac{1}{r}} \tag{2.2.15}$$

$$n_{i,t} = (1-s)\, n_{i,t-1} + h_{i,t} \tag{2.2.16}$$

Similarly, denote the Nash equilibrium strategy as $v_{i,t}\left(\{n_{i,t-1}\}\right)$.

By the format of the profit expression in equation (2.2.9), the term relevant to $n_{i,t-1}$ is separable to the terms related to decision variables $w_{i,t}$ and $v_{i,t}$. We reasonably conjecture that there exists an equilibrium where the wage bids and vacancies chosen are identical across firms regardless of the distribution of incumbent workers. Further, since the first term in equation (2.2.9) is linear in $n_{i,t-1}$, it should also be linear in the value function. Thus in our following analysis, we restrict to the type of self-consistent equilibrium described in the following proposition, the existence of which can be shown recursively.

Proposition 2.1. *There exists a symmetric equilibrium where*

$$w_{i,t}\left(h_t, \{n_{i,t-1}\}_{i=1}^{N}\right) = \frac{1}{N} W_t\left(h_t\right),\ v_{i,t}\left(\{n_{i,t-1}\}_{i=1}^{N}\right) = \frac{1}{N^{\frac{1}{r}}} v_t\left(n_{t-1}\right),\ \forall i = \{1, \cdots, N\}. \tag{2.2.17}$$

And the corresponding value function has the form

$$V_{i,t}\left(\{n_{i,t-1}\}_{i=1}^{N}\right) = V\left(n_{i,t-1}, n_{t-1}\right) = b n_{i,t-1} + f\left(n_{t-1}\right), \tag{2.2.18}$$

where b is a constant, and $f : [0,1] \to [0, \frac{z}{1-\beta}]$ is a generally nonlinear function.

Proof. See 2.A.1 □

Note that the linear term in the value function captures the direct benefit for a firm of carrying labor, while the second term captures the externality of the labor of competing firms on an individual firm. Focusing on the above type of equilibrium, we derive the policy functions of each stage as follows.

2.2.1.1 Equilibrium of Stage 2

Given the specific form of value function, the stage 2 problem for firm i can be simplified to

$$\max_{w_{i,t}} \left(z - \underline{w}\right) \frac{w_{i,t}}{W_t} h_t - w_{i,t} + \beta V\left((1-s)\, n_{i,t-1} + \frac{w_{i,t}}{W_t} h_t, (1-s)\, n_{t-1} + h_t\right) \tag{2.2.19}$$

which is further simplified to

$$\max_{w_{i,t}} \left(z - \underline{w} + \beta b\right) \frac{w_{i,t}}{W_t} h_t - w_{i,t} \tag{2.2.20}$$

subject to

$$W_t = w_{i,t} + \sum_{j \neq i} w_{j,t} \tag{2.2.21}$$

This objective function is concave in the choice variable, and therefore the optimal choice is characterized by the first order condition:

$$(z - \underline{w} + \beta b) \left(1 - \frac{w_i}{W_t}\right) \frac{h_t}{W_t} = 1. \tag{2.2.22}$$

Note that we can sum up the first order conditions of all firms $i = 1, \cdots, N$ and get

$$(z - \underline{w} + \beta b) (N - 1) \frac{h_t}{W_t} = N, \tag{2.2.23}$$

which leads us to the following lemma.

Lemma 2.1. *Given the value function with the specific format as in equation (2.2.18), and given any $h_t \in [0, u_t]$ as the outcome of stage 1. A symmetric Nash equilibrium at stage 2 exists where*

$$W_t (h_t) = \frac{N-1}{N} (z - \underline{w} + \beta b) h_t, \tag{2.2.24}$$

and

$$w_{i,t} (h_t) = \frac{N-1}{N^2} (z - \underline{w} + \beta b) h_t, \ \forall i. \tag{2.2.25}$$

2.2.1.2 Equilibrium of Stage 1

Given the symmetric equilibrium at stage 2, and the specific format of the value function, the optimization problem of firm i at stage 1 is equivalent to

$$\max_{v_{i,t}} \frac{1}{N} (z - \underline{w}) \cdot h_t - \frac{N-1}{N^2} (z - \underline{w} + \beta b) h_t - c \cdot v_{i,t} + \beta V \left((1-s) n_{i,t-1} + \frac{1}{N} h_t, (1-s) n_{t-1} + h_t \right), \tag{2.2.26}$$

and can be further simplified to

$$\max_{v_{i,t}} \frac{1}{N^2} (z - \underline{w} + \beta b) \cdot h_t - c \cdot v_{i,t} + \beta \cdot f ((1-s) n_{t-1} + h_t) \tag{2.2.27}$$

subject to

$$h_t = h (u_t, v_t) \tag{2.2.28}$$

$$v_t = \left(\sum_{i'=1}^{N} v_{i',t}^r \right)^{\frac{1}{r}} \tag{2.2.29}$$

$$v_{i,t} \geq 0 \tag{2.2.30}$$

Note that this problem is no longer guaranteed to be convex, which means the solution might not be characterized by the first order condition. However, by observing the objective function, the equilibrium is still symmetric.

Lemma 2.2. *Given the value function with the specific format as in equation (2.2.18), and given the stage 2 Nash equilibrium characterized in Lemma 2.1, a symmetric Nash equilibrium at stage 1 exists where*

$$v_t = v_t\left(n_{t-1}\right) \tag{2.2.31}$$

$$v_{i,t} = \frac{1}{N^{\frac{1}{r}}} v_t\left(n_{t-1}\right), \ \forall i \tag{2.2.32}$$

and

$$h_t = h_t\left(n_{t-1}\right) = h\left(u_t, v_t\right) = h\left(1 - \left(1 - s\right) n_{t-1}, v_t\left(n_{t-1}\right)\right). \tag{2.2.33}$$

Moreover, given $v_{j,t} = \frac{1}{N^{\frac{1}{r}}} v_t\left(n_{t-1}\right), \forall j \neq i$, $v_{i,t} = \frac{1}{N^{\frac{1}{r}}} v_t\left(n_{t-1}\right)$ is the solution to the optimization problem in (2.2.27).

Corollary 2.1. *In the special case where $r = 1$, the Nash equilibrium $v_t\left(n_{t-1}\right)$ in Lemma 2.2 is the simply the solution to*

$$\max_{v_t} \frac{1}{N^2}\left(z - \underline{w} + \beta b\right) \cdot h_t - c \cdot v_t + \beta \cdot f\left(\left(1 - s\right) n_{t-1} + h_t\right)$$

subject to

$$h_t = h\left(u_t, v_t\right).$$

In other words, all firms agree on the ideal level of aggregate v_t. Firm i would achieve v_t regardless of other firms' strategy $\{v_{j,t}\}$'s by always choosing $v_{i,t} = v_t - \sum_{j \neq i} v_{j,t}$ as best response. The symmetric equilibrium is where

$$v_{i,t} = \frac{1}{N} v_t, \ \forall i.$$

2.2.1.3 Self-consistent Equilibrium

Given the strategies in both stages, we can now express the value function of firm i as

$$\begin{aligned} V\left(n_{i,t-1}, n_{t-1}\right) =& \pi_{i,t} + \beta V\left(n_{i,t}, n_t\right) \\ =& \left(z - \underline{w}\right) \cdot n_{i,t} - w_{i,t} - c \cdot v_{i,t} + \beta V\left(n_{i,t}, n_t\right) \\ =& \left(z - \underline{w}\right)\left(1 - s\right) n_{i,t-1} + \frac{1}{N}\left(z - \underline{w}\right) h_t\left(n_{t-1}\right) - \frac{N-1}{N^2}\left(z - \underline{w} + \beta b\right) h_t\left(n_{t-1}\right) \end{aligned}$$

$$
\begin{aligned}
& - \frac{1}{N^{\frac{1}{r}}} c \cdot v_t\left(n_{t-1}\right) + \beta\left(b n_{i,t} + f\left(n_t\right)\right) \\
= & \left(z - \underline{w}\right)(1-s) n_{i,t-1} + \frac{1}{N}\left(z - \underline{w}\right) h_t\left(n_{t-1}\right) - \frac{N-1}{N^2}\left(z - \underline{w} + \beta b\right) h_t\left(n_{t-1}\right) \qquad (2.2.34) \\
& - \frac{1}{N^{\frac{1}{r}}} c \cdot v_t\left(n_{t-1}\right) + \beta\left(b\left((1-s) n_{i,t-1} + \frac{1}{N} h_t\left(n_{t-1}\right)\right) + f\left((1-s) n_{t-1} + h_t\left(n_{t-1}\right)\right)\right) \\
= & \left(z - \underline{w} + \beta b\right)(1-s) n_{i,t-1} \\
& + \left(\frac{1}{N^2}\left(z - \underline{w} + \beta b\right) h_t\left(n_{t-1}\right) - \frac{1}{N^{\frac{1}{r}}} c \cdot v_t\left(n_{t-1}\right) + \beta f\left((1-s) n_{t-1} + h_t\left(n_{t-1}\right)\right)\right)
\end{aligned}
$$

Recall that we have specified

$$
V\left(n_{i,t-1}, n_{t-1}\right) = b n_{i,t-1} + f\left(n_{t-1}\right), \forall n_{t-1} \in [0,1], n_{i,t-1} \in \left[0, n_{t-1}\right]. \qquad (2.2.35)
$$

Therefore, for the equilibrium and value function to be consistent, we have the following proposition.

Proposition 2.2. *If a value function of the form in (2.2.18) can be supported by an equilibrium, then*

$$
b = \frac{\left(z - \underline{w}\right)(1-s)}{1 - \beta(1-s)}, \qquad (2.2.36)
$$

and

$$
f\left(n_{t-1}\right) = \frac{1}{N^2} \frac{z - \underline{w}}{1 - \beta(1-s)} h_t\left(n_{t-1}\right) - \frac{1}{N^{\frac{1}{r}}} c \cdot v_t\left(n_{t-1}\right) + \beta f\left((1-s) n_{t-1} + h_t\left(n_{t-1}\right)\right), \qquad (2.2.37)
$$

where $h_t\left(n_{t-1}\right)$ and $v_t\left(n_{t-1}\right)$ are the equilibrium choices given f as characterized in Lemma 2.2.

Corollary 2.2. *If a value function of the form in (2.2.18) can be supported by an equilibrium, then*

$$
W_t\left(h_t\right) = \frac{N-1}{N} \frac{z - \underline{w}}{1 - \beta(1-s)} h_t. \qquad (2.2.38)
$$

This one-time payment scheme of wage premium is equivalent to a per-period wage scheme of

$$
wage = \underline{w} + \frac{N-1}{N}\left(z - \underline{w}\right).
$$

In other words, worker gets $\frac{N-1}{N}$ proportion of the surplus in the production process, regardless of the search frictions.

2.2.1.4 Numerical Approximation

We can numerically fit the function f through the recursive relationship in equation (2.2.37). For this process, we specify the functional form of the matching function $h(u,v)$ as the following form

$$h(u,v) = u \cdot \left(1 - e^{-\alpha v}\right), \tag{2.2.39}$$

which satisfy all conditions for h in 2.2.1. Figure 2.2 plots the approximated function f under different number of firms N.[1]

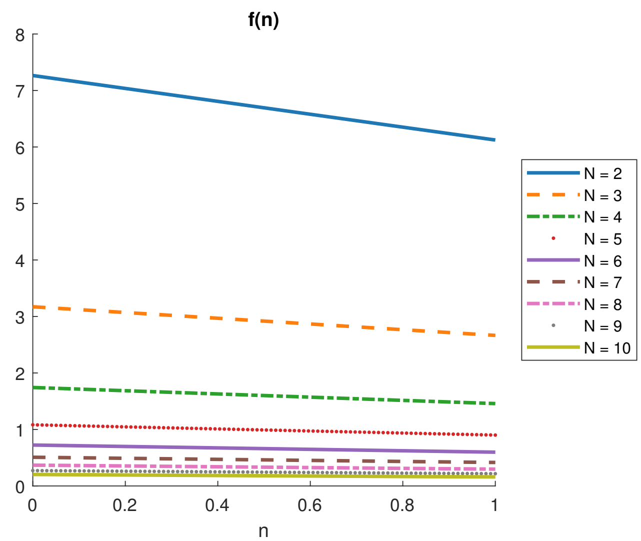


Figure 2.2: Nonlinear part of value function

There are several observations. Firstly, f is generally decreasing in n. In other words, conditional on the amount of labor in firm i, the more labor that the industry have in aggregate, the worse off firm i is. There is a negative externality from competing firms' labor stock. Secondly, f is higher when N is lower. Firms are generally better off when the industry is less competitive. Thirdly, the f function becomes flatter as N increases. In the extreme case where $N \to \infty$ and f becomes a constant zero function, firms lose the incentives to hire new labor and will use up their labor stock. This also happens when the vacancy cost is too high.

2.2.1.5 Steady State

The steady state of the dynamic model is reached if $n_t = n_{t-1}$, which is equivalent to

$$h(n_{t-1}) = s \cdot n_{t-1}. \tag{2.2.40}$$

[1] Choice of other parameters: $z = 1$, $\beta = 0.98$, $s = 0.1$, $c = 0.01$, $\alpha = 4$, $r = 0.9$, $\underline{w} = 0.4$.

We solve the steady states numerically under different parameter values. Figure 2.3 shows the wage and employment level at steady state as a function of the productivity z.[2] Wage is higher, while employment level is lower, when there are more firms on the market, holding other parameters constant. Moreover, both wage and employment level at steady state are more responsive to changes in productivity (have steeper slopes) when there are more firms. When there is only a small number of firms, productivity changes are mostly absorbed by the firms, instead of being reflected on changes in wage and employment level.

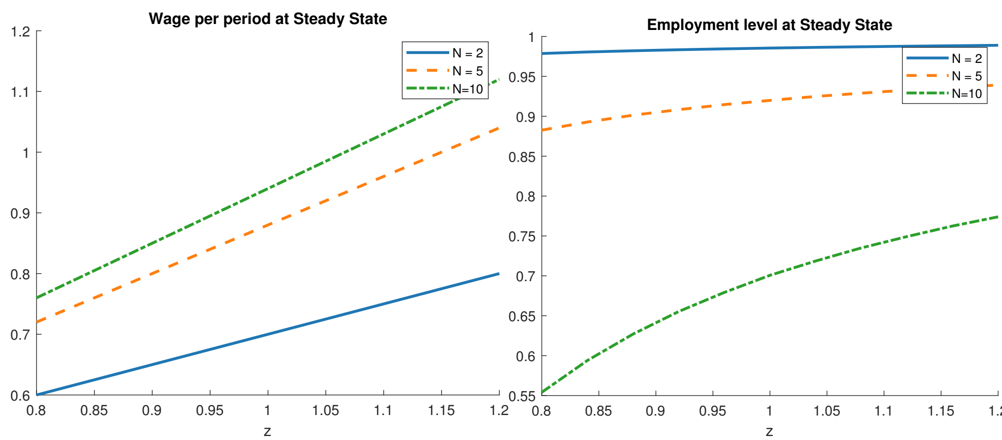


Figure 2.3: Steady state wage and employment level

2.2.2 Stochastic Case

We extend from the deterministic baseline model and assume that the productivity is uncertain over time. Specifically, productivities follow a Markov process where

$$z_{t+1} \sim F_z\left(z_t\right). \tag{2.2.41}$$

At the beginning of each period, the current productivity z_t is revealed. Given firms are risk neutral, their value functions can be expressed as

$$\begin{aligned} V_{i,t}\left(\{n_{i,t-1}\}_{i=1}^{N}, z_t\right) &= \pi_{i,t} + \beta E\left[V_{i,t+1}\left(\{n_{i,t}\}_{i=1}^{N}, z_{t+1}\right)|z_t\right] \\ &= \left(z_t - \underline{w}\right) n_{i,t} - w_{i,t} - cv_{i,t} + \beta E\left[V_{i,t+1}\left(\{n_{i,t}\}_{i=1}^{N}, z_{t+1}\right)|z_t\right], \end{aligned} \tag{2.2.42}$$

[2]Choice of other parameters: $\beta = 0.98$, $s = 0.1$, $c = 0.01$, $\alpha = 4$, $r = 0.9$, $\underline{w} = 0.4$.

where $w_{i,t}$ and $v_{i,t}$ are determined by Nash equilibrium. To simplify the problem, we focus on a special case of the Markov process which leads to a value function that has a similar format as in Proposition 2.1.

Proposition 2.3. *When the distribution of productivity shocks satisfy the linear relationship*

$$E\left[z_{t+1}|z_t\right] = \rho_0 + \rho z_t, \tag{2.2.43}$$

then the value function of firms can be expressed as

$$V\left(n_{i,t-1}, n_{t-1}, z_t\right) = b\left(z_t\right) n_{i,t-1} + f\left(n_{t-1}, z_t\right), \tag{2.2.44}$$

where the linear coefficient $b\left(z_t\right)$ has the linear form

$$b\left(z_t\right) = b_0 + b \cdot z_t, \tag{2.2.45}$$

and $f : [0,1] \times \mathbb{R} \to \mathbb{R}$ is a generally non-linear function. Moreover, this value function is supported by a symmetric equilibrium where

$$w_{i,t} = \frac{1}{N} W_t\left(h_t, z_t\right), \ \forall i \tag{2.2.46}$$

$$v_{i,t} = \frac{1}{N} v_t\left(n_{t-1}, z_t\right), \ \forall i \tag{2.2.47}$$

and

$$h_t = h_t\left(n_{t-1}, z_t\right). \tag{2.2.48}$$

We focus on the above type of equilibrium and assume that the productivities follow a mean reverting process

$$z_{t+1} = \rho z_t + (1-\rho)\bar{z} + \sigma\epsilon, \tag{2.2.49}$$

where $\rho \in [0,1]$, $\sigma \geq 0$, $\epsilon \sim N(0,1)$, and $\bar{z}$ is the unconditional mean of z_t. Under this specification, we can solve the equilibrium in each stage.

2.2.2.1 Equilibrium of Stage 2

Similar to equation (2.2.20), the problem of firm i at stage 2 can be expressed as

$$\max_{w_{i,t}} \left(z_t - \underline{w} + \beta E\left[b\left(z_{t+1}\right)|z_t\right]\right) \frac{w_{i,t}}{W_t} h_t - w_{i,t}. \tag{2.2.50}$$

Therefore, similar to Lemma 2.1, we have

$$W_t\left(h_t, z_t\right) = \frac{N-1}{N}\left(z_t - \underline{w} + \beta E\left[b\left(z_{t+1}\right)|z_t\right]\right) h_t, \tag{2.2.51}$$

$$w_{i,t}\left(h_t, z_t\right) = \frac{1}{N} W_t\left(h_t, z_t\right). \tag{2.2.52}$$

2.2.2.2 Equilibrium of Stage 1

The problem at stage 1 is

$$\max_{v_{i,t}} \frac{1}{N^2}\left(z_t - \underline{w} + \beta E\left[b\left(z_{t+1}\right)|z_t\right]\right)\cdot h_t - c\cdot v_{i,t} + \beta\cdot E\left[f\left(\left(1-s\right)n_{t-1}+h_t, z_{t+1}\right)|z_t\right] \quad (2.2.53)$$

subject to the same constraints (2.2.28-2.2.30). Similarly, a symmetric equilibrium exists for any given f and we denote the policy function as

$$v_t = v_t\left(n_{t-1}, z_t\right), \quad (2.2.54)$$

$$v_{i,t} = \frac{1}{N^{\frac{1}{r}}} v_t\left(n_{t-1}, z_t\right) \quad (2.2.55)$$

and the implied number of new hires as

$$h_t = h_t\left(n_{t-1}, z_t\right). \quad (2.2.56)$$

2.2.2.3 Self-consistent Equilibrium

We have

$$\begin{aligned} V\left(n_{i,t-1}, n_{t-1}, z_t\right) = &\left(z_t - \underline{w} + \beta E\left[b\left(z_{t+1}\right)|z_t\right]\right)\left(1-s\right)n_{i,t-1} \\ &+ \frac{1}{N^2}\left(z_t - \underline{w} + \beta E\left[b\left(z_{t+1}\right)|z_t\right]\right)h_t\left(n_{t-1}, z_t\right) - \frac{1}{N^{\frac{1}{r}}} c v_t\left(n_{t-1}, z_t\right) \\ &+ \beta\cdot E\left[f\left(\left(1-s\right)n_{t-1}+h_t, z_{t+1}\right)|z_t\right] \end{aligned} \quad (2.2.57)$$

By Proposition 2.3 and the distribution in (2.2.49), we have

$$E\left[b\left(z_{t+1}\right)|z_t\right] = b_0 + b\left(\rho z_t + \left(1-\rho\right)\bar{z}\right) \quad (2.2.58)$$

By requiring the value function to be self-consistent, we have the following proposition.

Proposition 2.4. *If the value function of the form in equations (2.2.44) and (2.2.45) under the distribution in (2.2.49) can be supported by an equilibrium, then*

$$b = \frac{1-s}{1-\beta\rho\left(1-s\right)}, \quad (2.2.59)$$

$$b_0 = \left(1-s\right)\left[\frac{\bar{z}-\underline{w}}{1-\beta\left(1-s\right)} - \frac{\bar{z}}{1-\beta\rho\left(1-s\right)}\right], \quad (2.2.60)$$

and f must satisfy that

$$\begin{aligned} f\left(n_{t-1}, z_t\right) = &\frac{1}{N^2}\left(\frac{z_t}{1-\beta\rho\left(1-s\right)} + \frac{\beta\bar{z}\left(1-s\right)\left(1-\rho\right)}{\left(1-\beta\left(1-s\right)\right)\left(1-\beta\rho\left(1-s\right)\right)} - \frac{\underline{w}}{1-\beta\left(1-s\right)}\right)h\left(n_{t-1}, z_t\right) \\ &- \frac{1}{N^{\frac{1}{r}}} c v_t\left(n_{t-1}, z_t\right) + \beta E\left[f\left(\left(1-s\right)n_{t-1} + h_t\left(n_{t-1}, z_t\right), z_{t+1}\right)|z_t\right] \end{aligned} \quad (2.2.61)$$

Corollary 2.3. *If the value function of the form in equations (2.2.44) and (2.2.45) under the distribution in (2.2.49) can be supported by an equilibrium, then*

$$W_t(h_t, z_t) = \frac{N-1}{N}\left(\frac{z_t}{1-\beta\rho(1-s)} + \frac{\beta\bar{z}(1-s)(1-\rho)}{(1-\beta(1-s))(1-\beta\rho(1-s))} - \frac{\underline{w}}{1-\beta(1-s)}\right)h_t, \tag{2.2.62}$$

This one-time payment scheme of wage premium is equivalent to a per-period wage scheme of

$$wage_t = \underline{w} + \frac{N-1}{N}(z_t - \underline{w}). \tag{2.2.63}$$

2.2.2.4 Numerical Approximation

We assume the same specific form of the matching function as in (2.2.39), and numerically approxiamte the value function through iterations. Figure 2.4 shows the 3-dimensional plot of the function $f(n, z)$.[3] As expected, f is decreasing in n and increasing in z.

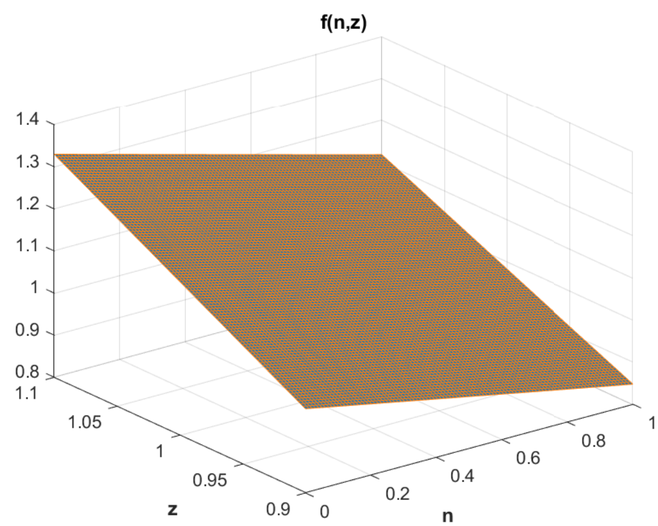


Figure 2.4: Nonlinear part of the value function (stochastic)

[3]Choice of other parameters: $\bar{z} = 1$,$N = 5$, $\beta = 0.98$, $s = 0.1$, $c = 0.01$, $\alpha = 4$, $r = 0.9$, $\underline{w} = 0.4$, $\rho = 0.996$, $\sigma = 0.01$.

2.3 Calibration

We refer to Shimer (2005) for calibration parameter and moments, but discretionarily choose the number of firms in our model. Table 2.1 shows the selected parameters and implied moments under $N = 10$.

		$N = 10$	Shimer (2005)
Common Parameters (quarterly)	β	0.98	0.9881
	s	0.1	0.1
	$\bar{z}$	1	1
	$\underline{w}$	0.4	0.4
	c	0.002	0.213
	ρ	0.996	0.996
	σ	0.01	0.0099
Other Parameters	r	1	NA
	α	3	
Moment	average job finding rate	0.4544	0.45

Table 2.1: Calibrated Parameters

We then see the implied volatility of unemployment and under each calibrated model. We run simulations for 1000 periods and remove the first 100 periods. We do this for 100 times to obtain the standard errors. All statistics are detrended by an HP filter with $\lambda = 10^5$, as is done in Shimer (2005). Table 2.2 shows the summary statistics of the simulated models. The labor market becomes significantly less volatile as we decrease the number of firms.

Note that although changing the level of competition itself will affect the volatility of unemployment, it actually changes the steady state wage and unemployment level at the same time. Therefore, once we adjust other parameter values to recalibrate the empirial moments after varying the number of firms, the volatility of unemployment does not vary a lot across different number of firms.

2.4 Extension

In the baseline model, we focused on the case where the allocation of workers is completely based on the relative ratio of wage bids. In other words, firms that bid higher get more work-

		$N=2$	$N=5$	$N=10$	Shimer (2005) Table 3	Shimer (2005) Table 1 (Real Data)
u (unemployment rate)	standard deviation	1.7×10^{-4} (5×10^{-6})	0.0010 (2.8×10^{-4})	0.0045 (0.0018)	0.009 (0.001)	0.190
	quarterly autocorrelation	0.8925 (0.0161)	0.8925 (0.0160)	0.8914 (0.0186)	0.939 (0.018)	0.936
f (job finding rate)	standard deviation	0.0017 (5×10^{-4})	0.0070 (0.0016)	0.0135 (0.0036)	0.01 (0.001)	0.118
	quarterly autocorrelation	0.8834 (0.0173)	0.8297 (0.0263)	0.5963 (0.0722)	0.878 (0.035)	0.908
z (labor productivity)	standard deviation	0.0214 (0.0015)	0.0214 (0.0015)	0.0214 (0.0015)	0.020 (0.003)	0.020
	quarterly autocorrelation	0.8924 (0.0144)	0.8924 (0.0144)	0.8924 (0.0144)	0.878 (0.035)	0.878

Table 2.2: Summary Statistics of Calibrated Model vs Real Data

ers, even if they posted fewer vacancies than others. In this section, we generalized the model and allow firms to attract marginally more workers by more actively searching. Specifically, the only modification we make on the baseline model is that unlike equation (2.2.7), we now have

$$h_{i,t}=\left[\lambda\frac{w_{i,t}}{W_t}+(1-\lambda)\frac{v_{i,t}}{\sum_{i'=1}^{N}v_{i',t}}\right]h_t, \tag{2.4.1}$$

where $\lambda\in[0,1]$. To avoid confusion, we focus on the case where $r=1$ so we can express $\frac{v_{i,t}}{\sum_{i'=1}^{N}v_{i',t}}=\frac{v_{i,t}}{v_t}$. We provide a very specific interpretation of the above allocation rule. Among all workers that are doing job searches, λ proportion of them applied to every single posting, while the remaining $(1-\lambda)$ randomly applied for only one posting. As a result, at stage 1, firms can already secure the $(1-\lambda)$ proportion of workers that applied exclusively since they do not have outside options. The remaining λ proportion of workers have not made their decisions and will wait till stage 2 to go to the firm that offers the highest wage. The parameter λ thus controls the important of wage over vacancies in attracting new workers. When $\lambda=1$, the model reduces to the baseline model in Section 2.2. On the other hand, when $\lambda=0$, the assignment of workers will completely be based on recruiting activities, which means firms have no incentives to commit to wages higher than the reservation wage. In this case, there will be no volatility in wages with respective to productivity shocks.

Proposition 2.5. *Given λ and other parameters of the model, there exists a symmetric equilibrium where each firm's value function has the form*

$$V\left(n_{i,t-1},n_{t-1},z_t\right)=\left(b_0+bz_t\right)n_{i,t-1}+f\left(n_{t-1},z_t\right), \tag{2.4.2}$$

where

$$b = \frac{1-s}{1-\beta\rho(1-s)}, \tag{2.4.3}$$

$$b_0 = (1-s)\left[\frac{\bar{z}-\underline{w}}{1-\beta(1-s)} - \frac{\bar{z}}{1-\beta\rho(1-s)}\right], \tag{2.4.4}$$

and $f : [0,1] \times \mathbb{R} \to \mathbb{R}$ *satisfies*

$$\begin{aligned} f(n,z) = {} & \left(\frac{1}{N} - \frac{N-1}{N^2}\lambda\right)\left(\frac{z_t}{1-\beta\rho(1-s)} + \frac{\beta\bar{z}(1-s)(1-\rho)}{(1-\beta(1-s))(1-\beta\rho(1-s))} - \frac{\underline{w}}{1-\beta(1-s)}\right) h_t(n,z) \\ & - c \cdot v_{i,t}(n,z) + \beta E\left[f((1-s)n + h(n,z), z_{t+1}) \,|z_t\right]. \end{aligned} \tag{2.4.5}$$

Moreover, the value function is supported by a symmetric subgame perfect equilibrium at each stage where

$$W_t(h_t) = \frac{N-1}{N}\left(\frac{z_t}{1-\beta\rho(1-s)} + \frac{\beta\bar{z}(1-s)(1-\rho)}{(1-\beta(1-s))(1-\beta\rho(1-s))} - \frac{\underline{w}}{1-\beta(1-s)}\right)\lambda h_t, \tag{2.4.6}$$

and $h_t(n_{t-1}, z_t)$ *and* $v_{i,t}(n_{t-1}, z_t)$ *come from the solution of*

$$\max_{v_{i,t}} (z_t - \underline{w}) h_{i,t} - w_{i,t} - c \cdot v_{i,t} + \beta E\left[V((1-s)n_{i,t-1} + h_{i,t}, (1-s)n_{t-1} + h_t, z_{t+1}) \,|z_t\right] \tag{2.4.7}$$

subject to

$$h_{i,t} = \left(\frac{1}{N}\lambda + \frac{v_{i,t}}{v_t}(1-\lambda)\right) h_t, \tag{2.4.8}$$

$$h_t = h(u_t, v_t), \tag{2.4.9}$$

$$v_t = \sum_{i'=1}^{N} v_{i,t}. \tag{2.4.10}$$

Corollary 2.4. *Given* λ *and an equilibrium as described in Proposition 2.5, the equilibrium wage scheme is equivalent to a per period wage of*

$$wage_t = \underline{w} + \frac{N-1}{N}\lambda(z_t - \underline{w}). \tag{2.4.11}$$

When $\lambda = 0$*, we always have*

$$wage_t = \underline{w}. \tag{2.4.12}$$

Note that the term $\frac{N-1}{N}\lambda$ is the implied bargaining power of workers. Therefore, with given N, we can use λ as an extra degree of freedom to match the wage level. The baseline model already provides a calibration of the $\lambda = 1$ case. We first provide the simulation results under the other extreme case $\lambda = 0$, and then another set of simulations where λ is calibrated. To calibrate λ, we match the implied bargaining power to 0.72, as selected in Shimer (2005). When $N = 10$, this implies $\lambda = 0.8$. In both cases, we still match the average job finding rate to be 0.45 and fix $c = 0.2$, which implies $\alpha = 0.28$ when $\lambda = 0$, and $\alpha = 1.35$ when $\lambda = 0.8$. All other parameters are the same as in Table 2.1.

Figure 2.5 and 2.6[4] first shows the average effect on the labor market by varying λ alone. Although increasing λ increases wage linearly, it decreases employment level and the job finding rate. The volatility of both unemployment and job finding rate are increasing in λ.

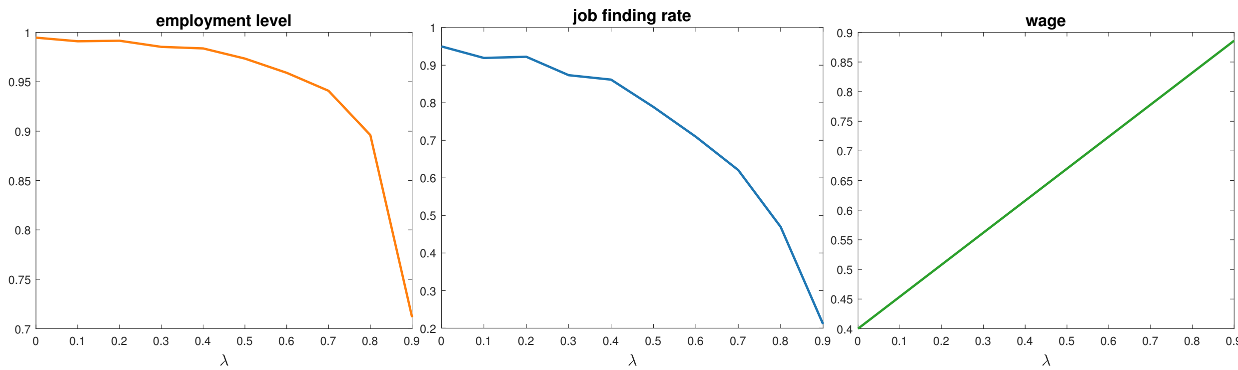


Figure 2.5: Comparative statics w.r.t. λ

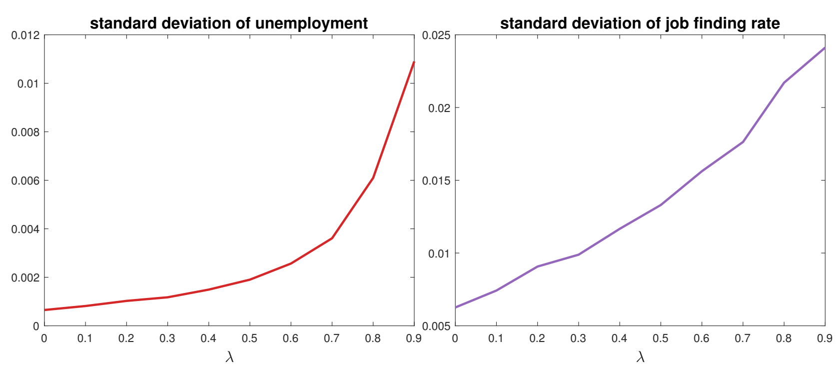


Figure 2.6: Volatility changes w.r.t λ

Table 2.3 shows the change in volatility by varying N under different calibration settings. In general, volatility is increasing in the level of competition. However, compared to the baseline case where $\lambda = 1$, the effect of N on volatility is less significant when $\lambda < 1$, as firms internalize the spillovers of posting vacancies.

[4]Choice of parameters: $N = 10$, $\lambda = 0.8$, $\alpha = 1.35$, $\bar{z} = 1$, $\beta = 0.98$, $s = 0.1$, $c = 0.2$, $r = 1$, $\underline{w} = 0.4$, $\rho = 0.996$, $\sigma = 0.01$.

		$N=2$	$N=5$	$N=10$	$N=20$
$\lambda=0$	u std	0.0056	0.0057	0.0058	0.0058
	f std	0.0165	0.02	0.0212	0.0219
$\lambda=0.8$	u std	0.0053	0.0060	0.0061	0.0062
	f std	0.0161	0.0203	0.0217	0.0225
$\lambda=1$	u std	1.7×10^{-4}	0.0010	0.0045	0.0049
	f std	0.0017	0.0070	0.0135	0.0141

Table 2.3: Volatility changes under different calibration settings

2.5 Conclusion

By combining the search friction of employment screening with the imperfect competition of the Shapley-Shubik model, the paper provides a tractable general equilibrium framework for examining the effects of labor market frictions and search externalities in a model that more closely approximates the structure of most modern industrial economies than does the simple Walrasian model of frictionless labor markets and perfect competition. We find that both imperfect competition and the labor allocation rule play important roles in determining the equilibrium dynamics of labor market moments including wage, unemployment rate, job finding rate and volatilities.

Specifically, we find that wage is increasing with the level of competition, whereas the employment level is decreasing. Moreover, the labor market volatility increases as the market becomes more competitive among firms. However, as the weight on wages in the labor allocation rule moves towards vacancy, not only does volatility decrease in general, the difference caused by different competition levels also becomes narrower. The implications from the model suggest that it is important to reconsider the current labor market paradigm before adopting a more conventional DMP model. An interesting future research direciton is to find an empirical correspondent to the weight parameter λ in the model. If it turns out that wages and the search on the intensive margin has an important weight, a more accurate measure of the competition level is needed.

On the other hand, based on the simulation we obtain from this model, there is clearly much work to be done to get the output of the calibrated model closer to the data. While the results we obtain are in the same ballpark as those obtained by Shimer, and support the conclusion that reduced competition can reduce the volatility of unemployment (consistent with the post WWII data), the overall fit of the model is not good. Some obvious issues to examine include the informational role of screening, and the possibility of additional external effects associated with screening, such as the lemons effect identified by Bolte, Immorlica, and Jackson (2020).

2.6 References

Bolte, L., Immorlica, N., & Jackson, M. O. (2020). The role of referrals in inequality, immobility, and inefficiency in labor markets. *Immobility, and Inefficiency in Labor Markets (January 1, 2020)*.

Diamond, P. A. (1982). Wage determination and efficiency in search equilibrium. *The Review of Economic Studies*, *49*(2), 217–227.

Mortensen, D. T., & Pissarides, C. A. (1994). Job creation and job destruction in the theory of unemployment. *The review of economic studies*, *61*(3), 397–415.

Pissarides, C. A. (2000). *Equilibrium unemployment theory*. MIT press.

Shapley, L., & Shubik, M. (1977). Trade using one commodity as a means of payment. *Journal of political economy*, *85*(5), 937–968.

Shimer, R. (2005). The cyclical behavior of equilibrium unemployment and vacancies. *American economic review*, *95*(1), 25–49.

Steinbaum, M. (2014). The value of search-and-matching models for the labor market. *The Washington Center for Equitable Growth*. Retrieved from https://equitablegrowth.org/value-search-matching-models-labor-market/

2.A Proofs

2.A.1 Proof of Proposition 2.1

The bound on the range of f comes from the property of the value function that $0 \leq V_{i,t} \leq \frac{z}{1-\beta}$. By Lemma 2.1, Lemma 2.2, and Proposition 2.2, it remains to show that a function $f : [0,1] \rightarrow \frac{z}{1-\beta}$ exists such that the implied policy function $h_t(n_{t-1})$ and $v_t(n_{t-1})$ from problem (2.2.27) are such that the recursive relationship of f in equation (2.2.37) is satisfied.

Define Γ as the functional operator such that

$$\Gamma(f)(n) = \frac{1}{N^2}\frac{z-\underline{w}}{1-\beta(1-s)}h(n) - \frac{1}{N^{\frac{1}{r}}}c\cdot v(n) + \beta f((1-s)n + h(n)), \forall n \in [0,1] \quad (2.A.1)$$

where $h(\cdot)$ and $v(\cdot)$ are obtained conditional on f by Lemma 2.2. The fix point of $\Gamma(\cdot)$ would satisfy the equilibrium condition. Since f is defined in a non-empty, compact space, and Γ is a continuous self map, by Kakutani's fixed point theorem there exists a fixed point f such that $\Gamma(f) = f$.

Chapter 3

Acquirer Innovation and the Venture Capital Market

Large firms develop technologies through both in-house research and development (R&D) and external acquisitions of small successful startups. The availability of potential acquisition targets is determined by past venture capital (VC) investment in startups. This paper develops a theory which connects corporate innovation with VC investment through corporate takeover activities of startups. We explore the mechanism where corporate decision makers use the level of VC investment to predict the acquisition opportunities in the near future, and make in-house R&D decisions accordingly. We show that increase in VC investment deters corporate internal R&D, and the deterrent effect is stronger for low-profit technologies. A strategic venture capitalist has more incentives to invest if corporate R&D can be more easily deterred, since it increases the demand to acquire their startups. The theory thus predicts high VC investment in technologies with lower profit than those firms invest in. The finding aligns with the empirical correlation between corporate R&D and VC investment across sectors.

3.1 Introduction

Venture capital (VC) spurs innovation in the economy by funding promising entrepreneurs and their startups. However, it is less obvious how VC investment affects innovation within large corporate firms (herein after firms), who are both competitors and potential acquirers of tech startups. This paper studies how VC investment affects R&D in large firms by affecting their acquisition opportunities. In other words, we study the "make-vs-buy" trade-off of technologies. We also solve for the optimal investment strategy of a venture capitalist in this environment. The paper provides implications for the cross-sectional and time-series observations in the VC market and corporate R&D.

The correlation between corporate R&D spending and VC investment varies across sectors. On one hand, both firms and VC are investing in "hot" markets/technologies, such as artificial intelligence, virtual reality, and autonomous cars. On the other hand, some areas funded heavily by VCs are not receiving much attention from large firms, who choose to be solely buyers of startups. Typical "VC's industries" include recreational goods, online platforms and mobile applications. For example, in 2012, Microsoft acquired Yammer, an enterprise social networking service company. In 2018, Microsoft made another acquisition of an artificial intelligence (AI) startup, Bonsai. Both targets were VC-backed. While Microsoft already had a research team in AI, it never had a competing product with Yammer. As a "tech giant", Microsoft seems to be using two different strategies when it comes to different technologies: "I won't make one on my own, but if somebody has done it, I'll buy it"[1] versus "I will try to make my own. But if somebody is doing better, I'll buy it."[2]

Typically, big firms target their R&D investment on projects with more potential and larger impact on the economy, while VC investment concentrates on technologies that are less valuable but still attractive.[3] This observation can be explained by a comparative advantage theory, where big firms and VC-backed startups are good at developing different types of technologies. However, such an explanation over-simplifies the strategic interactions between firms, startups and VCs, and therefore overlooks potential inefficiencies in the equilibrium allocation of resources. We provide a parsimonious model that focuses on the strategic interactions and generate plausible empirical implications.

We find that VC investment deters corporate in-house R&D, and the deterrent effect is stronger for low-profit technologies. Further, the VC market optimally invests in technologies with strong deterrence instead of simply pursuing high-profit technologies. The paper demonstrates that the strategic interaction between large firms and venture capitalists is crucial to generating the observed investment patterns in innovation.

The model features oligopolistic firms, venture capitalists (VCs) and entrepreneurs facing a technology (e.g., self-driving cars) with a given profit stream. A firm decides whether or not to invest in in-house R&D to develop the technology, which is risky. The entrepreneurs can found startups and develop the technology, if they are funded by a VC. A VC decides its investment level, and the total VC investment determines the size of the startup pool being funded. This then determines the firms' acquisition opportunities (probability to meet a target) in the next period. Firms who have the developed technology, either through successful R&D or acquisition, have access to the profit of the technology. Different from the conven-

[1] https://techcrunch.com/2012/07/19/microsoft-completes-its-1-2b-yammer-acquisition/

[2] Microsoft says the acquisition of Bonsai will serve to forward the kind of research the company has been pursuing. Source: https://techcrunch.com/2018/06/20/microsoft-is-buying-a-ai-startup-bonsai/

[3] The article "The Real Reason 'Stupid' Startups Raise So Much Money" provides an interesting analogy of the two types of technologies as painkillers and vitamins. Source: https://www.nirandfar.com/stupid-startups/

tional "make-or-buy" decision on whether to manufacture or purchase a certain product, the "buy" option for technologies is uncertain but predictable by current VC investment. When VC investment is high, the startup pool is large, and there are likely to be many successful startups in the next period. In this case, the probability of meeting a target is high, and the acquisition can be made at a low price. Therefore, firms have less incentives to do in-house R&D and will wait to acquire. Conversely, when VC investment is low, it is unlikely that the firm can acquire a successful startup in the next period. To increase its chance of obtaining the technology, then, the firm will invest in internal R&D.

The deterrent effect of VC investment on corporate R&D varies across technologies. Specifically, when the technology is highly profitable, firms will not risk missing out and will do in-house R&D regardless of the level of VC investment. But for relatively lower-profit technologies, firms will give up R&D more easily when observing high VC investment. From the VC's perspective, increasing investment will lead to a larger startup pool and thus more exit opportunities.[4] Moreover, higher investment has the extra benefit of deterring corporate R&D, and thus increasing the future demand for good startups. However, a large startup pool also leads to cheaper acquisition deals due to high supply. We show that the optimal level of VC investment can be non-increasing in the profit of the technology. When the profit of the technology is not too high, VC investment can effectively deter corporate R&D. In this case, VCs optimally invest a sufficient amount to keep firms from doing internal R&D. When the profit of the technology is high enough such that firms join the R&D race, VCs invest less, and the smaller startup pool serves as a back-up plan for corporate R&D failures.

The equilibrium thus appears as if large firms occupy high-profit technologies and "crowd out" VC investment to "leftover" technologies, whereas the actual mechanism behind the observed result is that VC investment is crowding out corporate R&D in these moderately profitable technologies.

In addition to explaining the investment level of VC and large firms in innovation activities, the model also provides a framework to study how the division of effort between corporate R&D and startups affects the efficiency of innovation. When startups are more productive than big firms in innovation, it is socially optimal for firms to completely outsource the innovation task to startups, which is not the equilibrium allocation for high-profit technologies, as discussed above. This implication is different from what the comparative advantage theory would suggest, where specialization is determined by productivities, and thus is always efficient. The model also preserves the inefficiency among firms due to over-investment and duplication cost from patent racing (Loury, 1979). However, this additional layer of inefficiency

[4] An "exit" of the VC refers to selling its stakes to new owners of the startup.

in specialization is due to a very different mechanism, and hence, is amenable to different kinds of remedies.

The rest of the paper is structured as follows. Section 3.2 provides more facts and summarizes related literature, as well as discusses their connections with the paper. Section 3.3 lays out the model and the general solution, and analyzes two special cases in detail. Section 3.4 extends the model with IPO exits. Section 3.5 provides some discussions and concerns of the model, and related research area. Section 3.6 concludes.

3.2 Background and Literature

It is common for large firms to acquire successful startups for the technology that they own. Technology can take various forms including patents, apps, online platforms, and even a team of talented people. In the hundreds of recently observed tech acquisitions,[5] the VC market plays a significant role in screening, funding, and monitoring the startups before they get acquired. While VC is commonly viewed as an intermediary between investors and startups, it also serves as an intermediary between acquirers and targets to some extent. In 2018, the US VC market invested a total of $136 billion in 9,845 startups. Although around 70% of the startups fail,[6] the ones that survived bring attractive returns to VCs when they exit. In this paper, startups are modeled as a continuum pool of entrepreneurs, while large firms are modeled as a finite number of discrete firms.

Acquisitions by large firms account for around 90% of exit counts, while IPO exits are much fewer in quantity but higher in value per exit. In the context of mergers and acquisitions (M&As), synergy is the driving force of business combinations. The reason that startups seek to be acquired could be attributed to the resources of the acquirers that can better commercialize the technology. In this paper, big firms are assumed to have previously developed the ability to profit from their technology as long as they have it, while startups cannot. In the extension of the model, we relax this assumption by positing that there is some probability that a successful startup that owns the technology will also obtain the profiting channel, in which case the startup can go public and be sold to the market at a higher price. Note that when all exits are through IPOs, the relationship between firms and startups is one of pure competition, which is of less interest. Therefore, in the baseline model, we focus on the case where all exits are through acquisitions.

[5]Facebook, Amazon, Microsoft, Google, and Apple have collectively made nearly 750 acquisitions. For more related examples and interesting stories, see cbinsights.com

[6]Source: CBInsights.com

VC also has a different investment pattern than corporate R&D. Ewens, Nanda, and Rhodes-Kropf (2018) documents the "spray-and-pray" investment approach, where VCs provide a little funding and limited governance to a large number of startups in various fields. In this paper, we assume that relative to the R&D department of big firms, each entrepreneur asks for a very small amount of funding, and has a relatively low probability of developing a technology successfully.

Despite being active in the M&A market, big tech firms continue to develop technologies on their own. In the year 2018, Amazon spent $28 billion on R&D.[7] The top 200 corporate R&D spenders in the US contribute over 80% of total business R&D ($368B), and over half of US national R&D ($553B).[8] Any change in innovation strategies within each of these "tech giants" can have a large impact on national and even global innovation activities.

In the literature, there has long been a debate on whether large firms or small companies are more innovative. The Schumpeterian hypothesis states that large firms in concentrated markets are more innovative because they can profit more from their new inventions, and therefore are the main generators of technological progress (Schumpeter, 1942). In fact, there is a monotonic relationship between firm size and R&D expenditure. In addition, large firms produce more patents in quantities and importance (Cohen, 2010). However, Kortum and Lerner (2000) finds evidence that VC-backed startups are more innovative than large firms in terms of patents per dolloar. From an efficiency perspective, the fact that firms are spending more on innovation while startups may be more productive in actually producing innovations implies a potential misallocation of resources. Therefore, to focus on the case where inefficiencies arise, the parameters chosen in numerical studies of the model are such that startups are more productive in innovation.

On the other hand, according to the survey by Hall and Lerner (2010), startups face higher costs of capital, which is only partially mitigated by the participation of VC. Big firms, on the other hand, can finance R&D internally. Therefore, the financial friction due to capital costs may lead to under-investment in innovation by startups. To capture this friction, the model allows for different discount factors of big firms and VCs. In the numerical analysis, the value of VC's discount factor is much lower than that of firms.

Moreover, not every area is suitable for VC intervention. In the survey of Metrick and Yasuda (2010), it is noted that VC should only invest in the type of innovation that "should" be done by small companies. In other words, VCs have the comparative advantage in innovation only in specific sectors. Nanda and Rhodes-Kropf (2013) suggests that such criterion would

[7] Labeled as "Technology and Content" on their annual report

[8] Source: Strategy&'s 2018 Global Innovation 1000 study and 2018 Global R&D Funding Forecast

select for riskier and more innovative startups in hot markets. Gompers (1995) finds that VCs concentrate investments in early stages and high technology companies. In this paper, we hold constant the innovative produtivities of firms and VCs across technologies, assume all parties are risk neutral, and incorporate no uncertainty. Even with these simplifying restrictions, however, the model predicts high VC investment in areas where big firms are not active, and in early stages of emerging markets. This finding suggests that the strategic interaction itself between big firms and the VC market likely plays an important role in determining the equilibrium allocation.

Phillips and Zhdanov (2017) finds strong positive association between VC investments and lagged M&A activity. There has also been discussion in the literature on whether there exists a connection between internal innovation and external knowledge acquisition by large firms. Bena and Li (2014), Phillips and Zhdanov (2013) and Levine (2017) find that large firms use the acquisitions of small firms as a substitute for internal innovation. Cassiman and Veugelers (2006) and Veugelers and Cassiman (1999) find that external acquisitions can be complementary to internal innovation. There is little evidence on how acquisitions affect firms' R&D decisions both ex-ante and ex-post (Hall, 1988). Pan, Huang, and Gopal (2019), Blonigen and Taylor (2000) and Krieger, Li, and Thakor (2018) study this issue in different industries and different settings.

In this paper, internally developed technology and externally acquired ones are perfectly substitutable. Different from the above research, we explicitly examine the mechanism where the amount of small firms is a choice variable of VCs, instead of exogenously given. Further, big firms and small startups differ in their innovative productivities and accesses to commercialization, instead of simply in sizes. Our result suggests that if VC investment is completely exogenous, there should be a negative correlation between firms' R&D and acquisitions, as well as a negative correlation between firms' R&D and VC investment. More importantly, if VC investment is strategically chosen, firms' R&D and VC investment (and acquisitions) should be positively correlated in general (driven by profit), with possible local negative association (driven by the deterrent effect). This result is different from that of Phillips and Zhdanov (2013) where acquisitions are perfectly procyclical.

Cunningham, Ederer, and Ma (2019) documents another potential incentive of acquirers, which is to kill the innovation within competing small firms. The possibility of killer acquisitions is assumed away in this paper through the lack of outside options of startups. If a startup is matched with a large firm but the firm chooses not to acquire it, the startup is not able to look for another buyer or keep the technology to itself. Relaxing this assumption gives even successful firms incentives to acquire startups so that other competing firms cannot access these startups. An interesting future study direction would be to model the successful development of technologies within startups and the acquisitions of startups as two separate steps, where both large firms and startups will face more complicated tradeoffs. Not only will the acquisition prices be set differently, but the occurrences of acquisitions will also be endogenously determined. In this case, acquisition is not only a channel to obtain technologies, but also a method to kill competitions.

The deterrent effect of VC investment on corporate R&D in this paper resembles the "crowding out" effect of government investment on private investment in the macroeconomics literature. Mansfield and Switzer (1984) finds through survey in the energy sector that government R&D does not simply replace but encourages private R&D. On the other hand, Wallsten (2000) finds that government R&D grants crowd out firm-financed R&D spending dollar for dollar. Mixed evidence is found in other work (Almus & Czarnitzki, 2003; Busom, 2000; González & Pazó, 2008; Popp & Newell, 2012). Nevertheless, the implication of the "crowding out" in this paper is different from that due to government interventions. When firms use either government subsidies or their own funds to conduct R&D, the innovative productivity should be the same. However, VC-backed startups have different productivities compared to corporate R&D. Further, the inefficiency brought by government crowding-out has a simpler solution than the one caused by the strategic interactions within the private sector.

According to the Schumpeterian paradigm (Schumpeter, 1942), more concentrated markets should encourage innovation since firms can better internalize the positive spillovers. Arrow (1962), on the other hand, argues that monopolies have lower incentives to innovate due to the replacement effect. Empirically, the evidence regarding this is mixed. Cohen (2010) finds no evidence that market structure could affect innovation activities. Aghion, Bloom, Blundell, Griffith, and Howitt (2005) documents an inverted-u shaped relationship between competition and innovation. David (2011), Thakor and Lo (2018), and Goettler and Gordon (2011) find mixed evidence in specific industries. In this paper, competition among firms will increase the aggregate investment in innovation due to patent racing, which follows the same intuition as in Loury (1979). This paper also specifies a framework for examining imperfect competition among venture capitalists, which is based on the Shapley-Shubik market game (Shapley & Shubik, 1977). This can be done by assuming that there are a finite number of active VCs in the market of each technology, rather than the competitive assumption of a continuum of VC firms.

3.3 Model

Time is discrete. At time $t = 0$, there exists a technology that can be developed into applications (e.g., the idea of self-driving cars). At $t = 0$, no firm possesses the developed technology, but the importance of the technology is revealed. The value of the technology is characterized by its profit stream $\{\pi_t\}_{t=0}^{\infty}$, which is based on the total consumer demand for the developed technology in each period. Moreover, we assume that any technology is short-lived, or will eventually not be profitable.

$$\lim_{t\to\infty} \pi_t = 0. \tag{3.3.1}$$

There are N identical firms , with discount factor β_F, in the market that are able to profit from the technology. In any period t, if there are k_t firms that currently own the developed technology, each of them will share the profit evenly and get $\frac{\pi_t}{k_t}$. Firms have innovative abilities for developing this technology. If a firm decides to develop the technology on its own, it has to invest a cost c in its R&D process. With a probability of p, the R&D will be successful. The outcome of R&D will be revealed at the beginning of the following period. Conditional on success, the firm will own the developed technology for the next period and thereafter. The R&D decision of firm $i \in \{1, \cdots, N\}$ can be described as $q_{i,t} \in [0, 1]$, which is the probability of investing in R&D at time t. Note that successful firms that already own the technology no longer need to make any R&D investment. In fact, they need not make any decisions regarding the technology any more and can simply continue to operate profitably.

There are M identical institutional investors (VCs), with discount factor β_V, that can fund entrepreneurs. By funding a startup (providing the amount necessary for an entrepreneur to launch a startup), the VC becomes a shareholder in the startup and owns α proportion of the startup (the remaining shares are held by the entrepreneur himself). When the startup is acquired, the VC will get α proportion of the proceeds from the acquisition.

There is also a potential pool of entrepreneurs in the market with the innovative skills to develop the technology. They can only start innovating if they are funded. Each entrepreneur needs a small amount of funding to launch a startup and start developing the technology. Unlike big firms, startups do not have access to the profit of the technology even if they succeed in developing it. This might be due to that big firms have better sales team, marketing team, and brand name recognition. Therefore, successful startups will seek to be acquired by big firms who still have not obtained the developed technology. The combined firm can then profit from the technology. The source of the synergy in this type of acquisitions is generated by the combination of the developed technology owned by the startups and the profiting channel owned by the big firms. Section 3.4 extends the model so that in each period, there is a chance that a successful startup will obtain the profiting channel and go for an IPO. In the baseline model, all exits are through acquisitions.

Successful startups are revealed at the beginning of the next period, and have the chance to meet with corporate acquirers. Let λ_t denote the acquisition opportunity (probability of meeting a successful startup) for a big firm at time t. Let I_t denote the size of the funded startup pool (or total capital invested into the pool) at time t. Each VC's investment jointly determines the size of the startup pool:

$$I_t = \sum_{i=0}^{M} I_{i,t}, \tag{3.3.2}$$

Each VC i has stakes in $\frac{I_{i,t}}{I_t}$ proportion of the startups in the pool. For simplicity, when an acquisition happens, we do not distinguish which VC is the actual investor of the target. Instead, each VC will receive an $\alpha\frac{I_{i,t}}{I_t}$ proportion of the payment. The relationship between the size of the pool and the acquisition opportunity is characterized by the mapping $g(\cdot)$:[9]

$$\lambda_{t+1} = g(I_t), \tag{3.3.3}$$

We assume the function $g(\cdot)$ defined on $[0,\infty)$ has the following functional form:

$$g(I) = 1 - \text{exp}(-\frac{I}{\phi N}), \tag{3.3.4}$$

where $\phi > 0$.

The above function $g(\cdot)$ satisfies the following properties. Firstly, $g' > 0$, acquisition opportunity is increasing in the size of the startup pool. The more startups you fund, the more are likely to be successful. Secondly, $g'' < 0$, there is diminishing return to the capital invested in startups. In other words, the more startups you fund, the lower the average likelihood of success. Thirdly, $g(0) = 0$ and $\lim_{I\to\infty} g(I) = 1$. Zero investment means definitive failure, and it is infinitely costly to guarantee success. Conditional on the specification of g, there is a one-to-one mapping between I_t and λ_{t+1} through equation (3.3.3). In other words, by observing the level of VC investment today, firms are able to predict their acquisition opportunities in the next period.

The timeline of the model is as follows. (1) At the beginning of a period t, the R&D outcome from the previous period is revealed privately within each firm. (2) Then, successful startups

[9] The "reduced-form" variable λ_t captures both the success probability of individual startups, as well as the matching/bidding process in the M&A market. Ceteris paribus, λ_{t+1} is only affected by I_t. If necessary, the model can be expanded such that the number of successful startups is directly determined by the level of VC investment. Then with finite potential acquirers and targets, acquisitions can happen according to a Nash-in-Nash bargaining process (Collard-Wexler, Gowrisankaran, & Lee, 2019), instead of the current setup with probabilistic matching and bilateral Nash bargaining.

begin meeting with acquirers. Acquisitions happen simultaneously. (3) Firms who currently own the technology start doing business and profiting from the technology, thus the number of successful firms k_t is revealed. (4) VCs simultaneously invest in the startup pool, and the size of the pool is observed by firms. (5) Firms simultaneously decide R&D spending, and the decisions are private.

Conditional on the profit stream and all other parameters, the number of currently successful firms k_t is the payoff-relevant state variable for decision making. For both firms and VC, their expected future payoffs are affected by historical actions only through the state variable k_t.

At each period t, firms are involved in the following costs and payoffs. First, a firm without the technology that meets with an acquisition target pays the cost of acquisition. Second, a firm with the technology receives the profit of this period $\frac{\pi_t}{k_t}$. Lastly, firm i pays the cost $q_{i,t}c$ for its R&D decision. Similarly, the cost and payoffs of a VC at time t is as follows. A VC receives its share of acquisition proceeds, and spends its investment in the startup pool. At the time of decision-making, firms and VCs know the current state variable k_t as well as the profit stream $\{\pi_{t'}\}_{t'=t}^{\infty}$. Therefore, firms and VCs make decisions by forming expectations on the future distribution of number of successful firms k_{t+1} (and implicitly k_{t+2} and onward) based on the binary distributions given the success probability and the acquisition opportunity, as well as the endogenous decision of other firms and VCs.

The cost of acquisition is determined by the following Nash Bargaining scheme. The synergy generated by the merger is the difference between the expected continuation value of a firm conditional on obtaining the technology and that on being unlucky. The acquirer will pay the target η proportion of this surplus, where η denotes the bargaining power of an individual startup.[10] In the model, it is possible that a firm who succeeded in R&D could still match with a startup. In this case, the surplus generated is zero and thus acquisition never happens or happens at zero cost. In other words, although the startup has successfully made the technology, what it made is completely redundant and hence brings no value to any potential matches. Therefore, the actual chance to exit for VCs decreases as there are more successful firms, even holding VC investment constant.

[10] Although η is treated as a constant parameter, the acquisition price is not constant with respect to number of competitors. When there are more incumbent successful firms, there are fewer acquirers. In this case, the surplus generated by acquisitions is lower, since the profit needs to be shared with the incumbent firms. Therefore, acquisition price increases with number of acquirers, although not through the channel of increasing startups' bargaining power. When the competition on the supply side increases (when λ_t is high), more acquisitions are likely to happen. This also reduces the surplus since the profit is expected to be shared among more successful firms. Therefore, acquisition price is decreasing in the number of (successful) startups. The model can also be modified such that η is varying with the competition among firms and startups.

3.3.1 Equilibrium

This section solves the above model by restricting to a symmetric equilibrium, i.e. all firms choose the same R&D decision, and all VCs choose the same investment level:

$$q_{i,t} = q_t, \forall i \in \{1, \cdots, N\} \tag{3.3.5}$$

$$I_{i,t} = \frac{1}{M} I_t, \forall i \in \{1, \cdots, M\}. \tag{3.3.6}$$

Denote k_t as the number of firms with the technology at time t.

3.3.1.1 The Firms' Problem

We first explore the partial equilibrium and study the firms' policy function under an exogenous VC investment strategy, i.e. VCs choose and commit to a deterministic investment strategy for a specific technology at the time when it emerges. By equation (3.3.3), it is equivalent that VCs commit to a strategy of acquisition opportunities from firms' perspectives. In other words, there exists a function $\Lambda()$ that is known to all firms at $t = 0$ such that

$$\lambda_t = \Lambda(k_{t-1} | \{\pi_{t'}\}_{t'=t}^{\infty}) \tag{3.3.7}$$

Note that constant λ is a special case that satisfies this assumption. This assumption is unnecessary when the profit stream is one-period: $\pi_t = 0, \forall t > 1$, since λ_1 is observed at $t = 0$ already. In Section 3.3.1.2, we will solve for the optimal policy of VCs so that their return is maximized, which is also a special case of Λ.

Define

$$\Pi_t = \Pi(k_t | \{\pi_{t'}\}_{t'=t}^{\infty}, \Lambda) \tag{3.3.8}$$

as the expected discounted total profit for a winner firm at time t, given the current number of winners, the profit stream, and the VC investment stream. Similarly, define

$$V_t = V(k_t | \{\pi_{t'}\}_{t'=t}^{\infty}, \Lambda) \tag{3.3.9}$$

as the value function of a firm without the technology at time t. The acquisition cost at time t is defined as

$$a_t = a(k_{t-1} | \{\pi_{t'}\}_{t'=t}^{\infty}, \Lambda). \tag{3.3.10}$$

Note that a is a function of the states of the previous period because the current states are not observed until after all acquisitions happen. Finally, define

$$q_t = q(k_t | \{\pi_{t'}\}_{t'=t}^{\infty}, \Lambda) \tag{3.3.11}$$

as the policy function, i.e. the probability to invest in R&D, for a firm without the technology at time t.

With q_t, the state transition of k_t and k_t^-, denoted as $S(k_{t+1}, k_{t+1}^- | k_t, k_t^-)$ is well defined.

$$S(k_t + j | k_t) \quad = \quad \binom{N - k_t}{j} h_t^j (1 - h_t)^{N - k_t - j}, \tag{3.3.12}$$

$$\tag{3.3.13}$$

for $j \in \{0, \cdots, N - k_t\}$, where

$$h_t = q_t p + (1 - q_t p)\lambda_{t+1}. \tag{3.3.14}$$

For simplicity, for a variable $x_t = x(k_t)$, denote

$$E_t[x_{t+1}] = \sum_{j=0}^{N-k_t} S(k_t + j | k_t) x(k_t + j) \tag{3.3.15}$$

Next, the conditional state transition process where firm i will gain the technology is given by

$$S^i(k_t + j + 1 | k_t) \quad = \quad \binom{N - k_t - 1}{j} h_t^j (1 - h_t)^{N - k_t - 1 - j}, \tag{3.3.16}$$

$$\tag{3.3.17}$$

where $j \in \{0, \cdots, N - k_t - 1\}$. Similarly, define a state transition process conditional on firm i not gaining the technology as

$$S^{-i}(k_t + j | k_t) \quad = \quad \binom{N - k_t - 1}{j} h_t^j (1 - h_t)^{N - k_t - 1 - j}, \tag{3.3.18}$$

where $j \in \{0, \cdots, N - k_t - 1\}$. Denote the related expected values by

$$E_t^i[x_{t+1}] \quad = \quad \sum_{j=0}^{N-k_t-1} S^i(k_t + j + 1 | k_t) x(k_t + j + 1) \tag{3.3.19}$$

$$E_t^{-i}[x_{t+1}] \quad = \quad \sum_{j=0}^{N-k_t-1} S^{-i}(k_t + j | k_t) x(k_t + j) \tag{3.3.20}$$

$$\tag{3.3.21}$$

We can now recursively define both value functions:

$$\Pi_t(k_t) \quad = \quad \frac{\pi_t}{k_t} + \beta_F E_t[\Pi_{t+1}] \tag{3.3.22}$$

$$\begin{aligned} V_t(k_t) &= -q_t c - \beta_F(1-pq_t)\lambda_{t+1}a_{t+1} \\ & \quad +\beta_F(h_t E_t^i[\Pi_{t+1}] + (1-h_t)E_t^{-i}[V_{t+1}]) \end{aligned} \tag{3.3.23}$$

By the Nash Bargaining assumption in the acquisition process, we have

$$a_t = \eta(E_{t-1}^i[\Pi_t] - E_{t-1}^{-i}[V_t]) \tag{3.3.24}$$

Proposition 3.1. *(Symmetric Nash equilibrium among firms given VC's strategy) Given VC's startegy Λ, and by restricting to a symmetric equilibrium across firms, the policy function q has the following form:*

$$q_{i,t} = q_t = q(k_t|\{\pi_{t'}\}_{t'=t}^{\infty}, \Lambda) = \begin{cases} 0 & \text{if } D(0) < c \\ 1 & \text{if } D(1) > c \\ q^*(k_t) & \text{otherwise} \end{cases} \tag{3.3.25}$$

where q^* is the solution to $D(q) = c$, and

$$\begin{aligned} D(q) &= D(q|\{\pi_{t'}\}_{t'=t}^{\infty}, \Lambda) \\ &= \beta_F p((1-\lambda_{t+1} + \eta\lambda_{t+1})(E_{t,q}^i[\Pi_{t+1}] - E_{t,q}^{-i}[V_{t+1}])) \end{aligned} \tag{3.3.26}$$

is the net payoff from R&D of an individual firm given other firms' R&D strategy q. And Π and V are defined as in equations (3.3.22) and (3.3.23).

Proof. See 3.A.1.1 for the proof and the derivation of $D(q)$. □

The model can now be solved numerically using backward induction given any $\{\pi_t\}_{t=0}^{\infty}$ and Λ, and other parameters $\{\beta_F, p, c, N, \eta\}$[11]. 3.A.2.1 illustrates the procedures of solving the problem numerically.

3.3.1.2 The VCs' Problem

We then discuss the partial equilibrium among VCs given firms' strategy Q, i.e., firms choose their R&D investment according to an exogenous function.

$$q_t = Q(k_t|\{\pi_{t'}\}_{t'=t}^{\infty}, \Lambda) \tag{3.3.27}$$

The next section discusses the general equilibrium where $Q(\cdot)$ is the actual optimal policy function of firms $q(\cdot)$. In this section, firms' strategy does not have to be optimized. For example, firms never do R&D ($Q \equiv 0$) is a valid instance of $Q(\cdot)$.

[11] α and ϕ are irrelevant to the firms' problem given Λ, but will be relevant in the VCs' problem

Define

$$R_t = R(k_t|\{\pi_{t'}\}_{t'=t}^{\infty}, Q) \tag{3.3.28}$$

as the expected total discounted net gain of VCs as a whole at time t. Let l_t be the number of acquisitions occurring at time t. Then the expected number of acquisition in the next period can be characterized by the following expression.

$$\begin{aligned} E_t^{\ell}[l_{t+1}] &= \sum_{j=0}^{N-k_t} \binom{N-k_t}{j} (q_t p)^j (1-q_t p)^{N-k_t-j} \sum_{l=0}^{N-k_t-j} \binom{N-k_t-j}{l} \lambda_{t+1}^l (1-\lambda_{t+1})^{N-k_t-j-l} l \\ &= \sum_{j=0}^{N-k_t} \binom{N-k_t}{j} (q_t p)^j (1-q_t p)^{N-k_t-j} (N-k_t-j)\lambda_{t+1} \\ &= \lambda_{t+1}(N-k_t)(1-q_t p) \\ &= \lambda_{t+1}(N-k_t)(1-Q(k_t)p) \end{aligned} \tag{3.3.29}$$

We have the recursive relationship

$$R(k_t) = -I_t + \beta_V(\alpha E_t^{\ell}[l_{t+1}]a_{t+1} + E_t[R(k_{t+1})]). \tag{3.3.30}$$

Given the belief that all VC's will play the symmetric equilibrium from $t+1$ onward, an individual VC i's expected continuation value at $t+1$ is $\frac{1}{M}R_{t+1}$. At time t, the Nash equilibrium also involves all other VC's investment

$$I_{-i,t} = \sum_{j \neq i} I_{j,t}. \tag{3.3.31}$$

We can then write VC i's problem as

$$\max_{I_{i,t}} R_t^i(I_{i,t}|I_{-i,t}) = -I_{i,t} + \beta_V\left(\frac{I_{i,t}}{I_{i,t}+I_{-i,t}}\alpha E_t^{\ell}[l_{t+1}|\lambda_{t+1}]a_{t+1} + \frac{1}{M}E_t[R_{t+1}|\lambda_{t+1}]\right) \tag{3.3.32}$$

subject to

$$\lambda_{t+1} = g(I_{i,t} + I_{-i,t}). \tag{3.3.33}$$

The closed-form solution to the above problem is not available, but we show the existence of a solution.

Proposition 3.2. *(Symmetric Nash equilibrium among VCs given firms' strategy) Given firms' strategy Q, there exists a symmetric equilibrium*

$$\lambda_{t+1}(k_t) = \lambda(k_t|\{\pi_{t'}\}_{t'=t}^{\infty}, Q) \tag{3.3.34}$$

among VCs such that

1. $I_{i,t}(k_t) = \frac{1}{M} I_t(k_t), \forall i, t$

2. $\lambda_{t+1}(k_t) = g(I_t(k_t))$

3. $I_{i,t}(k_t)$ *is the solution to the problem in (3.3.32) with* $I_{-i,t} = (M-1)I_{i,t}(k_t)$

4. $R(k_t)$ *is defined as in equation (3.3.30)*

Proof. See 3.A.1.2. □

Then we can numerically solve for the best response of VC i, and thus the symmetric equilibrium $\{I_t, \lambda_{t+1}\}$ at a given state k_t. Section 3.A.2.2 illustrates the procedures of solving $\lambda(\cdot)$ numerically by backward induction.

3.3.1.3 Markov Perfect Equilibrium

Based on the partial equilibria solved above, this section describes the general equilibrium of the market under the concept of Markov Perfect Equilibrium.

Definition 3.1. *Given model parameters* $\{N, M, \beta_F, \beta_V, c, p, \alpha, \eta, \phi\}$*, an equilibrium of the innovation market facing the technology* $\{\pi_t\}_{t=0}^{\infty}$ *is a 7-tuple*

$$[q_t(k_t), \lambda_{t+1}(k_t), I_t(k_t), \Pi_t(k_t), V_t(k_t), R_t(k_t), S(k_{t+1}|k_t)] \tag{3.3.35}$$

where:

1. $q_t(k_t)$ *solves the firms' problem as in Proposition 3.1 given* $\Lambda = \lambda$*.*

2. $\Pi_t(k_t)$ *and* $V_t(k_t)$ *captures the continuation value of firms given optimal R&D decision as in equations (3.3.22) and (3.3.23).*

3. $\lambda_{t+1}(k_t)$ *and* $I_t(k_t)$ *are the optimal VC's decision as in Proposition 3.2 given that at any specific choice of* λ, $Q = q_t(k_t, \lambda)$*.*

4. $R_t(k_t)$ *is the continuation value of VCs in aggregate as in equation (3.3.30).*

5. *The state variable* k_t *evolves according to the state transition probability* $S(k_{t+1}|k_t)$ *as in equation (3.3.12).*

The existence of the equilibrium follows from the existence of the VCs' solution given they take into account how their decisions affect firms as in condition 3. The proof is similar to that of Proposition 3.2, which relies on the property that the VC's problem can be converted to a bounded problem. We focus on the Markovian equilibrium where all strategies only depends on payoff-relevant states rather than other historical actions. This is appropriate in our setting since both technologies and venture capitalists are short-lived. However, in a highly concentrated market with long-lived players, it is worth considering potential collusion and other types of equilibria. Even under the Markovian restriction, the uniqueness of equilibrium is not guaranteed (see discussions in 3.A.1.1), which leaves room for further discussion about the possible industry dynamics regarding the development of a technology.

3.3.2 Efficiency

The efficient level of VC investment (I_t^e) and corporate R&D effort (q_t^e) can be solved from a benevolent planner's problem. Let the planner have discount factor β. If the technology is successfully developed at t, the planner receives π_t as payoff, otherwise the planner receives 0. At the same time, the planner internalizes all cost spent on the attempt to develop the technology ($I_t + (N - k_t)q_t c$). One immediate decision of the planner is that no investment will be made if at least one firm possesses the technology, i.e.

$$q_t^e = I_t^e = 0, \text{ if } k_t > 0. \tag{3.3.36}$$

In other words, any follow-up innovation activities by competitor firms or new startups in the general equilibrium are inefficient. These activities do not create additional social value, but incur a positive social cost. This is one form of over-investment due to "duplicated effort", and can be protected by exclusive patenting.

The continuation value of the planner conditional on success is straightforward:

$$\Pi_t^e = \sum_{t'=t}^{\infty} \beta^{t'-t} \pi_{t'}. \tag{3.3.37}$$

And the continuation value of the planner is therefore expressed as:

$$V_t^e = \max_{I_t, q_t} -I_t - Nq_t c + \beta((1 - (1 - h_{t+1})^N)\Pi_{t+1}^e + (1 - h_{t+1})^N V_{t+1}^e), \tag{3.3.38}$$

subject to

$$h_{t+1} = q_t p + (1 - q_t p)\lambda_{t+1} \tag{3.3.39}$$

$$\lambda_{t+1} = g(I_t) \tag{3.3.40}$$

The planner needs to assign innovation tasks to firms and startups. On one hand, the roles of q_t and I_t are similar, since they both increase h_{t+1}, which is the probability of an individual firm obtaining the technology. On the other hand, the cost structures of the two effects are different. The planner pays a linear cost in q_t, while there is a nonlinear relationship between I_t and λ_{t+1}. The following lemma describes an additional property of the relationship. The interpretation is that the successes of startups are independent with each other ex ante. In the context of a continuum of startups, any incremental investment contributes independently to the acquisition opportunities. In other words, we do not consider the setup where increased investment changes the riskiness of startups as suggested by Nanda and Rhodes-Kropf (2013).

Lemma 3.1. *(The successes of startups are independent with each other.) The function $g(\cdot)$ has the following property.*

$$g'(I) = (1 - g(I))g'(0),\ \forall I \geq 0, \tag{3.3.41}$$

or equivalently:

$$g(I_1 + I_2) = 1 - (1 - g(I_1))(1 - g(I_2)),\ \forall I_1, I_2 \geq 0. \tag{3.3.42}$$

Proof. The above two equations can be derived directly from the exponential functional form of $g(\cdot)$. □

For any given allocation (I_t, Nq_tc), the planner evaluates whether it is more efficient to replace a part or all of corporate R&D with investments in more startups, where this property becomes useful.

Definition 3.2. *Startups are* ***more productive*** *in innovation than big firms if*

$$g(Nc) \geq p. \tag{3.3.43}$$

Proposition 3.3. *(No firm R&D) When startups are more productive in innovation than big firms, the planner will never invest in corporate R&D.*

$$q_t^e = 0, \forall t, \forall \{\pi_t\}_{t=0}^{\infty},\ \text{if}\ g(Nc) \geq p. \tag{3.3.44}$$

Proof. See 3.A.1.3. □

The immediate implication of Proposition 3.3 is that any corporate internal innovation activities in the general equilibrium under $g(Nc) \geq p$ is inefficient. Under these conditions, the efficient allocation of resources is such that big firms completely outsource all innovation tasks to startups. In the strategic equilibrium (see the numeric solutions in Section 3.3.3), however, firms actively participate in the R&D race even if they are worse innovators. The friction between the equilibrium and the efficient allocation is that firms fail to internalize the gains of startups. When an acquisition happens, the deal value depends not only on the innovation costs of startups, but also on the profitability of the technology. As a result, for high profit technologies, firms have incentives to avoid the acquisition payment through their in-house R&D. Section 3.5.2 discusses this pattern in more detail.

The planner also needs to decide how much to invest within each party, conditional on the assignment. This model generates the same over-investment pattern driven by competition when $N > 1$ or $M > 1$, due to "duplication effort", as stated in Loury (1979). However, while firms over-invest due to competition, VCs may still under-invest even if competition is high due to other frictions. These result will be stated and proved in the two-period version (see Section 3.3.3.1) of the model, instead of the general version, for simplicity.

3.3.3 One-chance technology

One-chance technology is the type of technology with profit stream π_t such that

$$\pi_t = 0,\ \forall t > 1. \tag{3.3.45}$$

In other words, if a firm fails to obtain the technology before the production in $t = 1$ takes place, it will miss the opportunity to profit from it permanently. The only relevant parameter in decision making is thus π_1. Within this subsection, the simplified notation π is used for π_1. Similarly, λ is short for λ_1, q is short for $q_0(0)$.

For a one-chance technology π, the firms' R&D decision q given λ can be rewritten as

$$q(\pi,\lambda) = \begin{cases} 0 & \text{if } D(0,\lambda|\pi) < c \\ 1 & \text{if } D(1,\lambda|\pi) > c \\ q^*(\pi,\lambda) & \text{otherwise} \end{cases} \tag{3.3.46}$$

where q^* is the solution to $D(q,\lambda|\pi) = c$. And

$$\begin{aligned} D(q,\lambda|\pi) &= \beta p((1-\lambda+\eta\lambda)(E^i_{0,q}[\Pi_1] - E^{-i}_{0,q}[V_1])) \\ &= \beta p(1-\lambda+\eta\lambda)E^i_{0,q}[\Pi_1] \qquad (3.3.47) \\ &= \beta p(1-\lambda+\eta\lambda)\pi\sum_{j=0}^{N-1}\binom{N-1}{j}h^j(1-h)^{N-1-j}\frac{1}{j+1} \\ &= \beta p(1-\lambda+\eta\lambda)\pi\frac{1-(1-h)^N}{Nh} \qquad (3.3.48) \end{aligned}$$

where

$$h = h(q,\lambda) = qp + (1-qp)\lambda. \tag{3.3.49}$$

We can then further simplify the condition for each possible value of q.

Corollary 3.1. *(VC investment deters firm R&D) For any one-chance technology π, the firms' R&D decision $q(\pi, \lambda)$ is weakly decreasing in λ. Moreover, there exists thresholds $\underline{\lambda}(\pi), \overline{\lambda}(\pi) \in [0, 1]$, such that*

$$q(\pi,\lambda) = \begin{cases} 1 & \text{if } \lambda \in [0, \underline{\lambda}(\pi)] \\ \text{solution to } D(q,\lambda|\pi) = c & \text{if } \lambda \in (\underline{\lambda}(\pi), \overline{\lambda}(\pi)) \\ 0 & \text{if } \lambda \in [\overline{\lambda}(\pi), 1] \end{cases} \tag{3.3.50}$$

$$\frac{\partial q(\pi,\lambda)}{\partial \lambda} \begin{cases} = 0 & \text{if } \lambda \in [0, \underline{\lambda}(\pi)) \cup (\overline{\lambda}(\pi), 1] \\ < 0 & \text{if } \lambda \in (\underline{\lambda}(\pi), \overline{\lambda}(\pi)) \end{cases}, \tag{3.3.51}$$

although $\frac{\partial q(\pi,\lambda)}{\partial \lambda}$ is not well defined at $\underline{\lambda}(\pi)$ and $\overline{\lambda}(\pi)$.

Proof. See 3.A.1.4 for the mathematical proof and expressions for $\underline{\lambda}(\pi)$ and $\overline{\lambda}(\pi)$. □

The intuition of Corollary 3.1 is straightforward. There are two effects of increasing λ. First, the probability for firm i to meet a target is higher, which means the firm will be less worried about missing out on the technology. A higher λ makes the "Plan B: wait to buy" more attractive, and thus makes "Plan A: make now" less attractive. Therefore, each firm now has less incentive to invest in R&D since its alternatives are better. This effect is reflected in the term $(1 - \lambda + \eta\lambda)$. Note that this effect is weakened when targets have high bargaining power η. Secondly, the probability for other competing firms to meet a target is also higher, which means there will be more successful firms in expectation. Even if firm i obtains the technology, it will have to share the profit with more competitors, and thus get lower payoff. This further reduces the firm's incentive to invest in R&D in the first place. The second effect is reflected in $E^i_{0,q}[\Pi_1]$. See figure 3.1 for the plot of q as a function of λ.

The threshold $\underline{\lambda}(\pi)$ is the value of λ beyond which q starts to be smaller than 1. In other words, it is the size of the acquisition probability needed to deter firms from doing 100 percent R&D. Similarly, $\overline{\lambda}(\pi)$ is the threshold of λ beyond which q is always 0, which is the acquisition probability needed to deter firms from doing any R&D. As the technology becomes more profitable, it becomes more expensive for VCs to deter corporate R&D.

Corollary 3.2. *(More expensive to deter R&D in high profit technologies) Both $\underline{\lambda}(\pi)$ and $\overline{\lambda}(\pi)$ are weakly increasing in π.*

Proof. See 3.A.1.5. □

Next consider the optimal VC's investment for a one-chance technology π. Now aggregate VC's net gain as in equation 3.3.28 can be simplified as $R(\pi,\lambda)$. Rewrite the individual VC's problem (3.3.32) as:

$$\max_{I_i} R^i(I_i) = -I_i + \beta_V \frac{I_i}{I_i + I_{-i}} \alpha E_0^{\ell}[l_1|\lambda] a_1, \tag{3.3.52}$$

subject to

$$\lambda = g(I_i + I_{-i}) \tag{3.3.53}$$

For an individual VC, the benefits of increasing investment are as follows. (1) First, it will increase the aggregate size of the startup pool, resulting in higher λ. Therefore, there will be more acquisitions and thus exits. (2) Second, increasing I_i will increase VC i's share of the startup pool, conditional on other VCs' investment level. As a result, it will get a higher proportion of all proceeds from acquisitions. However, the marginal effect decreases as I_i gets larger. (3) Third, as a less obvious benefit, increasing I_i, and thus λ, can deter firm's R&D investment q, which further increases chances of acquisitions. (4) On the other hand, one drawback of increasing I_i is obvious, it results in higher cost. (5) Another less obvious drawback is that it increases the expected number of successful firms, and thus reduces the firms' payoff from obtaining the technology, and therefore results in lower selling prices of startups. Under the joint forces of the above effects, the function $R^i(I_i)$ is not well-behaved. Therefore, the optimization problem can not be solved by first order conditions alone.

Lemma 3.2. *$R^i(I_i|I_{-i},\pi)$ is a smooth function on $[0,\infty)$ except possibly on*

$$I_i = g^{-1}(\underline{\lambda}(\pi)) - I_{-i} \tag{3.3.54}$$

and

$$I_i = g^{-1}(\overline{\lambda}(\pi)) - I_{-i} \tag{3.3.55}$$

Proof. See 3.A.1.6 □

Figure 3.2 shows the relationship between a monopoly VC's (no effect (2)) net payoff and the choice of λ. By comparing figure 3.2 and 3.1, we can clearly see the loss of smoothness described above. Take the solid green line ($\pi = 2$) for example, when λ is lower than 0.76, q is always 1, the benefit from increasing λ comes solely from selling more to the firms whose R&D failed. Within the interval $[0.76, 0.83]$, increasing λ has the effect in point (3) above to discourage R&D, and therefore the benefit is more steeply increasing. Finally, as λ gets higher than 0.83, q is always 0 and the VC can no longer discourage R&D. After this point, the cost of increasing λ dominates all other benefits.

Proposition 3.4. *(Optimal VC investment as a function of profitability) The optimal level of VC investment implied by $\lambda(\pi)$ is in general discontinuous and non-monotone in π.*

Proof. See 3.A.1.7 □

Since the objective function is non-smooth, the maximizer of the function can thus jump between local optima. Such jump is inevitably discontinuous, and can be non-monotone as well. If two parallel technologies emerge, where one technology is slightly more profitable than the other, it is possible that the less profitable technology received more funding from a strategic and optimally-behaved VC investor. This might seem counter-intuitive at first. However, the logic follows from Corollary 3.1 that VCs can deter corporate R&D by increasing the size of the startup pool, plus the property of the VC's objective function that VCs will benefit from the extra exit opportunities when R&D is discouraged.

Figure 3.3 shows the acquisition opportunities as well as total capital invested by VCs as a function of the profitability of the technology, and both have a discontinuity. When a new technology is not so profitable, there are no incentives for firms to set up expensive R&D project teams. At the same time, VCs can invest in a handful of entrepreneurs and extract surplus from their exits. As the profit of the technology increases, VCs optimally increase their investment, while firms stay out of the technology and act solely as acquirers. Beyond the critical point, the profit of the technology is high enough that VCs cannot further restrain firms from doing R&D, or the cost of doing so is too high. For such technologies, firms are both makers and buyers, and VC's investment is also increasing in profitability. At the critical point, there exists multiple equilibria. In one equilibrium, firms do not do internal R&D, and VCs invest in a lot of startups. In the other equilibrium, firms do internal R&D, and VCs invest in fewer startups.

3.3.3.1 Market Structure and Efficiency

Lastly, as a continuation of the discussion about efficiency in Section 3.3.2, this section studies the efficiency of equilibrium within firms and VCs for one-chance technologies. In 3.3.2, it is described how inefficiency can arise in the equilibrium due to mis-allocation of the innovative task between the two parties. In fact, inefficiency still exists even within each party. Specifically, I compare the firms' problem with the planner's problem when VCs and startups are not present, and compare the optimal level of R&D chosen in both problems. Similarly, I compare the VCs' problem with the planner's problem when there is no corporate internal R&D, and compare the optimal level of investment chosen.

When startups are not present ($\lambda = 0$), the equilibrium firm's decision can be expressed as

$$q(\pi) = \begin{cases} 0 & \text{if } \beta_F p\pi < c \\ 1 & \text{if } \beta_F \pi \frac{1-(1-p)^N}{N} > c \\ q^*(\pi) & \text{otherwise} \end{cases} \tag{3.3.56}$$

where $q^*(\pi)$ is the solution to $\beta_F \pi(1-(1-pq)^N) = qNc$.

When startups are not present, the planner's decision on corporate innovation can be described by the following problem:

$$\max_{q\in[0,1]} -Nqc + \beta(1-(1-pq)^N)\pi. \tag{3.3.57}$$

The problem is a well-defined convex problem, and the solution is characterized by the following step function

$$q^e(\pi) = \begin{cases} 0 & \text{if } \beta p\pi < c \\ 1 & \text{if } \beta p\pi(1-p)^{N-1} > c \\ q^{e*}(\pi) & \text{otherwise} \end{cases} \tag{3.3.58}$$

where $q^{e*}(\pi)$ is the solution to $\beta p\pi(1-pq)^{N-1} = c$.

Proposition 3.5. *(Firms over-invest under competition) If $\beta_F = \beta$, firms over-invest in R&D,*[12]

$$q(\pi) \geq q^e(\pi), \forall \pi > 0, \tag{3.3.59}$$

where the equality is achieved for all $\pi > 0$ only when $N = 1$.

Proof. See 3.A.1.8. □

Now consider the case where there is no corporate R&D ($p = 0$), a VC's problem is simplified to

$$\max_{I_i} R^i(I_i) = -I_i + \beta_V \frac{I_i}{I_i + I_{-i}} \alpha E^\ell[l]a_1, \tag{3.3.60}$$

subject to

$$\lambda = g(I_i + I_{-i}) \tag{3.3.61}$$

In this case, the planner's problem is

$$\max_{I} -I + \beta(1-(1-\lambda)^N)\pi, \tag{3.3.62}$$

[12] In the numerical exercises of this paper, $\beta_F = \beta = 0.95$, this is based on the assumption that big tech firms have access to low cost (risk-free) financing.

subject to

$$\lambda = g(I). \tag{3.3.63}$$

Factors that lead VCs to under invest include high cost-of-capital, low ownership of startups, and low bargaining power of startups.

Proposition 3.6. *(VC under-invest under frictions) When M=1, and $\beta_V \leq \beta$, $\alpha \leq 1$, $\eta \leq 1$, the VC under-invest compared to the social planner.*[13]

$$I(\pi) \leq I^e(\pi), \forall \pi > 0, \tag{3.3.64}$$

where the equality is achieved for all $\pi > 0$ only when $\beta_V = \beta$, $\alpha = 1$, $\eta = 1$.

Proof. See 3.A.1.9. □

However, similar to the intuition in Proposition 3.5, VCs also tend to invest more as the VC industry becomes more competitive.

Proposition 3.7. *(VC over-invest under competition) When $\beta_V = \beta$, $\alpha = 1$, $\eta = 1$, and $M \geq 1$, VCs in aggregate over-invest.*

$$I(\pi) \geq I^e(\pi), \forall \pi > 0, \tag{3.3.65}$$

where the equality is achieved for all $\pi > 0$ only when $M = 1$.

Proof. See 3.A.1.10. □

3.3.4 Future technology

A future technology does not start profiting until sometime $T > 1$ in the future, i.e. its profit stream $\{\pi_t\}_{t=0}^{\infty}$ satisfies

$$\pi_t = 0, \forall t < T. \tag{3.3.66}$$

The phenomenon where the commercialization of a technology lags the invention of it has been discussed in Shleifer (1986), and could be explained by the coordination of multiple technologies.

For convenience, assume that such future technology is also short-lived:

$$\pi_t = 0, \forall t > T. \tag{3.3.67}$$

[13] In the numerical exercises, $\beta^V = 0.8$, which maps to the annual return rate of 20% of an average VC fund; $\alpha = 0.2$, which matches the share of VC of an average VC-backed startup; $\eta = 0.1$, which is selected to approximately match with the empirical acquisition/IPO value ratio.

Therefore, the relevant parameters in decision making are T and π_T. The decisions to be made are $\{q_t(k_t)\}_{t=0}^T$ and $\{\lambda_t(k_t)\}_{t=0}^T$.

If no investment is made by any party before $T-1$, the equilibrium at $T-1$ is exactly identical to a "one-chance technology". If positive investment is made at $t \leq T-1$, the investment level at $t+1$ is conditional on the number of successes realized at the beginning of $t+1$. Denote the unconditional distribution derived from equations (3.3.12)recursively as

$$P_t(k_t) = \sum_{k_{t-1}=0}^{N} S(k_t|k_{t-1})P_{t-1}(k_{t-1}). \tag{3.3.68}$$

With the initial condition

$$P_0(k_t) = \begin{cases} 1 & \text{if } k_t = 0 \\ 0 & \text{otherwise} \end{cases}, \tag{3.3.69}$$

all P_t's can now be computed by forward iteration. Then we can denote the expected R&D investment in each period as:

$$E_t^R = E[(N-k_t)q_t c] = \sum_{k_t=0}^{N} P_t(k_t)q_t(k_t)c, \tag{3.3.70}$$

and the expected VC investment in each period as:

$$E_t^V = E[I_t] = \sum_{k_t=0}^{N} P_t(k_t)I_t(k_t). \tag{3.3.71}$$

Further, denote

$$E_0^Q = \sum_{t=0}^{T} E[\frac{N-k_t}{N}q_t] \tag{3.3.72}$$

as the expected number of rounds of R&D attempts within each firm for the technology.

We use the backward induction method described in 3.A.2.2 to numerically solve the equilibrium R&D and VC investment in each period before T.

For firms, there is no direct benefit in having the technology ready before T, while investing early creates an extra opportunity cost due to time value of money. However, starting developing too late might result in failure and thus missing the technology. For technologies that are sufficiently profitable, firms have incentives to invest a few periods before the payoff date to increase the probability of success. Figure 3.5 plots E_0^Q under different T and π_T, without the participation of VCs. Moreover, for moderately profitable technologies, succeeding early has the implicit benefit of deterring later attempts. Due to this effect, firms start R&D at $T-2$ or earlier, even when q_{T-1} is lower than 1. Moreover, when q_{T-2} becomes higher as profit increases, the expected number of successful firms at $T-1$ also increases. As a result, the expected level of investment at $T-1$ might be lower than that of $T-2$. Potentially, the R&D race described here will induce inefficiency by hastening all investments in a technology that will payoff periods later, which is reminiscent of the inefficiency results form the patent racing literature. Figure 3.4 shows the R&D decision of firms on a future technology without the participation of VCs.

When VCs join the game, their investment has the effect of deterring corporate R&D, as described in Corollary 3.1. Figure 3.8 shows the equilibrium for a low profit technology. VCs take care of all investment,[14], and firm investment only happens in $T-1$. Note that the choice of firms of not investing in their own R&D is conditional on the observation that the VC market is sufficiently active.

For a more profitable future technology, such participation of VCs is not enough to keep firms from internal R&D. However, unlike for "one-chance technologies" where VCs have to let go of their shares in the market substantially once firms join the race, VCs facing a "future technology" have an additional means to deter corporate R&D: investing earlier. As a interesting result, VC investment and corporate R&D can happen in different time periods. Figure 3.9 shows the equilibrium for a higher-profit technology. In this case, VCs first invest in $T-2$, resulting in some M&As at the beginning of $T-1$. If enough firms have met a target, the remaining firms will do another round of R&D. If not enough matches happened, both VCs and firms will invest at $T-1$ to ensure success at T.

The deterrent effect of VC investment on corporate R&D is not uniform over time. Naturally, VCs have incentives to invest in the time periods where such impact is more effective. Figure 3.6 shows the effect on expected lifetime R&D effort E_0^Q by VC investment in different periods. In general, early VC investment can better deter corporate R&D, especially for smaller amount of investment. In addition, investing early implies more potential buyers, since most firms have not succeeded in their in-house R&D yet. As a result, without considering costs, VCs should adopt an investment strategy with early investment, or a combination of early and late investment. Under this case, the R&D race among firms would be alleviated.

On the other hand, the cost of investing early for VC is substantial. In the negotiation between startups and acquirers, the acquirers now have better outside options. Since the technology does not start profiting immediately, the firms could always wait for the next round of successful startups, or do their in-house R&D. As a result, acquisitions in earlier periods would occur at much lower prices. Technically, the bargaining power of startups have not

[14] Refer to the probability matrix in Figure 3.8c all λ and q are off-equilibrium except $\lambda(0,1)$.

changed, but the surplus that they have contributed is lower. In addition, the high cost of capital faced by VCs, reflected by low β_V, reduces their overall incentives to invest.

Figure 3.7 shows the time series of E_t^V and E_t^R under two future technologies with different terminal profit. We see a typical pattern that VC investment is ahead of R&D in both technologies. However, such effect only persists if the discount factor of VC is not too much lower than that of firms, and the cost of innovation within startups is significantly lower than that in firms.

3.4 Extension: IPO Exits

This section relaxes the assumption that all exits are through acquisitions and allow IPO exits for VC-backed startups. We keep the setup in the baseline model that there are N firms as potential acquirers, and each of them has a probability λ_t of meeting a target at time t. Additionally, there is a probability $\theta\lambda_t$ that a successful startup will also obtain the profiting channel and thus can operate independently. In this case, the startup will be sold to the market at the expected value of a successful firm through IPO. θ controls the ratio of IPO and acquisition exits. From the VC's perspective, the exit opportunities at t can be summarized by γ_t, which represents the probability of at least one exit.

$$\gamma_t = r(\lambda_t) = 1 - (1 - \lambda_t)^N (1 - \theta\lambda_t) \tag{3.4.1}$$

By investing in the startup pool, VCs jointly decide their exit opportunities in the next period:[15]

$$\gamma_{t+1} = G(I_t) \tag{3.4.2}$$

Note that in the special case where $\theta = 0$, the above notation is identical to the baseline model with $\varphi = N\phi$. By observing I_t, firms predict the acquisition opportunity will be

$$\lambda_{t+1} = r^{-1}(G(I_t)). \tag{3.4.3}$$

This mechanism is stronger when θ is low. On the other hand, as θ gets larger, startups become more independent and IPO becomes more likely, and λ_{t+1} will be low even when I_t is very high.[16]

[15] In the numerical exercise, the functional form of G is $G(I_t) = 1 - \exp(\frac{I_t}{\varphi})$, where $\varphi > 0$ is the parameter of the function.

[16] The value of θ can be calibrated by matching the exit counts through acquisitions and IPOs respectively, after reasonably selecting the number of potential acquirers. On the other hand, this parameter to some extent depends on the nature of the technology, which is not modeled in this paper. Intuitively, VCs should prefer technologies that allow startups to be more independent, holding other factors equal.

Denote k_t as the number of firms with the technology at time t, and k_t^- is the number of firms without. Note that due to the possibility of new entry through IPO, $k_t + k_t^- \geq N$. Then we can define the value functions similarly.

$$\Pi_t = \Pi(k_t, k_t^- | \{\pi_{t'}\}_{t'=t}^{\infty}, \Lambda) \tag{3.4.4}$$

$$V_t = V(k_t, k_t^- | \{\pi_{t'}\}_{t'=t}^{\infty}, \Lambda) \tag{3.4.5}$$

$$R_t = R(k_t, k_t^- | \{\pi_{t'}\}_{t'=t}^{\infty}, \Lambda) \tag{3.4.6}$$

The acquisition cost at time t is

$$a_t = a(k_{t-1}, k_{t-1}^- | \{\pi_{t'}\}_{t'=t}^{\infty}, \Lambda). \tag{3.4.7}$$

And the IPO value at time t is

$$o_t = o(k_{t-1}, k_{t-1}^- | \{\pi_{t'}\}_{t'=t}^{\infty}, \Lambda). \tag{3.4.8}$$

The policy function of firms can be defined as

$$q_t = q(k_t, k_t^- | \{\pi_{t'}\}_{t'=t}^{\infty}, \Lambda) \tag{3.4.9}$$

as the policy function, i.e. the probability to invest in R&D, for a firm without the technology at time t.

With the possibility of IPO, the new state transition distribution is as follows.

$$S(k_t + j, k_t^- - j | k_t, k_t^-) = (1 - \theta\lambda_{t+1}) \binom{k_t^-}{j} h_t^j (1 - h_t)^{k_t^- - j}, \tag{3.4.10}$$

$$S(k_t + j + 1, k_t^- - j | k_t, k_t^-) = \theta\lambda_{t+1} \binom{k_t^-}{j} h_t^j (1 - h_t)^{k_t^- - j}, \tag{3.4.11}$$

for $j \in \{0, \cdots, k_t^-\}$, where

$$h_t = q_t p + (1 - q_t p)\lambda_{t+1}. \tag{3.4.12}$$

Similarly, for a variable $x_t = x(k_t, k_t^-)$, denote

$$E_t[x_{t+1}] = \sum_{j=0}^{k_t^-} (S(k_t+j, k_t^- - j | k_t, k_t^-) x(k_t+j, k_t^- - j) + S(k_t+j+1, k_t^- - j | k_t, k_t^-) x(k_t+j+1, k_t^- - j) \tag{3.4.13}$$

Next, the conditional state transition process where firm i will gain the technology is given by

$$S^i(k_t + j + 1, k_t^- - 1 - j | k_t, k_t^-) = (1 - \theta\lambda_{t+1}) \binom{k_t^- - 1}{j} h_t^j (1 - h_t)^{k_t^- - 1 - j}, \tag{3.4.14}$$

$$S^i(k_t+j+2, k_t^- - 1 - j | k_t, k_t^-) \;=\; \theta\lambda_{t+1}\binom{k_t^- - 1}{j} h_t^j (1-h_t)^{k_t^- -1-j}, \tag{3.4.15}$$

where $j \in \{0, \cdots, k_t^- - 1\}$. Similarly, define a state transition process conditional on firm i not gaining the technology as

$$S^{-i}(k_t+j, k_t^- - j | k_t, k_t^-) \;=\; (1-\theta\lambda_{t+1})\binom{k_t^- - 1}{j} h_t^j (1-h_t)^{k_t^- -1-j}, \tag{3.4.16}$$

$$S^{-i}(k_t+j+1, k_t^- - j | k_t, k_t^-) \;=\; \theta\lambda_{t+1}\binom{k_t^- - 1}{j} h_t^j (1-h_t)^{k_t^- -1-j}, \tag{3.4.17}$$

where $j \in \{0, \cdots, k_t^- - 1\}$. Denote the related expected values by

$$\begin{aligned} E_t^i[x_{t+1}] \;=\; & \sum_{j=0}^{k_t^- -1} (S^i(k_t+j+1, k_t^- - 1 - j | k_t, k_t^-) x(k_t+j+1, k_t^- - 1 - j) \\ & + S^i(k_t+j+2, k_t^- - 1 - j | k_t, k_t^-) x(k_t+j+2, k_t^- - 1 - j) \end{aligned} \tag{3.4.18}$$

$$\begin{aligned} E_t^{-i}[x_{t+1}] \;=\; & \sum_{j=0}^{k_t^- -1} (S^{-i}(k_t+j, k_t^- - j | k_t, k_t^-) x(k_t+j, k_t^- - j) \\ & + S^{-i}(k_t+j+1, k_t^- - j | k_t, k_t^-) x(k_t+j+1, k_t^- - j) \end{aligned} \tag{3.4.19}$$

We can now recursively define both value functions:

$$\Pi_t(k_t, k_t^-) \;=\; \frac{\pi_t}{k_t} + \beta_F E_t[\Pi_{t+1}] \tag{3.4.20}$$

$$V_t(k_t, k_t^-) \;=\; -q_t c - \beta_F(1-pq_t)\lambda_{t+1}a_{t+1} \tag{3.4.21}$$

$$+\beta_F(h_t E_t^i[\Pi_{t+1}] + (1-h_t)E_t^{-i}[V_{t+1}]) \tag{3.4.22}$$

By the Nash Bargaining assumption in the acquisition process, we have

$$a_t = \eta(E_{t-1}^i[\Pi_t] - E_{t-1}^{-i}[V_t]). \tag{3.4.23}$$

Similarly, conditional on an IPO happening at time t, the state transition process is

$$S^o(k_t+j+1, k_t^- - j | k_t, k_t^-) = \binom{k_t^-}{j} h_t^j (1-h_t)^{k_t^- - j}. \tag{3.4.24}$$

Define

$$E_t^o[x_{t+1}] = \sum_{j=0}^{k_t^-} \binom{k_t^-}{j} S^o(k_t+j+1, k_t^- - j | k_t, k_t^-) x(k_t+j+1, k_t^- - j). \tag{3.4.25}$$

The IPO firm will be priced at:

$$o_t = E^o_{t-1}[\Pi_t]. \tag{3.4.26}$$

In a symmetric equilibrium across firms, the policy function q has the same form as in Proposition 3.1, with a different specification of $D(q)$.

$$\begin{aligned} D(q) \quad = \quad & \beta p \sum_{j=0}^{k_t^- - 1} S_t^i(j|k_t^-, q, \lambda_{t+1})[(1-\theta\lambda_{t+1})(\Pi_{t+1}(k_t + j + 1, k_t^- - j - 1) \\ & -\lambda_{t+1}(\Pi_{t+1}(k_t + j + 1, k_t^- - j - 1) - a_{t+1}(k_t, k_t^-)) \\ & -(1-\lambda_{t+1})V_{t+1}(k_t + j, k_t^- - j) \\ & +\theta\lambda_{t+1}(\Pi_{t+1}(k_t + j + 2, k_t^- - j - 1) \\ & -\lambda_{t+1}(\Pi_{t+1}(k_t + j + 2, k_t^- - j - 1) - a_{t+1}(k_t, k_t^-)) \\ & -(1-\lambda_{t+1})V_{t+1}(k_t + j + 1, k_t^- - j)] \end{aligned} \tag{3.4.27}$$

The aggregate value function of VCs can be recursively defined as

$$R_t = -I_t + \beta_V(\alpha E_t^\ell[l_{t+1}]a_{t+1} + \theta\lambda_{t+1}\alpha o_{t+1} + E_t[R_{t+1}]). \tag{3.4.28}$$

An individual VC i's problem is

$$\begin{aligned} \max_{I_{i,t}} R_t^i(I_{i,t}) \quad = \quad & -I_{i,t} + \beta_V(\frac{I_{i,t}}{I_{i,t} + I_{-i,t}}(\alpha E_t^\ell[l_{t+1}|\lambda_{t+1}]a_{t+1} + \theta\lambda_{t+1}\alpha o_{t+1}) \\ & + \frac{1}{M}E_t[R_{t+1}|\lambda_{t+1}]), \end{aligned} \tag{3.4.29}$$

subject to

$$\lambda_{t+1} = r^{-1}(G(I_{i,t} + I_{-i,t})) \tag{3.4.30}$$

See Section 3.A.2.3 and 3.A.2.4 for the numerical solutions in this setting.

3.5 Discussions

3.5.1 Spillovers

The model in this paper does not incorporate the positive spillovers of innovation, the presence of which may change the welfare and efficiency implications. One immediate modification on the model is to assume that the total social surplus of the technology is $R\pi$ (or $R\pi_t$, in the multi-period version), instead of simply π, where $R \geq 1$ measures the degree of the spillover effect. In this case, firms and VCs would be solving the same problem while the social planner correctly replaces π with $R\pi$ in her problem. As a result, the old "efficient" allocation would become under-invested for not internalizing all the benefit. Further, the "under-investment" of VC induced by low η, α, and β_V in the old allocation is aggravated in the presence of positive spillover. On the other hand, the "over-investment" in the old allocation induced by competition among firms and VCs to some degree alleviates the issue. Whether the combined effect leads to over- or under-investment depends on how large the spillover is, as well as the size of each individual effect.

Not all spillovers are well approximated by a multiplier on profit. Scotchmer (1991) documents the "sequential spillover" mechanism where the emergence of a technology, in the language of this model, is conditional on the successful invention of a previous technology. To take such mechanism into consideration, the social planner needs to evaluate the social benefit of a technology by incorporating all new technologies it might trigger, which is beyond the scope of this paper. On the other hand, the model specifies N exogenous firms for each technology which are eligible to profit from it. One can view these firms as the owners of a previous technology which is the building block of the current technology. However, in the firm's decision process, the benefit of becoming eligible for some other technologies in the future is not taken into account, which results in under-investment.

In the presence of positive spillovers, the need to fix the frictions causing under-investment in the baseline model is more urgent. One approach is to support entrepreneurs by providing them alternative and cheaper financing channels. This is in fact the growth strategy of many developing countries. Another way is to increase the independence of entrepreneurs, by providing them more exposure to the consumers. In the model, this will be reflected by more IPOs relative to acquisitions. However, these actions are not easy to implement, which is why the frictions are modeled as status quo.

The other solution to potential positive spillover effect is to induce even more over-investment in the baseline model. Without positive spillover, competition among firms leads to over-investment due to an implicit negative externality. When an incumbent firm is already in possession of a technology, any following success does not increase social surplus at all. Instead, the newly succeeded firm's surplus solely comes from reducing other firms' profit, which it does not internalize. Although VCs produce innovation with less cost within startups, their level of investment largely depends on the demand from the acquirers side. When the competition among VCs is not high enough, it is likely that they coordinate to invest just the right amount to stop firms from investing in R&D. As a result, the implied VC investment could be high even though VCs are less competitive, since their level of investment absorbs the competition among firms. As the VC industry becomes more competitive, the competition among

VCs becomes the main driver of over-investment. Therefore, maintaining a competitive environment in the high-tech industry and the VC industry seems to be a reasonable way to deal with positive spillovers.

3.5.2 The Hold-up Problem

It is well-established in economics theory that parties to an efficient trade refrain from cooperating due to the concerns of reduced bargaining power in the future. The model in this paper generates a similar pattern where firms would not give up their in-house R&D on high-profit technologies even if it is more efficient for them to completely outsource such tasks to smaller startups. However, this pattern is not driven by the fear of loss of bargaining power. On the contrary, the bargaining power of firms relative to startups is held as a constant in the model.

Breaking down the model carefully, we can see that the result will persist as long as the cost to produce a technology is not perfectly correlated with the value of the technology. As a technology gets more valuable, an acquisition of a successful startup in this technology becomes more expensive. Meanwhile the cost of in-house R&D is held at the same level, which provides incentives for firms to "DIY". On the corporate side, this friction can be alleviated when the bargaining power of startups is sufficiently low, which would, of course, reduce the incentives of startups and lead to severe under-investment in innovation by startups and VCs. In short, when startups are indeed more productive in innovation, the efficient (Hosius) split of surplus does not exist in this setup.

Another reason of the above hold-up problem is that acquisitions happen after the innovation outcome reveal. At this stage, firms have no incentives to acquire failed startups and will only be interested in successful ones. On the other hand, successful startups are scarce due to the high failure rate and uncertain innovation process. The entrepreneurs of these successful startups are well aware of the value of their success and will charge corporate buyers proportional to the value of the technology, instead of just taking their cost of innovation. An alternative M&A mechanism is where firms buy out all startups in the beginning and pay each of them just enough to cover the cost of innovation. Naturally, when a small fraction of these startups succeed, the corporate owners can inherit the technology. The above solution of writing a complete contract has been proven hard to do practically, which is why the hold-up problem exists in the first place. In addition, buying out startups at early stages creates a moral hazard problem, which may potentially destroy the comparative advantage of startups in innovation. Nevertheless, firms are indeed pursuing this path by becoming venture capitalists themselves. The next section 3.5.3 provides more details about this trend.

It is worth pointing out that the hold-up problem can be two-way. Recall that the main driving force of startups seeking to be acquired is that the big firms have better access to consumers, or profiting channel. Could it be that big firms have been "holding up" startups by occupying the profiting channel through acquiring successful startups in every possible technology? After all, the tech giants today were at one time startups. They were not born with the profiting channel they now possess, but gained it gradually and strategically. Due to the hold-up in profiting channels, VCs and startups have to exit through acquisitions instead of IPOs. Ewens and Farre-Mensa (2019) documents that recent regulatory change lead to higher bargaining power of startup founders against VCs, which allow them to stay private longer and grow larger before IPO. If the startups could obtain better profiting channels, they would be more independent, which could lead to a different market structure than the current one featuring oligopolistic tech giants.

3.5.3 Corporate Venture Capital

This paper does not distinguish corporate venture capital (CVC) with other VC or institutional investors, and they are all modeled as independent decision makers. Still, it is worth discussing the nature of CVCs and their objectives. During the past decade, more and more big firms entered the VC industry by starting their own venture arm[17]. As is mentioned in the previous discussion (3.5.2), CVC could be the potential solution to internalize the externalities that arise in the innovation process, as it combines the party with innovative ideas, the party with funding, and the party with the profiting channel.

On the other hand, big firms do not often acquire startups that are backed by their own venture capital arm (Guo, Lou, & Pérez-Castrillo, 2015). Instead, successful startups exit regularly through IPOs or acquisitions by other firms. In other words, CVC, to a greater extent, is a profit-maximizing subsidiary of its corporate head, rather than a channel of external R&D. Given the fact that big tech companies have significant cash holdings, it is not surprising that they seek high returns in the VC industry.

Nevertheless, the investment pattern of CVCs is distinctive from independent VCs (IVCs). In aggregate, CVCs are involved in much fewer deals than IVCs, while the value of each CVC deal is significantly higher than an average IVC deal. This pattern resembles our model assumption about the innovation production function of corporate internal R&D and startups. The investment strategy of CVCs may to some extent reflect the corporate expertise in selecting and monitoring R&D projects, i.e. they specialize in relatively high-cost and high-

[17] The most active ones include GV (Google), Intel Capital (Intel), M12 (Microsoft), Softbank Capital (Softbank).

success-probability projects, while IVCs adopt a "spray-and-pray" strategy on low-cost and low-success-probability projects. In observance of the recent increased activities and high returns of CVC, there could be potential synergy creation within CVC-backed startups in terms of innovative productivity that might be superior to both IVC-backed startups and corporate in-house R&D.

3.5.4 Risks

The VC industry is frequently labeled as "high-risk, high-return". From an empirical perspective , a VC takes huge risks in terms of whether an individual startup that it supports will be successful. This risk can be hedged or reduced by investing in a lot of startups. The model in this paper implicitly admits this mechanism by assuming a continuum of startups, and a one-to-one mapping between investment and exit opportunity. From a broader perspective, though, the VC also needs to take the risk of whether the technology/industry it is investing in will be profitable. The baseline model assumes away such risk so that VCs know exactly the profit stream of a technology at the beginning. Empirically, there are VCs that seem to hedge this risk by investing in many areas/technologies, while there are also VCs that specialize in very narrow industries. How large the risk of betting on the wrong technology is certainly depends on the VC's insight, as well as the economic environment.

The pattern that VCs invest in emerging technologies, as opposed to the "wait-and-see" strategy taken by big firms, is commonly interpreted as the willingness to take risks while the prospect of the technology is still uncertain. The model is able to generate a similar pattern without introducing any uncertainty. Instead, the result is driven purely by the strategic interactions between VCs and firms. The justification of the high return rate of VC in the model is that they provide the screening and monitoring service to startups and thus acquirers. However, it remains to be further investigated how much risk VCs are undertaking, and what composes the VC market's value-added to the economy.

3.6 Conclusion

This paper presents a model where VC investment affects corporate innovation decisions through its impact on future M&A opportunities. By observing the level of VC investment today, corporate decision makers will predict the opportunities of acquiring a successful startup in a specific technology in the near future, and decide whether it is worthwhile to invest in in-house R&D in the same technology. Different from the "make-vs-buy" desicion in other dimensions of corporate strategy, the tradeoff involving technology here is dynamic and uncertain, i.e. it is a "make-now-vs-maybe-buy-later" decision. For a moderately profitable technology, when VC investment is high enough,firms are willing to give up in-house R&D and rely completely on external acquisitions to obtain the technology. For a high-profit technology, firms will maintain their in-house R&D even if the VC investment is high. In this case, firms will seek to acquire externally only if their in-house R&D is not successful.

From the VC's perspective, the return on investment is higher if big firms are kept out of the R&D race. VCs therefore have incentives to increase investment to deter corporate R&D. Since the deterrence is more effective when the profitability of the technology is not too high, we might observe high VC investment, and active startups, in technologies that are not the most valuable. Broadly speaking, such startups will target an under-represented group of users, and develop a product to fulfill their special needs, so as to exploit the thin profit that firms do not bother to collect.

Further, although it is hard for VCs to deter corporate R&D for high-profit technologies, they can more effectively do so when the profiting horizon of the technology is longer. Specifically, VCs have the incentive to invest at the early stage of an emerging technology. It is worth pointing out the counterfactural predictions of the model. Without VC, firms will invest earlier. And cross-sectionally, firms will start innovating in the less profitable technologies without VCs and the acquisition opportunities they provide.

The bottom line of this paper is that there is interaction between VC investment and corporate innovation, and the two should not be viewed as independent. Such interaction will become stronger as the VC industry continues to grow, and if acquisitions remain the major exit channel for VC-backed startups.

3.7 References

Aghion, P., Bloom, N., Blundell, R., Griffith, R., & Howitt, P. (2005). Competition and innovation: An inverted-u relationship. *The Quarterly Journal of Economics*, *120*(2), 701–728.

Almus, M., & Czarnitzki, D. (2003). The effects of public r&d subsidies on firms' innovation activities: the case of eastern germany. *Journal of Business & Economic Statistics*, *21*(2), 226–236.

Arrow, K. J. (1962). Economic welfare and the allocation of resources for invention. In *The rate and direction of inventive activity: Economic and social factors* (pp. 609–626). Princeton University Press. Retrieved from `http://www.jstor.org/stable/j.ctt183pshc.26`

Bena, J., & Li, K. (2014). Corporate innovations and mergers and acquisitions. *The Journal of Finance*, *69*(5), 1923–1960.

Blonigen, B. A., & Taylor, C. T. (2000). R&d intensity and acquisitions in high-technology industries: evidence from the us electronic and electrical equipment industries. *The Journal of Industrial Economics*, *48*(1), 47–70.

Busom, I. (2000). An empirical evaluation of the effects of r&d subsidies. *Economics of innovation and new technology*, *9*(2), 111–148.

Cassiman, B., & Veugelers, R. (2006). In search of complementarity in innovation strategy: Internal r&d and external knowledge acquisition. *Management science*, *52*(1), 68–82.

Cohen, W. M. (2010). Fifty years of empirical studies of innovative activity and performance. In *Handbook of the economics of innovation* (Vol. 1, pp. 129–213). Elsevier.

Collard-Wexler, A., Gowrisankaran, G., & Lee, R. S. (2019). "nash-in-nash" bargaining: a microfoundation for applied work. *Journal of Political Economy*, *127*(1), 163–195.

Cunningham, C., Ederer, F., & Ma, S. (2019). Killer acquisitions. *Available at SSRN 3241707*.

David, J. M. (2011). Competition, innovation, and the sources of product quality and productivity growth. In *V intertic conference.*

Ewens, M., & Farre-Mensa, J. (2019). The deregulation of the private equity markets and the decline in ipos.

Ewens, M., Nanda, R., & Rhodes-Kropf, M. (2018). Cost of experimentation and the evolution of venture capital. *Journal of Financial Economics*, *128*(3), 422–442.

Goettler, R. L., & Gordon, B. R. (2011). Does amd spur intel to innovate more? *Journal of Political Economy*, *119*(6), 1141–1200.

Gompers, P. A. (1995). Optimal investment, monitoring, and the staging of venture capital. *The journal of finance*, *50*(5), 1461–1489.

González, X., & Pazó, C. (2008). Do public subsidies stimulate private r&d spending? *Research Policy*, *37*(3), 371–389.

Guo, B., Lou, Y., & Pérez-Castrillo, D. (2015). Investment, duration, and exit strategies for corporate and independent venture capital-backed start-ups. *Journal of Economics & Management Strategy*, *24*(2), 415–455.

Hall, B. H. (1988). The effect of takeover activity on corporate research and development. In *Corporate takeovers: Causes and consequences* (pp. 69–100). University of Chicago Press.

Hall, B. H., & Lerner, J. (2010). The financing of r&d and innovation. In *Handbook of the economics of innovation* (Vol. 1, pp. 609–639). Elsevier.

Kortum, S., & Lerner, J. (2000). Assessing the contribution of venture capital to innovation. *RAND Journal of Economics*, *31*(4), 674–692.

Krieger, J., Li, X., & Thakor, R. T. (2018). Find and replace: R&d investment following the erosion of existing products. *Available at SSRN 3240344*.

Levine, O. (2017). Acquiring growth. *Journal of Financial Economics*, *126*(2), 300–319.

Loury, G. C. (1979). Market structure and innovation. *Quarterly Journal of Economics*, *93*(3), 395–410.

Mansfield, E., & Switzer, L. (1984). Effects of federal support on company-financed r and d: the case of energy. *Management Science*, *30*(5), 562–571.

Metrick, A., & Yasuda, A. (2010). Venture capital and the finance of innovation. *VENTURE CAPITAL AND THE FINANCE OF INNOVATION, 2nd Edition, Andrew Metrick and Ayako Yasuda, eds., John Wiley and Sons, Inc*.

Nanda, R., & Rhodes-Kropf, M. (2013). Investment cycles and startup innovation. *Journal of Financial Economics*, *110*(2), 403–418.

Pan, Y., Huang, P., & Gopal, A. (2019). Storm clouds on the horizon? new entry threats and r&d investments in the us it industry. *Information Systems Research*.

Phillips, G. M., & Zhdanov, A. (2013). R&d and the incentives from merger and acquisition activity. *The Review of Financial Studies*, *26*(1), 34–78.

Phillips, G. M., & Zhdanov, A. (2017). *Venture capital investments and merger and acquisition activity around the world* (Tech. Rep.). National Bureau of Economic Research.

Popp, D., & Newell, R. (2012). Where does energy r&d come from? examining crowding out from energy r&d. *Energy Economics*, *34*(4), 980–991.

Schumpeter, J. A. (1942). *Capitalism, socialism and democracy*.

Scotchmer, S. (1991). Standing on the shoulders of giants: cumulative research and the patent law. *Journal of economic perspectives*, *5*(1), 29–41.

Shapley, L., & Shubik, M. (1977). Trade using one commodity as a means of payment. *Journal of political economy*, *85*(5), 937–968.

Shleifer, A. (1986). Implementation cycles. *Journal of Political Economy*, *94*(6), 1163–1190.

Thakor, R. T., & Lo, A. W. (2018). Competition and r&d financing: Evidence from the biopharmaceutical industry. *MIT Sloan Research Paper*.

Veugelers, R., & Cassiman, B. (1999). Make and buy in innovation strategies: evidence from belgian manufacturing firms. *Research policy*, *28*(1), 63–80.

Wallsten, S. J. (2000). The effects of government-industry r&d programs on private r&d: the case of the small business innovation research program. *The RAND Journal of Economics*, 82–100.

3.A Appendix

3.A.1 Proofs

3.A.1.1 Proof of Proposition 3.1

Consider the optimal decision of firm i, conditional on all other firms' strategy q, and other known states. From firm i's perspective, another firm will have the technology with probability:

$$h = h(q, \lambda_{t+1}) = pq + (1 - pq)\lambda_{t+1}. \tag{3.A.1}$$

Similar to equations (3.3.16), define the state transition process conditional on the success and failure of firm i, as well as the other firms' action q (embedded in h as in equation 3.A.1):

$$S_q^i(k_t + j + 1|k_t) = \binom{N - k_t - 1}{j} h^j (1-h)^{N-k_t-1-j}, \tag{3.A.2}$$

$$S_q^{-i}(k_t + j|k_t) = \binom{N - k_t - 1}{j} h^j (1-h)^{N-k_t-1-j}, \tag{3.A.3}$$

$$\tag{3.A.4}$$

where $j \in \{0, \cdots, N - k_t - 1\}$. Denote $E_{t,q}^i[x_{t+1}]$ and $E_{t,q}^{-i}[x_{t+1}]$ accordingly.

Since firm i's deviation is unobservable, we have:

$$a_{t+1,q} = \eta(E_{t,q}^i[\Pi_{t+1}] - E_t t, q^{-i}[V_{t+1}]) \tag{3.A.5}$$

The payoff of firm i for investing in R&D is:

$$D^{make} = \beta_F(pE_{t,q}^i[\Pi_{t+1}] + (1-p)\lambda_{t+1}(E_{t,q}^i[\Pi_{t+1}] - a_{t+1,q}) + (1-p)(1-\lambda_{t+1})E_{t,q}^{-i}[V_{t+1}]) \tag{3.A.6}$$

The payoff for not investing in R&D is:

$$D^{not\ make} = \beta_F(\lambda_{t+1}(E_{t,q}^i[\Pi_{t+1}] - a_{t+1,q}) + (1 - \lambda_{t+1})E_{t,q}^{-i}[V_{t+1}]) \tag{3.A.7}$$

The net payoff of investing in R&D as:

$$\begin{aligned} D(q) &= D^{make} - D^{not\ make} \\ &= \beta_F p(E_{t,q}^i[\Pi_{t+1}] - \lambda_{t+1}(E_{t,q}^i[\Pi_{t+1}] - a_{t+1,q}) - (1 - \lambda_{t+1})E_{t,q}^{-i}[V_{t+1}]) \\ &= \beta_F p((1 - \lambda_{t+1} + \eta\lambda_{t+1})(E_{t,q}^i[\Pi_{t+1}] - E_{t,q}^{-i}[V_{t+1}])) \end{aligned} \tag{3.A.8}$$

The best response for firm i is thus:

$$q^i(q) = \begin{cases} 0 & \text{if D(q)} < \text{c} \\ 1 & \text{if D(q)} > \text{c} \\ [0,1] & \text{if D(q)} = \text{c} \end{cases} \tag{3.A.9}$$

Therefore, by restricting to a symmetric equilibrium, we have

$$q_t(k_t|\{\pi_{t'}\}_{t'=t}^{\infty}, \Lambda) = \begin{cases} 0 & \text{if D}(0) < \text{c} \\ 1 & \text{if D}(1) > \text{c} \\ q^*(k_t|\{\pi_{t'}\}_{t'=t}^{\infty}, \Lambda) & \text{otherwise} \end{cases} \tag{3.A.10}$$

where q^* is the solution to $D(q) = c$.

Note that while the existence of solution q_t is guaranteed by the above expression, the uniqueness of the solution q^* is not guaranteed. 3.A.1.4 provides the proof that $D(q)$ is decreasing in q and thus q^* is unique under a simplified situation. However, generally, it is possible that multiple equilibria exist, in a dynamic way. For example, in one equilibrium, firms invest less in R&D in earlier periods and try harder later. In another equilibrium, firms invest more in R&D earlier and less in later periods. Such multiple equilibria can be eliminated if the profit stream is decreasing fast enough over time.

3.A.1.2 Proof of Proposition 3.2

The problem in (3.3.32) can be converted to a bounded problem:

$$\max_{\lambda_{i,t+1}\in[0,1)} -g^{-1}(\lambda_{i,t+1}) + \beta_V[\frac{\log(1-\lambda_{i,t+1})}{\log(1-\lambda_{t+1})}\alpha E_t^{\ell}[l_{t+1}|\lambda_{t+1}]a_{t+1} + \frac{1}{M}E_t[R_{t+1}|\lambda_{t+1}]], \tag{3.A.11}$$

subject to

$$\lambda_{t+1} \;\; = \;\; 1-(1-\lambda_{-i,t+1})(1-\lambda_{i,t+1}) \tag{3.A.12}$$

First, the objective function is continuous on $[0, 1)$. Since $\pi_t < \infty$, the objective function will converge to $-\infty$ as $\lambda_{i,t} \to 1$. Therefore, a solution to (3.A.11) always exists. In other words, there exists a mapping $\lambda_{i,t}(\lambda_{-i,t})$ defined on the support $[0, 1]$. By definition, we must have

$$\lambda_{i,t}(0) \geq 0, \text{ and } \lambda_{i,t}(1) < 1. \tag{3.A.13}$$

Therefore, there must exists $\lambda_{-i,t} \in [0, 1)$ such that

$$\lambda_{i,t}(\lambda_{-i,t}) = 1-(1-\lambda_{-i,t})^{\frac{1}{M-1}}. \tag{3.A.14}$$

This is the symmetric equilibrium solution.

3.A.1.3 Proof of Proposition 3.3

Denote the variable

$$x_t = \frac{I_t}{N}, \tag{3.A.15}$$

and the function

$$f(x_t) = g(I_t) = \lambda_{t+1}. \tag{3.A.16}$$

Since startups are more productive in innovation, we have

$$f(c) \geq p. \tag{3.A.17}$$

By the assumption in equation (3.3.42), we must have

$$f(x+y) = 1 - (1-f(x))(1-f(y)) \tag{3.A.18}$$

Given any choice (q_t, I_t), where $q_t > 0$, I show that the planner can achieve a higher h_{t+1} and thus higher net payoff by transferring all corporate R&D investment to startups.

Current corporate R&D investment per firm is $q_t c$, and current success probability per firm is

$$h_{t+1} = q_t p + (1-q_t p)\lambda_{t+1} = q_t p + (1-q_t p) f(x_t). \tag{3.A.19}$$

In the alternative allocation where all corporate R&D investment is transferred to startups, we have

$$h_{t+1} = f(x_t + q_t c). \tag{3.A.20}$$

It remains to show

$$f(x_t + q_t c) \geq q_t p + (1-q_t p) f(x_t). \tag{3.A.21}$$

By 3.A.17, we have

$$\begin{aligned} q_t p + (1-q_t p) f(x_t) &= f(x_t) + q_t p(1-f(x_t)) \\ &\leq f(x_t) + q_t f(c)(1-f(x_t)) \\ &\leq f(x_t) + f(q_t c)(1-f(x_t)) \end{aligned} \tag{3.A.22}$$

The last line comes from Jensen's inequality

$$q_t f(c) = (1-q_t) f(0) + q_t f(c) \leq f((1-q_t)0 + q_t c) = f(q_t c). \tag{3.A.23}$$

Then we have

$$\begin{aligned} q_t p + (1-q_t p) f(x_t) &\leq f(x_t) + f(q_t c)(1-f(x_t)) \\ &= 1-(1-f(x_t))(1-f(q_t c)) \\ &= f(x_t + q_t c) \end{aligned} \tag{3.A.24}$$

3.A.1.4 Proof of Corollary 3.1

Consider

$$D(q,\lambda|\pi) \quad = \quad \beta_F p(1-\lambda+\eta\lambda)\pi\frac{1-(1-h)^N}{Nh} \tag{3.A.25}$$

and

$$h = qp + (1-qp)\lambda. \tag{3.A.26}$$

Note that D is linear in π, redefine

$$D(q,\lambda|\pi) = \pi d(q,\lambda), \tag{3.A.27}$$

where

$$d(q,\lambda) = \beta_F p(1-\lambda+\eta\lambda)\frac{1-(1-h)^N}{Nh}. \tag{3.A.28}$$

Firstly, notice all terms in $d(q,\lambda)$ are positive. Secondly, the term $(1-\lambda+\eta\lambda)$ is decreasing in λ. Thirdly, the term h is increasing in λ.

Now we prove the last term $\frac{1-(1-h)^N}{Nh}$ is decreasing in h, and thus λ.

Consider the first order derivative of $\frac{1-(1-h)^N}{Nh}$:

$$\frac{d(\frac{1-(1-h)^N}{Nh})}{dh} = \frac{Nh(1-h)^{N-1}-[1-(1-h)^N]}{Nh^2} \tag{3.A.29}$$

Define the function

$$r(x) = 1-(1-x)^N, x\in[0,1]. \tag{3.A.30}$$

We have

$$r'(x) = N(1-x)^{N-1} > 0, \tag{3.A.31}$$

and

$$r''(x) = -N(N-1)(1-x)^{N-2} < 0. \tag{3.A.32}$$

Therefore, for any $x\in[0,1]$, we have

$$r(0) = r(x) + r'(x)(0-x) + r''(\hat{x})x^2 < r(x) - xr'(x). \tag{3.A.33}$$

Since $r(0)=0$, this implies

$$r(x) > xr'(x), \forall x\in[0,1]. \tag{3.A.34}$$

Therefore, by substituting $x=h$, we have

$$1-(1-h)^N > Nh(1-h)^{N-1} \tag{3.A.35}$$

Therefore,

$$\frac{\partial(\frac{1-(1-h)^N}{Nh})}{\partial h} = \frac{Nh(1-h)^{N-1} - [1-(1-h)^N]}{Nh^2} < 0. \tag{3.A.36}$$

As λ increases, the value of d decreases for all q, and thus requires a lower q to maintain $d(q,\lambda) = c/\pi$. In other words, q is decreasing in λ.

We now define the thresholds $\underline{\lambda}(\pi)$ and $\overline{\lambda}(\pi)$.

$$\underline{\lambda}(\pi) = \begin{cases} 0 & \text{if } d(1,0) < c/\pi \\ 1 & \text{if } d(1,1) > c/\pi \\ \text{solution to } d(1,\lambda) = c/\pi & \text{otherwise} \end{cases} \tag{3.A.37}$$

$$\overline{\lambda}(\pi) = \begin{cases} 0 & \text{if } d(0,0) < c/\pi \\ 1 & \text{if } d(0,1) > c/\pi \\ \text{solution to } d(0,\lambda) = c/\pi & \text{otherwise} \end{cases} \tag{3.A.38}$$

3.A.1.5 Proof of Corollary 3.2

From the expressions in 3.A.37 and 3.A.38, the result follows from the property that d is decreasing in λ and c/π is decreasing in π.

3.A.1.6 Proof of Lemma 3.2

$$\begin{aligned} R^i(I_i) &= -I_i + \beta(\frac{I_i}{I_i + I_{-i}}\alpha E_0^\ell[l_1|\lambda]a_1 \\ &= -I_i + \beta(\frac{I_i}{I_i + I_{-i}}\alpha N(1-qp)\lambda a(q,\lambda|\pi) \end{aligned} \tag{3.A.39}$$

We have

$$\frac{dR^i(I_i)}{dI_i} = \frac{\partial R}{\partial I_i} + \frac{\partial R}{\partial \lambda}\frac{\partial \lambda}{\partial I}\frac{\partial I}{\partial I_i} + \frac{\partial R}{\partial q}\frac{\partial q}{\partial \lambda}\frac{\partial \lambda}{\partial I}\frac{\partial I}{\partial I_i}, \tag{3.A.40}$$

where each term is well-defined everywhere except $\frac{\partial q}{\partial \lambda}$ at $\underline{\lambda}(\pi)$ and $\overline{\lambda}(\pi)$.

3.A.1.7 Proof of Proposition 3.4

From Lemma 3.2, the optimal value of I_i should be selected from the following candidates:

1. Corner solution $I_i = 0$

2. Internal solution where $\frac{dR^i(I_i)}{dI_i} = 0$, and $\frac{d^2R^i(I_i)}{dI_i^2} \leq 0$.
3. Critical points $g^{-1}(\underline{\lambda}(\pi)) - I_{-i}$ and $g^{-1}(\underline{\lambda}(\pi)) - I_{-i}$.

As π changes continuously, although each of the above candidates is continuous in π, the optimal point can jump from one candidate to another, thus lead to discontinuity. For example, when a jump from critical point $g^{-1}(\underline{\gamma}(\pi)) - I_{-i}$ to internal solution happens, firms can no longer afford deterring R&D. This example is also reflected in Figure 3.3.

3.A.1.8 Proof of Proposition 3.5

First, it is straightforward to show that $q(\pi) = q^e(\pi)$ when $N = 1$. We now focus on showing $q(\pi) \geq q^e(\pi)$ when $N > 1$.

Observe the conditions for $q = 0$ in both cases are identical, which means both parties start investing in R&D at the same threshold of π. On the other hand, the thresholds at which $q = 1$ are different in the two cases. In fact, it requires lower π for $q(\pi)$ to achieve 1 than $q^e(\pi)$

Define the function

$$r(x) = 1 - (1-x)^N, x \in [0,1]. \tag{3.A.41}$$

We have

$$r'(x) = N(1-x)^{N-1} > 0, \tag{3.A.42}$$

and

$$r''(x) = -N(N-1)(1-x)^{N-2} < 0. \tag{3.A.43}$$

Therefore, for any $x \in [0,1]$, we have

$$r(0) = r(x) + r'(x)(0-x) + r''(\hat{x})x^2 < r(x) - xr'(x). \tag{3.A.44}$$

Since $r(0) = 0$, this implies

$$r(x) > xr'(x), \forall x \in [0,1]. \tag{3.A.45}$$

Let $x = pq$, we have

$$1 - (1-pq)^N > Npq(1-pq)^{N-1}, \forall q \in [0,1]. \tag{3.A.46}$$

Specifically, when $q = 1$:

$$1 - (1-p)^N > Np(1-p)^{N-1}. \tag{3.A.47}$$

Therefore,

$$\beta\pi\frac{1-(1-p)^N}{N} > \beta p\pi(1-p)^{N-1}, \tag{3.A.48}$$

which implies that $q(\pi)$ achieves 1 faster than $q^e(\pi)$.

Now we compare the case where both qs are interior solutions. $q(\pi)$ is the solution to

$$\beta\pi r(pq) = qNc, \tag{3.A.49}$$

whereas $q^e(\pi)$ is the solution to

$$\beta\pi pqr'(pq) = qNc. \tag{3.A.50}$$

Since $r(pq) > pqr'(qp)$, we have

$$\beta\pi pq(\pi)r'(pq(\pi)) < \beta\pi r(pq(\pi)) = q(\pi)Nc, \tag{3.A.51}$$

which implies

$$\beta\pi pr'(pq(\pi)) < Nc. \tag{3.A.52}$$

Since r' is decreasing in q, the solution $q^e(\pi) \leq q(\pi)$.

3.A.1.9 Proof of Proposition 3.6

We have

$$a_1 \;=\; \eta E^i[\Pi] = \eta \sum_{j=0}^{N-1} \binom{N-1}{j} \lambda^j (1-\lambda)^{N-1-j} ((1-\theta\lambda)\frac{\pi}{j+1} + \theta\lambda\frac{\pi}{j+2}) \tag{3.A.53}$$

(3.A.54)

Therefore, we have

$$\begin{aligned}
E^\ell[l_1]a_1 &= \eta E^\ell[l_1]E^i[\Pi] \\
&\leq E^\ell[l_1]E^i[\Pi] \\
&= N\lambda \sum_{j=0}^{N-1} \binom{N-1}{j} \lambda^j (1-\lambda)^{N-1-j} \frac{\pi}{j+1} \\
&= N\lambda \sum_{j=0}^{N-1} \binom{N-1}{j} \lambda^j (1-\lambda)^{N-1-j} \frac{\pi}{j+1} \\
&= \pi \sum_{j=0}^{N-1} \frac{N!}{(j+1)!(N-1-j)!} \lambda^{j+1} (1-\lambda)^{N-1-j} \\
&= \pi \sum_{j'=1}^{N} \frac{N!}{j'!(N-j')!} \lambda^{j'} (1-\lambda)^{N-j'} \\
&= \pi(1-(1-\lambda)^N)
\end{aligned} \tag{3.A.55}$$

In a similar derivation, we can also show

$$\frac{\partial(\beta_V \alpha E^{\ell}[l_1]a_1)}{\partial\lambda} \leq \frac{\partial(\beta\pi(1-(1-\lambda)^N)}{\partial\lambda}. \tag{3.A.56}$$

In other words, the VC and the planner has the same marginal cost in increasing λ, but gets lower marginal benefit due to low β_V, low α, and low η. As a result, the VC will under-invest. $I(\pi) \leq I^e(\pi)$.

3.A.1.10 Proof of Proposition 3.7

The VC's problem is now written as

$$\max_{I_i} -I_i + \beta\pi\frac{I_i}{I_i + I_{-i}}(1-(1-\lambda)^N), \tag{3.A.57}$$

subject to

$$\lambda = g(I_i + I_{-i}) \tag{3.A.58}$$

The VC's optimal investment is characterized by the first order condition

$$-1 + \beta\pi\frac{I_i}{I}\frac{\partial(1-(1-\lambda)^N)}{\partial\lambda}\frac{\partial\lambda}{\partial I} + \beta\pi(1-(1-\lambda)^N)\frac{I-I_i}{I^2} = 0. \tag{3.A.59}$$

In a symmetric equilibrium, we have

$$-1 + \frac{1}{M}\beta\pi\frac{\partial(1-(1-\lambda)^N)}{\partial\lambda}\frac{\partial\lambda}{\partial I} + \frac{1}{I}\frac{M-1}{M}\beta\pi(1-(1-\lambda)^N) = 0. \tag{3.A.60}$$

Note that when $M=1$, this is identical to the planner's first order condition

$$-1 + \beta\pi\frac{\partial(1-(1-\lambda)^N)}{\partial\lambda}\frac{\partial\lambda}{\partial I} = 0. \tag{3.A.61}$$

Moreover, when $M > 1$ we have

$$\frac{\partial(1-(1-\lambda)^N)}{\partial\lambda}\frac{\partial\lambda}{\partial I} < \frac{1}{I}(1-(1-\lambda)^N), \tag{3.A.62}$$

the intuition of which follows from $r(x) > xr'(x)$ for a concave function $r(\cdot)$ where $r(0)=0$ (see 3.A.1.8 for more detail).

As a result, since both expressions are decreasing in I, we have $I(\pi) \geq I^e(\pi)$.

3.A.2 Procedures of Numeric Solutions

3.A.2.1 Numeric solution to firms' R&D decision

To obtain a starting point of the backward induction process, one can pick an arbitrarily small $\varepsilon > 0$, and find T such that

$$\sum_{t'=T}^{\infty} \beta_F \pi_{t'-T} < \varepsilon. \tag{3.A.63}$$

Note that the existence of such T is guaranteed by (3.3.1). Then we can approximate

$$\Pi_t = 0,\ V_t = 0\, \forall t \geq T. \tag{3.A.64}$$

The following steps illustrate the procedures in backward induction.

- Create 4 matrices for Π, V, Q, and Λ, each with dimension $(T+1) \times (N+1)$[18]. For convenience, the indexing of all matrices and vectors starts from 0. The element indexed by (t,k) of matrix X indicates $X_t(k|\{\pi_t\}_{t=0}^{\infty}, \Lambda)$ respectively.
- Fill in all elements of Λ with predetermined VC investment strategy. Fill in $\Pi(T,:)$, $V(T,:)$, and $Q(T,:,:)$ with zeros.
- For $t = T-1$ to 0 (step=-1) (backward induction, must be done in this order):

  For k in $\{0, \cdots, N\}$ (can be vectorized or parallelized):

  a) Solve for $q_t(k+j)$ according to Proposition 3.1, and fill it in $Q(t,k)$. Note that it is necessary to construct vectors of state transition probabilities as in equations (3.A.2) - (3.A.3) as an intermediate step. With $D(q)$ expressed as a function of the free variable q. Evaluate $D(0)$ and $D(1)$ to check for corner solutions. If solution is interior, any preferred package of `fsolve` or `bisect` can be used to obtain the solution[19].
  b) Construct vectors of state transition probabilities as in equations (3.3.12) and (3.4.14) - (3.3.18). Plug the above solution of q_t in equations 3.3.22 and 3.3.23. Fill in $\Pi(t,k)$ and $V(t,k)$ accordingly.

[18] A long profit stream and a large N will require a large memory space, and might not be computationally feasible

[19] This algorithm guarantees that a solution will be returned. However, it does not provide control over equilibrium selection in the case where multiple solutions exist.

3.A.2.2 Numeric solution to optimal VC investment

- Create 6 matrices for Π, V, Q ,I,Λ, R, each with dimension $(T+1)\times(N+1)$. Same as above, the indexing of all matrices and vectors starts from 0. The element indexed by (t,k) of matrix X (except Λ) indicates $X_t(k+j|\{\pi_t\}_{t=0}^{\infty},\Lambda_{t+1})$ respectively. And the element (t,k) in Λ indicates $\Lambda_{t+1}(k+j|\{\pi_t\}_{t=0}^{\infty})$.

- Fill in $\Pi(T,:,:)$, $V(T,:,:)$, $Q(T,:,:)$, $I(T,:,:)$, $R(T,:,:)$, and $\Lambda(T,:,:)$ with zeros.

- For $t=T-1$ to 0 (step=-1) (backward induction, must be done in this order):

  For k in $\{0,\cdots,N\}$ (can be vectorized or parallelized):

  a) Let $k_t=k$. Create a grid vector on $[0,1)$ representing possible values of $\lambda_{-i,t+1}$. For each $\lambda_{-i,t+1}$ (can be vectorized or parallelized):
     - Define a function of the variable $\lambda_{i,t+1}$ representing R_t^i, in which:
       * λ_{t+1} is computed according to equation (3.A.12).
       * By combining λ_{t+1} and Λ_{t+2} (represented by $\Lambda(t+1:,:)$), use the algorithm in 3.A.2.1 to solve q_t.
       * Plug in q_t and λ_{t+1} to equations (3.3.12) - (3.3.18) to obtain all state transition probabilities. Compute $E_t[R_{t+1}|\lambda_{t+1}]$ accordingly. Compute a_{t+1} from the newly computed state transition probabilities and pre-computed $\Pi(t+1,:,:)$, $V(t+1,:,:)$ based on equation (3.3.24).
       * Plug in q_t and λ_{t+1} to equation 3.3.29 and get $E_t^{\ell}[l_{t+1}]$.
       * Combine all terms in the objective function 3.A.11.
     - Use preferred optimizer[20] to obtain best response $\lambda_{i,t+1}(\lambda_{-i,t+1})$.

  b) Interpolate the best response function using preferred method, and select the symmetric equilibrium where

  $$(1-\lambda_{i,t+1})^M = 1-\lambda_{t+1} = (1-\lambda_{i,t+1})(1-\lambda_{-i,t+1}). \tag{3.A.65}$$

  c) From the equilibrium λ_{t+1}, solve and fill in the equilibrium λ_{t+1} in $\Lambda(t,k)$. Compute the corresponding I_t and fill in $I(t,k)$. Solve for the firms' equilibrium as in 3.A.2.1, and fill in $Q(t,k)$, $\Pi(t,k)$ and $V(t,k)$ accordingly. Compute R_t and fill in $R(t,k)$.

[20] The problem does not qualify for first order approach. Brute force search approach on the support is recommended.

3.A.2.3 Extension: Numeric solution to firms' R&D decision with IPO exits

Same as 3.A.2.1, we can pick an arbitrarily small $\varepsilon > 0$, and find T such that

$$\sum_{t'=T}^{\infty} \beta_F \pi_{t'-T} < \varepsilon. \tag{3.A.66}$$

Similar to equations (3.A.2) and (3.A.3), define

$$\begin{aligned}
S_q^i(k_t+j+1, k_t^- - 1 - j | k_t, k_t^-) &= (1-\theta\lambda_{t+1})\binom{k_t^- - 1}{j} h^j (1-h)^{k_t^- - 1 - j}, && (3.A.67)\\
S_q^i(k_t+j+2, k_t^- - 1 - j | k_t, k_t^-) &= \theta\lambda_{t+1}\binom{k_t^- - 1}{j} h^j (1-h)^{k_t^- - 1 - j}, && (3.A.68)\\
S_q^{-i}(k_t+j, k_t^- - j | k_t, k_t^-) &= (1-\theta\lambda_{t+1})\binom{k_t^- - 1}{j} h^j (1-h)^{k_t^- - 1 - j}, && (3.A.69)\\
S_q^{-i}(k_t+j+1, k_t^- - j | k_t, k_t^-) &= \theta\lambda_{t+1}\binom{k_t^- - 1}{j} h^j (1-h)^{k_t^- - 1 - j}, && (3.A.70)
\end{aligned}$$

The following steps illustrate the procedures in backward induction.

- Create 4 matrices for Π, V, Q, and Λ, each with dimension $(T+1)\times(N+1)\times(T+1)$[21]. For convenience, the indexing of all matrices and vectors starts from 0. The element indexed by (t,k,j) of matrix X indicates $X_t(k+j, N-k|\{\pi_t\}_{t=0}^{\infty}, \Lambda)$ respectively.
- Fill in all elements of Λ with predetermined VC investment strategy. Fill in $\Pi(T,:,:)$, $V(T,:,:)$, and $Q(T,:,:)$ with zeros.
- For $t = T-1$ to 0 (step=-1) (backward induction, must be done in this order):

    For k in $\{0,\cdots,N\}$ (can be vectorized or parallelized):

    For j in $\{0,\cdots,t\}$ (can be vectorized or parallelized):

    i. Solve for $q_t(k+j, N-k)$ according to Proposition 3.1, and fill it in $Q(t,k,j)$. Note that it is necessary to construct vectors of state transition probabilities as in equations 3.A.67 - 3.A.70 as an intermediate step. With $D(q)$ expressed as a function of the free variable q. Evaluate $D(0)$ and $D(1)$ to check for corner solutions. If solution is interior, any preferred package of `fsolve` or `bisect` can be used to obtain the solution[22].

[21] A long profit stream and a large N will require a large memory space, and might not be computationally feasible

[22] This algorithm guarantees that a solution will be returned. However, it does not provide control over equilibrium selection in the case where multiple solutions exist.

ii. Construct vectors of state transition probabilities as in equations 3.4.10 - 3.4.11 and 3.4.14 - 3.4.17. Plug the above solution of q_t in equations 3.3.22 and 3.3.23. Fill in $\Pi(t,k,j)$ and $V(t,k,j)$ accordingly.

3.A.2.4 Extension: Numeric solution to optimal VC investment with IPO exits

In this section, we specify the functional form of G to be

$$G(I_t) = 1 - \exp(-\frac{I_t}{\varphi}). \tag{3.A.71}$$

First, rewrite the VC's problem in (3.3.32) so that the choice variable is bounded:

$$\max_{\gamma_{i,t+1}\in[0,1]} -G^{-1}(\gamma_{i,t+1})+\beta_V[\frac{\log(1-\gamma_{i,t+1})}{\log(1-\gamma_{t+1})}(\alpha E_t^{\ell}[l_{t+1}|\lambda_{t+1}]a_{t+1}+\theta\lambda_{t+1}\alpha o_{t+1})+\frac{1}{M}E_t[R_{t+1}|\lambda_{t+1}]], \tag{3.A.72}$$

subject to

$$\gamma_{t+1} = 1-(1-\gamma_{-i,t+1})(1-\gamma_{i,t+1}) \tag{3.A.73}$$

$$\gamma_{t+1} = 1-(1-\lambda_{t+1})^N(1-\theta\lambda_{t+1}) \tag{3.A.74}$$

- Create 6 matrices for Π, V, Q ,I,Λ, R, each with dimension $(T+1)\times(N+1)\times(T+1)$. Same as above, the indexing of all matrices and vectors starts from 0. The element indexed by (t,k,j) of matrix X (except Λ) indicates $X_t(k+j, N-k|\{\pi_t\}_{t=0}^{\infty}, \Lambda_{t+1})$ respectively. And the element (t,k,j) in Λ indicates $\Lambda_{t+1}(k+j, N-k|\{\pi_t\}_{t=0}^{\infty})$.
- Fill in $\Pi(T,:,:)$, $V(T,:,:)$, $Q(T,:,:)$, $I(T,:,:)$, $R(T,:,:)$, and $\Lambda(T,:,:)$ with zeros.
- For $t = T-1$ to 0 (step=-1) (backward induction, must be done in this order):

  For k in $\{0,\cdots,N\}$ (can be vectorized or parallelized):

  For j in $\{0,\cdots,t\}$ (can be vectorized or parallelized):

  i. Let $k_t = k+j$, $k_t^- = N-k$.

  ii. Create a grid vector on $[0,1)$ representing possible values of $\gamma_{-i,t+1}$. For each $\gamma_{-i,t+1}$ (can be vectorized or parallelized):

  - Define a function of the variable $\gamma_{i,t+1}$ representing R_t^i, in which:
    * γ_{t+1} is computed according to equation (3.A.73), λ_{t+1} is computed according to equation (3.A.74).
    * By combining λ_{t+1} and Λ_{t+2} (represented by $\Lambda(t+1\ :,:,:)$), use the algorithm in 3.A.2.1 to solve q_t.

    * Plug in q_t and λ_{t+1} to equations 3.4.10 - 3.4.17 to obtain all state transition probabilities. Compute $E_t[R_{t+1}|\lambda_{t+1}]$ accordingly. Compute a_{t+1} and o_{t+1} from the newly computed state transition probabilities and pre-computed $\Pi(t+1,:,:)$, $V(t+1,:,:)$ based on equations 3.3.24, 3.4.26.
    * Plug in q_t and λ_{t+1} to equation 3.3.29 and get $E_t^{\ell}[l_{t+1}]$.
    * Combine all terms in the objective function 3.A.72.
  - Use preferred optimizer[23] to obtain best response $\gamma_{i,t+1}(\gamma_{-i,t+1})$.

iii. Interpolate the best response function using preferred method, and select the symmetric equilibrium where

$$(1-\gamma_{i,t+1})^M = 1-\gamma_{t+1} = (1-\gamma_{i,t+1})(1-\gamma_{-i,t+1}). \tag{3.A.75}$$

iv. From the equilibrium γ_{t+1}, solve and fill in the equilibrium λ_{t+1} in $\Lambda(t,k,j)$. Compute the corresponding I_t and fill in $I(t,k,j)$. Solve for the firms' equilibrium as in 3.A.2.1, and fill in $Q(t,k,j)$, $\Pi(t,k,j)$ and $V(t,k,j)$ accordingly. Compute R_t and fill in $R(t,k,j)$.

[23] The problem does not qualify for first order approach. Brute force search approach on the support is recommended.

3.A.3 Figures

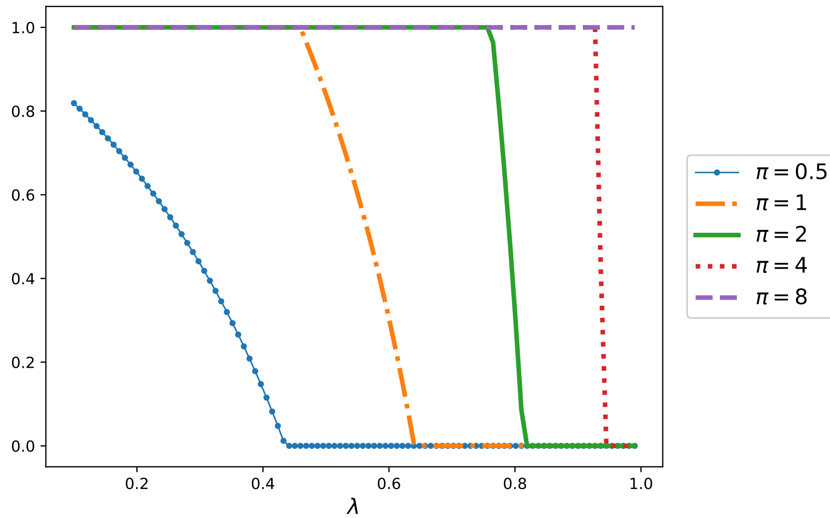


Figure 3.1: $q(\pi, \lambda)$[24]

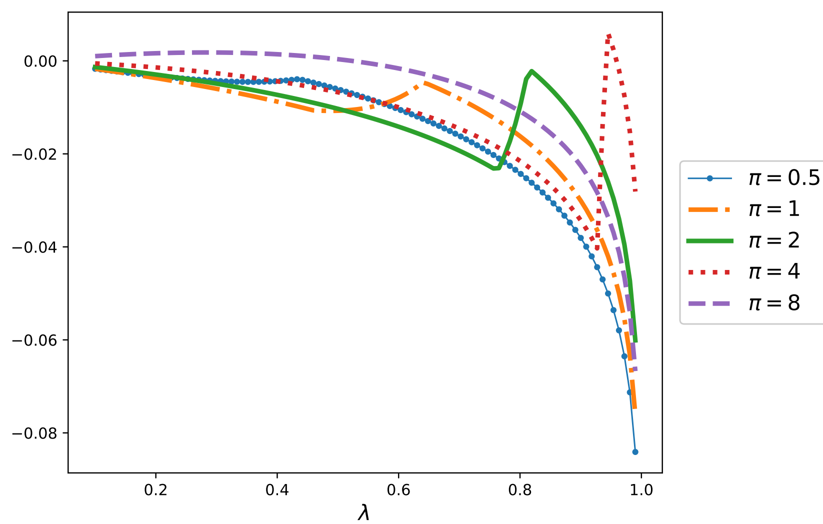
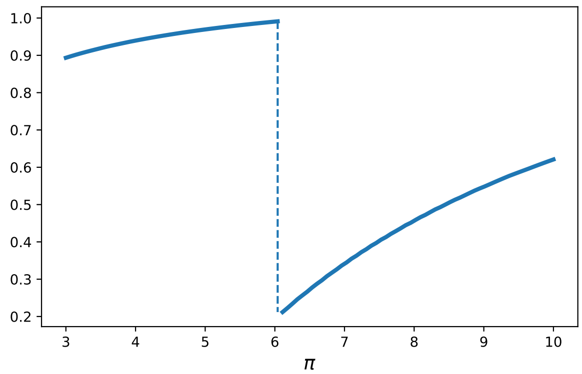


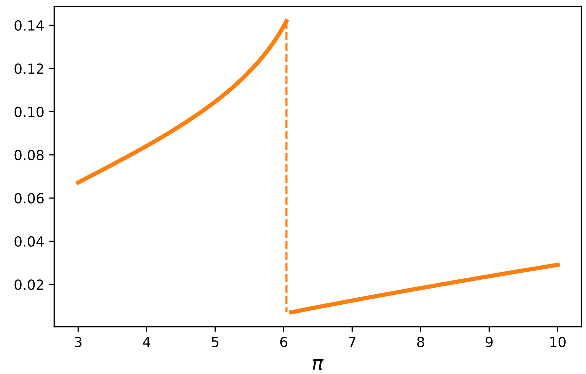
Figure 3.2: $R(\pi, \lambda)$[25]

[24]Choice of parameters: $N = 5, \beta_F = 0.95, p = 0.8, c = 0.1, \eta = 0.1$

[25]Choice of parameters: $N = 5, \beta_F = 0.95, \beta_V = 0.8, p = 0.8, c = 0.1, \eta = 0.1, \alpha = 0.2, \phi = 0.004$



(a) $\lambda(\pi)$



(b) $I(\pi)$

Figure 3.3: Optimal VC's investment[26]

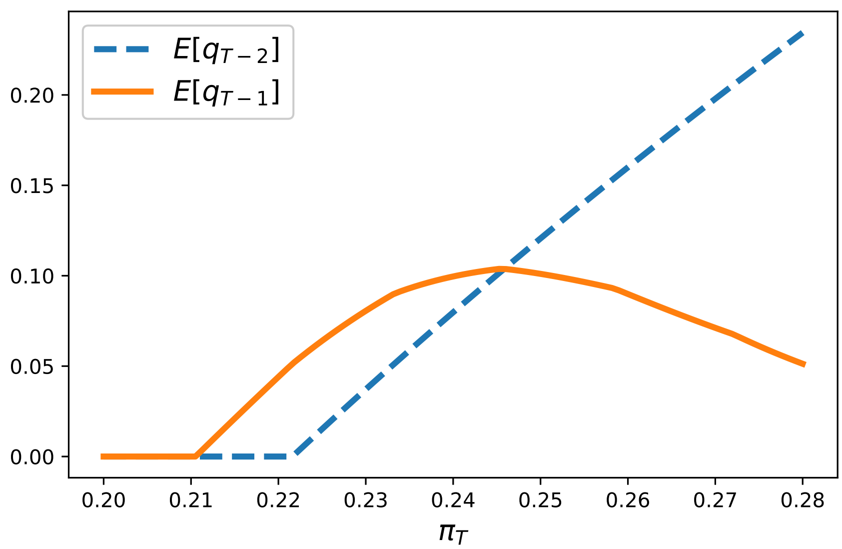


Figure 3.4: R&D investment for a future technology[27]

[26]Choice of parameters: $N = 5$, $M = 5$, $\beta_F = 0.95$, $p = 0.8$, $c = 0.1$, $\eta = 0.1$,$\alpha = 0.2$, $\phi = 0.004$
[27]Choice of parameters: $T = 2$, $N = 5$, $\beta_F = 0.95$, $p = 0.5$, $c = 0.1$
[28]Choice of parameters:$N = 5$, $\beta_F = 0.95$, $p = 0.5$, $c = 0.1$

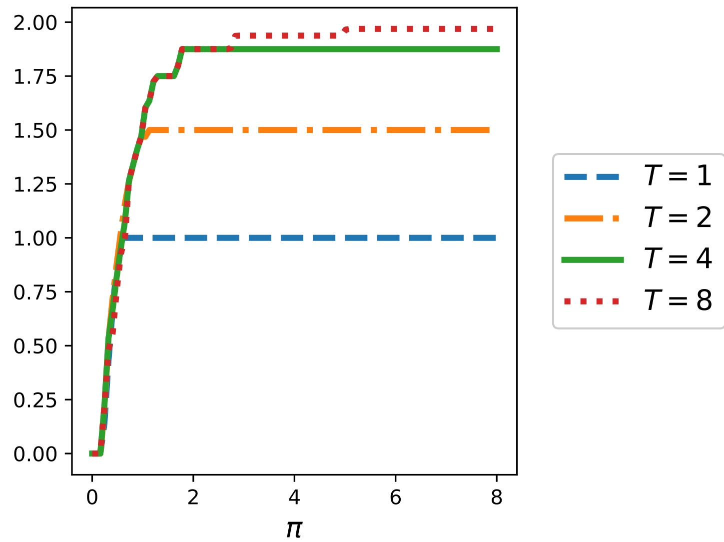


Figure 3.5: Expected average rounds of R&D for a future technology[28]

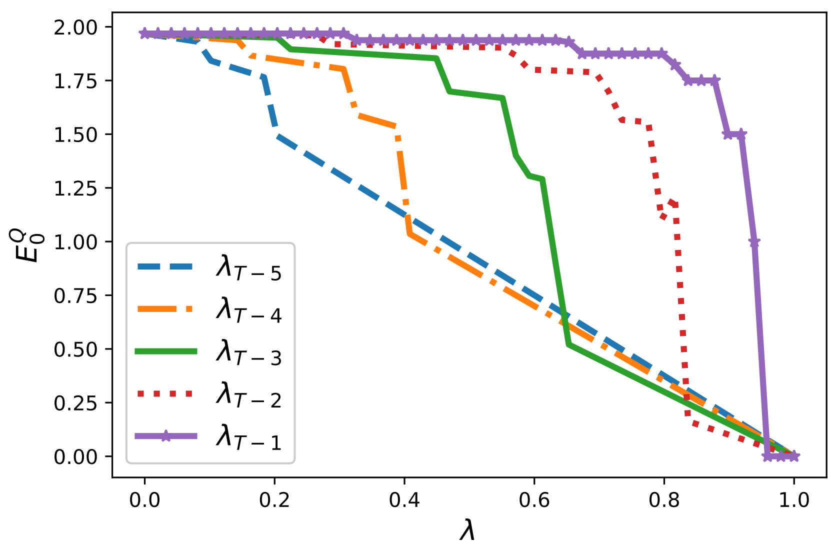


Figure 3.6: Effect of VC investment on lifetime R&D[29]

[29]Choice of parameters:$N = 5, \pi = 7, T = 10, \beta_F = 0.95, p = 0.5, c = 0.1$

[30]Choice of parameters: $T = 5$, $N = 5$, $M = 5$, $\beta_F = 0.95$, $\beta_V = 0.8$, $p = 0.8$, $c = 0.1$, $\eta = 0.1$,$\alpha = 0.2$, $\phi = 0.002$

[31] Choice of parameters: $\pi = 1, T = 2, N = 5, M = 5, \beta_F = 0.95, \beta_V = 0.8, p = 0.5, c = 0.1, \eta = 0.1$,$\alpha = 0.2$, $\phi = 0.002$

[32] Choice of parameters: $\pi = 2$, others same as 3.8.

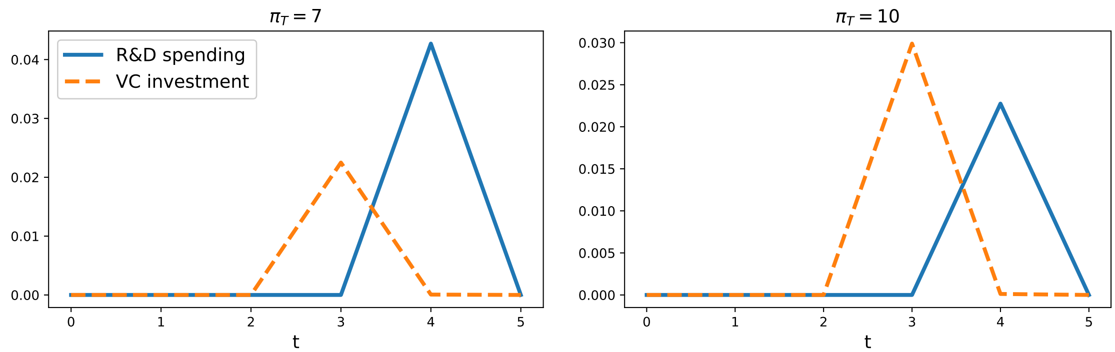
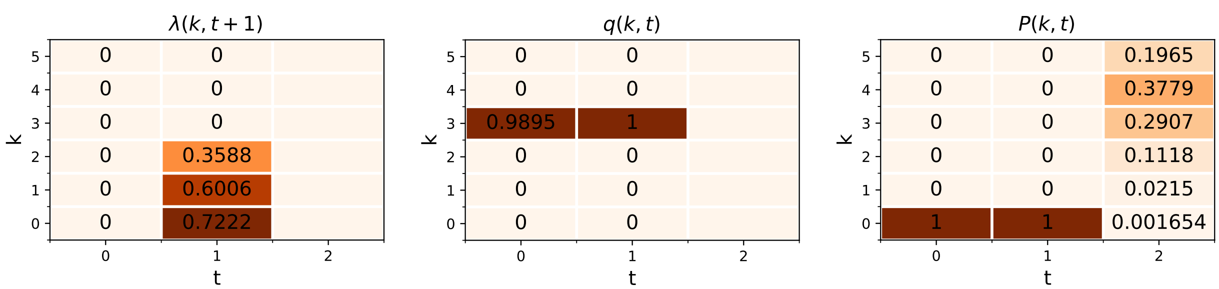
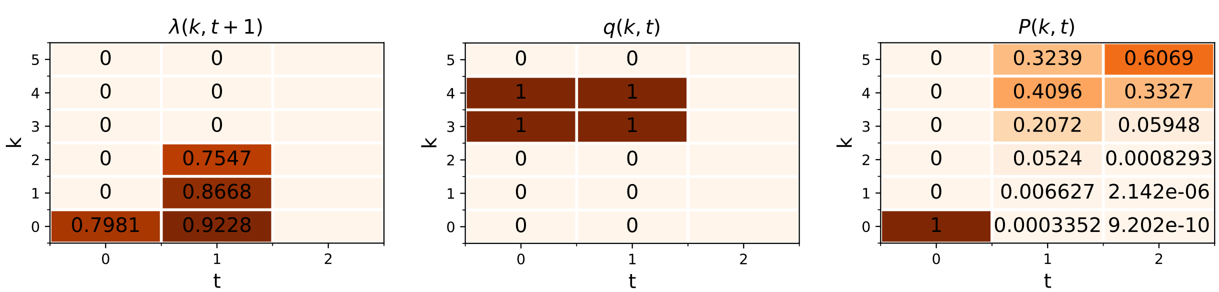


Figure 3.7: Time Series of Investment for future technologies[30]



(a) (b) (c)

Figure 3.8: Investment of future technology[31]



(a) (b) (c)

Figure 3.9: Investment of future technology[32]