Topics in Dynamic Programming and Economic Networks

Chien-Hsiang Yeh

July 2024

A thesis submitted for the degree of

Doctor of Philosophy of

The Australian National University

Declaration

Except where otherwise acknowledged, I certify that this thesis is my original work. The thesis is within the 100,000 word limit set by the Australian National University.

Signed: Chien Yeh

Chien-Hsiang Yeh

July 2024

To my parents, Shu-Jyun Li and Ching-Yu Yeh

Acknowledgements

Expressing gratitude is indeed challenging, but I want to convey my appreciation to my chair supervisor, Prof. John Stachurski. Thank you, John, for your continuous support, insightful suggestions, and encouragement. This thesis would not have been possible without John's enthusiasm for teaching, encouragement, thoughtfulness, patience, and hard work. My journey at the Australian National University (ANU) has been enriched by John's guidance, broad knowledge, inspiration, unconditional help, and support. I consider myself fortunate and honored to be one of John's students and to have the opportunity to work under his guidance on various projects.

I extend my sincere thanks to my panel supervisors, Assoc Prof. Timothy Kam and Prof. Fedor Iskhakov. Their guidance, wealth of knowledge, economic insights, and fantastic support have played a pivotal role in shaping this work. Working under their supervision has been an honor and a privilege.

I am also grateful to Dr. Tina Kao, Dr. Ruitian Lang, Assoc. Prof. Sephorah Mangin, Dr. Juergen Meinecke, Assoc. Prof. Idione Meneghel, Assoc. Prof. Chung Tran, and Dr. Thomas Yang for their invaluable comments, insightful discussions, and constructive feedback during my studies at ANU. Learning from the exceptional faculty members at the Research School of Economics has been a true privilege.

I want to convey my sincere appreciation to my office mate, Miss Qingsong Song, for the uplifting encouragement and the delightful friendship we share. I would also like to express my gratitude to my friends and PhD fellows: Dr. Sheung Chi Chow, Mr. Chaonuo Dai, Mr. Yunshan Dai, Mr. Shu Hu, Mr. Jiacheng Kang, Mr. Jiu Lian, Miss Haiyan Lin, Miss Dan Liu, Mr. Wending Liu, Mr. Jianhua Mei, Mr. Simon Mishricky, Dr. Ieng Man Ng, Mr. Fajar Oktiyanto, Miss Nisha Peng, Miss Van Anh Pham, Mr. Darapheak Tin, Mr. Kazuki Tomioka, Mr. Zeming Wang, Dr. Yurui Zhang, Miss Qiu Zhen, and

Miss Yanran Zhu. Their warm-hearted encouragement and sincere assistance have been invaluable throughout this journey.

A special mention of gratitude goes to my longtime friends, Mr. Ting-Wei Chen, Dr. Siao-Fong Li, and Mr. Daniel Wu, for their unwavering encouragement and sincere support.

Special thanks are due to my family for their boundless love and endless support. Their encouragement and the freedom they have given me are cherished beyond words.

Chien-Hsiang Yeh
July 2024

Abstract

Dynamic programming is a recursive method for finding optimal decision rules for sequential decision-making problems under uncertainty. It is the most fundamental tool for modern economics, addressing topics that include economic growth, optimal saving, asset pricing, business cycles, and fiscal and monetary policies. Complementing this, economic networks stand as another critical tool, recognising the dynamic interactions among a multitude of diverse agents. The network models provide a comprehensive analysis of how individual agent behaviours, incentives, and strategic interactions contribute to the larger systemic forces or shocks within aggregate economics.

This thesis addresses problems in dynamic programming and static economic networks. A key focus is on dynamic programming theory, specifically integrating both state-dependent and action-dependent discounting, as opposed to conventional constant discounting. The core contribution involves establishing a comprehensive optimality theory in dynamic programming in the context of eventual discounting.

Moreover, the research expands the concept of state-action-dependent discounting in dynamic programming to model-free reinforcement learning. The thesis demonstrates the convergence of Q-learning when subjected to the condition of eventual discounting.

In addition to dynamic programming, the thesis investigates a unified static network that includes production networks, network games, and financial networks. Our results show the existence and (almost sure) uniqueness of equilibrium in both eventually contracting and non-expansive frameworks.

Finally, employing dynamic programming techniques, the thesis constructs an endogenous production network characterised by a unique equilibrium, featuring robust characteristics such as a power-law distribution of firm sizes. This multifaceted exploration contributes insights to the understanding of the existence and uniqueness of equilibrium in dynamic programming and economic networks.

Table of Contents

CHAPTER 1

Overview of Thesis

This thesis studies problems in dynamic programming and static economic networks. Each chapter contributes to understanding the existence and uniqueness of equilibrium or fixed points within different economic paradigms, as well as providing insights into associated computational methodologies. The exploration is structured in four chapters, each of which is summarised below.

Chapter 2 extends discrete-time dynamic programming to account for state-action-dependent discounting, departing from the conventional constant discounting approach. This chapter is motivated by the fact that time preferences are influenced by subjective assessments and exogenous uncertainties. In particular, this chapter addresses a gap in the existing literature on action-dependent preferences from influential studies such as Uzawa (1968) and Becker and Mulligan (1997).

We show that the standard optimality results remain valid in the existence of "eventual discounting," where the expected multiplicative of discount factors eventually falls below one for any policies. The concept of eventual discounting is succinctly illustrated below. Considering state and action spaces, denoted by X and A respectively, and for all action policies $\sigma \colon \mathsf{X} \to \mathsf{A}$, we suppose that there exists an $n_\sigma \in \mathbb{N}$ such that

$$\sup_x \mathbb{E}_x \prod_{t=0}^{n_\sigma - 1} \beta(X_t, \sigma(X_t), X_{t+1}) < 1,$$

where $\beta \colon \mathsf{X} \times \mathsf{A} \times \mathsf{X} \to \mathbb{R}_+$ represents the function of state-action-dependent discount factors. The concept of eventual discounting allows the discount factors to exceed one with arbitrary probability, encompassing frameworks such as Hills and Nakata (2018) and Hubmer et al. (2021).

This condition of eventual discounting leads to the (eventual) contraction of policy operators and Bellman operators, ensuring standard optimality results and the convergences of key iterative methods in dynamic programming. Moreover, we address the necessity of eventual discounting for the existence of both policy values and a solution to the Bellman equation in compact state spaces with appropriate assumptions.

In addition, we extend these results to dynamic programming with unbounded rewards by utilising the Q-transform outlined in Ma et al. (2022). Even with unbounded rewards, the standard optimality of dynamic programming is maintained when discount factors are state-action-dependent and adhere to eventual discounting. Moreover, we establish the convergence of both the value function iteration and the action-value function iteration, which is the iteration of an expected action-value operator. Finally, we extend the scope of eventual discounting to applications involving risk-sensitive preferences.

In Chapter 3, we extend model-free learning algorithms by considering non-constant state-action-dependent discount factors. Specifically, we explore temporal difference learning methods, including Q-learning, SARSA, and double Q-learning, as introduced by Watkins (1989), Singh et al. (2000), and Hasselt (2010), respectively. We consider time discounting to be inherently governed by a parameterised function, where both exogenous states and subjective actions influence the discount factors. Similarly to Chapter 2, our framework allows the discount factors to exceed one with positive probability, provided that they follow the principle of eventual discounting. We show by Stochastic Approximation that eventual discounting is sufficient for ensuring the convergence of Q-learning, SARSA, and double Q-learning algorithms.

In Chapter 4, we explore a unified static network, a modification derived from Acemoglu et al. (2016b), aiming to provide an efficient and systematic method to verify the uniqueness of equilibrium across various static network models. Specifically, our focus is on exploring the equilibrium or solution of the network equation:

$$x = f(xW + \varepsilon).$$

Here, $x \in \mathbb{R}^N$ denotes a vector of economic states, $W \in \mathbb{R}^{N\times N}$ denotes the sensitivity matrix describing the structure of a network and the sensitivity extent of interactions between agents, $\varepsilon \in \mathbb{R}^N$ denotes the realisations of shocks, and $f = (f_1, \ldots, f_N)$ denotes the interaction functions specifying the total influences of agents in the networks and shocks. Each $f_j \colon \mathbb{R} \to \mathbb{R}$ is Lipschitz continuous with Lipschitz constant β_i for $j = 1, \ldots, N$.

From the literature of network games, production networks, and financial networks, we classify two types of networks: eventually contracting networks, characterised by $\rho(|W| \operatorname{diag} \beta) < 1$, and non-expansive networks, characterised by $\rho(W) = 1$ and $\beta = I$, where $\rho(A)$ denotes the spectral radius of matrix A.

We show the existence and (almost sure) uniqueness of equilibrium under both eventually contracting and non-expansive frameworks. Applying our findings, we establish that the clearing payment for the generalised Eisenberg-Noe model, studied by Liu et al. (2020), is almost surely unique under non-expansive conditions and always unique under eventually contracting conditions. Furthermore, our result serves as an alternative proof for the uniqueness of clearing payments in the Eisenberg-Noe model, without regularity conditions.

In addition, our exploration in Chapter 4 includes a discussion on computation methods and underscores the necessity of boundedness conditions when interaction functions are non-expansive. Moreover, by applying equilibrium in the study of systemic risk, we provide a measure to determine the key player whose removal from the network would have the most significant impact.

Chapter 5 attempts to explain the systemic deviation of Gibrat's Law, which posits that a firm's growth rate is independent of its size (Lucas, 1978, Simon and Bonini, 1958, Gabaix, 2009, Fu et al., 2005). This concept is pivotal in explaining firm-size distribution and the observed power law in firm sizes (Lucas, 1978). However, empirical tests often contradict this law, notably demonstrating that smaller firms exhibit greater volatility than larger ones, especially among downstream firms (Hall, 1987, Evans, 1987a,b, Almus, 2000, Calvo, 2006, Daunfeldt and Elert, 2013, Falk, 2008, Grazzi and Moschella, 2018, Oke, 2018, Aydogan and Donduran, 2019).

Motivated by these observations, Chapter 5 introduces a balanced and endogenous production chain where firms strategically choose between subcontracting and home production to minimise costs. This decision-making process is guided by the trade-off between intra-firm coordinate costs and inter-firm transaction costs, as proposed by Coase (1937) and Williamson (1979), where external transaction costs encourage firm growth while intra-firm coordination costs discourage it.

We establish the existence and uniqueness of equilibrium. Furthermore, we demonstrate that when assembly costs rise exponentially with tasks, the sizes of subcontractors downstream in a production chain follow a power law. Moreover, the breakdown of Gibrat's Law occurs when home producers experience higher technological volatility than subcontractors, resulting in greater size volatility for upstream firms. In addition, the model extends the balanced production network to a more realistic scenario where firms can choose an optimal number of suppliers during subcontracting. This extended model maintains a unique equilibrium and exhibits robust comparative statics.

In general, this thesis examines the existence and uniqueness of equilibrium or fixed points in both dynamic programming and static economic networks. Notably, the initial three chapters reveal a close relationship between the existence and uniqueness of fixed points and the spectral radii of relevant positive linear operators. Additionally, the convergence of associated computational methods is intricately influenced by these spectral radii. Lastly, the final chapter employs the techniques from network theory and dynamic programming to establish an endogenous production chain with a unique equilibrium.

CHAPTER 2

Dynamic Programming with State-Action-Dependent Discounting

2.1. Introduction

Dynamic programming or Markov decision processes is a framework for modelling and solving sequential decision-making problems. The conventional approach in dynamic programming assumes a constant discount factor to capture the trade-off between present and future rewards. However, in economics and finance, discount rates vary over time (Cochrane, 2011, Hills and Nakata, 2018). Constant discounting fails to explain the complexities of real-world decision-making, where agents have subjective time preferences or experience exogenous uncertainty in discount rates. For example, regarding shock-dependent discounts, Justiniano and Primiceri (2008) illustrate that the variance in discount factors accounts for a significant portion of consumption volatility. Albuquerque et al. (2016) shows that risk in time preference accounts for key asset pricing moments, such as equity premium, bond term premium, and weak correlation between stock returns and fundamentals.

Besides the uncertainty in exogenous state-dependent discounts, endogenous time preferences cannot be overlooked in economics. For instance, Choi et al. (2008) demonstrate that endogenously generated short-run international differences in subjective discounting, indicating increasing relative U.S. impatience, result in saving and current account imbalances that align with observed data patterns. Hashimzade et al. (2023) shows that, through a self-reinforcing redistribution mechanism, endogenous discounting can lead to a higher equilibrium interest rate and a much unequal wealth distribution, in comparison to the benchmark model with a constant discount rate. Moreover, Maeda and Nagaya (2023) show that the accelerated short-sighted consumption habit leads to earlier depletion of exhaustible resources.

This paper extends dynamic programming theory to account for state-action-dependent discounting. We develop a comprehensive dynamic programming theory by introducing the concept of eventual discounting, as proposed by Stachurski and Zhang (2021). Eventual discounting assumes that the expected multiplicative of discount factors is eventually less than one for any policy. We demonstrate that eventual discounting implies the eventual contraction of the Bellman operator, which also contracts in some weighted supremum norm. Under these conditions of eventual discounting, we demonstrate the existence of an optimal policy, validate Bellman's principle of optimality, and confirm that value function iteration, Howard policy iteration, and optimistic policy iteration converge to the desired value function or optimal policy.

In the existing literature, Stachurski and Zhang (2021) provides a complete theory of dynamic programming with state-dependent discount factors, while Sargent and Stachurski (2023) presents a theory of dynamic programming of state-action-dependent discounting in finite state and action spaces. Moreover, Toda (2021) and Toda (2023) use the Perov Contraction Theorem to prove that the Bellman operator is contracting in some weighted supremum norm and has a unique fixed point when the exogenous state space is finite, and discount factors are state-dependent.[1]

Addressing the gap in the literature concerning action-dependent time preferences, our exploration considers general state and action spaces with state-action-dependent discounting. In particular, we establish standard optimality results for dynamic programming with action-dependent discounting, building upon frameworks as outlined in Uzawa (1968) and Becker and Mulligan (1997), and integrating state-dependent stochastic discounting.

The study begins by examining the dynamic programming framework with bounded rewards. We demonstrate that the eventual discounting is sufficient for the existence and uniqueness of fixed points for both the Bellman operator and policy value operators. In particular, we show that the eventual discounting condition ensures these operators to eventually contract in the supremum norm or contract in some weighted supremum norm. This finding provides a robust foundation for the analysis of the convergence rates of these operators over time.

---

[1] The weighted function is an upper bound for the reward.

Furthermore, we explore the necessity of eventual discounting for the Markov decision process with a compact state space and strictly concave rewards. We show that eventual discounting is necessary to ensur the existence and uniqueness of policy values, which are fixed points to policy operators. Moreover, under appropriate assumptions, eventual discounting is necessary for the existence of the solution to the Bellman equation. In this context, we highlight that the spectral radii of operators for discounted conditional expectation primarily influence the convergence rate of the system. Notably, the eigenfunctions corresponding to these spectral radii are utilised as weighting vectors in weighted supremum norms.

In the extension, we extend our findings to dynamic programming with unbounded rewards using the Q-transform, following the approach outlined by Ma et al. (2022). We establish that the standard optimality of dynamic programming holds when the discount factors are state-action-dependent and eventually discounting, which extends Ma et al. (2022). We provide computational methods and prove that both value function iteration and action-value function iteration of an expected action-value operator converge to optimal value or action value. Furthermore, we explore risk-sensitive preferences under eventual discounting, which ensures the uniqueness of optimal value function and the existence of optimal policy.

Related literature. The related literature in the Markov decision process with state- or action-dependent discount factors includes Wei and Guo (2011), Minjárez-Sosa (2015), Wu et al. (2015), Wu and Zhang (2016) and Jasso-Fuentes et al. (2022).

About endogenous time preferences, Uzawa (1968) and Epstein and Hynes (1983) propose a theory in which impatience rises with consumption, suggesting that the rich are more likely to heavily discount future consumption. Although Uzawa time preference may be counterintuitive and not be supported by empirical evidence, it succeeds in theories of small open economics such as solving non-stationary issues (Schmitt-Grohé and Uribe, 2003, Guest and McDonald, 2001, Dutta and Yang, 2013). Helpman and Razin (1982) point out that when the rate of time preference remains lower than the interest rate, individuals rationally choose to accumulate foreign assets to finance their increasing consumption. The constant-discount representation of preferences cannot effectively capture this dynamic behaviour. Consequently, studies on small open economies commonly employ Uzawa-type endogenous discount factors (Obstfeld, 1990, Mendoza, 1991, Schmitt-Grohé and Uribe, 2003, Durdu et al., 2009, Bodenstein, 2011, 2013, Vasilev, 2022b). For instance, Vasilev (2022a) considers Uzawa time preference, where the higher level of real income today leads to a lower discount rate, to explain the propagation of cyclical fluctuations in Bulgaria. Durdu et al. (2009) use Uzawa endogenous time preference to model financial globalisation and the risk of Sudden Stop problems.

In contrast to Uzawa time preferences, Becker and Mulligan (1997) propose a framework that time patience is marginally increasing in future-oriented capital. Dutta and Yang (2013) establish endogenous discount factors such that marginal impatience is increasing in consumption (Uzawa type) and decreasing in future-oriented capital (Becker-Mulligan type). Their model is consistent with the empirical evidence from Australia that current consumption and turnover in future-oriented capital are positively correlated.

In empirical studies on time preferences and wealth, Lawrance (1991) reveals that the affluent exhibit greater patience than the less affluent from estimating the consumption Euler equations and the data in Panel Study of Income Dynamics. Huffman et al. (2019) examine heterogeneity in time preferences among elderly Americans and suggest that impatience correlates with lower wealth. Samwick (1998) indicates that time patience tends to increase with both income and age, by estimating the distribution of discount rates from the wealth data in Survey of Consumer Finances 1992. Cohen et al. (2020) provide a survey of the related literature.

A survey of dynamic programming with state-dependent discount rates can be found in Stachurski and Zhang (2021). In asset pricing and state-dependent discounts, since variation in asset returns is significantly due to variation in discount factors, asset pricing models consider stochastic discount factors depending on the state of consumption growth to include adjustments for risk (Lucas Jr, 1978, Rosenberg and Engle, 2002, Cochrane, 2009, Hansen and Renault, 2010). Campbell and Ammer (1993), Cochrane (2011) point out that variation in asset returns is predominantly due to variation in discount factors.

In literature of risk-sensitive preference, Hansen and Sargent (1995) develop a recursive dynamics of discounted costs for a linear, quadratic, and exponential Gaussian linear control model, introducing risk adjustment into the framework. Moreover, Bäuerle and Jaśkiewicz (2018) study a one-sector optimal growth model with unbounded shocks and rewards, in the framework of Hansen and Sargent (1995). They demonstrate the optimality equation for the non-expected utility and establish the Euler equation. Their analysis is based on an inequality involving associated random variables, a concept that is also utilised in the present paper. Weil (1993) develop a stochastic optimal consumption model with constant absolute risk aversion to study precautionary savings and the permanent income hypothesis. Backus et al. (2015) investigate a business cycle model incorporating aggregate risk and ambiguity. They observe that heightened uncertainty typically leads to a reduction in consumption.

If state and action spaces are finite, optimality and comparative statics can be established using fixed-point theory in complete lattices. Relevant techniques and insights can be found in works by Zhou (1994), Olszewski (2021), Balbus et al. (2022), Stachurski et al. (2022b), and Sargent and Stachurski (2023). Additionally, literature on the uniqueness of fixed-point problems for recursive preferences and dynamics includes Marinacci and Montrucchio (2010), Borovička and Stachurski (2020), Bloise et al. (2024), Ren and Stachurski (2021), and Christensen (2022).

The paper is structured as follows. Section 2.2 sets up the recursive decision process and presents its optimality results under eventual discounting. Section 2.3 presents the optimality of an eventually discounting Markov decision process. Section 2.4 studies the necessity of eventual discounting. Section 2.5 gives applications. Section 2.6 extends to unbounded rewards. Section 2.7 treats extensions in risk-sensitive preference.

2.2. Recursive Decision Process with Eventual Discounting

In this section, we introduce a framework of a recursive decision process and dynamic programming with state-action-dependent discounting. We show the optimality results and the convergences of the conventional computation methods under eventual discounting. The application of recursive decision processes in Markov decision processes is presented in the next section.

2.2.1. Preliminary. Let X be a metric space. Denote the family of Borel measurable and bounded real-valued functions on X as $mb\mathsf{X}$. Denote the family of real-valued continuous (resp., upper semicontinuous) and bounded functions on X as $cb\mathsf{X}$ (resp., $ub\mathsf{X}$). Define $\mathscr{B}(\mathsf{X})$ as the Borel σ-algebra on X. Given an everywhere positive function w on X, let $\| \cdot \|_w$ be the weighted supremum norm $\|v\|_w := \sup_{x \in \mathsf{X}} |v(x)|/w(x)$ for all $v \in m\mathsf{X}$ and let $\| \cdot \| := \| \cdot \|_{\mathbb{1}}$, where $\mathbb{1} \in mb\mathsf{X}$ is defined by $\mathbb{1}(x) \equiv 1$ for $x \in \mathsf{X}$. Let U be a metric space. A self-map F on U is called *globally stable* if F has a unique fixed point $u^* \in U$ and $F^k u \to u^*$ for all $u \in U$. Also, F is called *eventually contracting* if there is $k \in \mathbb{N}$ such that F^k is contracting on U. Throughout, $\leqslant$ is the pointwise order on $\mathbb{R}^{\mathsf{X}}$. Denote $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$.

2.2.2. Recursive Decision Process. Similar to Bertsekas (2022), consider a *generic Bellman equation*

$$v(x) = \sup_{a \in \Gamma(x)} B(x, a, v) \qquad (x \in \mathsf{X} \text{ and } v \in mb\mathsf{X}) \tag{2.1}$$

where

(i) X is a Borel space, referred to as the *state space*,[2]

(ii) A is a Borel space, referred to as the *action space*,

(iii) $\Gamma \colon \mathsf{X} \to \mathsf{A}$ is a nonempty correspondence, referred to as the *feasible correspondence*, such that $\Gamma(x)$ is a Borel measurable subset of A for all $x \in \mathsf{X}$, which defines

---

[2] A Borel space is a Borel subset of a complete and separable metric space.

- the set of the *feasible state-action pairs*

$$\mathsf{G} \coloneqq \{(x, a) \in \mathsf{X} \times \mathsf{A} \colon a \in \Gamma(x)\}, \text{ and}$$

- the set of *feasible stationary policies*

$$\Sigma \coloneqq \{\sigma \in \mathsf{A}^{\mathsf{X}} \colon \sigma \text{ is Borel-measurable and } \sigma(x) \in \Gamma(x) \text{ for all } x \in \mathsf{X}\},$$

(iv) a *value aggregator* $B \colon \mathsf{G} \times mb\mathsf{X} \to \mathbb{R}$ satisfies the monotonicity condition

$$v, w \in mb\mathsf{X} \text{ and } v \leqslant w \implies B(x, a, v) \leqslant B(x, a, w) \quad \text{for all } (x, a) \in \mathsf{G}. \tag{2.2}$$

A *recursive decision process* (RDP) is a tuple $\mathscr{R} = (\mathsf{X}, \mathsf{A}, \Gamma, B)$. We interpret $B(x, a, v)$ as the lifetime rewards, contingent on current state x and action a, using v to evaluate future states. We assume the following conditions for the primitives of an RDP throughout this section.

Condition 2.2.1.

(a) Γ is nonempty, continuous, and compact-valued, and

(b) $(x, a) \mapsto B(x, a, v)$ is bounded and Borel measurable on G for all $v \in mb\mathsf{X}$, and also continuous whenever $v \in cb\mathsf{X}$.

An RDP is called *regular* if it satisfies Condition 2.2.1. The regular conditions ensure that the dynamic programming is well-defined. In detail, if an RDP is regular, then the maximiser to the Bellman equation (2.1) exists when the function v within the value aggregator is continuous.[3]

Given $\sigma \in \Sigma$, a *policy operator* $T_\sigma \colon mb\mathsf{X} \to mb\mathsf{X}$ is defined by

$$T_\sigma v(x) \coloneqq B(x, \sigma(x), v) \qquad (x \in \mathsf{X} \text{ and } v \in mb\mathsf{X}).$$

Then, $T_\sigma v(x)$ returns the lifetime value of an action policy $\sigma \in \Sigma$, under state evaluation function $v \in mb\mathsf{X}$ at state $x \in \mathsf{X}$.

---

[3] In this paper, we call such assumptions for the existence of optimal solutions as regular conditions.

The *Bellman operator* $T\colon cb\mathsf{X} \to cb\mathsf{X}$ is defined by

$$Tv(x) \coloneqq \sup_{a \in \Gamma(x)} B(x, a, v) = \sup_{\sigma \in \Sigma} T_\sigma v(x) \qquad (x \in \mathsf{X} \text{ and } v \in cb\mathsf{X}),$$

which is the right-hand side of (2.1). Then, $Tv(x)$ returns the optimised lifetime values under state evaluation function $v \in cb\mathsf{X}$ at state $x \in \mathsf{X}$.

Given $v \in cb\mathsf{X}$, a policy $\sigma \in \Sigma$ is called v-*greedy* if

$$\sigma(x) \in \operatorname*{arg\,max}_{a \in \Gamma(x)} B(x, a, v) \qquad \text{for all } x \in \mathsf{X}.$$

A v-greedy policy optimises lifetime rewards given state evaluation function $v \in cb\mathsf{X}$. By the definition of greedy policies, σ is v-greedy if and only if $Tv = T_\sigma v$.

The regular condition 2.2.1 guarantees the existence of a greedy policy for any $v \in cb\mathsf{X}$, which is a fundamental requirement in dynamic programming. The following lemma shows that whenever $\mathscr{R}$ is regular, T_σ and T are well-defined on $mb\mathsf{X}$ and $cb\mathsf{X}$, respectively, and v-greedy policies exist for all $v \in cb\mathsf{X}$.

LEMMA 2.2.1. *If $\mathscr{R}$ is regular, then the following statements are true.*

(a) T_σ is a self-map on $mb\mathsf{X}$ for all $\sigma \in \Sigma$.

(b) T is a self-map in $cb\mathsf{X}$.

(c) For all $v \in cb\mathsf{X}$, there exists a v-greedy policy.

Lemma 2.2.1 implies that T_σ is a self-map on $mb\mathsf{X}$ for any $\sigma \in \Sigma$ whenever $\mathscr{R}$ is regular. It also implies that for any $v \in cb\mathsf{X}$ there exists a v-greedy policy. Moreover, since T is a self-map on $cb\mathsf{X}$, it ensures that the value function iteration algorithm, introduced in Section 2.2.4, is well-behaved.

2.2.3. Eventual Discounting. In this section, we present the assumptions regarding stochastic state-action-dependent discounting. Analogous to a standard contracting Markov decision process where a constant discount factor is strictly less than one, we admit the concept of eventual discounting outlined in Stachurski and Zhang (2021), Toda (2021) and Sargent and Stachurski (2023). By Gelfand's formula, denote $\rho(A)$ as the spectral radius of a bounded linear or monotone sublinear operator $A\colon mb\mathsf{X} \to mb\mathsf{X}$:

$$\rho(A) := \lim_{n\to\infty} \|A^n\|^{1/n}$$

where $\|A\|$ denotes the operator norm of A.[4]

Let $\mathscr{R} = (\mathsf{X}, \mathsf{A}, \Gamma, B)$. Assume that there is a $k\colon \mathsf{G}\times\mathsf{X} \to \mathbb{R}_+$ such that $(x, a) \mapsto k(x, a, B)$ is a Borel measurable function on G for all $B \in \mathscr{B}(\mathsf{X})$, and $B \mapsto k(x, a, B)$ is a Borel measure on X for all $(x, a) \in \mathsf{G}$. Suppose that the difference in value aggregators is bounded as follows:

$$B(x, a, v) - B(x, a, w) \leqslant \int_{\mathsf{X}} (v(x') - w(x'))k(x, a, \mathrm{d}x') \tag{2.3}$$

for all $(x, a) \in \mathsf{G}$ and $v, w \in mb\mathsf{X}$. Given $\sigma \in \Sigma$, let $L_\sigma\colon mb\mathsf{X} \to mb\mathsf{X}$ be a positive linear operator defined by[5]

$$L_\sigma h(x) := \int_{\mathsf{X}} h(x')k(x, \sigma(x), \mathrm{d}x') \qquad (x \in \mathsf{X}, h \in mb\mathsf{X}). \tag{2.4}$$

where we denote $L_\sigma h(x) := (L_\sigma(h))(x)$ to simplify notation. If the RDP is a Markov decision process, then $L_\sigma h$ is interpreted as the expected discounted present value of h. We then have[6]

$$\begin{aligned}
|T_\sigma v(x) - T_\sigma w(x)| &\leqslant |B(x, \sigma(x), v) - B(x, \sigma(x), w)| \\
&\leqslant \int_{\mathsf{X}} |v(x') - w(x')| k(x, \sigma(x), \mathrm{d}x') \\
&\leqslant L_\sigma |v - w|(x) \qquad\qquad (x \in \mathsf{X}, v, w \in mb\mathsf{X}).
\end{aligned}$$

We say that T_σ is *eventually discounting* if (2.3) holds and $\rho(L_\sigma) < 1$. Define the *eventual-discount factor* d_n^σ under policy $\sigma \in \Sigma$ by

$$d_n^\sigma := \sup_{x\in\mathsf{X}} L_\sigma^n \mathbb{1}(x) \qquad (n \in \mathbb{N}).$$

We will show that $\rho(L_\sigma) < 1$ if and only if $d_{n_\sigma}^\sigma < 1$ for some $n_\sigma \in \mathbb{N}$. The eventual discount factor generalises the constant discount factor $\beta < 1$. Moreover, it can be shown

---

[4] $\|A\| := \inf\{c \geqslant 0\colon \|Av\| \leqslant c\|v\| \text{ for all } v \in mb\mathsf{X}\}$.

[5] A linear operator operator L on $mb\mathsf{X}$ is positive if $v \in mb\mathsf{X}$ and $v \geqslant 0$ imply $Lv \geqslant 0$.

[6] We denote $L_\sigma|v - w|(x) = L_\sigma(|v - w|)(x)$ to simplify notation.

that $\rho(L_\sigma) < 1$ implies the global stability of policy operator T_σ. To further ensure that the Bellman operator T is globally stable, we introduce the following conditions with both finite and infinite feasible policies.[7]

ASSUMPTION 2.2.1.

(i) For any $\sigma \in \Sigma$, T_σ is eventually discounting, and

(ii) Σ is finite.

ASSUMPTION 2.2.2.

(a) $(x, a) \mapsto \int_{\mathsf{X}} h(x')k(x, a, \mathrm{d}x')$ is continuous and bounded on G for any $h \in cb\mathsf{X}$,

(b) for all $(x, a) \in \mathsf{G}$ and $v, w \in mb\mathsf{X}$, (2.3) holds, and

(c) $\rho(\hat{L}) < 1$, where $\hat{L} \colon mb\mathsf{X} \to mb\mathsf{X}$ is defined by

$$\hat{L}h(x) \coloneqq \sup_{\sigma \in \Sigma} L_\sigma h(x). \qquad (x \in \mathsf{X}, h \in mb\mathsf{X}) \tag{2.5}$$

If X and A are finite, Assumption 2.2.1 is sufficient for the optimality of dynamic programming. Since Assumption 2.2.1 implies that each policy value v_σ is bounded, we can find the optimal policy from finitely many feasible policies.

If Σ is infinite, we consider Assumption 2.2.2 to ensure the global stability of T. The assumption $\rho(\hat{L}) < 1$ implies that the Bellman operator is eventually contracting so that we can compute the optimal value function. Moreover, Bloise et al. (2024) imply that $\rho(\hat{L}) < 1$ is equivalent to that there exists an everywhere positive $f \in mb\mathsf{X}_+$ such that for some $\rho \in (0, 1)$ we have $\hat{L}f \leqslant \rho f$.[8] This implies that T is contractive in $\| \cdot \|_f$, as shown in the appendix.

Since it may be difficult to check eventual discounting for each policy, confirming the existence of $n_\sigma \in \mathbb{N}$ satisfying $d^\sigma_{n_\sigma} < 1$, especially when there are infinitely many policies, we introduce the following stricter assumptions of eventual discounting.

ASSUMPTION 2.2.3. There is an $\ell \colon \mathsf{X} \times \mathscr{B}(\mathsf{X}) \to \mathbb{R}_+$, satisfying that $x \mapsto \ell(x, B)$ is a Borel measurable function for all $B \in \mathscr{B}(\mathsf{X})$, and $B \mapsto \ell(x, B)$ is a Borel measure on X for all $x \in \mathsf{X}$, such that

---

[7] We sincerely thank Prof. Gaetano Bloise for pointing out the concept of monotone sublinear operators.

[8] See Claim 6 of Bloise et al. (2024).

(i) $L\colon mb\mathsf{X} \to mb\mathsf{X}$ is a positive linear operator,

(ii) $|B(x,a,v) - B(x,a,w)| \leqslant L|v-w|(x)$ for all $(x,a) \in \mathsf{G}$, and

(iii) $\rho(L) < 1$, where

$$Lh(x) := \int_{\mathsf{X}} h(x')\ell(x, \mathrm{d}x') \qquad (x \in \mathsf{X}, h \in mb\mathsf{X}). \tag{2.6}$$

We say that an RDP $\mathscr{R}$ is *eventually discounting* if either Assumption 2.2.1, 2.2.2, or 2.2.3 holds. We provide some useful properties regarding eventual discounting assumptions. The following lemma shows the relationship between the spectral radius $\rho(L_\sigma)$ and the expected multiplicative of eventual discount factor $d^\sigma_{n_\sigma}$.

Lemma 2.2.2. *If $\mathscr{R}$ is regular, and L_σ and $\hat{L}$ are defined as above, then the following statements are true.*

(a) For any $\sigma \in \Sigma$, $\rho(L_\sigma) < 1$ if and only if there is an $n_\sigma \in \mathbb{N}$ such that $d^\sigma_{n_\sigma} < 1$.

(b) For any $\sigma \in \Sigma$, $\rho(L_\sigma) = \lim_{n\to\infty}(d^\sigma_n)^{1/n}$.

(c) $\rho(\hat{L}) < 1$ implies $\rho(L_\sigma) < 1$ for all $\sigma \in \Sigma$.

Lemma 2.2.2 provides a method to compute $\rho(L_\sigma)$ by taking the limit of $(d^\sigma_n)^{1/n}$. Moreover, it implies that $\rho(L_\sigma) < 1$ and $d^\sigma_{n_\sigma} < 1$ (for some $n_\sigma \in \mathbb{N}$) are equivalent. The next example with state-dependent discounting, generalised from Chapter 10 of Stokey (1989), is a regular RDP.

Example 2.2.1 (One-Sector Optimal Growth). An economy contains many identical and infinitely lived households. There is a single good $y_t = F(k_t, \ell_t)$ produced by capital k_t and labour ℓ_t inputs. labour is supplied inelastically such that $\ell_t = 1$ for all t, and capital depreciates at a rate $\delta \in (0,1)$. Denote c_t as consumption and i_t as investment. Given stochastic discount, a social planner solves the problem

$$
\begin{aligned}
& \sup \mathbb{E}_{k_0,z_0} \sum_{t=0}^{\infty} \left(\prod_{i=0}^{t-1} \beta(z_i, c_i) \right) U(c_t) \\
\text{s.t.} \quad & c_t + i_t \leqslant z_t y_t, \\
& 0 \leqslant \ell_t \leqslant 1, \\
& k_{t+1} = (1-\delta)k_t + i_t \qquad (t \in \mathbb{N}_0), \\
& k_0 \geqslant 0 \quad \text{and} \quad z_0 \geqslant 0 \text{ given},
\end{aligned}
$$

where $\{z_t\}_{t\geqslant 0}$ is a sequence of exogenous shocks generated by a transition kernel Q on $(\mathsf{Z}, \mathscr{B}(\mathsf{Z}))$ with $\mathsf{Z} = [1, \bar{z}]$ for $1 < \bar{z} < \infty$. The state is $x = (k, z)$, and the action is $a = c$. Assume that F is continuously differentiable, strictly increasing, and strictly concave with

$$
\begin{aligned}
& F(0, \ell) = 0, \quad F_k(k, \ell) > 0, \quad F_\ell(k, \ell) > 0, \quad (k, \ell > 0) \\
& \lim_{k\to 0} F_k(k, 1) = \infty, \quad \lim_{k\to\infty} F_k(k, 1) = 0.
\end{aligned}
$$

Moreover, assume that U is continuous, β is continuous and strictly positive, and Q satisfies the Feller property such that $z \mapsto \int h(z')Q(z, \mathrm{d}z')$ is bounded and continuous for all $h \in cb\mathsf{Z}$. Let $\bar{k} > 0$ be such that $\bar{k} = \bar{z}F(\bar{k}, 1) + (1-\delta)\bar{k}$. Then, the set of maintainable capital stock is $[0, \bar{k}]$. Let the state space be $\mathsf{X} = [0, \bar{k}] \times [1, \bar{z}]$, the action space be $\mathsf{A} = [0, \bar{z}F(\bar{k}, 1)]$, the feasible correspondence be $\Gamma(k, z) = \{c \in \mathsf{A}\colon 0 \leqslant c \leqslant zF(k, 1)\}$. Hence, $(x, a) = ((k, z), c)$. Define the stochastic transition kernel by

$$
P((k, z), c, (\mathrm{d}k', \mathrm{d}z')) = Q(z, \mathrm{d}z')\mathbb{1}\{k' = (1-\delta)k + zF(k, 1) - c\}\mathrm{d}k'
$$

for all $((k, z), c, (k', z')) \in \mathsf{G} \times \mathsf{X}$, where $\mathbb{1}\{\cdot\}$ denotes an indicator function. Let the value aggregator B be

$$
B((k, z), c, v) = U(c) + \mathbb{E}_{(k,z)}[\beta(z, c)v(K', Z')] \qquad (((k, z), c, v) \in \mathsf{G} \times mb\mathsf{X}),
$$

where (K', Z') is the random variables generated by $P((k,z), c, \cdot)$. We can check that $\mathscr{R} = (\mathsf{X}, \mathsf{A}, \Gamma, B)$ is a regular RDP.[9] Moreover, we have

$$
\begin{aligned}
B((k,z), c, v) - B((k,z), c, w) &= \mathbb{E}_{(k,z)}[\beta(z,c)v(K',Z')] - \mathbb{E}_{(k,z)}[\beta(z,c)w(K',Z')] \\
&= \mathbb{E}_{(k,z)}\beta(z,c)(v(K',Z') - w(K',Z')).
\end{aligned}
$$

for all $((k,z),c) \in \mathsf{G}$ and $v, w \in mb\mathsf{X}$. Let $L_\sigma \colon mb\mathsf{X} \to mb\mathsf{X}$ be

$$
L_\sigma h(k,z) = \mathbb{E}^{\sigma}_{(k,z)} \beta(z, \sigma(k,z)) h(k', z') \qquad ((k,z) \in \mathsf{X}, h \in mb\mathsf{X}),
$$

where $\mathbb{E}^{\sigma}_{k,z}$ is the expectation conditioning on $(k_0, z_0) = (k,z)$ under transition kernel $P((k,z), \sigma(k,z), \cdot)$ defined above. Then, we have

$$
d_n^\sigma = \sup_{(k,z)\in\mathsf{X}} L_\sigma^n \mathbb{1}(k,z) = \sup_{(k,z)\in\mathsf{X}} \mathbb{E}^{\sigma}_{(k,z)} \prod_{t=0}^{n-1} \beta(Z_t, \sigma(K_t, Z_t)).
$$

Assume $\beta(z,c) \leqslant \bar{\beta}(z)$ for all k and c such that there exists an $n \in \mathbb{N}$:

$$
\sup_{(k,z)\in\mathsf{X}} \mathbb{E}^{\sigma}_{(k,z)} \prod_{t=0}^{n-1} \beta(Z_t, \sigma(K_t, Z_t)) \leqslant \sup_{(k,z)\in\mathsf{X}} \mathbb{E}_{(k,z)} \prod_{t=0}^{n-1} \bar{\beta}(Z_t) < 1. \tag{2.7}
$$

Define $L \colon mb\mathsf{X} \to mb\mathsf{X}$ by

$$
Lh(k,z) = \mathbb{E}_{(k,z)} \bar{\beta}(z) h(K', Z') \qquad ((k,z) \in \mathsf{X}, h \in mb\mathsf{X}).
$$

Hence, (2.7) implies $\sup_x L^n \mathbb{1}(x) < 1$ and then $\rho(L) < 1$. To this end, $\mathscr{R}$ is eventually discounting since Assumption 2.2.2 or 2.2.3 holds. □

2.2.4. Dynamic Programming. In this section, we define the optimal properties of dynamic programming and introduce dynamic programming algorithms, including value function iteration, Howard policy iteration, and optimistic policy iteration, to search for optimal policies,

Let $\mathscr{R} = (\mathsf{X}, \mathsf{A}, \Gamma, B)$ be an RDP. Given $\sigma \in \Sigma$, if the policy operator T_σ has a unique fixed point $v_\sigma \in mb\mathsf{X}$, then we call v_σ as σ-*value* function. We show in Section 2.2.5 that all policy operators have unique fixed points under eventual discounting assumptions,

---

[9] It is also a regular Markov decision process defined in Section 2.3.

which implies the existence and uniqueness of σ-value for any $\sigma \in \Sigma$. The *(optimal) value function* v^* of $\mathscr{R}$ is defined by

$$v^*(x) \coloneqq \sup_{\sigma \in \Sigma} v_\sigma(x) \qquad (x \in \mathsf{X}).$$

A policy $\sigma^* \in \Sigma$ is called *optimal* if $v_{\sigma^*} = v^*$; that is, if

$$v_{\sigma^*}(x) \geqslant v_\sigma(x) \qquad \text{for all } \sigma \in \Sigma \text{ and } x \in \mathsf{X}.$$

By the definition of T, we say that v satisfies the Bellman equations if $Tv = v$; that is v satisfies (2.1). We say that *Bellman's principle of optimality* holds if

$$\sigma \in \Sigma \text{ is optimal for } \mathscr{R} \iff \sigma \text{ is } v^*\text{-greedy}.$$

That is, Bellman's principle of optimality holds whenever $v_\sigma = v^* \iff Tv^* = T_\sigma v^*$. Bellman's principle of optimality ensures that we can find the optimal policy if we can compute v^*, and v^*-greedy policy exists.

The conventional dynamic programming algorithms are defined as follows (see, e.g., Bertsekas (2022)). Let $\mathscr{R}$ be regular. A sequence $\{v_k\}_{k \geqslant 0} \subset cb\mathsf{X}$ is called a *value function iteration* (VFI) if $v_{k+1} = Tv_k$ for $k \geqslant 0$ with any $v_0 \in cb\mathsf{X}$. VFI iterates the right-hand side of (2.1) and converges to v^* under appropriate assumptions.

To ensure the existence of greedy policies and the continuity of policy values in policy iteration algorithms, we introduce the following conditions. Note that if we know that greedy policies exist and are continuous for any value function, the following conditions are not required.[10]

Condition 2.2.2.

(a) A is a convex subset of a vector space,

(b) Γ is nonempty, continuous, compact-valued, and convex-valued,

(c) $(x, a) \mapsto B(x, a, v)$ is bounded and Borel measurable on G for all $v \in mb\mathsf{X}$, and

(d) $a \mapsto B(x, a, v)$ is strictly quasi-concave for $x \in \mathsf{X}$ all $v \in cb\mathsf{X}$.

---

[10] For example, state and action spaces are finite.

Define $\Sigma_C \subset \Sigma$ as the subset of all continuous policies. Let Condition 2.2.2 hold. Then, the optimal policy σ^* is continuous. It can be shown that $v^* = \sup_{\sigma \in \Sigma_C} v_\sigma$ and $Tv = \sup_{\sigma \in \sigma_C} T_\sigma v$ for all $v \in cb\mathsf{X}$. To this end, it suffices to restrict the policy iterations to continuous policies.

A sequence $\{\sigma_\kappa\}_{k \geqslant 0} \subset \Sigma_C$ is called a *Howard policy iteration* (HPI) if σ_{k+1} is v_{σ_k}-greedy for $k \geqslant 0$ and any $\sigma_0 \in \Sigma_C$. In each iteration, HPI evaluates the policy σ_k by v_{σ_k}, and if there is improvable state under v_{σ_k}, it updates policy σ_k to v_{σ_k}-greedy policy. Therefore, the policy values from HPI increase monotonically. Define the *Howard operator* $H \colon cb\mathsf{X} \to cb\mathsf{X}$ by

$$Hv = v_\sigma \text{ where } \sigma \text{ is } v\text{-greedy.}$$

The iteration, $\{H^k v_0\}_{k \geqslant 0}$ with $v_0 = v_\sigma$ for $\sigma \in \Sigma_C$, of the Howard operator H is an abstract version of HPI.

A sequence $\{v_k\}_{k \geqslant 0} \subset cb\mathsf{X}$ is called an *optimistic policy iteration* (OPI) if, fixing $m \in \mathbb{N}$, $v_{k+1} = T^m_{\sigma_k} v_k$ where σ_k is v_k-greedy for $k \geqslant 0$ with $v_0 = v_\sigma$ for any $\sigma \in \Sigma_C$. Compared to HPI, OPI approximates v_{σ_k} by $T^m_{\sigma_k} v_k$. Therefore, OPI can be more efficient if the calculation of v_{σ_k} is time-consuming, where $v_{\sigma_k} = \lim_{n \to \infty} T^n_{\sigma_k} v_{v_k}$, assuming that T_{σ_k} is globally stable. Thus, if $m \to \infty$, then OPI is the same as HPI; if $m = 1$, OPI is the same as VFI. Therefore, OPI can be seen as a combination of HPI and VFI.

To illustrate OPI, define optimistic policy operator $W \colon cb\mathsf{X} \to cb\mathsf{X}$ by

$$Wv = T^m_\sigma v \qquad (v \in cb\mathsf{X}), \tag{2.8}$$

where σ is the first policy among the list of all v-greedy policies. Then, $\{W^k v_0\}_{k \geqslant 0}$ with $v_0 = v_\sigma$ for $\sigma \in \Sigma_C$ is an abstract version of OPI.

2.2.5. Optimality. As discussed in the introduction, we extend the existing dynamic programming theory to incorporate the cases of action-dependent discounting, including action-dependent time preferences introduced by Uzawa (1968) and Becker and Mulligan (1997). In this section, we show the optimality results, including the existence of optimal policies and Bellman's Principle of Optimality, when an RDP is eventually discounting.

In addition, the computation algorithms, including VFI, HPI, and OPI, converge to the optimal value function or optimal policies.

In the following theorem, let $w \coloneqq \sum_{n=0}^{\infty} \hat{L}^n \mathbb{1}$ for $\| \cdot \|_w$.

Theorem 2.2.1. *Suppose that $\mathscr{R}$ is regular and either Assumption 2.2.1, 2.2.2 or 2.2.3 holds. Then, the following statements are true.*

- *(a) T_σ is eventually contracting on $m b\mathsf{X}$ for all $\sigma \in \Sigma$,*
- *(b) T is contractive on $(c b\mathsf{X}, \| \cdot \|_w)$ if Assumption 2.2.2 holds,*
- *(c) v^* is the unique solution to the Bellman equation in $c b\mathsf{X}$,*
- *(d) VFI converges to v^*,*
- *(e) Bellman's principle of optimality holds,*
- *(f) HPI converges to σ^*, and OPI converge to v^* if Assumption 2.2.1 holds,*
- *(g) Bellman's principle of optimality holds, and*
- *(h) at least one optimal (continuous) policy exists,*

If Condition 2.2.2 is satisfied and Assumption 2.2.2 or 2.2.3 holds, then

- *(α) HPI converges to σ^*, and*
- *(β) OPI converges to v^*,*

where all the iterated greedy policies to HPI and OPI are continuous.

Theorem 2.2.1 generalises the conventional dynamic programming theory to the case of state-action-dependent discounting. In detail, Theorem 2.2.1 shows that the optimal policy exists, Bellman's principle of optimality holds, VFI and OPI converge to v^*, and HPI converges to σ^*. Bellman's principle of optimality guarantees that we can first compute v^* by VFI or OPI and find σ^* by

$$\sigma^*(x) = \arg\max_a B(x, a, v^*) \qquad (x \in \mathsf{X}).$$

Theorem 2.2.1 shows that eventual discounting is sufficient for eventual contracting or global stability of T_σ or T. Additionally, Theorem 2.2.1 implies that T is contracting on

$(cb\mathsf{X}, \|\cdot\|_w)$, so we can analyze the convergence rate of VFI if we know the contraction modulus.

Moreover, the value function v^* is continuous. Observe that if there is a non-continuous function satisfying the Bellman equation, then it must be equal to or greater than the optimal value v^*, and there is no policy $\sigma\Sigma$ that its σ-value v_σ admits that function.[11] Since we are interested in v^*, which is continuous when the RDP is regular, we can restrict the iteration of VFI on the continuous functions.

In addition, if Condition 2.2.2 holds and $v \in cb\mathsf{X}$, then any policy $\sigma \in \Sigma$ is dominated by some continuous policy $\sigma_c \in \Sigma_C$ such that $T_\sigma v \leqslant T_{\sigma_c} v$. To this end, it also suffices to focus on the continuous policies for HPI and OPI. Note that Condition 2.2.2 is not necessary if we can ensure that the continuous greedy policy exists for any $v \in cb\mathsf{X}$. In particular, if action and state spaces are finite, then the greedy policy exists, so we do not need Condition 2.2.2.

2.2.6. A Generalised Blackwell's Condition. A generalised Blackwell's condition for the global stability of the Bellman operator T or policy operator T_σ is provided as follows.

Proposition 2.2.1. *Let $\mathcal{T}$ be an order-preserving self-map on $U \subset mb\mathsf{X}$. If there exists a positive linear operator G on $mb\mathsf{X}$ such that $\rho(G) < 1$ and*

$$\mathcal{T}(v + c) \leqslant \mathcal{T}v + Gc \quad \text{for all } c, v \in U \text{ with } c \geqslant 0,$$

then $\mathcal{T}$ is eventually contracting on U.

By Proposition 2.2.1, if we can check that, for any $\sigma \in \Sigma$, there exists a positive linear operator L_σ on $mb\mathsf{X}$ such that $\rho(L_\sigma) < 1$ and

$$T_\sigma(v + c)(x) = B(x, \sigma(x), v + c) \leqslant B(x, \sigma(x), v) + L_\sigma c(x) = T_\sigma v(x) + L_\sigma c(x)$$

for all $x \in \mathsf{X}$ and $c, v \in mb\mathsf{X}$ with $c \geqslant 0$, then T_σ is globally stable on U and has a unique fixed point $v_\sigma \in U$.

---

[11] That is, if $Tw = w$ for some non-continuous $w \in mb\mathsf{X}$, then $w \geqslant v^*$ and $w > v_\sigma$ for any $\sigma \in \Sigma$.

2.3. Markov Decision Process with Eventual Discounting

In this section, we focus on a Markov decision process with state-action-dependent discounting and bounded rewards.

2.3.1. Markov Decision Process. A (generalised) Markov decision process is an RDP such that the value aggregator is separated into reward and expected future contination value. In detail, a *(generalised) Markov decision process* $\mathscr{M}$ is an RDP $(\mathsf{X}, \mathsf{A}, \Gamma, B)$ such that

$$\begin{aligned} B(x, a, v) &:= \int_{\mathsf{X}} [r(x, a, x') + \beta(x, a, x')v(x')] P(x, a, \mathrm{d}x') \\ &= \mathbb{E}_{x,a}[r(x, a, X') + \beta(x, a, X')v(X')] \end{aligned} \tag{2.9}$$

for all $(x, a) \in \mathsf{G}$ and $v \in mb\mathsf{X}$, where

- $r \colon \mathsf{G} \times \mathsf{X} \to \mathbb{R}$ is a Borel measurable function, referred to as the *reward*,
- $\beta \colon \mathsf{G} \times \mathsf{X} \to \mathbb{R}_+$ be a Borel measurable and everywhere positive function, referred to as the *discount factors*, and
- $P \colon \mathsf{G} \times \mathsf{X} \to \mathbb{R}_+$ is a *stochastic kernel* on X contingent on current state and action; that is, $B \mapsto P(x, a, B)$ is a Borel probability measure on X for all $(x, a) \in \mathsf{G}$, and $(x, a) \mapsto P(x, a, B)$ is a Borel measurable function on G for all $B \in \mathscr{B}(\mathsf{X})$.

To simplify notation, denote β_σ, P_σ and r_σ as $\beta_\sigma(x, x') := \beta(x, \sigma(x), x')$, $P_\sigma(x, \cdot) := P(x, \sigma(x), \cdot)$, and $r_\sigma(x) := \mathbb{E}_x^\sigma r(x, \sigma(x), X')$, respectively, for all x, x' and any $\sigma \in \Sigma$, where $\mathbb{E}_x^\sigma$ denotes the expectation under P_σ transition kernel conditioning on x. Also, denote $r(x, a) := \mathbb{E}_{x,a} r(x, a, X')$ for any $(x, a) \in \mathsf{G}$. Given an MDP $\mathscr{M}$ and $\sigma \in \Sigma$, the policy operator T_σ following (2.9) becomes

$$T_\sigma v(x) := r_\sigma(x) + \int_{\mathsf{X}} \beta_\sigma(x, x') v(x') P_\sigma(x, \mathrm{d}x') \qquad (x \in \mathsf{X}, v \in mb\mathsf{X}).$$

If T_σ has a unique fixed point $v_\sigma \in mb\mathsf{X}$, then iteration implies $v_\sigma = T_\sigma v_\sigma = T_\sigma^n v_\sigma$. Letting $n \to \infty$, if it converges, we have

$$v_\sigma(x) := \mathbb{E}_x^\sigma \left[\sum_{t=0}^{\infty} \left(\prod_{i=0}^{t-1} \beta_\sigma(X_i, X_{i+1}) \right) r_\sigma(X_t) \right] \qquad (x \in \mathsf{X}). \tag{2.10}$$

where $\prod_{i=0}^{-1} \beta_i^\sigma = 1$ by convention and $\{X_t\}_{t\in\mathbb{N}_0}$ is a stochastic process such that $X_0 = x$ and X_{t+1} is generated by $P_\sigma(X_t, \cdot)$ for all $t \in \mathbb{N}_0$. Moreover, the Bellman operator becomes

$$Tv(x) = \sup_{a\in\Gamma(x)} \{r(x,a) + \mathbb{E}_{x,a}\beta(x,a,X')v(X')\} \qquad (x \in \mathsf{X}, v \in cb\mathsf{X}).$$

We introduce the following regular conditions to an MDP.

Condition 2.3.1.

(i) Γ is nonempty, compact-valued, and continuous.

(ii) $(x,a) \mapsto r(x,a)$ is continuous and bounded on G.

(iii) β is bounded and strictly positive,

(iv) $(x,a) \mapsto \int_{\mathsf{X}} f(y)\beta(x,a,y)P(x,a,\mathrm{d}y)$ is bounded and continuous on G for every $f \in cb\mathsf{X}$.

We say that an MDP $\mathscr{M}$ is *regular* if condition 2.3.1 is satisfied. Condition 2.3.1 ensures that $\mathscr{M}$ is a regular RDP and then the greedy policies exist by Lemma 2.2.1.

Lemma 2.3.1. *If an MDP $\mathscr{M}$ is regular, then it is a regular RDP, and the following statements are true.*

(a) T_σ is a self-map on $mb\mathsf{X}$ for all $\sigma \in \Sigma$.

(b) T and T_σ are self-maps on $cb\mathsf{X}$ for all $\sigma \in \Sigma_C$.

(c) For all $v \in cb\mathsf{X}$, there exists a v-greedy policy.

Analogous to Condition 2.2.1, to ensure the continuity of optimal policies or greedy policies, we introduce the conditions with concavity.

Condition 2.3.2.

(i) A is a convex subset of a vector space,

(ii) Γ is nonempty, continuous, compact-valued, and convex-valued.

(iii) $(x,a) \mapsto r(x,a)$ is continuous and bounded on G, and $a \mapsto r(x,a)$ is strictly concave for all $x \in \mathsf{X}$.

(iv) β is bounded and strictly positive,

(v) $(x, a) \mapsto \int_{\mathsf{X}} f(y)\beta(x, a, y)P(x, a, \mathrm{d}y)$ is bounded and continuous on G for every $f \in cb\mathsf{X}$,

(vi) $a \mapsto \int_{\mathsf{X}} f(y)\beta(x, a, y)P(x, a, \mathrm{d}y)$ is concave for all $x \in \mathsf{X}$ and $f \in cb\mathsf{X}$.

Condition 2.3.2 guarantees that continuous v-greedy policies exist for all $v \in cb\mathsf{X}$, and HPI and OPI are well-behaved.

2.3.2. Eventual Discounting. Let $\mathscr{M}$ be an MDP. To this end, (2.3) becomes

$$B(x, a, v) - B(x, a, w) \leqslant \int_{\mathsf{X}} (v(x') - w(x'))\beta(x, a, x')P(x, a, \mathrm{d}x')$$

for all $(x, a) \in \mathsf{G}$ and $v, w \in mb\mathsf{X}$. The corresponding $L_\sigma \colon mb\mathsf{X} \to mb\mathsf{X}$ is defined by

$$L_\sigma h(x) := \int_{\mathsf{X}} h(x')\beta(x, \sigma(x), x')P(x, \sigma(x), \mathrm{d}x') = \mathbb{E}_x^\sigma \beta_\sigma(x, X')h(X') \tag{2.11}$$

for all $x \in \mathsf{X}$ and $h \in mb\mathsf{X}$, which returns the expected discounted present value of h. We say that T_σ is *eventually discounting* if $\rho(L_\sigma) < 1$, or equivalently, there is an $n_\sigma \in \mathbb{N}$ such that

$$d_{n_\sigma}^\sigma := \sup_x \mathbb{E}_x^\sigma \prod_{t=0}^{n_\sigma - 1} \beta_\sigma(X_t, X_{t+1}) < 1.$$

By definitions, we have $d_1^\sigma = \sup_x L_\sigma \mathbb{1}(x)$ and $d_{n_\sigma}^\sigma = \sup_x L_\sigma^{n_\sigma} \mathbb{1}(x)$ by iteration. Define $d_n \colon \mathsf{X} \to \mathbb{R}_+$ by

$$\begin{aligned} d_n(x) := \sup_{\sigma_1 \in \Sigma} \Bigg\{ \mathbb{E}_x^{\sigma_1} \beta_{\sigma_1}(X_0, X_1) \sup_{\sigma_2 \in \Sigma} \Bigg\{ \mathbb{E}_{X_1}^{\sigma_2} \beta_{\sigma_2}(X_1, X_2) \sup_{\sigma_3 \in \Sigma} \Bigg\{ \cdots \\ \cdots \sup_{\sigma_n \in \Sigma} \left\{ \mathbb{E}_{X_{n-1}}^{\sigma_n} \beta_{\sigma_n}(X_{n-1}, X_n) \right\} \cdots \Bigg\} \Bigg\} \Bigg\}. \end{aligned} \tag{2.12}$$

for all $x \in \mathsf{X}$. That is, $d_1(x) = \sup_{\sigma_1} \mathbb{E}_x^{\sigma_1} \beta_{\sigma_1}(X_0, X_1)$, and $d_n(x) = \sup_{\sigma_n} \mathbb{E}_x^{\sigma_n} \beta_{\sigma_n}(x, X') d_{n-1}(X')$ for $n \geqslant 2$ and $x \in \mathsf{X}$. We will show that $\sup_x d_n(x) < 1$, for some $n \in \mathbb{N}$, implies the global stability of T. Define $\hat{L} \colon mb\mathsf{X} \to mb\mathsf{X}$ by

$$\hat{L}h(x) := \sup_{\sigma \in \Sigma} L_\sigma h(x) \qquad (x \in \mathsf{X}, h \in mb\mathsf{X}.)$$

The operator $\hat{L}h$ returns the highest expected discounted present value of h among all feasible policies that determine subjective discount factors. Thus, iteration yields $d_n(x) =$

$\hat{L}^n \mathbb{1}(x)$ for all $x \in \mathsf{X}$. Similar to the assumptions in 2.2, we consider the following assumptions for eventual discounting.

Assumption 2.3.1. Σ is finite, and T_σ is eventually discounting for any $\sigma \in \Sigma$.

Assumption 2.3.2. Either $\rho(\hat{L}) < 1$ or $\sup_x d_n(x) < 1$ for some $n \in \mathbb{N}$.

Assumption 2.3.3. There is an $\ell \colon \mathsf{X} \times \mathscr{B}(\mathsf{X}) \to \mathbb{R}_+$, satisfying that $x \mapsto \ell(x, B)$ is a Borel measurable function for all $B \in \mathscr{B}(\mathsf{X})$, and $B \mapsto \ell(x, B)$ is a Borel measure on X for all $x \in \mathsf{X}$, such that

$$L_\sigma h \leqslant Lh$$

for any $h \in mb\mathsf{X}_+$, and $\rho(L) < 1$, where L is defined by

$$Lh(x) = \int_{\mathsf{X}} h(x') \ell(x, \mathrm{d}x')$$

for all $x \in \mathsf{X}$.

We say that $\mathscr{M}$ is *eventually discounting* if either Assumption 2.3.1, 2.3.2, or 2.3.3 is satisfied. Eventual discounting implies that the expected discounted present value of the future reward is bounded above and converges to zero as the time approaches infinity, in the sense that $L_\sigma^n r_\sigma \to 0$ as $n \to \infty$ for all $\sigma \in \Sigma$. Assumption 2.3.3 is a sufficient condition to Assumption 2.3.1 and 2.3.2. Also, $\rho(\hat{L}) < 1$ or $\sup_x d_n(x) < 1$ for some $n \in \mathbb{N}$ implies $\rho(L_\sigma) < 1$ and that T_σ is eventually discounting for all $\sigma \in \Sigma$.

Lemma 2.3.2. *If $\mathscr{M}$ is regular and Assumption 2.3.3 holds, then Assumption 2.3.1 and 2.3.2 hold.*

Example 2.3.1 (Firm valuation with stochastic interest rates). [12] Assume that the state space X is finite and follows a stochastic kernel $P \colon \mathsf{X} \times \mathsf{X} \to \mathbb{R}_+$. Suppose that the discount factors are

$$\beta_t := \frac{1}{1 + r_t} \qquad (t \in \mathbb{N})$$

---

[12] We do not consider action in this example, so it is not an MDP. However, this example provides some intuition for the assumption of eventual discounting through linear algebra.

where $r_t = r(X_t)$ denotes the (real) interest rates following a stochastic process. If $\pi_t = \pi(X_t)$ is the profit at time t, then the expected present value of the firm is

$$v(x) = \mathbb{E}_x \sum_{t=0}^{\infty} \left(\prod_{i=0}^{t} \beta_i \right) \pi_t,$$

given current state $X_0 = x$. Let $A(x, x') = P(x, x')/(1 + r(x))$ for all $(x, x') \in \mathsf{X} \times \mathsf{X}$. If $\rho(A) < 1$, then Assumption 2.3.3 is satisfied and $v = \pi + Av$ (Sargent and Stachurski, 2023). $\square$

Example 2.3.2 (Uzawa Time Preferences). Mendoza (1991), Schmitt-Grohé and Uribe (2003), Vasilev (2022a,b), and Izadi and Lamsoo (2022) study a small open economy where a representative household has Uzawa preference, characterised by the richer being more impatient than the poor. Uzawa preference has the merits of stabilising a small open economy and generating a non-degenerate distribution of wealth (Guest and McDonald, 2001). The household chooses consumption c_t and working hours h_t to maximise the utility

$$\begin{aligned}
&\mathbb{E}_0 \sum_{t=0}^{\infty} \theta_t U(c_t, h_t) \\
&U(c, h) = \frac{(c - \nu^{-1} h^{\nu})^{1-\gamma}}{1 - \gamma} \\
&\theta_{t+1} = b(c_t, h_t)\theta_t, \quad t \geqslant 0 \qquad \text{with } \theta_0 = 1 \\
&b(c, h) = (1 + c - \nu^{-1} h^{\nu})^{-\psi},
\end{aligned}$$

where $\nu > 1$ is the labour supply elasticity, $\psi > 0$ is the elasticity of discount factor with respect to component $1 + c - \nu^{-1} h^{\nu}$, and $\gamma > 1$ measures the degree of relative risk aversion. The composite commodity $c_t - \nu^{-1} h^{\nu}$ is assumed to be positive, thereby bounding utility from above.[13] We can see that $b(c, h) < 1$ when $c > \nu^{-1} h^{\nu}$. Hence, the feasible correspondence is the subset of $\{(c, h) \colon c > \nu^{-1} h^{\nu}\}$ such that the discount factors are strictly less than one. To solve the problem by discretization in programming, we can assume there is $\varepsilon > 0$ such that $c - \nu^{-1} h^{\nu} \geqslant \varepsilon > 0$, ensuring $b(c, h) \leqslant (1 + \varepsilon)^{-\psi} < 1$ for

---

[13] Otherwise, since $\gamma > 1$, if $c \leqslant \nu^{-1} h^{\nu}$, then $U(c, h) \to \infty$ as $c \uparrow \nu^{-1} h^{\nu}$. Then, households can arbitrarily increase utility. Since the domain, $c \neq \nu^{-1} h^{\nu}$, is open, there is no maximiser in this case.

any (c, h), so we have $\sup_x d_1(x) < 1$, defined by (2.12), whence Assumption 2.3.2 or 2.3.3 holds.[14] □

Example 2.3.3. Durdu et al. (2009) consider a small open economy where a representative household chooses the optimal consumption c_t and maximises the preference,

$$\mathbb{E}_0 \sum_{t=0}^{\infty} \exp\left\{ - \sum_{\tau=0}^{t-1} \psi \ln(1+c_\tau) \right\} \frac{c_t^{1-\gamma}}{1-\gamma},$$

where $\psi > 0$ is the elasticity of the rate of time preference with respect to $1+c$. Since $\beta_t = \exp\{-\psi \ln(1+c_t)\} < 1$ for $c_t > 0$, if we discretise the consumption space and consider inner solutions, then we have $c \geqslant \varepsilon$ and $\beta(c) \leqslant 1/(1+\varepsilon)^\psi$ for some $\varepsilon > 0$. □

Example 2.3.4 (Uzawa Time Preferences with Stochastic Discounting). This example illustrates both the Uzawa time preference discussed by Durdu et al. (2009) and the concept of state-dependent discounting introduced by Hubmer et al. (2021):[15]

$$\beta_t = \beta(c_t, Z_t) = Z_t \exp\{-\psi \ln(1+c_t)\},$$

where c_t is consumption and Z_t is an AR(1) process follows

$$Z_{t+1} = \rho_Z Z_t + (1-\rho_Z)\mu_Z + \sigma_\varepsilon \varepsilon_{t+1} \qquad \{\varepsilon_t\} \overset{IID}{\sim} N(0,1) \tag{2.13}$$

with $\rho_Z = 0.992, \mu_Z = 0.944, \sigma_\varepsilon = 0.0006$. Let Q be the transition kernel of Z_t. They discretise the process onto a grid of $N = 15$ states by Tauchen's method which allows us to write the operator L, defined in Assumption 2.3.3, as a matrix

$$L_{ij} = \beta(x_i) Q(x_i, x_j), \qquad 1 \leqslant i, j \leqslant N.$$

The spectral radius of matrix L is 0.9469, computed by Stachurski and Zhang (2021). Then, since $\beta_t \leqslant Z_t$, there exists $n \in \mathbb{N}$ such that

$$\mathbb{E}_0 \beta_0 \beta_1 \cdots \beta_{n-1} \leqslant \mathbb{E}_0 Z_0 Z_1 \cdots Z_{n-1} < 1$$

---

[14] On the other hand, the steady state satisfies $\beta(c,h)(1+r) = 1$ so that $\beta(c,h) < 1$, where $r > 0$ is the real interest rate. For example, Vasilev (2022a) calibrate the parameters such that $b(c,h) = 0.982$. If we solve the model around the steady state, the discount factor is bounded away from one.

[15] The literature of state-dependent discounting with AR(1) process includes Hills and Nakata (2018), Hills et al. (2019) and Nakata (2016).

for any consumption path. □

Example 2.3.5 (Becker-Mulligan Time Preferences). Becker and Mulligan (1997) propose the endogenous time preference that increases in future-oriented capital. Stern (2006) considers Becker-Mulligan time preferences in an optimal growth model:

$$\begin{aligned} &\sup_{\{c_t, s_t, k_t\}_{t \geqslant 0}} \sum_{t=1}^{\infty} \prod_{i=1}^{t-1} \beta(s_i) U(c_t) \\ &\text{s.t. } c_t + \pi s_t + k_t \leqslant f(k_{t-1}), \\ &\qquad c_t \geqslant 0, s_t \geqslant 0, k_t \geqslant 0, \text{ for all } t \in \mathbb{N}_0 \end{aligned}$$

where k_0 is given, $\beta \colon [0, \infty) \to (0, \infty)$ is continuous, concave and strictly increasing, π is the price of s_t, and k_t denotes capital.[16] Stern (2006) assumes that f is continuous and strictly increasing, and there exists a $k_m \geqslant 0$ such that $\beta(k_m/\pi) < 1$ and $f(k) < k$ whenever $k > k_m$. Hence, we have $\beta(s) \leqslant \beta(k_m/\pi) < 1$ for all $s \in [0, k_m/\pi]$, so Assumption 2.2.1 is satisfied.

Erol et al. (2011) also consider a similar model:

$$\begin{aligned} &\sup_{\{c_t, k_{t+1}\}_{t \geqslant 1}} \sum_{t=0}^{\infty} \prod_{i=1}^{t} \beta(k_i) U(c_t) \\ &\text{s.t. } c_t + k_{t+1} \leqslant f(k_t), \\ &\qquad c_t \geqslant 0, k_t \geqslant 0, \qquad \text{for all } t \in \mathbb{N}_0 \end{aligned}$$

where k_0 is given. Erol et al. (2011) assumes that $\beta \colon [0, \infty) \to (0, \infty)$ is continuous, differentiable, strictly increasing and $\sup_{k>0} \beta(k) < 1$, so Assumption 2.3.2 is satisfied.[17]

□

Example 2.3.6 (Becker-Mulligan Preference with Stochastic Discounting). This example modifies the Becker-Mulligan Preference of Erol et al. (2011) to incorporate stochastic

---

[16] The time preference $\beta(s_t)$ is interpreted as the degree to which generation t cares for generation $t+1$, while variable s_t represents actions or resources that the parent could take to strengthen the relationship with her child.

[17] Erol et al. (2011) also assumes that k has a compact support.

uncertainty. Consider a model from Example 2.3.5:

$$\begin{aligned} &\sup_{\{c_t,k_{t+1}\}_{t\geqslant 1}} \mathbb{E}_0 \sum_{t=0}^{\infty} \left(\prod_{i=1}^{t} Z_i \beta(k_i) \right) U(c_t) \\ &\text{s.t. } c_t + k_{t+1} \leqslant f(k_t), \\ &\qquad c_t \geqslant 0, k_t \geqslant 0, \qquad \text{for all } t \in \mathbb{N}_0 \end{aligned}$$

where k_0 is given, and $\{Z_t\}_{t\geqslant 0}$ is an AR(1) process satisfying that there exists $n \in \mathbb{N}$ such that $\mathbb{E}_0 Z_1 \cdots Z_n < 1$. Similar to Erol et al. (2011), assumes that $\beta\colon [0,\infty) \to (0,\infty)$ is continuous, differentiable, strictly increasing and $\sup_{k>0} \beta(k) \leqslant 1$. To this end, we have

$$\mathbb{E}_0 Z_1 \beta(k_1) Z_2 \beta(k_2) \cdots Z_n \beta(k_n) \leqslant \mathbb{E}_0 Z_1 Z_2 \cdots Z_n < 1.$$

for any $\{k_t\}_{t\geqslant 0}$, which implies Assumption 2.3.2 and 2.3.3. □

2.3.3. Optimality. In this section, we demonstrate the optimality of an MDP under eventual discounting. Our results indicate that the standard methods for computing optimal values and policies apply to state-action-dependent dynamic programming frameworks, such as Uzawa or Becker-Mulligan time preferences, even when stochastic discounting is considered, as Example 2.3.4 and Example 2.3.6.

THEOREM 2.3.1. *Let $\mathscr{M}$ be regular. If either Assumption 2.3.1, 2.3.2 or 2.3.3 holds, then the following statements are true.*

- *(a) T_σ is eventually contracting on $mb\mathsf{X}$ with fixed point v_σ for all $\sigma \in \Sigma$,*
- *(b) T is contracting on $(cb\mathsf{X}, \|\cdot\|_w)$ if Assumption 2.3.2 holds,*
- *(c) v^* is the unique solution to the Bellman equation in $cb\mathsf{X}$,*
- *(d) VFI converges to v^*,*
- *(e) HPI converges to σ^*, and OPI converge to v^* if Assumption 2.3.1 holds,*
- *(f) Bellman's principle of optimality holds, and*
- *(g) at least one optimal (continuous) policy exists,*

If Condition 2.3.2 is satisfied and Assumption 2.3.2 or 2.3.3 holds, then

- *(α) HPI converges to σ^*, and*

(β) OPI converges to v^*,

where all the iterated greedy policies to HPI and OPI are continuous.

Theorem 2.3.1 generalises the traditional constant-discounting dynamic programming theory to the case of state-action-dependent discounting. In particular, it establishes that if the MDP is regular and eventually discounting, then several key properties hold: the existence of an optimal policy, the validity of Bellman's principle of optimality, and convergence of VFI to v^*.

Moreover, Theorem 2.3.1 affirms that the optimality of Example 2.3.4 of Uzawa preferences or Example 2.3.6 of Becker-Mulligan time preferences holds true. In addition, it asserts that the optimal policy can be derived by the v^*-greedy policy, where v^* is computed by VFI.

In the following lemma, we illustrate that the assumption of eventual discounting may be necessary for ensuring the existence of a solution to the Bellman equation under appropriate conditions.

Lemma 2.3.3. *Let $\mathscr{M}$ be regular, and X be compact. Suppose that there exists $f \in cb\mathsf{X}_+$ such that $r_\sigma(x) \geqslant f(x) > 0$ for all $x \in \mathsf{X}$ and $\sigma \in \Sigma$. If T has a fixed point $v \in cb\mathsf{X}_+$, then $\rho(\hat{L}) < 1$ and $\sup_x d_n(x) < 1$.*

Therefore, considering a compact state space, if the MDP admits strictly positive rewards $r(x, a, x') \geqslant f(x) > 0$ for $(x, a, x') \in \mathsf{G} \times \mathsf{X}$ and $f \in cb\mathsf{X}_+$, then the eventual discounting condition, $\sup_x d_n(x) < 1$ for some $n \in \mathbb{N}$, is necessary for the existence of a solution to the Bellman equation (i.e., T admits a fixed point in $cb\mathsf{X}_+$.) This implies that the spectral radius condition, $\rho(\hat{L}) < 1$, is necessary for a well-behaved dynamic programming. If $\rho(\hat{L}) \geqslant 1$, Lemma 2.3.3 suggests that VFI may not converge, and there would be no solution to the Bellman equation or maximised lifetime value.

Furthermore, Theorem 2.3.1 implies that T and T_σ are contractive in some weighted supremum norm, enabling analysis of convergence rate. The following lemma provides the contraction modulus for T and T_σ for continuous policy $\sigma \in \Sigma_C$. In the following lemma, assuming eventual discounting, let

$$w_\sigma(x) := \mathbb{E}_x^\sigma \sum_{t=0}^{\infty} \prod_{i=0}^{t-1} \beta_\sigma(X_i, X_{i+1}) \quad (x \in \mathsf{X}) \quad \text{and} \quad \lambda_\sigma := \sup_{x \in \mathsf{X}} \frac{w_\sigma(x) - 1}{w_\sigma(x)} < 1$$

for all $\sigma \in \Sigma_C$. Let

$$w := \sum_{n=1}^{\infty} \hat{L}^n \mathbb{1} \qquad \text{and} \qquad \lambda := \sup_{x \in \mathsf{X}} \frac{w(x) - 1}{w(x)} < 1.$$

Corollary 2.3.1. *If $\mathscr{M}$ is regular, X is compact, and either Assumption 2.3.2 or 2.3.3 holds, then T_σ is contractive on $(cb\mathsf{X}, \|\cdot\|_{w_\sigma})$ with modulus λ_σ for all $\sigma \in \Sigma_C$, and T is contractive on $(cb\mathsf{X}, \|\cdot\|_w)$ with modulus λ. Moreover, $L_\sigma w_\sigma \leqslant \lambda_\sigma w_\sigma$ for all $\sigma \in \Sigma_C$ and $\hat{L} w \leqslant \lambda w$.*

2.4. Necessity of Eventual Discounting

In this section, we examine the necessity of eventual discounting for both the existence and uniqueness of the policy values. Specifically, we demonstrate the necessity of the spectral radius conditions to the existence and uniqueness of fixed points of policy operators in the context of a compact state space. Since the convergence of HPI and OPI relies on the global stability of policy operators, eventual discounting for policy operators plays an essential role in MDPs. Furthermore, we establish that the convergence rate of T_σ corresponds to the spectral radius of the expected discounting operator L_σ, where the corresponding eigenfunction serves as the weight for the weighted supremum norm.

2.4.1. An MDP with a Compact State Space. In this section, we analyze an MDP with a compact state space and positive rewards. Let $\mathscr{M} = (\mathsf{X}, \mathsf{A}, \Gamma, B)$ be an MDP, where value aggregator B follows (2.9). Let $mb\mathsf{X}_+$ (resp., $cb\mathsf{X}_+$) denote the set of everywhere positive functions in $mb\mathsf{X}$ (resp., $cb\mathsf{X}$.) Throughout this section, we consider the following regular conditions.

Condition 2.4.1.

(i) X is compact, and A is a convex subset of a vector space.

(ii) β and P are continuous.

(iii) $(x, a) \mapsto r(x, a)$ is strictly positive and continuous on G, and $a \mapsto r(x, a)$ is strictly concave for all $x \in \mathsf{X}$.

(iv) Γ is nonempty, continuous, compact-valued and convex-valued.

(v) $(x, a) \mapsto \int_{\mathsf{X}} f(y)\beta(x, a, y)P(x, a, \mathrm{d}y)$ is continuous for all $f \in cb\mathsf{X}$.

(vi) $a \mapsto \int_{\mathsf{X}} f(y)\beta(x, a, y)P(x, a, \mathrm{d}y)$ is concave for all $x \in \mathsf{X}$ and $f \in cb\mathsf{X}$.

Condition 2.4.1 is strict since it may be challenging to show Condition 2.4.1 (iii) and (vi). In some simple cases, Condition 2.4.1 (vi) holds if β and P are only state-dependent. Another possible condition is that β does not depend on future states, P is not action-dependent, and $a \mapsto \beta(x, a) \int f(y)P(x, \mathrm{d}y)$ is concave for all $x \in \mathsf{X}$ and $f \in cb\mathsf{X}$.

Condition 2.4.1 is similar to Condition 2.3.2. The main differences are that X is compact and r is strictly positive, which can be achieved by scaling up a bounded reward. Consequently, a continuous v-greedy policy exists for any $v \in cb\mathsf{X}$. Moreover, T and T_σ are self-maps on $cb\mathsf{X}_+$ for $\sigma \in \Sigma_C$.

Lemma 2.4.1. *If Condition 2.4.1 holds, then the following statements are true.*

(a) T_σ is a self-map on $mb\mathsf{X}_+$ for all $\sigma \in \Sigma$.

(b) T and T_σ are self-maps on $cb\mathsf{X}_+$ for all $\sigma \in \Sigma_C$.

(c) For all $v \in cb\mathsf{X}_+$, there exists a continuous v-greedy policy.

An MDP $\mathscr{M}$ is *ergodic* if for each policy $\sigma \in \Sigma$ the induced P_σ-Markov chain is ergodic/irreducible. The next assumption guarantees that $\mathscr{M}$ is ergodic. Given $\sigma \in \Sigma$, let P_σ^n be defined by $P_\sigma^1 = P_\sigma$ and $P_\sigma^n = \int P_\sigma(x, \mathrm{d}y)P_\sigma^{n-1}(\mathrm{d}y, z)$ for all $(x, z) \in \mathsf{X} \times \mathsf{X}$ and $i \in \mathbb{N}$.

Assumption 2.4.1. For all $\sigma \in \Sigma$, there exists an $n \in \mathbb{N}$ such that the transition density P_σ^n is positive everywhere.

Assumption 2.4.1 implies that the induced Markov chain is irreducible since P_σ^n is positive everywhere, and then all states will be visited eventually.

Fix $\sigma \in \Sigma_C$. Denote $k_\sigma(x, \mathrm{d}x') := \beta_\sigma(x, x')P_\sigma(x, \mathrm{d}x')/\bar{\beta}_\sigma(x)$ for all $(x, x') \in \mathsf{X} \times \mathsf{X}$, where

$$\bar{\beta}_\sigma(x) := \int \beta_\sigma(x, x')P_\sigma(x, \mathrm{d}x').$$

Note that since β is strictly positive, $\bar{\beta}$ is strictly positive. We can write L_σ as

$$L_\sigma v(x) = \bar{\beta}_\sigma(x) \int v(x')k_\sigma(x, \mathrm{d}x') \qquad (x \in \mathsf{X}, v \in cb\mathsf{X}_+).$$

Denote e_σ as the eigenvector of L_σ corresponding to its spectral radius satisfying $L_\sigma e_\sigma = \rho(L_\sigma)e_\sigma$. We show in the appendix that L_σ^2 is a compact operator such that $e_\sigma \in cb\mathsf{X}$ exists and is everywhere positive. The following proposition demonstrates that the spectral radius and the corresponding eigenfunction dominate the contraction of the policy operator, in the sense that $\rho(L_\sigma) < 1$ if and only if T_σ is contracting on $(cb\mathsf{X}_+, \|\cdot\|_{e_\sigma})$ with modulus $\rho(L_\sigma)$. We consider the following eventual discounting assumption.

Assumption 2.4.2.

(a) T_σ is eventually discounting for all $\sigma \in \Sigma$.

(b) $\rho(\bar{L}) < 1$, where $\bar{L} \colon cb\mathsf{X}_+ \to cb\mathsf{X}_+$ is defined by

$$\bar{L}v := \sup_{\sigma \in \Sigma_C} L_\sigma v$$

for $v \in cb\mathsf{X}_+$.

Given the regular Conditions 2.4.1, we can restrict the feasible policy set to the continuous policies when searching for greedy policies. In this context, we focus on the sublinear operator $\bar{L}$. Then, the maximised expected multiplicative of discount factors, $\bar{L}^n \mathbb{1}$, becomes

$$d_{C,n}(x) := \sup_{\sigma_1 \in \Sigma_C} \left\{ \mathbb{E}_x^{\sigma_1} \beta_{\sigma_1}(X_0, X_1) \sup_{\sigma_2 \in \Sigma_C} \left\{ \mathbb{E}_{X_1}^{\sigma_2} \beta_{\sigma_2}(X_1, X_2) \sup_{\sigma_3 \in \Sigma_C} \left\{ \cdots \right.\right.\right.$$
$$\left.\left.\left. \cdots \sup_{\sigma_n \in \Sigma_C} \left\{ \mathbb{E}_{X_{n-1}}^{\sigma_n} \beta_{\sigma_n}(X_{n-1}, X_n) \right\} \cdots \right\} \right\} \right\}.$$

The eventual discounting $\rho(\bar{L}) < 1$ is equivalent to $\sup_x d_{C,n}(x) < 1$. The following proposition shows that the eventual discounting of T_σ is necessary for the existence and uniqueness of v_σ.

Proposition 2.4.1. *If Assumption 2.4.1 and Condition 2.4.1 hold, then for all $\sigma \in \Sigma_C$ the following statements are equivalent.*

(a) $\rho(L_\sigma) < 1$.

(b) There exists an $n_\sigma \in \mathbb{N}$ *such that* $d^\sigma_{n_\sigma} < 1$.

(c) T_σ *is globally stable on* $cb\mathsf{X}_+$.

(d) T_σ *is contracting on* $(cb\mathsf{X}_+, \| \cdot \|_{e_\sigma})$ *with modulus* $\rho(L_\sigma)$.

(e) T_σ *has a fixed point in* $cb\mathsf{X}_+$.

(f) T_σ *has a unique fixed point in* $cb\mathsf{X}_+$.

Moreover, if $\rho(L_\sigma) \geqslant 1$, *then* T_σ *has no fixed point in* $cb\mathsf{X}_+$.

Proposition 2.4.1 shows that $\rho(L_\sigma) < 1$ is necessary for the existence and uniqueness of v_σ, the fixed point of T_σ. If, in addition, we have $\sup_x d_{C,n}(x) < 1$ for some $n \in \mathbb{N}$, then the MDP is eventually discounting, ensuring that the optimal properties described in Theorem 2.3.1 hold true. We summarise these results in the following theorem, which can be considered as a corollary of Theorem 2.2.1. To apply Theorem 2.2.1, we use the eigenvectors as the weighting vector for the weighted supremum norm.

In the following theorem, let $w := \sum_{n=0}^\infty \bar{L}^n \mathbb{1}$ and $\lambda := \sup_x (w(x) - 1)/w(x)$.

Theorem 2.4.1. *Let Condition 2.4.1 hold. If* $\mathscr{M}$ *admits* $r_\sigma \geqslant f \in cb\mathsf{X}_+$ *for all* $\sigma \in \Sigma$, *and* T *has a fixed point in* $cb\mathsf{X}_+$, *then* $\rho(\bar{L}) < 1$. *If Assumption 2.4.1 and 2.4.2 hold, then the following statements are true.*

(a) T_σ *is eventually contracting on* $mb\mathsf{X}_+$ *for all* $\sigma \in \Sigma$, *and* T_σ *is contracting with modulus* $\rho(L_\sigma)$ *on* $(cb\mathsf{X}_+, \| \cdot \|_{e_\sigma})$ *for all* $\sigma \in \Sigma_C$,

(b) T *is contracting on* $(cb\mathsf{X}_+, \| \cdot \|_w)$ *with modulus* λ,

(c) v^* *is the unique solution to the Bellman equation in* $cb\mathsf{X}_+$,

(d) HPI converges to σ^*,

(e) VFI converges to v^*,

(f) OPI converges to v^*,

(g) Bellman's principle of optimality holds, and

(h) at least one optimal continuous policy exists.

where all the iterated greedy policies to HPI and OPI are continuous.

2.5. Applications

In this section, we apply the main results to optimal growth and optimal default problems.

2.5.1. Optimal Growth. In this section, we consider optimal growth problems with eventual discounting.

Example 2.5.1 (One-sector Optimal Growth, Continued). Let $\mathscr{M}$ be the MDP defined in Example 2.2.1. The Bellman equation is

$$v(k,z) = \sup_{c\in[0,zF(k,1)]} \Bigg\{ U(c) + \beta(z,c)\int_{\mathsf{X}} v(k',z')Q(z,\mathrm{d}z')\mathbb{1}\{k'=(1-\delta)k+zF(k,1)-c\}\mathrm{d}k' \Bigg\}.$$

Let

$$d_n(x) \coloneqq \sup_{\sigma_1\in\Sigma}\Big\{\mathbb{E}^{\sigma_1}_{k,z}\beta(Z_0,\sigma_1(K_0,Z_0)) \sup_{\sigma_2\in\Sigma}\Big\{\mathbb{E}^{\sigma_2}_{K_1,Z_1}\beta(Z_1,\sigma_2(K_1,Z_1)) \sup_{\sigma_3\in\Sigma}\Big\{\cdots \cdots \sup_{\sigma_n\in\Sigma}\Big\{\mathbb{E}^{\sigma_n}_{K_{n-1},Z_{n-1}}\beta(Z_{n-1},\sigma_n(K_{n-1},Z_{n-1}))\Big\}\cdots\Big\}\Big\}\Big\}.$$

for all $x \in \mathsf{X}$. Assume that there exists an $n \in \mathbb{N}$ such that $\sup_x d_n(x) < 1$. Then, Assumption 2.2.2 holds, and T_s is eventually discounting for all $\sigma \in \Sigma$. Theorem 2.3.1 shows that the optimality of dynamic programming holds and the related dynamic programming algorithms converge to the unique optimal value. □

Example 2.5.2 (Uzawa Time Preference with Stochastic Discounting, Continued). This example continues Example 2.3.4 by modifying the optimal saving problem in Hubmer et al. (2021) to the case of Uzawa endogenous time preference. Following Hubmer et al.

(2021), suppose that the policy operator is

$$T_\sigma v(x,z) = u(R(x,z)x + y(x,z) - \sigma(x)) \\ + \beta(R(x,z)x + y(x,z) - \sigma(x), z) \int v(\sigma(x), z')Q(z, \mathrm{d}z')$$

where $x \in \mathsf{X} := \mathbb{R}_+$ is the present asset, z is the exogenous shocks generated by Q as defined by Example 2.3.4, $R(x,z)$ is the gross return rates on asset holdings, $y(x,z)$ is the labour net income, $R(x,z)x + y(x,z) - \sigma(x) = c$ is the consumption, and $\sigma(x)$ is the asset or saving leaving to the next period. Assume that β and $\{Z_t\}$ are defined as Example 2.3.4. Let the utility function be

$$u(c) := \frac{c^{1-\gamma}}{1-\gamma} \qquad (\gamma > 1).$$

The feasible correspondence is

$$\Gamma(x,z) := \{x' \in \mathbb{R} \colon \bar{x} \leqslant x' \leqslant R(x,z)x + y(x,z)\}$$

To solve the problem numerically, we discretise both X and Z into finite grid points. To this end, the reward function is bounded, and the continuity assumptions in Condition 2.3.1 are satisfied. As discussed in Example 2.3.4, the MDP is eventually discounting. Therefore, all of the conclusions in Theorem 2.3.1 hold. ☐

2.5.2. Optimal Default. This example considers an optimal saving problem with default following Arellano (2008), Hatchondo et al. (2009), Yue (2010), Hatchondo et al. (2016) and Ma et al. (2022). We assume state-action-dependent discounting. A country with current assets w_t chooses between continuing to participate in international financial markets and defaulting. Let $y_t = y(Z_t, \xi_t)$ be output, where $\{Z_t\}$ is a Markov process and $\{\xi_t\}$ is an IID shock. Assume that default results in permanent exclusion from financial markets which yields the lifetime value

$$v^d(y,z) = \mathbb{E}_z \sum_{t=0}^{\infty} \prod_{i=0}^{t-1} \beta^d(Z_i) u(y_t),$$

where $\beta^d(z)$ represents the stochastic discount dependent on state z given defaulting. The value of continued participation in the financial market is

$$v^c(w,y,z) = \sup_{-b \leqslant w' \leqslant R(w+y)} u(w+y-w'/R) + \beta(z,w')\mathbb{E}_z v(w',Y',Z')$$

where $b > 0$ is a constant borrowing constraint, $\beta(z,w')$ is the discount factor depends on state z and wealth w', in the spirit of Becker-Mulligan time preferences in Example 2.3.5, and v is the value function satisfying

$$v(w,y,z) = \max\{v^d(y,z), v^c(w,y,z)\}.$$

The Bellman equation is

$$\begin{aligned} v(w,y,z) = \sup_{\substack{\delta \in \{0,1\} \\ -b \leqslant w' \leqslant R(w+y)}} & \left\{ \delta \left[\mathbb{E}_z \sum_{t=0}^{\infty} \prod_{i=0}^{t-1} \beta^d(Z_i) u(y_t) \right] \right. \\ & \left. + (1-\delta)\Big[u(w+y-w'/R) + \beta(z,w')\mathbb{E}_z v(w',Y',Z')\Big] \right\}. \end{aligned}$$

Let $\mathsf{X} = W \times Y \times Z$ where W, Y and Z are domains of w_t, y_t and z_t, respectively. Assume that u is continuous, and either u is bounded, or z_t and ξ_t have compact supports. Suppose that β^d and β are continuous, bounded, strictly positive. Assume there is $n \in \mathbb{N}$ such that

$$\sup_{(w,y,z)\in\mathsf{X}} \mathbb{E}_z \prod_{t=0}^{n-1} \beta^d(Z_t) < 1,$$

which implies the convergence of $\mathbb{E}_z \sum_{t=0}^{\infty} \prod_{i=0}^{t-1} \beta^d(Z_i) u(y_t)$. Moreover, let $\sigma(w,y,z) = (\sigma_\delta(w,y,z), \sigma_{w'}(w,y,z))$ be the policy of action (δ, w') given state $(w,y,z) \in \mathsf{X}$. Let L_σ be

$$\begin{aligned} L_\sigma h(w,y,z) &:= \mathbb{E}_{w,y,z} \beta(z, \sigma_{w'}(w,y,z)) h(\sigma_{w'}(w,y,z), Y', Z') \\ &\geqslant (1-\delta)\mathbb{E}_z \beta(z, \sigma_{w'}(w,y,z)) h(\sigma_{w'}(w,y,z), Y', Z') \end{aligned}$$

for all $(w,y,z) \in \mathsf{X}$, $\delta \in \{0,1\}$, and $h \in mb\mathsf{X}$. Suppose that $\rho(\hat{L}) < 1$, where $\hat{L}h = \sup_{\sigma \in \Sigma} L_\sigma h$ for $h \in mb\mathsf{X}$. Therefore, v^d is bounded and continuous, and the MDP is eventually discounting, so all conclusions in Theorem 2.3.1 hold.

2.5.3. Asset Pricing. In this section, we apply the eventual discounting concept to an asset pricing problem. Since there is no action space in this problem, it is not a dynamic programming model. However, it provides insight into how eventual discounting influences the convergence of the price function.

There is an ex-dividend contract that trades at prices Π_t and pays dividend D_t. To this end, purchasing this contract at t and selling at $t+1$ pays $\Pi_{t+1} + D_{t+1}$. Let the Lucas stochastic discount factor be

$$M_{t+1} = \bar{\beta} \left(\frac{C_{t+1}}{C_t} \right)^{-\gamma},$$

where $\bar{\beta} > 0$ is a constant discount factor measuring the impatience of the agent. Given the absence of arbitrage, the price at time t must satisfy

$$\Pi_t = \mathbb{E}_t M_{t+1} (\Pi_{t+1} + D_{t+1}). \tag{2.14}$$

Let $\{X_t\}_{t \geqslant 0}$ be a Q-Markov process on a compact state space X. Suppose that the dividend growth obeys

$$\ln \frac{D_{t+1}}{D_t} = \mu_d + X_t + \sigma_d \eta_{d,t+1}$$

where $\{\eta_{d,t}\}_{t \geqslant 0}$ is IID and standard normal. Moreover, consumption growth obeys

$$\ln \frac{C_{t+1}}{C_t} = \mu_c + X_t + \sigma_c \eta_{c,t+1}$$

where $\{\eta_{c,t}\}_{t \geqslant 0}$ is IID and standard normal. Let $V_t := \Pi_t / D_t$ be the price-dividend ratio. Then, we obtain

$$\begin{aligned} V_t = \frac{\Pi_t}{D_t} &= \mathbb{E}_t \left[M_{t+1} \frac{D_{t+1}}{D_t} \left(\frac{\Pi_{t+1}}{D_{t+1}} + 1 \right) \right] \\ &= \mathbb{E}_t \left[\bar{\beta} \exp(-\gamma \mu_c + \mu_d + (1-\gamma) X_t - \gamma \sigma_c \eta_{c,t+1} + \sigma_d \eta_{d,t+1}) \left(V_{t+1} + 1 \right) \right]. \end{aligned} \tag{2.15}$$

Conditioning on $X_t = x$, (2.15) yields the value function

$$v(x) = \int_{\mathsf{X}} \bar{\beta} \exp\left(-\gamma \mu_c + \mu_d + (1-\gamma) x + \frac{\gamma^2 \sigma_c^2 + \sigma_d^2}{2} \right) (1 + v(x')) Q(x, \mathrm{d}x')$$

for all $x \in \mathsf{X}$ and $v \in cb\mathsf{X}$. Define $A \colon cb\mathsf{X} \to cb\mathsf{X}$ by

$$Af(x, x') := \int_{\mathsf{X}} \bar{\beta} \exp\left(-\gamma \mu_c + \mu_d + (1-\gamma) x + \frac{\gamma^2 \sigma_c^2 + \sigma_d^2}{2} \right) f(x') Q(x, \mathrm{d}x')$$

for all $x \in \mathsf{X}$ and $f \in cb\mathsf{X}$. We define the corresponding discount function

$$\beta(x) := \bar{\beta} \exp\left(-\gamma\mu_c + \mu_d + (1-\gamma)x + \frac{\gamma^2\sigma_c^2 + \sigma_d^2}{2}\right)$$

for all $x \in \mathsf{X}$. Now, if $x \mapsto \beta(x) \int_{\mathsf{X}} f(x')Q(x, \mathrm{d}x')$ is continuous for any $f \in cb\mathsf{X}$, then a version of Theorem 2.3.1 without policy (or Lemma 2.8.14) shows that A has a unique fixed point if $\rho(A) < 1$.[18] Furthermore, if Q admits a continuous density and there is $n \in \mathbb{N}$ such that Q^n is everywhere positive, then a version of Proposition 2.4.1 without policy shows that A has a unique fixed point if and only if $\rho(A) < 1$.

2.5.3.1. Incomplete Market with Subjective Discounting. This section considers the asset pricing with heterogeneous expectations in Harrison and Kreps (1978). Investors are risk-neutral. Suppose that there are finitely many investor classes, denoted by set A. Agents in each class $a \in \mathsf{A}$ have a subjective probability distribution $P_a \colon \mathsf{X} \times \mathsf{X} \to \mathbb{R}_+$. Fix the random states $\{X_t\} \subset \mathsf{X}$. We further assume that agents in class $a \in \mathsf{A}$ have subjective time preferences such that $\beta_t = \beta_a(X_t, X_{t+1})$ for $a \in \mathsf{A}$. Let $\Pi_t = \pi(X_t)$ and $D_t = d(X_t)$ for all t. Harrison and Kreps (1978) shows that the price scheme is consistent if and only if[19]

$$\Pi_t = \max_{a \in \mathsf{A}} \mathbb{E}^a[\beta_t(\Pi_{t+1} + D_{t+1})].$$

The Bellman equation is

$$\begin{aligned}
\Pi_t &= \max_{a \in \mathsf{A}} \mathbb{E}^a[\beta_t(\Pi_{t+1} + D_{t+1})]. \\
\pi(x) &= \max_{a \in \mathsf{A}} \mathbb{E}^a[\beta_a(x, X')(\pi(X') + d(X'))] \\
&= \max_{a \in \mathsf{A}} \int_{\mathsf{X}} \beta_a(x, x')[\pi(x') + d(x')] P_a(x, \mathrm{d}x')
\end{aligned}$$

for $x \in \mathsf{X}$. Define $r(x, a) := \int_{\mathsf{X}} \beta_a(x, x')\pi(x')P_a(x, \mathrm{d}x')$ and

$$T_a\pi(x) := r(x, a) + \int_{\mathsf{X}} \beta_a(x, x')\pi(x')P_a(x, \mathrm{d}x')$$

---

[18] Alternatively, consider that there is a single feasible policy for applying Theorem 2.3.1.

[19] A price is called consistent when it prevents any investor from achieving an excessive expected return through adroit and legitimate speculation.

for all $a \in \mathsf{A}$ and $x \in \mathsf{X}$. If r is bounded and continuous, and the MDP is eventually discounting that either Assumption 2.3.1, 2.3.2 or 2.3.3 holds, then Theorem 2.3.1 shows that the optimal price π^* is unique, the optimality of dynamic programming holds, and the VFI will converge to π^*.

2.6. Unbounded Rewards

In this section, we study an eventually discounting MDP with an unbounded reward function, which generalises the main results in Section 2.3.

2.6.1. MDP with Unbounded Rewards. Throughout this section, we assume the following regular conditions with unbounded reward $r\colon \mathsf{G} \to \mathbb{R} \cup \{-\infty, \infty\}$.[20]

Condition 2.6.1.

(i) Γ is nonempty and compact-valued.

(ii) r is u.s.c..

(iii) β is bounded, continuous, and strictly positive.

(iv) $(x, a) \mapsto \int_{\mathsf{X}} f(y)\beta(x, a, y)P(x, a, \mathrm{d}y)$ is bounded and continuous on G whenever $f \in cb\mathsf{X}$.

If Condition 2.6.1 is satisfied, define spaces $\mathcal{V}$ and $\mathcal{G}$ by

$$\begin{aligned}\mathcal{V} := \{v\colon \mathsf{X} \to \mathbb{R} \cup \{-\infty\}\colon\, & v \text{ is u.s.c., } v/\kappa \text{ is bounded above,} \\ & \text{and } (x, a) \mapsto \mathbb{E}_{x,a} v(X') \text{ is bounded below}\}, \\ \mathcal{G} := \{g\colon \mathsf{G} \to \mathbb{R}\colon\, & g \text{ is u.s.c., bounded below, and } \|g\|_\kappa < \infty\}.\end{aligned}$$

The regular conditions, with the assumptions below, ensure the existence of a greedy policy. In this section, we say that an MDP is *regular* if Condition 2.6.1 holds.

Define the maximal reward function $\bar{r}$ and the (conditionally) expected maximal reward function $\hat{r}$ as

$$\bar{r} := \sup_{a \in \Gamma(x)} r(x, a) \quad (x \in \mathsf{X}) \quad \text{and} \quad \hat{r}(x, a) := \mathbb{E}_{x,a}\bar{r}(X') \quad ((x, a) \in \mathsf{G}). \tag{2.16}$$

---

[20] The conditions for the case of measurable functions are presented in the appendix.

We introduce a mild assumption that $\hat{r}$ is bounded below and $\bar{r}$ is bounded above by some function following Ma et al. (2022).

Assumption 2.6.1.

(i) There exist a $\kappa \colon \mathsf{X} \to [1, \infty)$ and a constant $d \geqslant 0$ such that $\bar{r}(x) \leqslant d\kappa(x)$ and $a \mapsto \mathbb{E}_{x,a}\kappa(X')$ is continuous for all $x \in \mathsf{X}$. Moreover, if Condition 2.6.1 holds, then κ is also continuous.

(ii) There exist $n \in \mathbb{N}$, $\alpha \in \mathbb{R}$ and a linear operator $L \colon \mathcal{V} \to \mathcal{V}$ such that $\rho(L) < 1$, $\|L\| < \infty$, $\|L^n\| < 1$, $\alpha \in (0, 1/\|L^n\|^{1/n})$ and

$$\mathbb{E}_{x,a}\beta(x, a, X')\kappa(X')L^t\mathbb{1}(X') \leqslant \alpha\kappa(x)L^{t+1}\mathbb{1}(x)$$

for all $(x, a) \in \mathsf{G}$ and $t \in \{0, 1, \ldots, n-1\}$.

(iii) $\hat{r}$ is bounded below.

If $\kappa \equiv \mathbb{1}$ and $\alpha = 1$, then Assumption 2.6.1 (ii) is the same as Assumption 2.2.3. Hence, Assumption 2.2.3 extends the eventual contracting of Assumption 2.2.3.[21] Also, Assumption 2.6.1 implies that the pair $(\mathcal{G}, \|\cdot\|_\kappa)$ is a Banach space.

Recall that the Bellman equation is

$$v(x) = \sup_{a \in \Gamma(x)} \{r(x, a) + \mathbb{E}_{x,a}[\beta(x, a, X')v(X')]\}.$$

Define the *action-value function* $g(x, a) \coloneqq \mathbb{E}_{x,a}\beta(x, a, X')v(X')$ for any $(x, a) \in \mathsf{G}$, which is the expected discounted future value conditioning on (x, a), with state evaluation function v. The Bellman equation can be written as $v(x) = \sup_{a \in \Gamma(x)}\{r(x, a) + g(x, a)\}$. Changing (x, a) to (x', a'), multiplying $\beta(x, a, x')$, and taking expectation yield

$$g(x, a) = \mathbb{E}_{x,a}\left[\beta(x, a, X') \sup_{a' \in \Gamma(X')} \{r(X', a') + g(X', a')\}\right] \tag{2.17}$$

for all $(x, a) \in \mathsf{G}$. Define the *expected value operator* $E \colon \mathcal{V} \to \mathcal{G}$ by

$$Ev(x, a) \coloneqq \mathbb{E}_{x,a}\beta(x, a, X')v(X') \qquad ((x, a) \in \mathsf{G}, v \in \mathcal{V}).$$

---

[21] Alternatively, we can consider Assumption 2.3.2 in this framework.

Define the *maximum value operator* $M\colon \mathcal{G} \to \mathcal{V}$ by

$$Mg(x) = \sup_{a \in \Gamma(x)} \{r(x,a) + g(x,a)\} \qquad (x \in \mathsf{X}, g \in \mathcal{G}).$$

We say that a policy $\sigma \in \Sigma$ is *g-greedy* if $r(x, \sigma(x)) + g(x, \sigma(x)) = Mg(x)$ for all $x \in \mathsf{X}$. Let g^* be the solution to (2.17). We will show that the value function is given by $v^*(x) = \sup_{a \in \Gamma(x)}\{r(x,a) + g^*(x,a)\}$ for all $x \in \mathsf{X}$, and a policy $\sigma \in \Sigma$ is optimal if and only if σ is g^*-greedy.

Define the *expected value Bellman operator* R on $\mathcal{G}$ by

$$Rg(x,a) := \mathbb{E}_{x,a}\left[\beta(x,a,X') \sup_{a' \in \Gamma(X')} \{r(X',a') + g(X',a')\}\right] \quad ((x,a) \in \mathsf{G}, g \in \mathcal{G}).$$

The *action-value function iteration* (AFI) is the iteration $\{g_t\}_{t \geqslant 0} \subset \mathcal{G}$ such that $g_{t+1} = Rg_t$ for all $t \in \mathbb{N}_0$ with $g_0 \in \mathcal{G}$. The following theorem indicates that the optimality results hold for an eventual discounting MDP with unbounded rewards.

Theorem 2.6.1. *If $\mathscr{M}$ is regular and Assumption 2.6.1 holds, then the following statements are true.*

(a) v^* and v_σ are well-defined for all $\sigma \in \Sigma$,

(b) R is globally stable on $(\mathcal{G}, \|\cdot\|_\kappa)$, and T is globally stable on $(\mathcal{V}, \|\cdot\|)$,

(c) R admits a unique fixed point g^* in $\mathcal{G}$, and T admits a unique fixed point v^* in $\mathcal{V}$,

(d) $v^* = Mg^* \in \mathcal{V}$ and $g^* = Ev^* \in \mathcal{G}$,

(e) VFI converges to v^*, AFI converges to g^*,

(f) at least one optimal policy exists, and

(g) a feasible policy is optimal if and only if it is g^*-greedy if and only it is v^*-greedy.

Theorem 2.6.1 demonstrates that if an MDP with unbounded reward satisfies the eventual discounting conditions, then AFI converges to g^*, and a g^*-greedy policy is optimal.

2.6.2. Application in Optimal Savings. An agent solves an optimal savings problem with borrowing constraint:

$$\begin{aligned} &\sup \mathbb{E} \sum_{t=0}^{\infty} \prod_{i=0}^{t-1} \beta_i u(c_t) \\ \text{s.t. } &0 \leqslant c_t \leqslant w_t \\ &w_{t+1} = R_{t+1}(w_t - c_t) + y_{t+1} \\ &(w_0, y_0) \text{ given.} \end{aligned} \tag{2.18}$$

Here $\beta_t \in (0, \infty)$ is the discount factor, $u \colon \mathbb{R}_+ \to \mathbb{R} \cup \{-\infty, \infty\}$ is a utility function, c_t denotes consumption, $w_t \geqslant 0$ denotes wealth, $y_t \geqslant 0$ denotes non-financial income, and R_t denotes the gross rate of return on financial income. Assume that asset return, income, and discount factors satisfy

$$R_t = R(\varepsilon_t), \quad y_t = y(z_t, \varepsilon_t), \quad \text{and} \quad \beta_t = \beta(z_t, z_{t+1})$$

where β is strictly positive, bounded and continuous, z_t is a Markov process with state space Z and ε_t is an IID shock that could be vector-valued. We first note that Condition 2.6.1 is satisfied. The Bellman equation is

$$v(w, z) = \sup_{0 \leqslant c \leqslant w} \{u(c) + \mathbb{E}_z[\beta(z, Z')v(R(\varepsilon')(w - c) + y(Z', \varepsilon'), Z')]\}$$

To this end, the state is $x = (w, z) \in \mathbb{R}_+ \times \mathsf{Z}$ and the action is $a = c \in \Gamma(x) = [0, w]$. Suppose that u is u.s.c., increasing, $\inf_z \mathbb{E}u(y(Z', \varepsilon')) > -\infty$, and there exist $p > 0$ and $q > 1$ such that

$$u(c) \leqslant pc + q \text{ for all } c > 0.$$

Define the weighting function by $\kappa(x) = pw + q$ for all $x = (w, z)$ and set $d = 1$ for Assumption 2.6.1. To ensure Assumption 2.6.1, assume

$$\sup_z \mathbb{E}_z \beta(z, Z') y(Z', \varepsilon') < \infty, \quad \text{and} \quad \mathbb{E}R(\varepsilon') > 1, \tag{2.19}$$

In addition, assume that there is $m \in \mathbb{N}$ such that

$$\sup_{z \in \mathsf{Z}} \mathbb{E}_z \beta_0 \beta_1 \beta_2 \cdots \beta_{m-1} (\mathbb{E}R(\varepsilon'))^m < 1.$$

This assumption is the eventual discounting for Assumption 2.6.1 (b). Define operator L by $Lh(x) := \mathbb{E}_z \beta(z, Z')h(X')\mathbb{E}R(\varepsilon')$ for all $x \in \mathsf{X}$.

LEMMA 2.6.1. *If $\mathscr{M}$ follows (2.18) and all the above corresponding assumptions hold, then there is large enough $q > 1$ such that $\kappa(x) = pw + q$ and*

$$\mathbb{E}_{x,a}\beta(x, a, X')\kappa(X')L^n \mathbb{1}(x') \leqslant \kappa(x)L^{n+1}\mathbb{1}(x)$$

for all $x \in \mathsf{X}$ and all $n \in \{1, \dots, m-1\}$. Moreover, Assumption 2.6.1 (a) and (c) hold.

By Lemma 2.6.1, it remains to show $\rho(L) < 1$ to ensure Assumption 2.6.1 (b). Since $\sup_{z \in \mathsf{Z}} \mathbb{E}_z \beta_0\beta_1\beta_2 \cdots \beta_{m-1}(\mathbb{E}R(\varepsilon'))^m < 1$ for some $m \in \mathbb{N}$, we have $\|L^m \mathbb{1}\| < 1$ and then $\rho(L) < 1$. Therefore, Assumption 2.6.1 is satisfied, and then all conclusions of Theorem 2.6.1 hold.

2.7. Extension: Risk-Sensitive Preferences

We study the risk-sensitive preference models with state-action-dependent discounting in this section, where the agent is risk-averse in future utility and future consumption. Let an MDP satisfy the following regular conditions.

CONDITION 2.7.1.

- (i) $(x, a) \mapsto r(x, a)$ is u.s.c.,
- (ii) β is bounded and strictly positive,
- (iii) Γ is nonempty, compact-valued, and continuous,
- (iv) $(x, a) \mapsto (-\beta(x, a)/\theta) \ln \int \exp(-\theta h(y))P(x, a, \mathrm{d}y)$ is u.s.c. for any $h \in c b\mathsf{X}$,

Let $\theta > 0$ be the agents' risk-sensitive coefficient. For any feasible policy $\sigma \in \Sigma$ and Borel measurable function $v \colon \mathsf{X} \to \mathbb{R} \cup \{-\infty\}$, let

$$T_\sigma v(x) := r(x, \sigma(x)) - \frac{\beta(x, \sigma(x))}{\theta} \ln \mathbb{E}_x^\sigma e^{-\theta v(X')} \qquad (x \in \mathsf{X}). \tag{2.20}$$

The *σ-value function* is defined by

$$v_\sigma(x) = \limsup_{x \to \infty} T_\sigma^n \bar{r}(x) \qquad (x \in \mathsf{X}).$$

The associated Bellman equation is

$$v(x) = \sup_{a \in \Gamma(x)} \left\{ r(x,a) - \frac{\beta(x,a)}{\theta} \ln \mathbb{E}_{x,a} e^{-\theta v(X')} \right\} \quad (x \in \mathsf{X}). \tag{2.21}$$

Suppose that the state process evolves following

$$x_{t+1} = f(x_t, a_t, \varepsilon_{t+1}), \tag{2.22}$$

where f is Borel measurable, and $\{\varepsilon_t\}$ is an IID process taking values in $\mathbb{R}^m$. For each t, let $\varepsilon_t = (\varepsilon_{1t}, \dots, \varepsilon_{mt})$.

Assumption 2.7.1.

- (i) $r \colon \mathsf{G} \to \mathbb{R}$ and $\beta \colon \mathsf{G} \to \mathbb{R}$ are increasing in x,
- (ii) r_σ and β_σ are increasing for any $\sigma \in \Sigma$
- (iii) $f(x, a, \varepsilon')$ is increasing in (x, ε'), and $f(x, \sigma(x), \varepsilon')$ is increasing in (x, ε') for any $\sigma \in \Sigma$,
- (iv) $\Gamma(x_1) \subset \Gamma(x_2)$ if $x_1 \leqslant x_2$,
- (v) $\varepsilon_{1t}, \dots, \varepsilon_{mt}$ are independent for each t.

Let $g(x,a) \coloneqq -\beta(x,a)/\theta \log \mathbb{E}_{x,a} \exp(-\theta v(X'))$ for all $(x,a) \in \mathsf{G}$. Let $\mathcal{G}$ be the u.s.c. functions g on G such that $\|g\| < \infty$ and g is increasing in X. Similar to (2.17), the transformed Bellman equation is

$$g(x,a) = -\frac{\beta(x,a)}{\theta} \ln \mathbb{E}_{x,a} \exp\left(-\theta \sup_{a' \in \Gamma(x')} \{ r(X', a') + g(X', a') \} \right). \tag{2.23}$$

for $(x,a) \in \mathsf{G}$ and $g \in \mathcal{G}$. Define the *risk-sensitive expected value Bellman operator* R by letting $Rg(x,a)$ be the right-hand side of (2.23) for any $g \in \mathcal{G}$ and $(x,a) \in \mathsf{G}$.

Assumption 2.7.2. $\bar{r}$ is bounded above and $\hat{r}$ is bounded below, where $\bar{r}$ is defined by (2.16) and $\hat{r}$ is defined by

$$\hat{r}(x,a) \coloneqq -\frac{1}{\theta} \ln \mathbb{E}_{x,a} \exp(-\theta \bar{r}(X')) \qquad ((x,a) \in \mathsf{G}).$$

Define $\bar{\beta}(x) \coloneqq \sup_{a \in \Gamma(x)} \beta(x, a)$ for all $x \in \mathsf{X}$. Suppose that there exists an $\ell \colon \mathsf{X}^2 \to \mathbb{R}_+$ such that $x \mapsto \ell(x, B)$ is a measurable function for all $B \in \mathscr{B}(\mathsf{X})$, $B \mapsto \ell(x, B)$ is a measure on X, and $L \colon mb\mathsf{X} \to mb\mathsf{X}$ is defined by

$$Lh(x) \coloneqq \int \ell(x, \mathrm{d}x') h(x')$$

for $x \in \mathsf{X}$ and $h \in mb\mathsf{X}$.

Assumption 2.7.3. $\rho(L) < 1$, and $\int_{\mathsf{X}} h(x') \bar{\beta}(x) P(x, a, \mathrm{d}x') \leqslant Lh(x)$ for all $(x, a) \in \mathsf{G}$ and $h \in mb\mathsf{X}_+$.

If Assumption 2.7.1, 2.7.2, and 2.7.3 are satisfied, the optimality results follow from the following theorem.

Theorem 2.7.1. *If Condition 2.7.1, Assumption 2.7.1, 2.7.2, and 2.7.3 hold, then the following statements are true.*

- *(a) $R\mathcal{G} \subset \mathcal{G}$ and R is eventually contracting on $(\mathcal{G}, \| \cdot \|)$,*
- *(b) R admits a unique fixed point g^* in $\mathcal{G}$,*
- *(c) v^* is well defined and*

$$v^*(x) = \sup_{a \in \Gamma(x)} \{ r(x, a) + g^*(x, a) \} \quad g^*(x, a) = -\frac{\beta(x, a)}{\theta} \log \mathbb{E}_{x,a} e^{-\theta v^*(X')},$$

- *(d) AFI converges to g^*,*
- *(e) at least one optimal policy exists, and*
- *(f) a feasible policy is optimal if and only if it is g^*-greedy.*

2.8. Appendix

We quote some useful theorems in this section, including the generalised Contraction Mapping Theorem, the Maximum Theorem, and Gelfand's formula. The following contraction mapping theorem can be found in Chapter 4 of Cheney (2001).

THEOREM 2.8.1 (Generalised Contraction Mapping Theorem). *If F is a self-map in a metric space X and F^m is contractive for some $m \in \mathbb{N}$, then F has a unique fixed point, which is the limit of every sequence $\{F^k x\}$ for arbitrary $x \in X$.*

Let X and Y be topological spaces. In the following theorem, given a correspondence $\varphi \colon X \to Y$, denote $\mathrm{Gr}(\varphi) \coloneqq \{(x, y) \colon x \in X,\, y \in \varphi(x)\}$.

THEOREM 2.8.2 (Berge Maximum Theorem). *Let $\varphi \colon X \to Y$ be a continuous correspondence between topological spaces with nonempty compact values. Suppose $f \colon Gr(\varphi) \to \mathbb{R}$ is continuous. Define the value function $m \colon X \to \mathbb{R}$ by*

$$m(x) \coloneqq \max_{y \in \varphi(x)} f(x, y),$$

and the correspondence $\mu \colon X \to Y$ of maximisers by

$$\mu(x) \coloneqq \{y \in \varphi(x) \colon f(x, y) = m(x)\}.$$

Then:

- *(a) The value function m is continuous.*
- *(b) The* arg max *correspondence μ has nonempty compact values.*
- *(c) If either f has a continuous extension to all $X \times Y$ or Y is Hausdorff, then the* arg max *correspondence μ is upper semicontinuous.*

See Chapter 17 of Aliprantis and Border (2006) and Hernández-Lerma and Lasserre (2012a) for the detail of the Maximum theorem.

Let E be a Banach space. A closed convex set $K \subset E$ is called a *cone* if $x \in K$ implies $tx \in K$ for $t \geqslant 0$, and $x, -x \in K$ implies $x = \theta$, where θ denotes the zero element of E.

A cone K is solid if it has an interior point; it is normal if $\theta \preceq x \preceq y$ implies $\|x\| \leqslant M\|y\|$ for $x, y \in K$, where $\preceq$ denotes partial order on E, and M is a constant.[22] The cone of non-negative functions in $mb\mathsf{X}$ or $cb\mathsf{X}$ is both solid and normal.

Theorem 2.8.3 (Local Spectral Radius). *Let A be a positive operator in a space E with a cone K. If the cone K is solid and normal, and f is an interior element of K, then*

$$\rho(A) = \lim_{n\to\infty} \|A^n f\|^{1/n}.$$

See Chapter 9 of Krasnosel'skii et al. (2012) for the proof of Theorem 2.8.3.

2.8.1. Proofs in Section 2.2.

Proof of Lemma 2.2.1. Let Condition 2.2.1 hold. Clearly, since $\sigma \in \Sigma$ is measurable, Condition 2.2.1 implies that T_σ is a self-map on $mb\mathsf{X}$. We next show that T is a self-map on $cb\mathsf{X}$. Fix $v \in cb\mathsf{X}$. Since $(x, a) \mapsto B(x, a, v)$ is bounded and continuous on G, and Γ is continuous and compact-valued, it follows from the (measurable) Maximum theorem that $x \mapsto \sup_{a\in\Gamma(x)} B(x, a, v) = Tv(x)$ is continuous and the correspondence $\tau \colon \mathsf{X} \to \mathsf{A}$ defined by

$$\tau(x) = \operatorname*{arg\,max}_{a\in\Gamma(x)} B(x, a, v)$$

is nonempty, u.s.c, and compact-valued and admits a Borel measurable selector σ satisfying $\sigma(x) \in \tau(x)$ for all $x \in \mathsf{X}$. Then, σ is a Borel measurable v-greedy policy. Since $Tv(x) = B(x, a, v)$ for $a \in \tau(x)$, Tv is bounded. Therefore, Tv is a self-map on $cb\mathsf{X}$. $\square$

Proof of Lemma 2.2.2. Let $\mathscr{R}$ be regular. Fix $\sigma \in \Sigma$ and let the linear operator L_σ be defined by (2.4). It follows from Theorem 1.5.5 of Bühler and Salamon (2018) that $\rho(L_\sigma) := \lim_{n\to\infty} \|L_\sigma^n\|^{1/n}$ always exists and is bounded above by $\|L_\sigma\|$. By Theorem 2.8.3, since (i) L_σ is a positive linear operator on $mb\mathsf{X}$, (ii) the positive cone in $mb\mathsf{X}$ is solid and normal under the pointwise partial order, and (iii) $\mathbb{1}$ lies interior to the positive cone in

---

[22] The definition of normal cone is equivalent to that there exists $\delta > 0$ such that $\|x+y\| \geqslant \delta$ for $x, y \in K$ and $\|x\| = \|y\| = 1$.

$mb\mathsf{X}$, we obtain

$$\rho(L_\sigma) = \lim_{n\to\infty} \|L_\sigma^n \mathbb{1}\|^{1/n} = \lim_{n\to\infty} \left\{ \sup_x |L_\sigma^n \mathbb{1}(x)| \right\}^{1/n} = \lim_{n\to\infty} (d_n^\sigma)^{1/n}, \tag{2.24}$$

where $\sup_x L_\sigma^n \mathbb{1}(x) = d_n^\sigma$ for all $n \in \mathbb{N}$ is shown by iteration. To show (a), suppose $\rho(L_\sigma) < 1$. Then, (2.24) implies that there is an $n_\sigma \in \mathbb{N}$ such that $d_{n_\sigma}^\sigma < 1$. Conversely, suppose $d_{n_\sigma}^\sigma < 1$ for some $n_\sigma \in \mathbb{N}$. Denote $\ell_\sigma(x, \mathrm{d}x') = k(x, \sigma(x), \mathrm{d}x')$. Since any $n \in \mathbb{N}$ can be written uniquely as $n = kn_\sigma + i$ for some $k, i \in \mathbb{N}_0$ with $i < n_\sigma$, we have, for sufficiently large n,

$$\begin{aligned} d_n^\sigma &= \sup_x L_\sigma^n \mathbb{1}(x) = \sup_x L_\sigma^{kn_\sigma} L_\sigma^{n-kn_\sigma} \mathbb{1}(x) \\ &= \sup_x \int_{\mathsf{X}} \ell_\sigma^{kn_\sigma}(x, \mathrm{d}x') \int_{\mathsf{X}} \mathbb{1}(x'') \ell_\sigma^{n-kn_\sigma}(x', \mathrm{d}x'') \\ &\leqslant \sup_x \int_{\mathsf{X}} \ell_\sigma^{kn_\sigma}(x, \mathrm{d}x') \left(\sup_{x'} \int_{\mathsf{X}} \ell_\sigma^{n-kn_\sigma}(x', \mathrm{d}x'') \right) \\ &= d_{kn_\sigma}^\sigma d_{n-kn_\sigma}^\sigma. \end{aligned} \tag{2.25}$$

Then, we have

$$(d_n^\sigma)^{1/n} = (d_{kn_\sigma}^\sigma d_{n-kn_\sigma}^\sigma)^{1/n} \leqslant (d_{n_\sigma}^\sigma)^{k/n} (d_i^\sigma)^{1/n} \leqslant (d_{n_\sigma}^\sigma)^{k/n} (M)^{1/n}.$$

where $M = \max\{d_1^\sigma, \dots, d_{n_\sigma - 1}^\sigma\}$. Since $k/n \to 1/n_\sigma$ as $n \to \infty$, the right-hand side converges to $(d_{n_\sigma}^\sigma)^{1/n_\sigma} < 1$ as $n \to \infty$. Hence, $\rho(L_\sigma) < 1$. Since σ is arbitrarily chosen, the statements hold for all $\sigma \in \Sigma$. Part (c) follows from $\rho(L_\sigma) = \lim_{n\to\infty} \|L_\sigma^n \mathbb{1}\|^{1/n}$ with the fact $L_\sigma^n \mathbb{1} \leqslant \hat{L}^n \mathbb{1}$ for $n \in \mathbb{N}$ by iteration. ☐

LEMMA 2.8.1. *Let L_σ be the operator defined by (2.4). If $\mathscr{R}$ is regular, and $\rho(L_\sigma) < 1$ for all $\sigma \in \Sigma$, then T_σ is eventually contracting, globally stable on $mb\mathsf{X}$, and has a unique fixed point $v_\sigma \in mb\mathsf{X}$ for all $\sigma \in \Sigma$. Moreover, if $\sigma \in \Sigma$ is continuous, then $v_\sigma \in cb\mathsf{X}$.*

PROOF OF LEMMA 2.8.1. Let L_σ be the operator defined by (2.4). Suppose that $\mathscr{R}$ is regular and $\rho(L_\sigma) < 1$ for all $\sigma \in \Sigma$. Fix $\sigma \in \Sigma$. Let $v, w \in mb\mathsf{X}$. Then, we have

$$|T_\sigma v(x) - T_\sigma w(x)| \leqslant L_\sigma |v - w|(x) \qquad (x \in \mathsf{X}).$$

Since L_σ is order-preserving, iteration implies $|T_\sigma^n v - T_\sigma^n w| \leqslant L_\sigma^n |v-w|$ for all $n \in \mathbb{N}$. Since $\rho(L_\sigma) < 1$, Lemma 2.2.2 implies that there is an $n_\sigma \in \mathbb{N}$ such that $d_{n_\sigma}^\sigma < 1$. Since $d_{n_\sigma}^\sigma = \sup_x L_\sigma^{n_\sigma} \mathbb{1}(x)$, we have

$$L_\sigma^{n_\sigma} h(x) \leqslant \|h\| L_\sigma^{n_\sigma} \mathbb{1} \leqslant d_{n_\sigma}^\sigma \|h\| \qquad (h \in mb\mathsf{X}.)$$

Therefore, we have

$$|T_\sigma^{n_\sigma} v - T_\sigma^{n_\sigma} w| \leqslant L_\sigma^{n_\sigma} |v-w| \leqslant d_{n_\sigma}^\sigma \|v-w\|.$$

Taking supremum over X on the left, we have $\|T_\sigma^{n_\sigma} v - T_\sigma^{n_\sigma} w\| \leqslant d_{n_\sigma}^\sigma \|v-w\|$. We conclude that $T_\sigma^{n_\sigma}$ is a contracting map with modulus $d_{n_\sigma}^\sigma$. Finally, the generalised Contraction Mapping theorem shows that T_σ is globally stable and has a unique fixed point $v_\sigma \in mb\mathsf{X}$. Next, if $\sigma \in \Sigma$ is continuous, then T_σ is a self-map on $cb\mathsf{X}$. Then, T_σ is globally stable on both $cb\mathsf{X}$ and $mb\mathsf{X}$. Since $cb\mathsf{X} \subset mb\mathsf{X}$ and v_σ is unique in both $mb\mathsf{X}$ and $cb\mathsf{X}$, we have $v_\sigma \in cb\mathsf{X}$. □

Lemma 2.8.2. *Let L_σ be the operator defined by (2.4). If $\mathscr{R}$ is regular, and Assumption 2.2.2 holds, then $\rho(L_\sigma) < 1$ and T_σ has a unique fixed point and is eventually contracting and globally stable on $mb\mathsf{X}$ for all $\sigma \in \Sigma$.*

Proof of Lemma 2.8.2. Suppose that $\mathscr{R}$ is regular, and Assumption 2.2.2 holds. Fix $\sigma \in \Sigma$. By definition, we have $L_\sigma \mathbb{1} \leqslant \hat{L}\mathbb{1}$, which implies $L_\sigma^2 \mathbb{1} \leqslant \hat{L} L_\sigma \mathbb{1} \leqslant \hat{L}^2 \mathbb{1}$. Iteration yields $L_\sigma^n \mathbb{1} \leqslant \hat{L}^n \mathbb{1}$ for all $n \in \mathbb{N}$. Since $\rho(\hat{L}) < 1$, Theorem 2.8.3 implies that there is an $m \in \mathbb{N}$ such that $\|\hat{L}^m\| < 1$.[23] Since $\|L_\sigma^m \mathbb{1}\| \leqslant \|\hat{L}^m \mathbb{1}\| \leqslant \|\hat{L}^m\| < 1$, we have $\rho(L_\sigma) = \lim_{n\to\infty} \|L_\sigma^n \mathbb{1}\|^{1/n} < 1$. Fix $v, v' \in mb\mathsf{X}$. Since $T_\sigma v(x) - T_\sigma v'(x) \leqslant L_\sigma(v-v')(x)$ for $x \in \mathsf{X}$, we have $|T_\sigma(x)v - T_\sigma v'(x)| \leqslant L_\sigma |v-v'|(x)$ for all $x \in \mathsf{X}$. Iteration implies $|T_\sigma^m v - T_\sigma^m v'| \leqslant L_\sigma^m |v-v'| \leqslant \|v-v'\| L_\sigma^m \mathbb{1}$. Taking supremum, we have $\|T_\sigma^m v - T_\sigma^m v'\| \leqslant \|v-v'\| \|L_\sigma^m \mathbb{1}\|$. Then, T_σ^m is contracting, so the generalised Contraction Theorem concludes the other statements. □

Lemma 2.8.3. *If $\mathscr{R}$ is regular and Assumption 2.2.2 holds, then T has a unique fixed point and is eventually contracting and globally stable on $cb\mathsf{X}$.*

---

[23] See also Claim 5 of Bloise et al. (2024) for the Gelfand's formula of a monotone sublinear operator.

Proof of Lemma 2.8.3. Suppose that $\mathscr{R}$ is regular and Assumption 2.2.2 holds. Then, T is a self-map on $cb\mathsf{X}$. Fix $v, v' \in cb\mathsf{X}$. We have

$$\begin{aligned} |Tv(x) - Tv'(x)| &\leqslant \sup_{\sigma} |T_\sigma v(x) - T_\sigma v'(x)| \\ &\leqslant \sup_{\sigma} |L_\sigma (v - v')(x)| \\ &\leqslant \sup_{\sigma} L_\sigma |v - v'|(x) = \hat{L}|v - v'|(x) \end{aligned}$$

for all $x \in \mathsf{X}$. Then, iteration yields

$$|T^n v(x) - T^n v'(x)| \leqslant \hat{L}^n |v - v'|(x)$$

for all $x \in \mathsf{X}$ and $n \in \mathbb{N}$, which implies $\|T^n v - T^n v'\| \leqslant \|\hat{L}^n \mathbb{1}\| \|v - v'\| \leqslant \|\hat{L}^n\| \|v - v'\|$. Since $\rho(\hat{L}) < 1$, and's formula implies that there exists $n \in \mathbb{N}$ such that $\|\hat{L}^n\| < 1$. Therefore, $\hat{L}^n$ is contracting, and then the generalised Contraction Mapping shows that $\hat{L}$ has a unique fixed point and globally stable. ☐

Lemma 2.8.4. *If $\mathscr{R}$ is regular, and Assumption 2.2.2 holds, then T is contracting on $(cb\mathsf{X}, \|\cdot\|_w)$, where $w := \sum_{n=0}^{\infty} \hat{L}^n \mathbb{1}$.*

Proof of Lemma 2.8.4. Suppose that $\mathscr{R}$ is regular, and Assumption 2.2.2 holds. Then, $\rho(\hat{L}) < 1$. Let $\{v_n\}_{n \geqslant 0}$ be a monotonically increasing sequence in $mb\mathsf{X}$ such that $v_n \uparrow v \in mb\mathsf{X}$ pointwise. Then, since $\hat{L}$ and L_σ preserve orders, the Dominated Convergence Theorem implies that

$$\begin{aligned} \lim_{n \to \infty} \hat{L} v_n(x) &= \sup_n \sup_{\sigma \in \Sigma} L_\sigma v_n(x) \\ &= \sup_{\sigma \in \Sigma} \sup_n \int_{\mathsf{X}} v_n(x) k(x, \sigma(x), \mathrm{d}x') \\ &= \sup_{\sigma \in \Sigma} \int_{\mathsf{X}} \sup_n v_n(x) k(x, \sigma(x), \mathrm{d}x') \\ &= \sup_n \int_{\mathsf{X}} v(x) k(x, \sigma(x), \mathrm{d}x') = \hat{L} v(x) \end{aligned}$$

for $x \in \mathsf{X}$. Therefore, Assumption 8 of Bloise et al. (2024) holds. It then follows from Claim 6 of Bloise et al. (2024) that there are $\lambda \in (0,1)$ and everywhere positive $w \in mb\mathsf{X}$

such that $\hat{L}w \leqslant \lambda w$. Note that we can construct w by $w = \sum_{n=0}^{\infty} \hat{L}^n \mathbb{1}$.[24] Fix $v, v' \in cb\mathsf{X}$. Then, we have

$$\begin{aligned} |Tv(x) - Tv'(x)| &= |\sup_{\sigma} T_\sigma v(x) - \sup_{\sigma} T_\sigma v'(x)| \\ &\leqslant \sup_{\sigma} |T_\sigma v(x) - T_\sigma v'(x)| \\ &\leqslant \sup_{\sigma} L_\sigma(|v - v'|)(x) = \hat{L}(|v - v'|)(x) \\ &\leqslant \hat{L}(\|v - v'\|_w w)(x) = \|v - v'\|_w \hat{L}(w)(x) \\ &\leqslant \|v - v'\|_w \lambda w(x) \end{aligned}$$

for $x \in \mathsf{X}$. Dividing w and taking supremum, it implies that $\|Tv - Tv'\|_w \leqslant \lambda \|v - v'\|_w$. ☐

PROPOSITION 2.8.1. *If Assumption 2.2.1 holds, then HPI converges to σ^* in finitely many steps, OPI converges to v^*, VTI converges to v^*, v^* is the unique solution to Bellman equation, Bellman's principle of optimality holds, and at least one optimal policy exists.*

PROOF OF PROPOSITION 2.8.1. Suppose that Assumption 2.2.1 holds. Then, T_σ is globally stable for all $\sigma \in \Sigma$. Since Σ is finite, then it follows from Chapter 8 or Chapter 9 of Sargent and Stachurski (2023) that HPI converges to σ^* in finitely many steps, and the other statements hold. ☐

LEMMA 2.8.5. *If $\mathscr{R}$ is regular and Assumption 2.2.3 holds, then T is eventually contracting and globally stable on $cb\mathsf{X}$ and has a unique fixed point $\bar{v} \in cb\mathsf{X}$, and T_σ is eventually contracting and globally stable on $mb\mathsf{X}$, and have unique fixed points $v_\sigma \in mb\mathsf{X}$ for all $\sigma \in \Sigma$.*

---

[24] See the proof of Claim 6, Bloise et al. (2024).

Proof of Lemma 2.8.5. Let $\mathscr{R}$ be regular and Assumption 2.2.3 hold. Lemma 2.2.1 implies that T is a self-map on $cb\mathsf{X}$. Fix $v, w \in cb\mathsf{X}$. Then, we have

$$\begin{aligned} |Tv(x) - Tw(x)| &= \left| \sup_\sigma T_\sigma v(x) - \sup_\sigma T_\sigma w(x) \right| \\ &\leqslant \sup_\sigma |T_\sigma v(x) - T_\sigma w(x)| \\ &\leqslant \sup_\sigma |B(x, \sigma(x), v) - B(x, \sigma(x), w)| \\ &\leqslant L|v - w|(x) \qquad (x \in \mathsf{X}). \end{aligned}$$

Hence, we obtain $|Tv - Tw| \leqslant L|v - w|$. Since L preserves order, iteration gives

$$|T^n v - T^n w| \leqslant L^n |v - w| \leqslant L^n \mathbb{1} \|v - w\| \leqslant \|L^n \mathbb{1}\| \|v - w\|$$

for $n \in \mathbb{N}$. Taking the supremum on the left, we have $\|T^n v - T^n w\| \leqslant \|L^n \mathbb{1}\| \|v - w\|$ for all $n \in \mathbb{N}$. Since L is a positive linear operator, we have $\rho(L) = \lim_{n\to\infty} \|L^n \mathbb{1}\|^{1/n}$ by Theorem 9.1 of Krasnosel'skii et al. (2012) as the proof of Lemma 2.2.2. Since $\rho(L) < 1$, there exists an $m \in \mathbb{N}$ such that $\|L^m \mathbb{1}\| < 1$. We conclude that $\|T^m v - T^m w\| \leqslant \|L^m \mathbb{1}\| \|v - w\|$ for all $v, w \in cb\mathsf{X}$. Therefore, T^m is contracting, so the generalised Contracting Mapping theorem shows that T is globally stable on $cb\mathsf{X}$ and has a unique fixed point $\bar{v} \in cb\mathsf{X}$. The proof for T_σ is similar. □

Lemma 2.8.6. *If $\mathscr{R}$ is regular, T_σ is globally stable on $mb\mathsf{X}$ for all $\sigma \in \Sigma$, and T is globally stable on $cb\mathsf{X}$, then the following statements are true.*

- *(a) If $\bar{v}$ is the unique fixed point of T on $cb\mathsf{X}$, then $v^* = \bar{v}$ and there exists a $\sigma^* \in \Sigma$ such that $v^* = v_{\sigma^*}$.*
- *(b) If T has a non-continuous fixed point $\bar{v} \in mb\mathsf{X}$, then $v^* < \bar{v}$ and $v_\sigma < \bar{v}$ for all $\sigma \in \Sigma$.*

Proof of Lemma 2.8.6. Let $\mathscr{R}$ be regular and suppose that T_σ is globally stable on $mb\mathsf{X}$ for all $\sigma \in \Sigma$, and T is globally stable on $cb\mathsf{X}$. We first show (a). Fix $\sigma \in \Sigma$ and let $\bar{v}$ be the unique fixed point of T on $cb\mathsf{X}$. Since $\bar{v}$ is the fixed point of T and $T\bar{v} = \sup_{\sigma\in\Sigma} T_\sigma \bar{v}$, we obtain $\bar{v} = T\bar{v} \geqslant T_\sigma \bar{v}$. Since T_σ is order-preserving, iteration gives

$\bar{v} \geqslant T_\sigma \bar{v} \geqslant \cdots \geqslant T_\sigma^n \bar{v}$ for $n \in \mathbb{N}$. In addition, since T_σ is globally stable, we have $\bar{v} \geqslant v_\sigma$ as $n \to \infty$, which implies $\bar{v} \geqslant v^*$ by taking supremum over Σ.

Conversely, Lemma 2.2.1 implies that there exists a $\sigma^* \in \Sigma$ such that $T_{\sigma^*}\bar{v} = T\bar{v}$. Then, we have $T_{\sigma^*}\bar{v} = \bar{v}$, whence $\bar{v} = v_{\sigma^*}$ by the uniqueness of fixed point of T_{σ^*}. By the definition of v^*, we obtain $v^* \geqslant v_{\sigma^*} = \bar{v}$. We conclude that $v^* = \bar{v}$ and an optimal policy σ^* exists.

To show (b), assume that there is a non-continuous $\bar{v} \in mb\mathsf{X}$ such that $T\bar{v} = \bar{v}$. Toward contradiction, suppose that there is $\sigma \in \Sigma$ such that $v_\sigma \geqslant \bar{v}$. Then, we have $v^* \geqslant v_\sigma \geqslant \bar{v}$. Moreover, since $\bar{v}$ is a fixed point of T, we have $\bar{v} = T\bar{v} \geqslant T_\sigma \bar{v}$ for any $\sigma \in \Sigma$. Iteration implies $\bar{v} \geqslant T_\sigma^n \bar{v}$, so $\bar{v} \geqslant v_\sigma$ as $n \to \infty$ by the global stability of T_σ. It implies that $\bar{v} \geqslant v^*$. Then, we obtain $\bar{v} = v^*$, which is continuous by (a), contradicting with non-continuity. Therefore, we must have $v_\sigma < \bar{v}$ for any $\sigma \in \Sigma$. The same argument also implies $v^* < \bar{v}$. □

Lemma 2.8.7. *If Condition 2.2.2 holds, then the following statements are true.*

(a) T and T_σ are self-maps on $cb\mathsf{X}$ for all $\sigma \in \Sigma_C$, and

(b) for any $v \in cb\mathsf{X}$ there exists a continuous v-greedy policy.

Proof of Lemma 2.8.7. Let $\mathscr{R}$ be regular and Assumption 2.2.2 holds. Then, if $\sigma \in \Sigma$ is continuous and $v \in cb\mathsf{X}$, then $x \mapsto T_\sigma(x) = B(x, \sigma(x), v)$ is continuous. Hence, T_σ is a self-map on $cb\mathsf{X}$. Since Γ is convex-valued and compact-valued, and $a \mapsto B(x, a, v)$ is strictly quasi-concave for any $x \in \mathsf{X}$ and $v \in cb\mathsf{X}$, it follows from the Maximum theorem that the Tv is continuous and then T is a self-map on $cb\mathsf{X}$. Moreover, for any $v \in cb\mathsf{X}$, the maximiser correspondence $x \mapsto \arg\max_{a \in \Gamma(x)} B(x, a, v)$ is single-valued and continuous, whence a continuous v-greedy policy exists. □

Lemma 2.8.8 (Bellman's Principle of Optimality). *Suppose that $\mathscr{R}$ is regular and T_σ is globally stable for all $\sigma \in \Sigma$. If v^* is a fixed point to T, then Bellman's principle of optimality holds.*

Proof of Lemma 2.8.8. Let all the stated assumptions hold. We want to show: $\sigma \in \Sigma$ is optimal (i.e., $v_\sigma = v^*$) if and only if σ is v^*-greedy. Suppose that $\sigma \in \Sigma$ is v^*-greedy: $T_\sigma v^* = T v^*$. Since $T v^* = v^*$, we obtain $T_\sigma v^* = v^*$ so that $v_\sigma = v^*$. Conversely, suppose that $\sigma \in \Sigma$ is optimal: $v_\sigma = v^*$. Since T_σ has a unique fixed point v_σ, we obtain $v^* = T_\sigma v^*$. Since $T v^* = v^*$, we have $T_\sigma v^* = v^* = T v^*$, so that σ is v^*-greedy. $\square$

Lemma 2.8.9 (HPI). *Suppose that Condition 2.2.2 holds, T_σ is globally stable on $c b\mathsf{X}$ with fixed point $v_\sigma \in c b\mathsf{X}$ for all $\sigma \in \Sigma_C$, and T is globally stable on $c b\mathsf{X}$. If $v^* \in c b\mathsf{X}$ is the unique fixed point of T on $c b\mathsf{X}$, then $\{H^n v_{\sigma_0}\}_{n \geqslant 0}$ converges to v^* for any $\sigma_0 \in \Sigma_C$, and v^* is the unique fixed point of H.*

Proof of Lemma 2.8.9. Let all the stated assumptions hold. Let $\{\sigma_k\}_{k \geqslant 0} \subset \Sigma_C$ be such that $\sigma_0 \in \Sigma_C$ and $v_{\sigma_k} = H v_{\sigma_{k-1}}$ for all $k \in \mathbb{N}$; that is, $\sigma_k \in \Sigma_C$ satisfies $T_{\sigma_k} v_{\sigma_{k-1}} = T v_{\sigma_{k-1}}$. Note that for any $k \geqslant 0$, the continuity of σ_k implies that T_{σ_k} is a self-map on $c b\mathsf{X}$, v_{σ_k} is continuous, and then a continuous v_{σ_k}-greedy policy σ_{k+1} exists by Lemma 2.8.7. By definition, $T_{\sigma_k} v_{\sigma_{k-1}} = T v_{\sigma_{k-1}} \geqslant T_{\sigma_{k-1}} v_{\sigma_{k-1}} = v_{\sigma_{k-1}}$. Applying T_{σ_k} on both sides repeatedly, since T_{σ_k} is order-preserving, iteration yields $T^n_{\sigma_k} v_{\sigma_{k-1}} \geqslant T_{\sigma_k} v_{\sigma_{k-1}} \geqslant T v_{\sigma_{k-1}} \geqslant v_{\sigma_{k-1}}$. Taking n to infinity, the global stability of the policy operator implies that $v_{\sigma_k} \geqslant T v_{\sigma_{k-1}} \geqslant v_{\sigma_{k-1}}$ for all $k \geqslant 0$. Since T is order-preserving, we have $T v_{\sigma_k} \geqslant T^2 v_{\sigma_{k-1}}$ for all $k \in \mathbb{N}$ and then $v_{\sigma_k} \geqslant T v_{\sigma_{k-1}} \geqslant T^2 v_{\sigma_{k-2}}$. Induction yields $v_{\sigma_k} \geqslant T^k v_{\sigma_0}$ for $k \in \mathbb{N}$. Since $v^* \geqslant v_{\sigma_k}$ by definition, we obtain $v^* \geqslant v_{\sigma_k} \geqslant T^k v_{\sigma_0}$. Taking $k \to \infty$, since T is globally stable with unique fixed point v^*, we have $v_{\sigma_k} \to v^*$ as $k \to \infty$.

Next, let σ^* be v^* greedy policy. Since $T_{\sigma^*} v^* = T v^* = v^*$ and T_{σ^*} has a unique fixed point v_{σ^*}, we have $v_{\sigma^*} = v^*$ and $H v^* = v_{\sigma^*} = v^*$, whence v^* is a fixed point of H. To see that v^* is the unique fixed point of H, let $\bar{v}$ be the fixed point of H. Then, we have $\bar{v} = H \bar{v} = v_\sigma$ where σ is $\bar{v}$-greedy. It implies that $T \bar{v} = T v_\sigma = T_\sigma v_\sigma = v_\sigma = \bar{v}$. Since v^* is the unique fixed point of T by global stability, we have $\bar{v} = v^*$. $\square$

Lemma 2.8.10. *Let $V_u := \{v \in c b\mathsf{X} \colon T v \geqslant v\}$. If Condition 2.2.2 holds, then W, defined by (2.8), is an order-preserving self-map on V_u, and we have*

$$v \in V_u \Longrightarrow T v \leqslant W v \leqslant T^m v.$$

PROOF OF LEMMA 2.8.10. Let Condition 2.2.2 hold. Then, T and T_σ are order-preserving self-maps on $cb\mathsf{X}$ for all $\sigma \in \Sigma_C$, and v-greedy continuous policy exists for all $v \in cb\mathsf{X}$ by Lemma 2.8.7. Fix any $v \in V_u$ and let $\sigma \in \Sigma_C$ be such that $T_\sigma v = Tv$. Since $T_\sigma u \leqslant Tu$ for any $u \in cb\mathsf{X}$, T and T_σ are order-preserving and $v \leqslant Tv$, we see that $Wv \in V_u$:

$$Wv = T_\sigma T_\sigma^{m-1} v \leqslant T T_\sigma^{m-1} v \leqslant T T_\sigma^{m-1} Tv = T T_\sigma^{m-1} T_\sigma v = TWv.$$

Then, W is a self-map on V_u. Also since T_σ^m is order-preserving, W is order-preserving. Next, regarding the first inequality, since T_σ is order-preserving, $v \leqslant Tv$, and $T_\sigma v = Tv$, we have

$$T_\sigma^{m-1} v \leqslant T_\sigma^{m-1} Tv = T_\sigma^{m-1} T_\sigma v = Wv.$$

Repeating the same iteration, we have $T_\sigma^{m-j} v \leqslant Wv$ for $j < m$. In particular, for $j = m-1$, we have $T_\sigma v \leqslant Wv$. Since $T_\sigma v = Tv$, we obtain $Tv \leqslant Wv$. For the second inequality, since $T_\sigma v \leqslant Tv$ by definition, and T and T_σ are order-preserving, we have $Wv = T_\sigma^m v \leqslant T^m v$. □

LEMMA 2.8.11 (OPI). *If Condition 2.2.2 holds and T is globally stable on $cb\mathsf{X}$ with unique fixed point $v^* \in cb\mathsf{X}$, then the OPI iteration $\{v_k\}$ converges to v^* with $v_0 = v_\sigma$ for some $\sigma \in \Sigma_C$.*

PROOF OF LEMMA 2.8.11. Let all the stated assumptions hold. Pick $\sigma \in \Sigma_C$ and let $v_0 = v_\sigma$. Let $\{v_k\}_{k \geqslant 0}$ be the OPI iteration. First, claim that

$$T^k v_0 \leqslant W^k v_0 \leqslant T^{km} v_0 \qquad (k \in \mathbb{N}).$$

Since $v_0 = T_\sigma v_0 \leqslant T v_0$, $v_0 \in V_u$. Hence, it follows from Lemma 2.8.10 that the claim holds for $k = 1$. Suppose that the claim holds for some $k \in \mathbb{N}$. Since all operators are order-preserving and self-map on V_u, Lemma 2.8.10 with $W^k v_0, T^{km} v_0 \in V_u$ implies the iteration:

$$T^{k+1} v_0 \leqslant T W^k v_0 \leqslant W W^k v_0 \leqslant W T^{km} v_0 \leqslant T^{(k+1)m} v_0.$$

Therefore, the claim holds for all $k \in \mathbb{N}$ by induction. Now, since $T^k v \to v^*$ as $k \to \infty$, the above claim implies that $W^k v_0 \to v^*$. Since OPI iteration follows $v_k = W^k v_0$, we conclude that $\{v_k\}$ converges to v^*. $\square$

Proof of Theorem 2.2.1. Let $\mathscr{R}$ be regular. If Assumption 2.2.1 holds, then Lemma 2.8.1 implies (a), and Proposition 2.8.1 yields the other statements. If Assumption 2.2.2 holds, then Lemma 2.8.2 implies (a), and Lemma 2.8.3 implies that T is globally stable. Moreover, Lemma 2.8.4 implies (b). If Assumption 2.2.3 holds, then Lemma 2.8.5 shows that T and T_σ are eventually contracting on $cb\mathsf{X}$ and $mb\mathsf{X}$, respectively, for all $\sigma \in \Sigma$. Suppose that either Assumption 2.2.2, or 2.2.3 holds. Part (c), (d), and (g) follow from Lemma 2.8.6 with global stability of T. Part (f) follows from Lemma 2.8.8 with (c). Part (α) follows from Lemma 2.8.9 with (c). Part (β) follows from Lemma 2.8.11 with (c). $\square$

Proof of Proposition 2.2.1. Let the stated assumptions hold. Fix $v, w \in U \subset mb\mathsf{X}$. Then, by the properties of $\mathcal{T}$, we have

$$\mathcal{T}v = \mathcal{T}(v + w - w) \leqslant \mathcal{T}(w + |v - w|) \leqslant \mathcal{T}w + G|v - w|.$$

Exchanging the roles of v and w, we obtain $|\mathcal{T}v - \mathcal{T}w| \leqslant G|v - w|$. Then, by the same proof in Lemma 2.8.5, $\mathcal{T}$ is eventually contracting on U. $\square$

Lemma 2.8.12. *If Condition 2.2.2 holds, T_σ is globally stable on $mb\mathsf{X}$ for all $\sigma \in \Sigma$, and T is globally stable on $cb\mathsf{X}$, then $v^* = \sup_{\sigma \in \Sigma_C} v_\sigma$ and $Tv = \sup_{\sigma \in \Sigma_C} T_\sigma v$ for all $v \in cb\mathsf{X}$.*

Proof of Lemma 2.8.12. Let all the assumptions hold. Fix $v \in cb\mathsf{X}$. By Lemma 2.8.7, there is a continuous $\sigma' \in \Sigma$ such that $Tv = T_{\sigma'} v$, which implies that $Tv = \sup_{\sigma \in \Sigma} T_\sigma v = T_{\sigma'} v \leqslant \sup_{\sigma \in \Sigma_C} T_\sigma v$. Moreover, since Lemma 2.8.6 implies that $v^* \in cb\mathsf{X}$ is a fixed point of T, and Lemma 2.8.7 implies that there exists $\sigma^* \in \Sigma_C$ satisfying $v^* = Tv^* = T_{\sigma^*} v^*$, we have $v^* = v_{\sigma^*}$ and then $\sup_{\sigma \in \Sigma} v_\sigma = v^* = v_{\sigma^*} \leqslant \sup_{\sigma \in \Sigma_C} v_\sigma$. $\square$

2.8.2. Proofs in Section 2.3.

Proof of Lemma 2.3.1. Let $\mathscr{M}$ be an MDP satisfying Condition 2.3.1. Since $(x, a) \mapsto B(x, a, v) = r(x, a) + \int_{\mathsf{X}} v(x')\beta(x, a, x')P(x, a, \mathrm{d}x')$ is continuous and bounded on G for

$v \in c b\mathsf{X}$ by assumption, Condition 2.2.1 holds. Clearly, the value aggregator B satisfies monotonicity condition (2.2). Therefore, $\mathscr{M}$ is regular. Lemma 2.2.1 concludes the remaining statements. $\square$

LEMMA 2.8.13. *If Condition 2.3.2 holds, then the following statements are true.*

(a) T and T_σ are self-maps on $cb\mathsf{X}$ for all $\sigma \in \Sigma_C$, and

(b) for any $v \in cb\mathsf{X}$ there exists a continuous v-greedy policy.

PROOF OF LEMMA 2.8.13. The statements follow from Lemma 2.8.7 and that Condition 2.3.2 implies Condition 2.2.2. $\square$

PROOF OF LEMMA 2.3.2. Let $\mathscr{M}$ be regular and Assumption 2.3.3 hold. Then, we have $L_\sigma \mathbb{1} \leqslant L\mathbb{1}$ for all $\sigma \in \Sigma$. Pick $\sigma \in \Sigma$. Since L_σ and L are order-preserving, iteration implies $L_\sigma^n \mathbb{1} \leqslant L^n \mathbb{1}$ for all $n \in \mathbb{N}$. Let $\{X_t\}$ be a P_σ-Markov process with $X_0 = x$ and $\beta_t = \beta(X_t, \sigma(X_t), X_{t+1})$ for all $t \in \mathbb{N}_0$. Then, iteration implies

$$L_\sigma^n \mathbb{1}(x) = \mathbb{E}_x^\sigma\{\beta_0 \beta_1 \cdots \beta_{n-1}\} \leqslant L^n \mathbb{1}(x) \leqslant \|L^n \mathbb{1}\|$$

for $x \in \mathsf{X}$ and $n \in \mathbb{N}$. Taking supremum over X, we have $d_n^\sigma = \|L_\sigma^n \mathbb{1}\| \leqslant \|L^n \mathbb{1}\|$ for all $n \in \mathbb{N}$. Since $\lim_{n\to\infty} \|L^n \mathbb{1}\|^{1/n} = \rho(L) < 1$, there exists a $n \in \mathbb{N}$ satisfying $\|L^n \mathbb{1}\| < 1$, which implies $d_n^\sigma < 1$ for all $\sigma \in \Sigma$. Moreover, we have

$$\rho(L_\sigma) = \lim_{n\to\infty} \|L_\sigma^n \mathbb{1}\|^{1/n} \leqslant \lim_{n\to\infty} \|L^n \mathbb{1}\|^{1/n} = \rho(L) < 1.$$

Now, since $\hat{L}\mathbb{1} = \sup_\sigma L_\sigma \mathbb{1} \leqslant L\mathbb{1}$, iteration yields $\hat{L}^n \mathbb{1} \leqslant L^n \mathbb{1}$ for $n \in \mathbb{N}$. Then, the same argument implies $\rho(\hat{L}) \leqslant \rho(L) < 1$. $\square$

LEMMA 2.8.14. *If $\mathscr{M}$ is regular and Assumption 2.3.2 holds, then T_σ is eventually contracting and globally stable on $mb\mathsf{X}$, and v_σ defined by (2.10) is the unique fixed point of T_σ in $mb\mathsf{X}$ for any $\sigma \in \Sigma$.*

PROOF OF LEMMA 2.8.14. Suppose that $\mathscr{M}$ is regular and Assumption 2.3.2 holds. Then, Assumption 2.2.2 holds. Fix $\sigma \in \Sigma$. Then, it follows from Lemma 2.8.2 and Lemma 2.2.2 that T_σ is globally stable and eventually discounting and has a unique fixed point. Next, we check that v_σ, defined by (2.10), is a fixed point of T_σ. Let $\{X_t\}_{t\geqslant 0}$ be a stochastic process generated by P_σ with $X_0 = x$. Let $\beta_t = \beta(X_t, \sigma(X_t), X_{t+1})$ for $t \in \mathbb{N}_0$. Since $v_s = T_\sigma v_s = T_\sigma^n v_s$, we have

$$\begin{aligned}
v_s(x) &= r_\sigma(x) + \mathbb{E}_x^\sigma \beta_0 v_s(X_1) \\
&= r_\sigma(x) + \mathbb{E}_x^\sigma \beta_0 [r_\sigma(X_1) + \mathbb{E}_{X_1}^\sigma \beta_1 v_s(X_2)] \\
&= \cdots \\
&= \mathbb{E}_x^\sigma \left[\sum_{t=0}^{n} \prod_{i=0}^{t-1} \beta_i r_\sigma(X_t) \right] + \mathbb{E}_x^\sigma \beta_0 \beta_1 \cdots \beta_n v_s(X_{n+1}) \\
&= \lim_{n\to\infty} \mathbb{E}_x^\sigma \left[\sum_{t=0}^{n} \prod_{i=0}^{t-1} \beta_i r_\sigma(X_t) \right] = v_\sigma(x), \qquad (x \in \mathsf{X})
\end{aligned}$$

where we use

$$\begin{aligned}
\lim_{n\to\infty} \mathbb{E}_x^\sigma \beta_0 \beta_1 \cdots \beta_n v_s(X_{n+1}) &\leqslant \lim_{n\to\infty} \left(\sup_x \mathbb{E}_x^\sigma \beta_0 \beta_1 \cdots \beta_n \|v_s\| \right) \\
&\leqslant \lim_{\substack{n\to\infty,\\ n=kn_\sigma+i,\\ i<n_\sigma, i,k\in\mathbb{N}}} \left(\sup_x \mathbb{E}_x^\sigma \prod_{t=0}^{kn_\sigma-1} \beta_t \mathbb{E}_{X_{kn_\sigma}}^\sigma \prod_{j=kn_\sigma}^{kn_\sigma+i-1} \beta_j \|v_s\| \right) \\
&\leqslant \left(\lim_{k\to\infty} (d_{n_\sigma}^\sigma)^k \right) \sup_{i<n_\sigma} d_i^\sigma \|v_s\| = 0.
\end{aligned}$$

Therefore, $v_s = v_\sigma$. ☐

PROOF OF THEOREM 2.3.1. Let $\mathscr{M}$ be regular. Lemma 2.3.1 implies that $\mathscr{M}$ is a regular RDP. Suppose that either Assumption 2.3.1, 2.3.2 or 2.3.3 holds. Lemma 2.8.14 implies (a). Since Assumption 2.3.1 implies Assumption 2.2.1, Assumption 2.3.2 implies Assumption 2.2.2, Assumption 2.3.3 implies Assumption 2.2.3, and Condition 2.3.2 implies Condition 2.2.2, the statements follow from Theorem 2.2.1. ☐

PROOF OF LEMMA 2.3.3. Suppose that $\mathscr{R}$ is regular, $r_\sigma \geqslant f \in cb\mathsf{X}_+$ for all $\sigma \in \Sigma$, and T has a fixed point $v \in cb\mathsf{X}_+$. Since v is the fixed point of T, we have

$$v = Tv = \sup_{\sigma\in\Sigma}\{r_\sigma + L_\sigma v\} \geqslant \sup_{\sigma\in\Sigma}\{f + L_\sigma v\} = f + \hat{L}v.$$

Then, we have

$$\hat{L}v(x) \leqslant v(x) - f(x) \leqslant \left(\sup_{y \in \mathsf{X}} \frac{v(y) - f(y)}{v(y)}\right) v(x) =: \lambda v(x)$$

for all $x \in \mathsf{X}$. Since β and v are everywhere positive, $\hat{L}v$ is everywhere positive. Then, we have $f(x) < v(x)$ and $(v(x) - f(x))/v(x) < 1$ for all $x \in \mathsf{X}$. Moreover, since X is compact, and f and v are continuous, we have $\lambda \in (0, 1)$. Now, since $\hat{L}v \leqslant \lambda v$ and the proof of Lemma 2.8.4 implies that Assumption 8 of Bloise et al. (2024) holds, Claim 6 of Bloise et al. (2024) implies $\rho(\hat{L}) < 1$. Since $\rho(\hat{L}) < 1$, we have $\lim_{n \to \infty} \hat{L}^n \mathbb{1} = 0$. Therefore, since iteration yields $d_n = \hat{L}^n \mathbb{1}$ for $n \in \mathbb{N}$, there is $n \in \mathbb{N}$ such that $\sup_x d_n(x) = \|\hat{L}^n \mathbb{1}\| < 1$. ☐

Proof of Corollary 2.3.1. Suppose that $\mathscr{M}$ is regular, X is compact, and either Assumption 2.3.2 or 2.3.3 holds. Let $\sigma \in \Sigma_C$. Then, T_σ is a self-map on $c b\mathsf{X}$. Theorem 2.3.1 implies that T is globally stable on $(c b\mathsf{X}, \|\cdot\|)$, and T_σ is globally stable on $(c b\mathsf{X}, \|\cdot\|)$ and has a unique fixed point $v_\sigma \in c b\mathsf{X}$. If we consider $r \equiv \mathbb{1}$, then the policy value is $w_\sigma(x) = \mathbb{E}_x^\sigma \sum_{t=0}^{\infty} \prod_{i=0}^{t-1} \beta_i^\sigma$ for all $x \in \mathsf{X}$, where $\beta_t^\sigma = \beta_\sigma(X_t, X_{t+1})$ for $t \in \mathbb{N}_0$, given a stochastic process $\{X_t\}$ generated by P_σ with $X_0 = x$. Since β is bounded and strictly positive, and $\prod_{i=1}^{-1} \beta_i^\sigma = 1$, we have $w_\sigma \gg \mathbb{1}$. Since w_σ is the fixed point of T_σ when $r = \mathbb{1}$, w_σ is bounded and

$$w_\sigma(x) = 1 + \int w_\sigma(x')\beta_\sigma(x, x')P_\sigma(x, \mathrm{d}x') \qquad (x \in \mathsf{X}).$$

Rewriting this equation, we have for all $x \in \mathsf{X}$

$$\begin{aligned} L_\sigma w_\sigma(x) &= \int w_\sigma(x')\beta_\sigma(x, x')P_\sigma(x, \mathrm{d}x') \\ &= w_\sigma(x) - 1 \leqslant \left(\sup_y \frac{w_\sigma(y) - 1}{w_\sigma(y)}\right) w_\sigma(x) =: \lambda_\sigma w_\sigma(x). \end{aligned}$$

Since w_σ is bounded and continuous with $w_\sigma \gg \mathbb{1}$, and X is compact, we obtain $\lambda_\sigma < 1$. Therefore, by fixing $v, v' \in m b\mathsf{X}$, we have

$$\begin{aligned} |T_\sigma(x)v - T_\sigma v'(x)| &\leqslant L_\sigma(|v - v'|)(x) \leqslant L_\sigma(\|v - v'\|_{w_\sigma} w_\sigma)(x) \\ &\leqslant \|v - v'\|_{w_\sigma} \lambda_\sigma w_\sigma(x) \end{aligned}$$

for all $x \in \mathsf{X}$. Dividing $w_\sigma(x)$ and taking supremum over X, we have $\|T_\sigma v - T_\sigma v'\|_{w_\sigma} \leqslant \lambda_\sigma \|v - v'\|_{w_\sigma}$, so T_s is contractive with modulus λ_σ. Similarly, since T is globally stable on $(cb\mathsf{X}, \|\cdot\|)$, T admits a fixed point $w \in cb\mathsf{X}$ when $r \equiv \mathbb{1}$. Then, we have

$$w = \sup_\sigma(\mathbb{1} + L_\sigma w) = \mathbb{1} + \sup_\sigma L_\sigma w = \mathbb{1} + \hat{L}w = \cdots = \sum_{n=0}^{\infty} \hat{L}^n \mathbb{1}.$$

where the last equality follows from the iteration of $w = \mathbb{1} + \hat{L}w$ and the fact $\lim_{n\to\infty} \hat{L}^n \mathbb{1} = 0$ and $\lim_{n\to\infty} \hat{L}^n w \leqslant \|w\| \lim_{n\to\infty} \hat{L}^n \mathbb{1} = 0$. Also, $w \gg \mathbb{1}$ since $\beta \gg 0$. In addition, since w is continuous, and X is compact, we have $\lambda := \sup_{x\in\mathsf{X}}(w(x) - 1)/w(x) < 1$ and $\hat{L}w = w - \mathbb{1} \leqslant \lambda w$. Fix $v, v' \in cb\mathsf{X}$. We have

$$|Tv - Tv'| \leqslant \hat{L}(|v - v'|) \leqslant \hat{L}(\|v - v'\|_w w) \leqslant \|v - v'\|_w \lambda w,$$

which implies $\|Tv - Tv'\|_w \leqslant \lambda \|v - v'\|_w$ for all $v, v' \in cb\mathsf{X}$. ☐

2.8.3. Proofs in Section 2.4. Let $\sigma \in \Sigma_C$. Let $\{X_t^\sigma\}_{t\geqslant 0}$ be a k_σ-Markov process. Suppose that $\{X_t^\sigma\}_{t\geqslant 0}$ admits a stationary distribution π_σ. Let $L_1(\pi_\sigma)$ be the set of Borel measurable functions $g \colon \mathsf{X} \to \mathbb{R}$ such that

$$\|g\| := \int |g(x)| \pi_\sigma(\mathrm{d}x) < \infty.$$

For $f, g \in L_1(\pi_\sigma)$, we write $f \geqslant g$ if $f(x) \geqslant g(x)$ for π_σ-almost all $x \in \mathsf{X}$. We write $f \gg g$ if $f(x) > g(x)$ for π_σ-almost all $x \in \mathsf{X}$. Define $\mathscr{G}_\sigma$ to be all $f \in L_1(\pi_\sigma)$ such that $f \gg 0$.

PROOF OF LEMMA 2.4.1. The proof is identical to Lemma 2.8.13 with the fact that r is positive everywhere. ☐

LEMMA 2.8.15. *If Assumption 2.4.1 and Condition 2.4.1 hold, then for all* $\sigma \in \Sigma_C$ *the following statements are true.*

(a) k_σ *is continuous and bounded,*

(b) there exists an $m \in \mathbb{N}$ *such that* k_σ^m *is everywhere positive,*

(c) there exists a unique stationary density π_σ *for* k_σ *on* X*, and*

(d) π_σ *is everywhere positive and continuous on* X

Proof of Lemma 2.8.15. Let the stated assumptions hold. Fix $\sigma \in \Sigma_C$. Since β, P and σ are continuous, β_σ and P_σ are continuous on a compact set X so that they are bounded. Then, k_σ is continuous and bounded, which shows (a). Let $\{X_t^\sigma\}_{t \geqslant 0}$ be the state process such that

$$\mathbb{P}\{X_{t+1}^\sigma \in B | X_t^\sigma = x\} = \int_B k_\sigma(x, \mathrm{d}y)$$

for every $x \in \mathsf{X}$ and Borel set $B \subset \mathsf{X}$. Since Assumption 2.4.1 holds and β is strictly positive, the Markov chain induced by P_σ is irreducible, which implies that the Markov chain induced by k_σ is also irreducible. Hence, there exists an $m \in \mathbb{N}$ such that k_σ^m is everywhere positive, which shows (b), and then $\{X_t^\sigma\}$ is irreducible. It then implies that (c): there exists a unique stationary distribution π_σ: $X_t^\sigma \overset{d}{=} \pi_\sigma$. Finally it follows from the proof of Lemma C1 of Borovička and Stachurski (2020) that (d): π_σ is everywhere positive and continuous. ☐

Lemma 2.8.16. *If Assumption 2.4.1 and Condition 2.4.1 hold, then the following statements are true for all $\sigma \in \Sigma_C$:*

(a) L_σ is a bounded linear operator on $L_1(\pi_\sigma)$.

(b) $L_\sigma g$ is continuous for $g \in \mathscr{G}_\sigma$.

(c) $L_\sigma g \geqslant 0$ when $g \geqslant 0$ and $L_\sigma g \in \mathscr{G}_\sigma$ when $g \in \mathscr{G}_\sigma$.

(d) L_σ is irreducible and L_σ^2 is compact.

(e) $\rho(L_\sigma) > 0$ and there exists a continuous function $e_\sigma \in \mathscr{G}_\sigma$ such that $L_\sigma e_\sigma = \rho(L_\sigma) e_\sigma$.

(f) L_σ is order preserving on $L_1(\pi_\sigma)$.

Proof of Lemma 2.8.16. Let Assumption 2.4.1 and Condition 2.4.1 hold. Fix $\sigma \in \Sigma_C$. For (a) and (b), since k_σ is continuous and bounded by Lemma 2.8.15, and $\bar{\beta}_\sigma$ is bounded and everywhere positive, the result follows from the proof in Lemma C2 of Borovička and Stachurski (2020). For (c), the first claim is obvious, and the second follows from that β is everywhere positive. For (d), the proof is identical to Lemma C3 of Borovička and

Stachurski (2020) with continuity of π_σ and k_σ from Lemma 2.8.15. Part (e) and (f) are identical to Lemma 8.4 of Stachurski et al. (2022a). □

Lemma 2.8.17. *If Assumption 2.4.1 and Condition 2.4.1 hold, then* $\|T_\sigma v - T_\sigma w\|_{e_\sigma} \leqslant \rho(L_\sigma)\|v-w\|_{e_\sigma}$ *for all* $v, w \in cb\mathsf{X}_+$ *and all* $\sigma \in \Sigma_C$.

Proof of Lemma 2.8.17. Let Assumption 2.4.1 and Condition 2.4.1 hold. Fix $\sigma \in \Sigma_C$. Lemma 2.8.16 implies that there is $e_\sigma \in cb\mathsf{X}_+$ such that $L_\sigma e_\sigma = \rho(L_\sigma)e_\sigma$. Then, for $v, w \in cb\mathsf{X}_+$ we have

$$\begin{aligned} |T_\sigma v(x) - T_\sigma w(x)| &= \left|\bar\beta_\sigma(x)\int k_\sigma(x, \mathrm{d}x')(v(x') - w(x'))\right| \\ &\leqslant \bar\beta_\sigma(x)\int k_\sigma(x,\mathrm{d}x')|v(x')-w(x')| \\ &\leqslant \bar\beta_\sigma(x)\int k_\sigma(x,\mathrm{d}x')e_\sigma(x')\sup_{x'}\frac{|v(x')-w(x')|}{e_\sigma(x')} \qquad (2.26) \\ &= \bar\beta_\sigma(x)\int k_\sigma(x,\mathrm{d}x')e_\sigma(x')\|v-w\|_{e_\sigma} \\ &= \|v-w\|_{e_\sigma} L_\sigma e_\sigma(x) = \|v-w\|_{e_\sigma}\rho(L_\sigma)e_\sigma(x). \end{aligned}$$

Dividing both sides with $e_\sigma(x)$ and taking the supremum, we have $\|T_\sigma v - T_\sigma w\|_{e_\sigma} \leqslant \rho(L_\sigma)\|v-w\|_{e_\sigma}$ for all $v, w \in cb\mathsf{X}_+$. □

Proof of Proposition 2.4.1. Let Assumption 2.4.1 and Condition 2.4.1 hold. Part (a) and (b) are equivalent by Lemma 2.2.2. Fix $\sigma \in \Sigma_C$. Then, we have $r_\sigma \in cb\mathsf{X}_+$ and $T_\sigma v = r_\sigma + L_\sigma v$ for $v \in cb\mathsf{X}_+$. Since L_σ is irreducible and L_σ^2 is compact by Lemma 2.8.16, it follows from Theorem 3.1 of Stachurski et al. (2022b) that part (a) and (c) are equivalent, and T_σ has no fixed point in $cb\mathsf{X}_+$ if $\rho(L_\sigma) \geqslant 1$. Then, part (e) and (f) implies part (a), and clearly part (c) implies (e) and (f). Suppose that $\rho(L_\sigma) < 1$. Lemma 2.8.17 shows that $\|T_\sigma v - T_\sigma w\|_{e_\sigma} \leqslant \rho(L_\sigma)\|v-w\|_{e_\sigma}$ for all $v, w \in cb\mathsf{X}_+$, whence a contraction map so that (a) implies (d). Finally, if that T_σ is a contraction map in $\|\cdot\|_{e_\sigma}$, then it is globally stable. □

Lemma 2.8.18. *If Condition 2.4.1 holds and* $\rho(\bar L) < 1$, *then* T *is eventually contracting, where* $\bar L v := \sup_{\sigma\in\Sigma_C} L_\sigma v$ *for* $v \in cb\mathsf{X}_+$. *Moreover,* T *is contracting on* $(cb\mathsf{X}, \|\cdot\|_w)$ *with modulus* λ, *where* $w := \sum_{n=0}^\infty \bar L^n \mathbb{1}$ *and* $\lambda = \sup_x (w(x)-1)/w(x)$.

Proof of Lemma 2.8.18. Suppose that Condition 2.4.1 holds and $\rho(\bar{L}) < 1$. Since for all $v \in cb\mathsf{X}_+$, there exists a v-greedy policy $\sigma \in \Sigma_C$, we gave $Tv = T_\sigma v \leqslant \sup_{\sigma \in \Sigma_C} T_\sigma v$. Since $Tv \geqslant \sup_{\sigma \in \Sigma} T_\sigma v$ by definition, we have $Tv = \sup_{\sigma \in \Sigma_C} T_\sigma v$. Then, the similar proof for Lemma 2.8.3 shows that $\|T^n v - T^n v'\| \leqslant \|\bar{L}^n \mathbb{1}\| \|v - v'\|$ for any $v, v' \in cb\mathsf{X}_+$. Since $\rho(\bar{L}) < 1$, there is $m \in \mathbb{N}$ such that $\|\bar{L}^m \mathbb{1}\| \leqslant 1$. Then, T^m is contractive. Finally, the method in the proof of Corollary 2.3.1 shows that T is contracting on $(cb\mathsf{X}, \|\cdot\|_w)$ with modulus λ. ☐

Proof of Theorem 2.4.1. Let Condition 2.4.1 hold. Lemma 2.8.12 implies $Tv = \sup_{\sigma \in \Sigma_C} T_\sigma v$ for $v \in cb\mathsf{X}_+$. If $r \equiv f \in cb\mathsf{X}_+$, and T has a fixed point in $cb\mathsf{X}_+$, then the argument in Lemma 2.3.3, with $\bar{L}$ instead of $\hat{L}$, shows $\rho(\bar{L}) < 1$. Suppose that Assumption 2.4.1 and 2.4.2 hold. Then, Proposition 2.4.1 implies (a). Lemma 2.8.18 implies (b). Finally, the proof for Theorem 2.3.1 yields the other statements with $Tv = \sup_{\sigma \in \Sigma_C} T_\sigma v$ for $v \in cb\mathsf{X}_+$. ☐

2.8.4. Proofs in Section 2.6. We also consider the following regular conditions in this section.

Condition 2.8.1.

(i) Γ is nonempty, compact-valued.

(ii) $a \mapsto r(x, a)$ is u.s.c. for all $x \in \mathsf{X}$.

(iii) β is bounded and strictly positive.

(iv) $a \mapsto \int_{\mathsf{X}} f(y) \beta(x, a, y) P(x, a, y) \mathrm{d}y$ is continuous on $\Gamma(x)$ for all $x \in \mathsf{X}$ and for $f \in mb\mathsf{X}$.

If Condition 2.8.1 is satisfied, we consider the value function spaces $\mathcal{V}$ and $\mathcal{G}$ defined by

$$\mathcal{V} := \{v \colon \mathsf{X} \to \mathbb{R} \cup \{-\infty\} \colon v \in m\mathsf{X},\ v/\kappa \text{ is bounded above,}$$

$$\text{and } (x, a) \mapsto \mathbb{E}_{x,a} v(x') \text{ is bounded below}\},$$

$$\mathcal{G} := \{g \colon \mathsf{G} \to \mathbb{R} \colon g \in m\mathsf{G}, g \text{ is bounded below}, \|g\|_\kappa < \infty,$$

$$\text{and } a \mapsto g(x, a) \text{ is u.s.c. on } \Gamma(x) \text{ for all } x \in \mathsf{X}\}$$

where $\kappa \geqslant \mathbb{1}$ is a real-valued function on X, which is further defined in Assumption 2.6.1. We say that an MDP is *regular* if either one of Condition 2.8.1 or 2.6.1 holds.

Lemma 2.8.19. *Suppose that Assumption 2.6.1 holds. If $\mathscr{M}$ is regular, then $(\mathcal{G}, \|\cdot\|_\kappa)$ is a Banach space.*

Proof of Lemma 2.8.19. Suppose that Assumption 2.6.1 holds, and $\mathscr{M}$ is regular. If Condition 2.6.1 holds, then $(\mathcal{G}, \|\cdot\|_\kappa)$ is complete follows from Ma et al. (2022). Suppose that Condition 2.8.1 holds. Let $B_\kappa(\mathsf{G})$ be the space of Borel measurable real-valued functions f on G satisfying $\|f\|_\kappa < \infty$. Since $(B_\kappa(\mathsf{G}), \|\cdot\|_\kappa)$ is a Banach space, it suffices to show that $\mathcal{G}$ is closed in $B_\kappa(\mathsf{G})$. Let $\{g_n\} \subset \mathcal{G}$ such that $\|g_n - g\|_\kappa \to 0$. Clearly, $\|g\|_\kappa$ is finite. We next show that $a \mapsto g(x,a)$ is u.s.c. for all $x \in \mathsf{X}$. Fix $x' \in \mathsf{X}$. For all $a_0 \in \Gamma(x')$ and $y > g(x', a_0)$, let $\varepsilon = y - g(x', a_0)$. Since $\|g_n - g\|_\kappa \to 0$, for all $\delta > 0$ there exist $N \in \mathbb{N}$ such that for all x and $a \in \Gamma(x)$ we have

$$|g_N(x,a) - g(x,a)| < \kappa(x)\delta.$$

We choose δ such that $\kappa(x')\delta < \varepsilon/3$. Since $g_N(x, \cdot)$ is u.s.c. on $\Gamma(x)$ for all $x \in \mathsf{X}$, there exists a neighborhood U of a_0 such that for all $a \in U$

$$g_N(x', a) < g_N(x', a_0) + \varepsilon/3.$$

Hence, the previous inequalities imply

$$\begin{aligned} g(x', a) &< g_N(x', a) + \kappa(x')\delta < g_N(x', a_0) + \kappa(x')\delta + \varepsilon/3 \\ &< g(x', a_0) + 2\kappa(x')\delta + \varepsilon/3 < g(x', a_0) + \varepsilon = y. \end{aligned}$$

for all $a \in U$. Therefore, $a \to g(x', a)$ is u.s.c.. Since x' is arbitrarily picked, $a \mapsto g(x,a)$ is u.s.c. on $\Gamma(x)$ for all $x \in \mathsf{X}$. We conclude that $\mathcal{G}$ is closed in $B_\kappa(\mathsf{G})$ and then $(\mathcal{G}, \|\cdot\|_\kappa)$ is complete. □

Lemma 2.8.20 (Well-defined Value). *If Assumption 2.6.1 hold, then $v_\sigma(x)$ and $v^*(x)$ are well-defined in $\mathbb{R} \cup \{-\infty\}$ for all $x \in \mathsf{X}$ and $\sigma \in \Sigma$.*

Proof of Lemma 2.8.20. Suppose that Assumption 2.6.1 holds. Fixing $x \in \mathsf{X}$ and $\sigma \in \Sigma$. Since $\bar{r}(x) \leqslant d\kappa(x)$ and $\mathbb{E}_{x,a}\beta(x,a,X')\kappa(X')L^n\mathbb{1}(X') \leqslant \alpha\kappa(x)L^{n+1}\mathbb{1}(x)$ for all $x \in \mathsf{X}$ and $n \geqslant 0$, iteration implies that for all $t \in \mathbb{N}$ we have

$$\begin{aligned}\mathbb{E}_x \prod_{i=0}^{t-1} \beta(x_i, \sigma(x_i), x_{i+1}) r(x_t, \sigma(x_t)) &\leqslant \mathbb{E}_x \prod_{i=0}^{t-1} \beta(x_i, \sigma(x_i), x_{i+1}) \bar{r}(x_t) \\ &\leqslant \mathbb{E}_x \prod_{i=0}^{t-1} \beta(x_i, \sigma(x_i), x_{i+1}) d\kappa(x_t) \\ &\leqslant d\mathbb{E}_x \prod_{i=0}^{t-2} \beta(x_i, \sigma(x_i), x_{i+1}) \alpha\kappa(x_{t-1}) L\mathbb{1}(x_{t-1}) \\ &\leqslant d\alpha^t \kappa(x)(L^t\mathbb{1})(x).\end{aligned}$$

Since there is an $n \in \mathbb{N}$ such that $\alpha^n \|L^n\| < 1$ by assumption, we have

$$\begin{aligned}v_\sigma(x) &= \mathbb{E}_x \sum_{t=0}^{\infty} \prod_{i=0}^{t-1} \beta(x_i, \sigma(x_i), x_{i+1}) r(x_t, \sigma(x_t)) \\ &\leqslant \sum_{t=0}^{\infty} d\alpha^t \|L^t\mathbb{1}\| \kappa(x) \leqslant d\kappa(x) \frac{\sum_{t=0}^{n-1} \alpha^t \|L^t\|}{1 - \alpha^n \|L^n\|} < \infty.\end{aligned}$$

Therefore, $v_\sigma(x)$ is well-defined in $\mathbb{R} \cup \{-\infty\}$ for all $x \in \mathsf{X}$ and $\sigma \in \Sigma$. Moreover, since the upper bound holds for all $\sigma \in \Sigma$, the definition of v^* implies that $v^*(x)$ is bounded above and then well-defined for all $x \in \mathsf{X}$. □

Lemma 2.8.21. *If $\mathscr{M}$ is regular and Assumption 2.6.1 hold, then $M\mathcal{G} \subset \mathcal{V}$, $E\mathcal{V} \subset \mathcal{G}$, $T\mathcal{V} \subset \mathcal{V}$ and $R\mathcal{G} \subset \mathcal{G}$.*

Proof of Lemma 2.8.21. Let Condition 2.8.1 and Assumption 2.6.1 hold (the proof for Condition 2.6.1 is similar, see also Ma et al. (2022) for $R\mathcal{G} \subset \mathcal{G}$.) We first show that $E\mathcal{V} \subset \mathcal{G}$. Let $v \in \mathcal{V}$. Then, v/κ is bounded above and $\mathbb{E}_{(\cdot,\cdot)}v(X')$ is bounded below. Similar to Lemma 8.3.7 of Hernández-Lerma and Lasserre (2012b), we can show that $a \mapsto \mathbb{E}_{x,a}\beta(x,a,X')v(X')$ is u.s.c. on $\Gamma(x)$ for all $x \in \mathsf{X}$. In detail, fix $x \in \mathsf{X}$ and let $\{a_n\}_{t \geqslant 0} \subset \Gamma(x)$ be such that $a_n \to a \in \Gamma(x)$. Since v/κ is bounded above, there exists an $m \in \mathbb{R}$ such that $v \leqslant m\kappa$. Hence, $v - m\kappa$ is non-positive, so there is a non-increasing sequence of bounded measurable functions $\{v_m^k\}$ such that $v_m^k \downarrow v_m$.[25] Then, Assumption 2.6.1 implies that, for all k,

$$\begin{aligned}\limsup_{n\to\infty} \int v_m(x')\beta(x,a_n,x')P(x,a_n,\mathrm{d}x') &\leqslant \limsup_{n\to\infty} \int v_m^k(x')\beta(x,a_n,x')P(x,a_n,\mathrm{d}x') \\ &= \int v_m^k(x')\beta(x,a,x')P(x,a,\mathrm{d}x')\end{aligned}$$

Letting $k \to \infty$, the Monotone Convergence theorem yields that

$$\limsup_{n\to\infty} \int v_m(x')\beta(x,a_n,x')P(x,a_n,\mathrm{d}x') \leqslant \int v_m(x')\beta(x,a,x')P(x,a,\mathrm{d}x').$$

To this end, since $x \in \mathsf{X}$ is arbitrary, $a \mapsto \mathbb{E}_{x,a}\beta(x,a,X')v_m(X') = Ev_m(x,a)$ is u.s.c. for all $x \in \Gamma(x)$. It implies that $Ev(x,\cdot)$ is u.s.c. on $\Gamma(x)$. Furthermore, since $v/\kappa \leqslant m$, Assumption 2.6.1 implies that, for all $(x,a) \in \mathsf{G}$,

$$Ev(x,a) = \mathbb{E}_{x,a}\beta(x,a,X')v(X') \leqslant \mathbb{E}_{x,a}\beta(x,a,X')m\kappa(X') \leqslant m\kappa(x)\alpha L\mathbb{1}(x).$$

Hence, Ev/κ is bounded above by $m\alpha\|L\|$. Moreover, to see that Ev is bounded below, let $B \coloneqq \{x \in \mathsf{X} \colon v(x) < 0\}$. Since v/κ is bounded above, we have

$$\int_{\mathsf{X}\setminus B} v(x')P(x,a,\mathrm{d}x') \leqslant \int_{\mathsf{X}\setminus B} m\kappa(x')P(x,a,\mathrm{d}x') \leqslant m\|\kappa\|.$$

for all $(x,a) \in \mathsf{G}$. Since $\mathbb{E}_{(\cdot,\cdot)}v(X')$ is bounded below, there is $e \in \mathbb{R}$ such that

$$\int_B v(x')P(x,a,\mathrm{d}x') + \int_{\mathsf{X}\setminus B} v(x')P(x,a,\mathrm{d}x') \geqslant e$$

for all $(x,a) \in \mathsf{G}$. The above two inequalities imply

$$\int_B v(x')P(x,a,\mathrm{d}x') \geqslant e - \int_{\mathsf{X}\setminus B} v(x')P(x,a,\mathrm{d}x') \geqslant e - m\|\kappa\|,$$

---

[25] If Condition 2.6.1 holds, then κ is continuous and v is u.s.c., implying $v - m\kappa$ is u.s.c.. Then, there is a $\{v_m^k\} \subset cb\mathsf{X}$ such that $v_m^k \downarrow v_m$.

for all $(x, a) \in \mathsf{G}$. Since β is bounded, we see that $Ev(\cdot,\cdot) = \mathbb{E}_{(\cdot,\cdot)}\beta(\cdot,\cdot,X')v(X')$ is also bounded below: for all $(x, a), \in \mathsf{G}$

$$\begin{aligned}\mathbb{E}_{x,a}\beta(x,a,X')v(X') &= \int_B \beta(x,a,x')v(x')P(x,a,\mathrm{d}x') + \int_{\mathsf{X}\setminus B} \beta(x,a,x')v(x')P(x,a,\mathrm{d}x') \\ &\geqslant \|\beta\|(e - m\|\kappa\|).\end{aligned}$$

Since Ev is bounded below and Ev/κ is bounded above, we conclude that $\|Ev\|_\kappa < \infty$, which implies $Ev \in \mathcal{G}$ and then $E\mathcal{V} \subset \mathcal{G}$.

Next, we show that $M\mathcal{G} \subset \mathcal{V}$. Let $g \in \mathcal{G}$. Since $r(x,\cdot)$ and $g(x,\cdot)$ are u.s.c. on $\Gamma(x)$ for all $x \in \mathsf{X}$, Proposition D.5 of Hernández-Lerma and Lasserre (2012a) implies that $x \mapsto Mg(x) = \sup_{a\in\Gamma(x)}\{r(x,a) + g(x,a)\}$ is measurable. Moreover, Assumption 2.6.1 implies that for all $x \in \mathsf{X}$

$$\begin{aligned}Mg(x) &= \sup_{a\in\Gamma(x)}\{r(x,a) + g(x,a)\} \leqslant \sup_{a\in\Gamma(x)}\{r(x,a)\} + \sup_{a\in\Gamma(x)}\{g(x,a)\} \\ &\leqslant d\kappa(x) + \|g\|_\kappa \kappa(x).\end{aligned}$$

Therefore, Mg/κ is bounded above. Also, since $\|g\|_\kappa < \infty$ and κ is bounded, we have $|g(\cdot)| \leqslant \|g\|_\kappa \kappa(\cdot) < \infty$ so that g is bounded below by some $\underline{g} \in \mathbb{R}$. Then, we have, for all $(x, a) \in \mathsf{G}$,

$$\begin{aligned}\mathbb{E}_{x,a}Mg(X') &= \mathbb{E}_{x,a}\sup_{a'\in\Gamma(X')}\{r(X',a') + g(X',a')\} \geqslant \mathbb{E}_{x,a}\sup_{a'\in\Gamma(X')}\{r(X',a') + \underline{g}\} \\ &= \mathbb{E}_{x,a}\bar{r}(X') + \underline{g} = \hat{r}(x,a) + \underline{g}.\end{aligned}$$

Since $\hat{r}$ is bounded below by assumption, $\mathbb{E}_{(\cdot,\cdot)}Mg(X')$ is bounded below. Therefore, we have $Mg \in \mathcal{V}$ and then $M\mathcal{G} \subset \mathcal{V}$. Now, since $T = ME$ and $R = EM$, we obtain that $T\mathcal{V} = ME\mathcal{V} \subset M\mathcal{G} \subset \mathcal{V}$ and $R\mathcal{G} = EM\mathcal{G} \subset E\mathcal{V} \subset \mathcal{G}$. ☐

Lemma 2.8.22. *Let $\mathscr{M}$ be regular and Assumption 2.6.1 hold. If $\bar{g}$ is the unique fixed point of R, then $\bar{v} = M\bar{g}$ is the unique fixed point of T and $\bar{g} = E\bar{v}$. In addition, R is globally stable on $\mathcal{G}$ if and only if T is globally stable on $\mathcal{V}$.*

PROOF OF LEMMA 2.8.22. Let $\mathscr{M}$ be regular and Assumption 2.6.1 hold. Let $\bar{g}$ be unique the fixed point of R. Since $T = ME$ and $R = EM$, we have $M\bar{g} = MR\bar{g} = MEM\bar{g} = TM\bar{g}$, so that $M\bar{g}$ is a fixed point of T. Suppose that $v \neq M\bar{g}$ is a fixed point of T. Then, since v is a fixed point of T, we obtain $Ev = ETv = EMEv = REv$ implies that Ev is a fixed point of R. Since $\bar{g}$ is the unique fixed point of R, we must have $\bar{g} = Ev$ and then $M\bar{g} = MEv = Tv = v$. Therefore, $M\bar{g}$ must be the unique fixed point of T.

The above statement also shows $\bar{g} = E\bar{v}$. For the second statement, observe the iteration $T^n v = (ME)^n v = M(EM)^{n-1}Ev = MR^{n-1}Ev$ for any $v \in \mathcal{V}$. Since R is globally stable and $Ev \in \mathcal{G}$, we have $R^{n-1}Ev \to \bar{g}$ as $n \to \infty$. Hence, $T^n v \to M\bar{g}$ as $n \to \infty$ for any $v \in \mathcal{V}$. Similarly, we can show that the global stability of T implies the global stability of R. □

LEMMA 2.8.23. *If $\mathscr{M}$ is regular and Assumption 2.6.1 hold, then R is eventually contracting and globally stable on $(\mathcal{G}, \|\cdot\|_\kappa)$, and T is globally stable on $(\mathcal{V}, \|\cdot\|)$.*

PROOF OF LEMMA 2.8.23. Suppose that $\mathscr{M}$ is regular and Assumption 2.6.1 hold. Lemma 2.8.21 shows that $R\mathcal{G} \subset \mathcal{G}$ and $T\mathcal{V} \subset \mathcal{V}$. Fix $g, h \in \mathcal{G}$. Assumption 2.6.1 (b) implies that for all $(x, a) \in \mathsf{G}$ we have

$$\begin{aligned}
|Rg(x,a) - Rh(x,a)| &= \left| \mathbb{E}_{x,a}\beta(x,a,X') \left[\sup_{a' \in \Gamma(X')} \{r(X',a') + g(X',a')\} \right.\right. \\
&\qquad \left.\left. - \sup_{a' \in \Gamma(X')} \{r(X',a') + h(X',a')\} \right] \right| \\
&\leqslant \mathbb{E}_{x,a} \left(\beta(x,a,X') \sup_{a' \in \Gamma(X')} |g(X',a') - h(X',a')| \right) \\
&\leqslant \mathbb{E}_{x,a}\beta(x,a,X') \|g - h\|_\kappa \kappa(X') \\
&\leqslant \|g - h\|_\kappa \kappa(x) \alpha L \mathbb{1}(x).
\end{aligned}$$

Then, iteration implies

$$\begin{aligned}|R^2g(x,a) - R^2g(x,a)| &\leqslant \mathbb{E}_{x,a}\beta(x,a,X') \sup_{a'\in\Gamma(X')} |Rg(X',a') - Rg(X',a')| \\ &\leqslant \mathbb{E}_{x,a}\beta(x,a,X')\|g-h\|_\kappa \kappa(X')\alpha L\mathbb{1}(X') \\ &\leqslant \|g-h\|_\kappa \kappa(x)\alpha^2 L^2\mathbb{1}(x).\end{aligned}$$

Induction yields

$$|R^n g(x,a) - R^n h(x,a)| \leqslant \|g-h\|_\kappa \kappa(x)\alpha^n L^n \mathbb{1}(x)$$

for all $(x,a) \in \mathsf{G}$ and $n \in \mathbb{N}$. Dividing $\kappa(x)$ on the both sides and taking supremum, we obtain

$$\|R^n g - R^n h\|_\kappa \leqslant \|g-h\|_\kappa \alpha^n \|L^n \mathbb{1}\| \leqslant \|g-h\|_\kappa \alpha^n \|L^n\|.$$

Since there exists an $n \in \mathbb{N}$ such that $\|L^n\| < 1$ and $\alpha < 1/\|L^n\|^{1/n}$, R is eventually contracting on $(\mathcal{G}, \|\cdot\|_\kappa)$, whence it is also globally stable by the generalised Banach contraction mapping theorem. Finally, Lemma 2.8.22 implies that T is globally stable. □

LEMMA 2.8.24. *If $\mathscr{M}$ is regular and Assumption 2.6.1 hold, then a v-greedy policy and a g-greedy policy exist for all $v \in \mathcal{V}$ and $g \in \mathcal{G}$.*

PROOF OF LEMMA 2.8.24. Let Condition 2.8.1 and Assumption 2.6.1 hold. Let $v \in \mathcal{V}$. Condition 2.8.1 and the proof of Lemma 2.8.21 implies that $a \mapsto r(x,a) + \mathbb{E}_{x,a}\beta(x,a,C')v(C')$ is u.s.c. for all $x \in \mathsf{X}$. Since $\Gamma(x)$ is compact for all $x \in \mathsf{X}$, the v-greedy policy exists by the Maximum theorem. Let $g \in \mathcal{G}$. Then, $a \mapsto r(x,a) + g(x,a)$ is u.s.c. for all $x \in \mathsf{X}$, so the g-greedy policy exists by the Maximum theorem. The proof for Condition 2.6.1 is similar. □

LEMMA 2.8.25. *If $\mathscr{M}$ is regular and Assumption 2.6.1 hold, and $\bar{g}$ is a fixed point of R, then $v^* = M\bar{g}$, $\bar{g} = Ev^*$, and an optimal policy exists. Moreover, the following statements are equivalent.*

- *(a) a policy $\sigma \in \Sigma$ is optimal,*
- *(b) σ is $\bar{g}$-greedy, and*

(c) σ is v^*-greedy.

Proof of Lemma 2.8.25. Let $\mathscr{M}$ be regular and Assumption 2.6.1 hold. Lemma 2.8.22 implies that $\bar{v} = M\bar{g}$ is a unique fixed point of T, so $\bar{v} = T\bar{v} \geqslant T_\sigma \bar{v}$ for all $\sigma \in \Sigma$. Fix $\sigma \in \Sigma$. Let $\{X_t\}$ be a P_σ-Markov process and $\beta_t^\sigma = \beta(X_t, \sigma(X_t), X_{t+1})$ for all $t \in \mathbb{N}_0$. Then, since $\bar{v} \geqslant T_\sigma \bar{v}$ and $\bar{g} = M\bar{v}$ by Lemma 2.8.22, iteration implies that for all $x \in \mathsf{X}$ and $\sigma \in \Sigma$ we have

$$
\begin{aligned}
\bar{v}(X_0) &\geqslant T_\sigma \bar{v}(X_0) = r(X_0, \sigma(X_0)) + \mathbb{E}_{X_0,\sigma(X_0)} \beta_0 \bar{v}(X_1) \\
&\geqslant r(X_0, \sigma(X_0)) + \mathbb{E}_{X_0,\sigma(X_0)} \beta_0^\sigma [r(X_1, \sigma(x_1)) + \mathbb{E}_{X_1,\sigma(X_1)} \beta_1^\sigma \bar{v}(X_2)] \\
&\geqslant \dots \\
&\geqslant \mathbb{E}_{X_0,\sigma(X_0)} \sum_{t=0}^{N} \prod_{i=0}^{t-1} \beta_i^\sigma r(X_t, \sigma(X_t)) + \mathbb{E}_{X_0,\sigma(X_0)} \beta_0^\sigma \beta_1^\sigma \cdots \beta_{N-1}^\sigma \mathbb{E}_{X_N,\sigma(X_N)} \beta_N^\sigma \bar{v}(X_{N+1}) \\
&= \mathbb{E}_{X_0,\sigma(X_0)} \sum_{t=0}^{N} \prod_{i=0}^{t-1} \beta_i^\sigma r(X_t, \sigma(X_t)) + \mathbb{E}_{X_0,\sigma(X_0)} \beta_0^\sigma \beta_1^\sigma \cdots \beta_{N-1}^\sigma \bar{g}(X_N, \sigma(x_N)).
\end{aligned}
\tag{2.27}
$$

Now, Assumption 2.6.1 implies

$$
\begin{aligned}
|\mathbb{E}_{X_0,\sigma(X_0)} \beta_0^\sigma \beta_1^\sigma \cdots \beta_{N-1}^\sigma \bar{g}(X_N, \sigma(X_N))| &\leqslant \mathbb{E}_{X_0,\sigma(X_0)} \beta_0^\sigma \beta_1^\sigma \cdots \beta_{N-1}^\sigma |\bar{g}(X_N, \sigma(X_N))| \\
&\leqslant \mathbb{E}_{X_0,\sigma(X_0)} \beta_0^\sigma \beta_1^\sigma \cdots \mathbb{E}_{X_{N-1},\sigma(X_{N-1})} \beta_{N-1}^\sigma \|g\|_\kappa \kappa(X_N) \\
&\leqslant \|g\|_\kappa \mathbb{E}_{X_0,\sigma(X_0)} \beta_0^\sigma \beta_1^\sigma \cdots \beta_{N-2}^\sigma \kappa(X_{N-1}) \alpha L \mathbb{1}(X_{N-1}) \\
&\leqslant \dots \\
&\leqslant \|\bar{g}\|_\kappa \kappa(X_0) \alpha^N L^N \mathbb{1}(X_0) \\
&\leqslant \|\bar{g}\|_\kappa \kappa(X_0) \|\alpha^N L^N \mathbb{1}\|.
\end{aligned}
$$

Since there exists an $n \in \mathbb{N}$ satisfying $\alpha^n \|L\|^n < 1$, letting $N = pn + q$ with $p, q \in \mathbb{N}_0$ and $q < n$, we obtain

$$
\alpha^N \|L^N \mathbb{1}\| \leqslant (\alpha^{np} \|L^{np}\|) \alpha^q \|L^q\| \leqslant (\alpha^n \|L^n\|)^p \max_{q<n} \{\alpha^q \|L^q\|\} \to 0 \text{ as } n(N) \to \infty.
$$

Therefore, letting $N \to \infty$, (2.27) and Lemma 2.8.20 imply $\bar{v} \geqslant v_\sigma$ for all $\sigma \in \Sigma$. Hence, we have $\bar{v} \geqslant \sup_\sigma v_\sigma = v^*$. Next, since $\bar{g} = E\bar{v}$ and $\bar{g}$-g „reedy policy exists by Lemma

2.8.24, there exists σ^* such that

$$\begin{aligned}\bar{v}(x) &= M\bar{g}(x) = r(x, \sigma^*(x)) + g(x, \sigma^*(x)) \\ &= r(x, \sigma^*(x)) + \mathbb{E}_{x,\sigma^*(x)}\beta(x, \sigma^*(x), x')\bar{v}(x') = T_{\sigma^*}\bar{v}(x).\end{aligned} \tag{2.28}$$

for all $x \in \mathsf{X}$. Hence, the same iteration of (2.27) with $\bar{v} = T_{\sigma^*}\bar{v}$ implies $\bar{v} = v_{\sigma^*} \leqslant v^*$. We conclude that $\bar{v} = v^*$. The above arguments also show that a g^*-greedy policy σ^* is optimal: $v^* = v_{\sigma^*}$ ((b) $\Rightarrow$ (a)), and an optimal policy exists. We now show that part (a) implies (b). We write $M_\sigma g(x) = r(x, \sigma(x)) + g(x, \sigma(x))$ for $x \in \mathsf{X}$, $g \in \mathcal{G}$ and $\sigma \in \Sigma$. Since $\bar{g} = Ev^*$, if σ^* is optimal, then we have

$$M_{\sigma^*}\bar{g} = M_{\sigma^*}Ev^* = T_{\sigma^*}v^* = T_{\sigma^*}v_{\sigma^*} = v_{\sigma^*} = v^* = Tv^* = MEv^* = M\bar{g}.$$

Hence, σ^* is $\bar{g}$-greedy. We next show that a policy is $\bar{g}$-greedy if and only if it is $M\bar{g}$-greedy. If $\sigma \in \Sigma$ is $\bar{g}$-greedy: $M_\sigma\bar{g} = M\bar{g}$, then $TM\bar{g} = MEM\bar{g} = MR\bar{g} = M\bar{g} = M_\sigma\bar{g} = M_\sigma R\bar{g} = M_\sigma EM\bar{g} = T_\sigma M\bar{g}$, whence σ is $M\bar{g}$-greedy. Conversely, if σ is $M\bar{g}$-greedy: $T_\sigma M\bar{g} = TM\bar{g}$, then $M_\sigma\bar{g} = M_\sigma R\bar{g} = M_\sigma EM\bar{g} = T_\sigma M\bar{g} = TM\bar{g} = MEM\bar{g} = MR\bar{g} = M\bar{g}$, whence σ is $\bar{g}$-greedy. Similarly, the same method shows that a policy is v^*-greedy if and only if it is Ev^*-greedy. ☐

PROOF OF THEOREM 2.6.1. Let Condition 2.8.1 and Assumption 2.6.1 hold. Part (a) follows from 2.8.20. Part (b) follows from Lemma 2.8.23. Part (c), (d), (f), and (g) follow from Part (b), Lemma 2.8.22, and 2.8.25. Part (e) follows from Part (b) and (c). ☐

PROOF OF LEMMA 2.6.1. Let all the stated assumptions hold. We show the statement by induction. First, observe that

$$\hat{r}(x, a) = \mathbb{E}_z u(R(\varepsilon')(w - c) + y(\varepsilon', Z')) \geqslant \mathbb{E}_z u(y(Z', \varepsilon')) > -\infty,$$

$$\bar{r}(x) = \sup_{0 \leqslant c \leqslant w} u(c) = u(w) \leqslant pw + q =: \kappa(x).$$

The last equation implies that we can set $d = 1$ in Assumption 2.6.1. Also, observe that

$$\mathbb{E}_{x,a}\kappa(X') = p\mathbb{E}_z(R(\varepsilon')(w - c) + y(Z', \varepsilon')) + q$$

is continuous in $(x,a) = (w,z,c)$. Therefore, Assumption 2.6.1 (a) and (c) hold. Next, note that

$$\begin{aligned}\frac{\mathbb{E}_{x,a}\beta(x,a,X')\kappa(X')}{\kappa(x)} &= \frac{\mathbb{E}_z\beta(z,Z')(pW'+q)}{pw+q} \\ &= \frac{\mathbb{E}_z\beta(z,Z')[p(R(\varepsilon')(w-c)+y(Z',\varepsilon'))+q]}{pw+q} \\ &\leqslant \frac{\mathbb{E}_z\beta(z,Z')[p(R(\varepsilon')w+y(Z',\varepsilon'))+q]}{pw+q}.\end{aligned}$$

Since the right-hand side is a monotone function of w and then achieves the supremum at either $w=0$ or $w=\infty$, we have

$$\frac{\mathbb{E}_{x,a}\beta(x,a,X')\kappa(x')}{\kappa(x)} \leqslant \max\left\{\frac{\mathbb{E}_z\beta(z,Z')[py(Z',\varepsilon')+q]}{q}, \mathbb{E}_z\beta(z,Z')\mathbb{E}R(\varepsilon')\right\}.$$

where the last inequality follows from that ε_t is IID. Then, since $q>1$ can be arbitrarily large, (2.19) implies

$$\frac{\mathbb{E}_z\beta(z,Z')[py(Z',\varepsilon')+q]}{q} \to \mathbb{E}_z\beta(z,Z')$$

as $q\to\infty$. Since $\mathbb{E}R(\varepsilon')>1$, we have

$$\frac{\mathbb{E}_{x,a}\beta(x,a,X')\kappa(X')}{\kappa(x)} \leqslant \mathbb{E}_z\beta(z,Z')\mathbb{E}R(\varepsilon') =: L\mathbb{1}(x)$$

for large enough $q>1$. Hence, we have $\mathbb{E}_{x,a}\beta(x,a,X')\kappa(X')L^0\mathbb{1}(x') \leqslant \kappa(x)L\mathbb{1}(x)$ for large enough $q>1$. Assume the induction hypothesis that $\mathbb{E}_{x,a}\beta(x,a,X')\kappa(X')L^n\mathbb{1}(X') \leqslant$

$\kappa(x)L^{n+1}\mathbb{1}(x)$ for some $n \in \{1, \dots, m-1\}$. Then, iteration yields

$$\begin{aligned}
&\frac{\mathbb{E}_{x,a}\beta(x,a,X')\kappa(x')L^{n+1}\mathbb{1}(X')}{\kappa(x)} \\
&\leqslant \frac{\mathbb{E}_z\beta(z,Z')[p(R(\varepsilon')w + y(Z',\varepsilon')) + q]\mathbb{E}_{Z'}\beta_1\beta_2\dots\beta_{n+1}(\mathbb{E}R(\varepsilon'))^{n+1}}{pw+q} \\
&\leqslant \max\left\{\frac{\mathbb{E}_z\beta(z,Z')[py(Z',\varepsilon') + q]\mathbb{E}_{Z'}\beta_1\beta_2\dots\beta_{n+1}(\mathbb{E}R(\varepsilon'))^{n+1}}{q},\right. \\
&\qquad\qquad\qquad \left.\mathbb{E}_z\beta(z,Z')\beta_1\beta_2\cdots\beta_{n+1}(\mathbb{E}R(\varepsilon'))^{n+2}\right\} \\
&\to \max\{\mathbb{E}_z\beta(z,Z')\beta_1\beta_2\cdots\beta_{n+1}(\mathbb{E}R(\varepsilon'))^{n+1}, \\
&\qquad\qquad\qquad \mathbb{E}_z\beta(z,Z')\beta_1\beta_2\cdots\beta_{n+1}(\mathbb{E}R(\varepsilon'))^{n+2}\} \\
&\leqslant L^{n+2}\mathbb{1}(x),
\end{aligned}$$

where the first inequality follows from the definitions of κ and $L^{n+1}\mathbb{1}$, the second inequality follows from that the supremum is at either $w = 0$ or $w = \infty$, the limit follows from taking q arbitrarily large, and the last inequality uses $\mathbb{E}R(\varepsilon') > 1$. Therefore, induction implies that there is large enough $q(n) > 1$ such that

$$\mathbb{E}_{x,a}\beta(x,a,X')\kappa(X')L^n\mathbb{1}(X') \leqslant \kappa(x)L^{n+1}\mathbb{1}(x)$$

for all $x \in \mathsf{X}$ and all $n \in \{1, \dots, m-1\}$. □

2.8.5. Proofs in Section 2.7.

Lemma 2.8.26. *If Condition 2.7.1 and and Assumption 2.7.2 holds, then $R\mathcal{G} \subset \mathcal{G}$.*

Proof of Lemma 2.8.26. Let Condition 2.7.1 and Assumption 2.7.2 hold. Let $g \in \mathcal{G}$. We first show that $\|Rg\| < \infty$. Since $g \geqslant -\|g\|$, we have

$$\begin{aligned}
Rg(x,a) &\geqslant -\frac{\beta(x,a)}{\theta}\ln\mathbb{E}_{x,a}\exp\left(-\theta\sup_{a'\in\Gamma(X')}\{r(X',a') - \|g\|\}\right) \\
&= -\frac{\beta(x,a)}{\theta}\ln\mathbb{E}_{x,a}\exp\left(-\theta(\bar{r}(X') - \|g\|)\right) = \beta(x,a)(\hat{r}(x,a) - \|g\|).
\end{aligned}$$

Hence, since $\hat{r}$ is bounded below, Rg is also bounded below. Similarly, we can show that $Rg(x,a) \leqslant \beta(x,a)(\hat{r}(x,a) + \|g\|)$ for all $(x,a) \in \mathsf{G}$. Since β is bounded and $\bar{r}$ is bounded above, $\hat{r}$ is bounded above and then Rg is bounded above. Therefore, we have $\|Rg\| < \infty$.

Next, since g and r are u.s.c., and Γ is compact-valued and u.s.c., the map $x \mapsto h(x) := \sup_{a \in \Gamma(x)}\{r(x,a) + g(x,a)\}$ is u.s.c. by the Maximum theorem. Now, note that if $x_1 \leqslant x_2$ in X, then $x_1' = f(x_1, a, \varepsilon') \leqslant f(x_2, a, \varepsilon') = x_2'$ and then $\Gamma(x_1') \subset \Gamma(x_2')$. Since r and g are increasing in x, we have

$$h(x_1') = \sup_{a' \in \Gamma(x_1')} \{r(x_1', a') + g(x_1', a')\} \leqslant \sup_{a' \in \Gamma(x_2')} \{r(x_2', a') + g(x_2', a') = h(x_2').$$

Therefore, we have $-(1/\theta)\ln \mathbb{E}_{x_1,a} \exp(-\theta h(x_1')) \leqslant -(1/\theta)\ln \mathbb{E}_{x_2,a} \exp(-\theta h(x_2'))$. Since in addition β is increasing in x, we see that Rg is increasing in x. We conclude that $R\mathcal{G} \subset \mathcal{G}$. ☐

Lemma 2.8.27. *If Condition 2.7.1, Assumption 2.7.1, 2.7.2, and 2.7.3 hold, then R is eventually contracting on $(\mathcal{G}, \|\cdot\|)$ and has a unique fixed point in $\mathcal{G}$.*

Proof of Lemma 2.8.27. Let Condition 2.7.1, Assumption 2.7.1, 2.7.2, and 2.7.3 hold. We show that there is an $n \in \mathbb{N}$ such that R^n satisfies Blackwell's condition. Let $g \in \mathcal{G}$ and $K \geqslant 0$. Fix $(x,a) \in \mathsf{G}$. Observe that

$$\begin{aligned} R(g+K)(x,a) &= \frac{-\beta(x,a)}{\theta} \ln \mathbb{E}_{x,a} \exp\left(-\theta \sup_{a' \in \Gamma(X')} \{r(X',a') + g(X',a') + K\}\right) \\ &= Rg(x,a) + \beta(x,a)K \leqslant Rg(x,a) + KL\mathbb{1}(x). \end{aligned}$$

Define the function $\varphi(t) = -1/\theta \ln \mathbb{E}_{x,a} \exp(-\theta t)$ for random variable $t = t(X')$. Since φ is monotonically increasing, iteration implies

$$\begin{aligned} R^2(g+K)(x,a) &= \beta(x,a)\varphi\left(\sup_{a' \in \Gamma(X')} \{r(X',a') + R(g+K)(X',a')\}\right) \\ &= \beta(x,a)\varphi\left(\sup_{a' \in \Gamma(X')} \{r(X',a') + Rg(X',a') + \beta(X',a')K\}\right) \\ &\leqslant \beta(x,a)\varphi\left(\sup_{a' \in \Gamma(X')} \{r(X',a') + Rg(X',a')\} + K \sup_{a' \in \Gamma(X')} \{\beta(X',a')\}\right). \end{aligned}$$

Let $X = \sup_{a'\in\Gamma(X')}\{r(X',a') + Rg(X',a')\}$ and $Y = K\sup_{a'\in\Gamma(X')}\{\beta(X',a')\}$. Since $Rg(x,a)$, $r(x,a)$, $\beta(x,a)$, and $\Gamma(x)$ are increasing in x, and $x' = f(x,a,\varepsilon')$ is increasing in x and ε' for independent ε', we have $\operatorname{Cov}(e^{-\theta X}, e^{-\theta Y}|(x,a)) \geqslant 0$. Therefore, since $\mathbb{E}_{x,a}e^{-\theta X}e^{-\theta Y} \geqslant \mathbb{E}_{x,a}e^{-\theta X}\mathbb{E}_{x,a}e^{-\theta Y}$, we have[26]

$$\begin{aligned}\varphi(X+Y) &= \frac{-1}{\theta}\ln \mathbb{E}_{x,a}e^{-\theta X - \theta Y}\\ &\leqslant \frac{-1}{\theta}\ln\left(\mathbb{E}_{x,a}e^{-\theta X}\mathbb{E}_{x,a}e^{-\theta Y}\right) = \varphi(X) + \varphi(Y).\end{aligned}$$

We then have

$$\begin{aligned}&R^2(g+K)(x,a)\\ &\quad\leqslant \beta(x,a)\varphi\left(\sup_{a'\in\Gamma(X')}\{r(X',a') + Rg(X',a')\}\right) + \beta(x,a)\varphi\left(K\sup_{a'\in\Gamma(X')}\{\beta(X',a')\}\right)\\ &\quad= R^2 g(x,a) + \beta(x,a)\varphi\left(K\sup_{a'\in\Gamma(X')}\{\beta(X',a')\}\right).\end{aligned}$$

Now, since $t \mapsto \exp(-\theta t)$ is convex, Jensen inequality implies

$$\varphi(Y) = \frac{-1}{\theta}\ln \mathbb{E}_{x,a}e^{-\theta Y} \leqslant \frac{-1}{\theta}\ln e^{-\theta \mathbb{E}_{x,a}Y} = \mathbb{E}_{x,a}Y.$$

Therefore, we obtain

$$\begin{aligned}R^2(g+K)(x,a) &\leqslant R^2 g(x,a) + K\mathbb{E}_{x,a}\beta(x,a)\sup_{a'\in\Gamma(X')}\beta(X',a')\\ &\leqslant R^2 g(x,a) + KL\bar{\beta}(x),\end{aligned}$$

where L and $\bar{\beta}$ are defined in Assumption 2.7.3. With the same argument, the induction shows that $R^n(g+K)(x,a) \leqslant R^n g(x,a) + KL^{n-1}\bar{\beta}(x)$ for all $n \in \mathbb{N}$. Since $\rho(L) < 1$ implies that there is $m \in \mathbb{N}$ such that $\|L^{m-1}\bar{\beta}\| < 1$, we have

$$R^m(g+K) \leqslant R^m g + K\|L^{m-1}\bar{\beta}\|.$$

---

[26] See also Lemma 3 of Bäuerle and Jaśkiewicz (2018).

Hence, R^m satisfies the Blackwell's condition and is a contraction map. Then, since $\mathcal{G}$ is a Banach space, R admits a unique fixed point in $\mathcal{G}$ by the generalised Contracting Mapping theorem. □

LEMMA 2.8.28. *If Condition 2.7.1, Assumption 2.7.1, 2.7.2, and 2.7.3 hold, then for all constant* $c \in \mathbb{R}$ *we have* $v_\sigma(x) = \limsup_{n\to\infty} T_\sigma^n \bar{r}(x) = \limsup_{n\to\infty} T_\sigma^n(\bar{r}+c)(x)$ *for all* $x \in \mathsf{X}$ *and* $\sigma \in \Sigma$.

PROOF OF LEMMA 2.8.28. Let Condition 2.7.1, Assumption 2.7.1, 2.7.2, and 2.7.3 hold. Fix $\sigma \in \Sigma$ and constant $K \geqslant 0$. Since $\bar{r} + K \geqslant \bar{r}$, the monotonicity of T_σ implies $\limsup_{n\to\infty} T_\sigma^n(\bar{r}+K) \geqslant \limsup_{n\to\infty} T_\sigma^n \bar{r}$. Next, similar to the iteration in Lemma 2.8.27, we have, for all $x \in \mathsf{X}$,

$$\begin{aligned}
T_\sigma(\bar{r}+K)(x) &= r_\sigma(x) - \frac{\beta_\sigma(x)}{\theta} \ln \mathbb{E}_x^\sigma e^{-\theta(\bar{r}+K)(X')} \\
&= r_\sigma(x) - \frac{\beta_\sigma(x)}{\theta} \ln e^{-\theta K} \mathbb{E}_x^\sigma e^{-\theta \bar{r}(X')} \\
&= r_\sigma(x) - \frac{\beta_\sigma(x)}{\theta} \ln \mathbb{E}_x^\sigma e^{-\theta \bar{r}(X')} + \beta_\sigma(x) K = T_\sigma \bar{r}(x) + \beta_\sigma(x) K,
\end{aligned}$$

where $r_\sigma(x) = r(x, \sigma(x))$ and $\beta_\sigma(x) = \beta(x, \sigma(x))$ for all $x \in \mathsf{X}$. Since r_σ, β_σ, $x' = f(x, \sigma(x), \varepsilon')$ are increasing in x, we see that $T_\sigma \bar{r}(x)$ is increasing in x. Then, it follows from the argument in Lemma 2.8.27 and Assumption 2.7.1 that $\mathbb{E}_x^\sigma e^{-\theta T_\sigma \bar{r}(X')} e^{-\theta \beta_\sigma(X')K} \geqslant \mathbb{E}_x^\sigma e^{-\theta T_\sigma \bar{r}(X')} \mathbb{E}_x^\sigma e^{-\theta \beta_\sigma(X')K}$, so iteration yields that for all $x \in \mathsf{X}$

$$\begin{aligned}
T_\sigma^2(\bar{r}+K)(x) &= r_\sigma(x) - \frac{\beta_\sigma(x)}{\theta} \ln \mathbb{E}_x^\sigma e^{-\theta T_\sigma(\bar{r}+K)(X')} \\
&= r_\sigma(x) - \frac{\beta_\sigma(x)}{\theta} \ln \mathbb{E}_x^\sigma e^{-\theta (T_\sigma \bar{r}(X') + \beta_\sigma(X')K)} \\
&= r_\sigma(x) - \frac{\beta_\sigma(x)}{\theta} \ln \mathbb{E}_x^\sigma e^{-\theta T_\sigma \bar{r}(X')} e^{-\theta \beta_\sigma(X')K} \\
&\leqslant r_\sigma(x) - \frac{\beta_\sigma(x)}{\theta} \ln \mathbb{E}_x^\sigma e^{-\theta T_\sigma \bar{r}(X')} \mathbb{E}_x^\sigma e^{-\theta \beta_\sigma(X')K} \\
&= T_\sigma^2 \bar{r}(x) - \frac{\beta_\sigma(x)}{\theta} \ln \mathbb{E}_x^\sigma e^{-\theta \beta_\sigma(X')K} \\
&\leqslant T_\sigma^2 \bar{r}(x) + K \mathbb{E}_x^\sigma \beta_\sigma(x) \beta_\sigma(X') \\
&\leqslant T_\sigma^2 \bar{r}(x) + K L \bar{\beta}(x)
\end{aligned}$$

where the second inequality follows from Jensen inequality as iteration in Lemma 2.8.27, and the last inequality follows from Assumption 2.7.3. Then, using the same induction, we have

$$T_\sigma^n(\bar{r}+K)(x) \leqslant T_\sigma^n \bar{r}(x) + KL^{n-1}\bar{\beta}(x).$$

Hence, since $\rho(L) < 1$ by Assumption 2.7.3, there exists $m \in \mathbb{N}$ such that $\|L^m \bar{\beta}\| < 1$. Then, we have $\limsup_{n\to\infty} L^n \bar{\beta}(x) \leqslant \limsup_{n\to\infty} \|L^n \bar{\beta}\| \to 0$ for all $x \in \mathsf{X}$. Therefore, we have $\limsup_{n\to\infty} T_\sigma^n(\bar{r}+K)(x) \leqslant \limsup_{n\to\infty} T_\sigma^n \bar{r}(x)$ for all $x \in \mathsf{X}$. Then, we have $\limsup_{n\to\infty} T_\sigma^n(\bar{r}+K)(x) = \limsup_{n\to\infty} T_\sigma^n \bar{r}(x)$ for all $x \in \mathsf{X}$ when $K \geqslant 0$. Similarly, we can show the statement when $K < 0$, where the above inequalities are reversed. ☐

PROOF OF THEOREM 2.7.1. Let Condition 2.7.1, Assumption 2.7.1, 2.7.2, and 2.7.3 hold. Part (a) follows from 2.8.26 and 2.8.27. Part (b), (c), (e), and (f) of optimality follow from the proof of Theorem 5.3 of Ma et al. (2022) and Lemma 2.8.27 and 2.8.28. Part (d) follows from (a) and (b). ☐

CHAPTER 3

Q-Learning with State-Action-Dependent Discounting

3.1. Introduction

Reinforcement learning algorithms, such as Q-learning introduced by Watkins (1989), learn optimal actions from interaction between agents and environments. In reinforcement learning, a discount factor is a meta-parameter of learning performance and is often set to be less than one constant. However, in economics and finance, it is well-known that discount factors vary over time and are state-action-dependent discounting (Cochrane, 2011, Hills and Nakata, 2018). For example, asset pricing models consider a stochastic discount factor, which is a function of risk preference, for the trade-off between current and future state-dependent consumption levels (Lucas Jr, 1978, Campbell and Ammer, 1993, Rosenberg and Engle, 2002, Cochrane, 2009, 2011, Hansen and Renault, 2010).

To apply Q-learning in economic or finance models, it is essential to generalise constant discounting to stochastic state-action-dependent discounting. In particular, we are interested in the case that discount factors may be greater than one with a positive probability. It has been noted by Nakata (2016), Hills and Nakata (2018) and Hubmer et al. (2021) that if a discount factor is defined as the reciprocal of gross interest rate, then it exhibits dynamic behaviour and may occasionally exceed one.

In the existing literature, Yoshida et al. (2013) show the convergence of Q-learning with state-dependent discount factors and provide a framework to optimise the state-dependent discount function, demonstrating superior performance compared to a constant discount factor by simulation. Similarly, Sharma et al. (2021) explore the convergence of Q-learning and SARSA algorithms with both state- and action-dependent discount factors. However, both of them assume that discount factors are strictly less than one, whence the Bellman operator and the Q-factor Bellman operator are evidently contraction maps.

To address this gap, this paper shows the convergence of three model-free reinforcement learning algorithms under state-action-dependent discounting, including Q-learning, on-policy learning SARSA, and double Q-learning algorithms (Watkins and Dayan, 1992, Singh et al., 2000, Hasselt, 2010). Following the assumptions of state-dependent discounting outlined in Stachurski and Zhang (2021), we assume that discount factors are eventually discounting: for all feasible action policies σ there exists an $n_\sigma \in \mathbb{N}$ such that

$$\sup_x \mathbb{E}_x \prod_{t=0}^{n_\sigma - 1} \beta(X_t, \sigma(X_t), X_{t+1}) < 1,$$

where β denotes the discount factor and is a function of states X_t and actions/policies $\sigma(X_t)$.[1] Eventual discounting indicates that the expected multiplicative of discount factors is eventually less than one for any policy.

Our condition does not rule out $\beta_t = \beta(X_t, \sigma(X_t), X_{t+1}) > 1$ with positive probability. On the other hand, the conventional proofs of the convergences for the Q-learning, SARSA, and double Q-learning rely on the fact that the discount factor is strictly less than one, which yields a contractive Bellman operator (Watkins and Dayan, 1992, Tsitsiklis, 1994, Bertsekas and Tsitsiklis, 1995, Singh et al., 2000, Hasselt, 2010).

In addition, we analyze the contraction of the Bellman operator in the weighted norm of the eigenvector corresponding to the spectral radius of appropriate matrices governing discounting. This property provides convergence rates and error bounds for dynamic programming algorithms under state-action-dependent discounting. Moreover, this feature pins down the bound for optimal Q-factor.

Finally, we extend the Stochastic Approximation algorithm and Q-learning to incorporate concave operators. Since a concave operator may have a unique fixed point and be globally stable, Stochastic Approximation with a concave operator converges whenever it is bounded. This provides an alternative method to prove the convergence of Q-learning.

Related literature - Regarding action-dependent discounting, endogenous time preferences or Uzawa time preferences are broadly studied in small open economies such as Obstfeld

---

[1] $\{X_t\}_{t \geqslant 0}$ is a Markov process defined in Section 3.2.1.

(1990), Mendoza (1991), Schmitt-Grohé and Uribe (2003), Vasilev (2022a), Durdu et al. (2009).

The related literature in Markov decision processes with state- or action-dependent discount factors includes Wei and Guo (2011), Minjárez-Sosa (2015), Wu et al. (2015), Wu and Zhang (2016), Jasso-Fuentes et al. (2022). The literature about dynamic programming theory with state-dependent stochastic discounts in economics and finance includes Stachurski and Zhang (2021), Toda (2021), Sargent and Stachurski (2023), Toda (2023).

The applications of Q-learning in economics are as follows. Park and Ryu (2022) study suppliers' collusion behaviours concerning supply chain ethics and transparency with Q-learning agents. Through simulations, they show that suppliers tend to exhibit low levels of ethics and transparency. Neuneier (1997) utilises Q-learning to study the optimal asset allocation. Waltman and Kaymak (2008) use Q-learning to model the learning behaviour of firms in repeated Cournot oligopoly games and find that Q-learning firms learn to collude with each other. Calvano et al. (2020) find that Q-learning artificial intelligence autonomously learns to charge supracompetitive prices in an oligopoly model of repeated price competition. Charpentier et al. (2021) conduct a comprehensive survey on the applications of reinforcement learning in economics and finance.

The paper is structured as follows. Section 3.2.1 introduces the framework and reinforcement learning algorithms. Section 3.3 presents the main results of the convergence of Q-leaning algorithms. Section 3.5 discusses the stability of the algorithm. Section 3.6 extends Stochastic Approximation and Q-learning to incorporate concavity.

3.2. Q-learning, SARSA, and Double Q-learning

In this section, we introduce the Markov decision process and the model-free learning algorithms, including Q-learning, SARSA, and double Q-learning.

3.2.1. Background. A finite *Markov decision process (MDP)* is a tuple $(\mathsf{X}, \mathsf{A}, \Gamma, \beta, P, r)$ satisfying that (i) X and A are finite state space and action space, respectively, (ii) Γ is a nonempty correspondence from $\mathsf{X} \to \mathsf{A}$, referred to as the feasible correspondence, which defines the feasible state-action pairs $\mathsf{G} \coloneqq \{(x, a) \in \mathsf{X} \times \mathsf{A} \colon a \in \Gamma(x)\}$, (iii) $\beta \colon \mathsf{G} \times \mathsf{X} \to \mathbb{R}_+$ is a function of discount factors, (iv) $r \colon \mathsf{G} \times \mathsf{X} \to \mathbb{R}$ is a function of rewards, and (v) a stochastic kernel $P \colon \mathsf{G} \times \mathsf{X} \to \mathbb{R}_+$ satisfies $\sum_{x' \in \mathsf{X}} P(x, a, x') = 1$ for all $(x, a) \in \mathsf{G}$.

Given an MDP $(\mathsf{X}, \mathsf{A}, \Gamma, \beta, P, r)$, the set of *feasible policies* is defined by

$$\Sigma \coloneqq \{\sigma \in \mathsf{A}^{\mathsf{X}} \colon \sigma(x) \in \Gamma(x), \quad \forall x \in \mathsf{X}\}.$$

An MDP is *recurrent* or *ergodic* if the Markov chain corresponding to every deterministic stationary policy $\sigma \in \Sigma$ consists of a single recurrent class.[2] If an MDP is recurrent, then for every policy $\sigma \in \Sigma$ the induced Markov chain will eventually visit every state; that is, every state is visited infinitely often.

To any stationary policy $\sigma \in \Sigma$ and initial state x, consider a trajectory process $X \coloneqq \{X_t\}_{t \in \mathbb{N}_0}$ taking values in state space X and controlled by policies $\sigma(X) \coloneqq \{\sigma(X_t)\}_{t \in \mathbb{N}_0}$ such that X_{t+1} is generated by $P(X_t, \sigma(X_t), \cdot)$ for all $t \in \mathbb{N}_0$. The processes X and $\sigma(X)$ are well-defined on some measurable space $(\Omega, \mathscr{F}, \mathbb{P})$ satisfying $\mathbb{P}(X_{t+1} = x' | \mathscr{F}_t) = P(X_t, \sigma(X_t), x')$ almost surely for any $x' \in \mathsf{X}$ and $t \in \mathbb{N}_0$, where $\mathscr{F}_t \coloneqq \sigma\{X_0, \dots, X_t, a_0, \dots\}$ denotes the information set up to time t.

Let r_t be the reward drawn from a fixed reward distribution $r : \mathsf{G} \times \mathsf{X} \to \mathbb{R}$ such that $\mathbb{E}[r_t | (x, a, x') = (X_t, \sigma(X_t), X_{t+1})] = r(x, a, x')$, where the conditional expectation of r_t is $r(x, a, x')$ conditioning on $X_t = x, \sigma(X_t) = a$ and being governed by underlying state transition $X_{t+1} = x'$.

---

[2] A state x is recurrent if it is eventually visited or returned to.

The σ-value function v_σ is defined by

$$v_\sigma(x) \coloneqq \mathbb{E}_x^\sigma \left[\sum_{t=0}^{\infty} \prod_{i=0}^{t-1} \beta_i r_t \right]$$

where $\beta_i \coloneqq \beta(X_i, \sigma(X_i), X_{i+1})$ for all $i \in \mathbb{N}_0$, $\mathbb{E}_x^\sigma$ denotes expectation conditioning on $\{x_0 = x\}$ with transition probability measure $P(X_t, \sigma(X_t), \cdot)$ for $t \in \mathbb{N}_0$, and $\prod_{i=0}^{-1} \beta_i \coloneqq 1$ by convention. The maximum total reward or *value function* is defined by

$$v^*(x) \coloneqq \sup_{\sigma \in \Sigma} v_\sigma(x) \qquad (x \in \mathsf{X}).$$

A policy $\sigma^* \in \Sigma$ is *optimal* if $v_{\sigma^*} = v^*$. The *Bellman equation* follows

$$v(x) = \max_{a \in \Gamma(x)} \left\{ r(x,a) + \sum_{x' \in \mathsf{X}} \beta(x,a,x') v(x') P(x,a,x') \right\} \qquad (x \in \mathsf{X})$$

where $r(x,a) \coloneqq \sum_{x' \in \mathsf{X}} r(x,a,x') P(x,a,x')$ for all $(x,a) \in \mathsf{G}$. Given $v \in \mathbb{R}^{\mathsf{X}}$, a policy $\sigma \in \Sigma$ is *v-greedy* if

$$\sigma(x) \in \operatorname*{arg\,max}_{a \in \Gamma(x)} \left\{ r(x,a) + \sum_{x' \in \mathsf{X}} \beta(x,a,x') v(x') P(x,a,x') \right\} \qquad (x \in \mathsf{X}).$$

The *Bellman operator* $T \colon \mathbb{R}^{\mathsf{X}} \to \mathbb{R}^{\mathsf{X}}$ is defined by

$$Tv(x) \coloneqq \max_{a \in \Gamma(x)} \left\{ r(x,a) + \sum_{x' \in \mathsf{X}} \beta(x,a,x') v(x') P(x,a,x') \right\} \qquad (x \in \mathsf{X}).$$

3.2.2. Eventual Discounting. In this section, we introduce the main assumptions imposed on state-action-dependent discount factors. Define the expected multiplicative of discount factors d_n by

$$d_n \coloneqq \max_{\sigma \in \Sigma} \max_{x \in \mathsf{X}} \left\{ \mathbb{E}_x^\sigma \prod_{t=0}^{n-1} \beta(X_t, \sigma(X_t), X_{t+1}) \right\}. \tag{3.1}$$

We say that the MDP is *eventually discounting* if there is $n \in \mathbb{N}$ such that $d_n < 1$. In other words, the MDP is eventually discounting if the expected multiplicative of discount factors is eventually smaller than one for any policy.

We briefly discuss the sufficient conditions for eventual discounting. Denote function $B\colon \mathsf{G} \times \mathbb{R}^{\mathsf{X}} \to \mathbb{R}$ as

$$B(x, a, v) := r(x, a) + \sum_{x' \in \mathsf{X}} P(x, a, x')\beta(x, a, x')v(x') \qquad ((x, a, v) \in \mathsf{G} \times \mathbb{R}^{\mathsf{X}}).$$

Suppose that there exists a $|\mathsf{X}| \times |\mathsf{X}|$ matrix L such that

$$|B(x, a, v) - B(x, a, w)| \leqslant \sum_{x' \in \mathsf{X}} L(x, x')|v(x') - w(x')| \qquad ((x, a) \in \mathsf{G}). \tag{3.2}$$

for all $v, w \in \mathbb{R}^{\mathsf{X}}$. Let $\rho(L) := \max\{|\lambda| : \lambda \text{ is an eigenvalue of } L\}$ be the *spectral radius* of L. If the spectral radius of L is less than one, we can show that it is eventually discounting; that is, $\rho(L) < 1$ implies $d_n < 1$ for some $n \in \mathbb{N}$. One possible assumption is to choose L to be

$$L_m(x, x') := \max_{a \in \Gamma(x)} \beta(x, a, x')P(x, a, x') \qquad ((x, x') \in \mathsf{X}^2) \tag{3.3}$$

and assume $\rho(L_m) < 1$.

Lemma 3.2.1. *If L_m defined by (3.3) satisfies $\rho(L_m) < 1$, then there exists an $n \in \mathbb{N}$ such that $d_n < 1$.*

Given σ, define the $|\mathsf{X}| \times |\mathsf{X}|$ matrix L_σ by

$$L_\sigma(x, x') := \beta(x, \sigma(x), x')P(x, \sigma(x), x')$$

for all $(x, x') \in \mathsf{X} \times \mathsf{X}$. Induction implies

$$\sum_{x'} L_\sigma^n(x, x') = \mathbb{E}_x^\sigma \prod_{t=0}^{n-1} \beta(x_t, \sigma(X_t), X_{t+1}) \qquad (x \in \mathsf{X}).$$

Then, we can show that $\rho(L_\sigma) < 1$ if and only if there is $n \in \mathbb{N}$ such that $d_n^\sigma < 1$, where

$$d_n^\sigma := \max_x \mathbb{E}_x^\sigma \prod_{t=0}^{n-1} \beta(X_t, \sigma(X_t), X_{t+1}).$$

Moreover, $d_n < 1$ for some $n \in \mathbb{N}$ implies

$$\max_\sigma d_n^\sigma = \max_\sigma \max_x \sum_{x'} L_\sigma^n(x, x') < 1.$$

Hence, we can further show that the eventual discounting is equivalent to

$$\max_{\sigma} \rho(L_\sigma) < 1.$$

Example 3.2.1. In economics and finance, discount factors are frequently determined as the reciprocals of gross real interest rates: $\beta_t = 1/(1+i_t)$, where $i_t \in \mathbb{R}$ is the real interest rate. It is well-known that real interest rates could be negative when there is zero lower bound for nominal interest rates (Hills and Nakata, 2018, Nakata, 2016, Hubmer et al., 2021). Hills et al. (2019) and Hubmer et al. (2021) consider an AR(1) process: $\beta_t = Z_t$, where $\{Z_t\}$ follows

$$Z_{t+1} = \rho_Z Z_t + (1-\rho_Z)\mu_Z + \sigma_\varepsilon \varepsilon_{t+1} \qquad \{\varepsilon_t\} \overset{IID}{\sim} N(0,1). \tag{3.4}$$

Hubmer et al. (2021) calibrate $\rho_Z = 0.992, \mu_Z = 0.944, \sigma_\varepsilon = 0.0006$. They discretize the process onto a grid of $N = 15$ states by Tauchen's method which allows us to write the operator L defined in (3.3) as

$$L_{ij} = \beta(x_i)P(x_i, x_j), \qquad 1 \leqslant i, j \leqslant N.$$

The spectral radius of matrix L is 0.9469 computed by Stachurski and Zhang (2021).

As summarised in the following proposition, if an MDP is eventually discounting, then the optimality of dynamic programming holds that a v^*-greedy policy is optimal. Moreover, we can compute the value function v^* by the value function iteration of $\{T^k v\}_{k\geqslant 0}$ for any $v \in \mathbb{R}^{\mathsf{X}}$. The proof for the following proposition can be found in Sargent and Stachurski (2023).

Proposition 3.2.1. *For an eventually discounting MDP,*

(i) the value function v^* is the unique solution to Bellman's equation in $\mathbb{R}^{\mathsf{X}}$,

(ii) $\lim_{k\to\infty} T^k v = v^*$ for all $v \in \mathbb{R}^{\mathsf{X}}$, and

(iii) a feasible policy is optimal if and only if it is v^*-greedy.

3.2.3. Q-learning. For each $v \in \mathbb{R}^{\mathsf{X}}$, the *Q-factor*, $Q \in \mathbb{R}^{\mathsf{G}}$, corresponding to $v \in \mathbb{R}^{\mathsf{X}}$ is defined by

$$Q(x,a) = r(x,a) + \sum_{x' \in \mathsf{X}} \beta(x,a,x')v(x')P(x,a,x') \qquad (x,a) \in \mathsf{G}.$$

Denote Q^* as the Q-factor corresponding to value function v^*:

$$Q^*(x,a) = r(x,a) + \sum_{x' \in \mathsf{X}} \beta(x,a,x')v^*(x')P(x,a,x') \qquad (x,a) \in \mathsf{G}.$$

The Bellman equation implies that

$$v^*(x) = \max_{a \in \Gamma(x)} Q^*(x,a) \qquad (x \in \mathsf{X}).$$

The goal of the Q-learning algorithm is to learn $Q^*(x,a)$ for all state-action pairs. Denote $\{x_t\}_{t \geqslant 0}$ as the realisation of the Markov process where x_{t+1} is generated by $P(x_t, a_t, \cdot)$ for $t \in \mathbb{N}$, given $\{a_t\}_{t \geqslant 0}$. The Q-learning iterates the vector $Q_t \in \mathbb{R}^{\mathsf{G}}$ following the rule:

$$Q_{t+1}(x_t,a_t) = (1 - \alpha_t(x_t,a_t))Q_t(x_t,a_t) + \alpha_t(x_t,a_t)\left[r_t + \beta_t \max_{b \in \Gamma(x_{t+1})} Q_t(x_{t+1},b)\right], \tag{3.5}$$

for $t \in \mathbb{N}_0$, where x_t, a_t, r_t, and α_t are the state, action, reward, and step size at time step t. In detail, given realised $\{(x_t,a_t)\}_{t \in \mathbb{N}_0}$, $\alpha_t(x,a) \in [0,1]$ is a step-size coefficient that $\alpha_t(x,a) = 0$ for all $(x,a) \neq (x_t,a_t)$; that is, α_t is set to zero for those are outside of the support of (x_t,a_t). Moreover, r_t is a random sample of reward that $\mathbb{E}[r_t|(x,a,x') = (x_t,a_t,x_{t+1})] = r(x,a,x')$, β_t is a random sample of discount factor that $\mathbb{E}[\beta_t|(x,a,x') = (x_t,a_t,x_{t+1})] = \beta(x,a,x')$, and x_{t+1} is a random successor state generated by $P(x_t,a_t,\cdot)$. In each iteration of Q-learning, Q_t is updated, with learning rate α_t, by the observed reward r_t and optimal Q-value at observed state x_{t+1} multiplied by the discount factor β_t.

3.2.4. SARSA. We follow the terminology in Singh et al. (2000). While Q-learning updates Q-factors greedily by selecting the action with the maximum Q-value for the next state, denoted as $\max_{b \in \Gamma(x_{t+1})} Q_t(x_{t+1},b)$, *state–action–reward–state–action* (SARSA) algorithm adheres to a specified policy for updating Q-factors. The SARSA update rule is given by:

$$Q_{t+1}(x_t,a_t) = [1-\alpha_t(x_t,a_t)]Q_t(x_t,a_t) + \alpha_t(x_t,a_t)\left[r_t + \beta_t Q_t(x_{t+1},a_{t+1})\right], \tag{3.6}$$

where a_{t+1} is determined by a learning policy, and $\alpha_t(x,a) \in [0,1]$ with $\alpha_t(x,a) = 0$ for $(x,a) \neq (x_t,a_t)$. Thus, SARSA is categorised as an on-policy algorithm, and its convergence is contingent upon the chosen learning policy.

A *learning policy* π is a set of probabilities $\Pr(\cdot|x,t,Q,n_t(x))$ such that action $a \in \mathsf{A}$ is selected with probability $\Pr(a|x,t,Q,n_t(x))$, given the state x, time step t, current estimate Q of the (optimal) Q-value, and the number $n_t(x)$ of visits to state x up to time t. A learning policy is *greedy* if it consistently selects the action with the highest current Q-value.

Furthermore, a learning policy π is a *greedy-in-the-limit-with-infinite-exploration* (GLIE) learning policy if it satisfies the following criteria:

(a) Infinite Visits: Each state is visited infinitely often, and each action is executed infinitely often in every state.

(b) Greedy in the Limit: As $t \to \infty$, the learning policy is greedy with respect to the Q-value function with probability 1.

An example of a GLIE learning policy is *ε-greedy* exploration: at time step t in state x, selects the greedy action with probability $1-\varepsilon_t(x)$, and selects a random exploration action with probability $\varepsilon_t(x)$, where $\varepsilon_t(x) \coloneqq c/n_t(x)$ for $0 < c < 1$.

A learning policy is *restricted rank-based randomized (RRR)* if it selects actions probabilistically according to the ranks of Q-values: $\Pr(a|x,t,Q) = P^R(\rho(Q,x,a))$, where $\rho(Q,x,a)$ represents the rank of action a in state x based on its action value $Q(x,a)$, and $P^R\colon \{1,\dots,|A|\} \to \mathbb{R}$ maps action ranks to probabilities such that $P^R(1) \geqslant P^R(2) \geqslant \cdots \geqslant P^R(|A|)$ and $\sum_{i=1}^{|A|} P^R(i) = 1$. For example, Q-learning is a SARSA iteration with an RRR learning policy where $P^R(1) = 1$ and $P^R(i) = 0$ for $i = 2,\dots,|A|$.

We say that Q_t is generated by SARSA iteration ((3.6)) with a GLIE (RRR) learning policy π if a_{t+1} is selected by a GLIE (RRR) learning policy with $Q = Q_t$ in $\Pr(\cdot|x,t,Q,n_t(x))$.

The key difference between GLIE and RRR learning policies is that a GLIE learning policy has a decaying exploration such that the learning policy converges to the greedy policy over time, while an RRR learning policy maintains persistent exploration. Therefore, Q_t generated by SARSA with a GLIE learning policy will converge to optimal Q-value Q^*, while Q_t may not converge to Q^* if the SARSA iteration follows an RRR learning policy.

Given a learning policy with ranking by Q-value, $\rho(Q, x, a)$, and ranking probability, P^R, let $\bar{Q}$ be the corresponding Q-factor satisfying

$$\bar{Q}(x,a) = r(x,a) + \sum_{x' \in \mathsf{X}} P(x,a,x')\beta(x,a,x') \sum_{a' \in \mathsf{A}} P^R(\rho(\bar{Q}, x', a'))\bar{Q}(x', a')$$

for $(x, a) \in G$. Given constant discounting $\beta(x,a,x') \equiv \beta$, Singh et al. (2000) prove that Q_t value computed by SARSA with a GLIE learning policy converges to optimal Q^*, and Q_t updated with an RRR learning policy converges to the Q-factor function $\bar{Q}$ corresponding to that learning policy.

On the other hand, the ranking policy can depend on (x, a) pair only without referencing Q-factor. Let $\Pi \coloneqq \{f \colon \mathsf{A} \to \{1, 2, \dots, |A|\} \colon f \text{ is a bijection}\}$ denote the set of permutations of A, where each $f \in \Pi$ ranks actions. A *restricted policy* $\bar{\pi} \colon \mathsf{X} \to \Pi$ ranks actions in each state without action value Q. That is, $\bar{\pi}(x, a) \coloneqq \bar{\pi}(x)(a)$ is the assigned rank of action a in state x. Given ranking probability P^R, $P^R(\bar{\pi}(x, a))$ is the probability that the policy selects action a in state x under restricted policy $\bar{\pi}$ that ranks a. The Q-factor of a restricted policy $\bar{\pi}$, denoted $\bar{Q}^{\pi}$, satisfies

$$\bar{Q}^{\bar{\pi}}(x,a) = r(x,a) + \sum_{x' \in \mathsf{X}} P(x,a,x')\beta(x,a,x') \sum_{a' \in \mathsf{A}} P^R(\bar{\pi}(x', a'))\bar{Q}^{\bar{\pi}}(x', a')$$

for $(x, a) \in G$. We say a restricted policy $\bar{\pi}$ is *optimal* under probabilities of ranks P^R if

$$\bar{\pi} \in \operatorname*{arg\,max}_{\pi \in \Pi} \bar{Q}^{\pi}.$$

Also, the *greedy restricted policy* for a Q-value function Q is $\bar{\pi}(x, a) = \rho(Q, x, a)$, which ranks actions with their corresponding Q-values. Singh et al. (2000) shows that the greedy restricted policy with respect to $\bar{Q}$ (estimated by SARSA with an RRR learning policy) is an optimal restricted policy.

3.2.5. Double Q-learning. Double Q-learning is designed to address the overestimation of Q-values that occurs in standard Q-learning. Given random variables $X_1, \ldots, X_n$, since $\mathbb{E}(\max_i\{X_i\}) \geqslant \max_i \mathbb{E}(X_i)$ by Jensen inequality, Q-learning is known to overestimate optimal values during experiments. That is, $\max_b Q_t(x_{t+1}, b)$ is an estimate for $\mathbb{E}\{\max_b Q_t(x_{t+1}, b)\}$, rather than for $\max_b \mathbb{E}[Q_t(x_{t+1}, b)]$, which is the desired value. Here, the expectation is the average over all possible runs of the same experiments, and Q_t iteration is a sample mean that approximates Q^*. To avoid overestimation, Hasselt (2010) introduces double Q-learning, as outlined in Algorithm 1.

There are two Q-functions in double Q-learning: Q^A and Q^B. Each Q-function, say Q^A, is updated with another Q-value, Q^B, with Q^A-greedy action a^*. Since Q^A and Q^B update with different sets of experiment samples, Q^B is an unbiased estimate for Q-value at Q^A-greedy action. Hasselt (2010) illustrates that since $\mathbb{E}[Q^B(x', a^*)] \leqslant \max_a \mathbb{E}[Q^A(x', a)]$, double Q-learning may actually underestimate the action value.

Hasselt (2010) shows the convergence of double Q-learning assuming a constant discount factor $\beta < 1$. As illustrated in Algorithm 1, we assume that β is a random variable observed at the time that the Q-functions are updated. Moreover, we assume that β is governed by a parameterised function such that there exists eventual discounting.

Algorithm 1: Double Q-learning

```
Initialise $Q^A, Q^B, x$
repeat
    Choose $a$, based on $Q^A(x,\cdot)$ and $Q^B(x,\cdot)$. Observe $r$, $\beta$, $x'$
    Choose (e.g. random) either UPDATE(A) or UPDATE(B)
    if UPDATE(A) then
        Define $a^* = \arg\max_a Q^A(x', a)$
        $Q^A(x,a) \leftarrow Q^A(x,a) + \alpha(x,a)\big(r + \beta Q^B(x', a^*) - Q^A(x,a)\big)$
    else if UPDATE(B) then
        Define $b^* = \arg\max_a Q^B(x', a)$
        $Q^B(x,a) \leftarrow Q^B(x,a) + \alpha(x,a)\big(r + \beta Q^A(x', b^*) - Q^B(x,a)\big)$
    end
    $x \leftarrow x'$
until end
```

3.3. Main Results

This section presents the assumptions and main results for the convergence of Q-learning, SARSA, and double Q-learning with state-action-dependent discount factors.

3.3.1. Assumptions and Convergences. We first introduce the standard Robbins-Monro conditions for the Stochastic Approximation algorithm.

Let $\mathscr{F}_t$ be the information field up to time t:

$$\mathscr{F}_t = \sigma\{x_0, \dots, x_t, a_0, \dots, a_t, \alpha_0, \dots, \alpha_t, Q_0, \dots, Q_t, r_0, \dots, r_{t-1}, \beta_0, \dots, \beta_{t-1}\} \tag{3.7}$$

Let Ω be the sample space of all possible trajectories of $\{(x_t, a_t, r_t, \beta_t)\}_{t\in\mathbb{N}}$ and $\mathscr{F} = \bigotimes_{t\in\mathbb{N}_0} \mathscr{F}_t$. Let $\mathbb{P}$ be the probability measure on $(\Omega, \mathscr{F})$. For trajectory $\omega \in \Omega$ and $(x, a) \in \mathsf{G}$, let $\mathcal{T}_{x,a}(\omega) \subset \mathbb{N}$ be the set of times at which an update of $Q_t(x, a)$ is performed.

Assumption 3.3.1. The following conditions hold:

(a) X and A are finite;

(b) the stepsizes $\{\alpha_t\}_{t\in\mathbb{N}_0}$ is a sequence of random variables on $(\Omega, \mathbb{P}, \mathscr{F})$ such that $\alpha_t(x, a) \in [0, 1]$, $\alpha_t(x, a) = 0$ for $t \notin \mathcal{T}_{x,a}(\omega)$ and

$$\sum_{t\in\mathcal{T}_{x,a}(\omega)} \alpha_t(x, a) = \infty, \qquad \sum_{t\in\mathcal{T}_{x,a}(\omega)} \alpha_t^2(x, a) < \infty$$

for all $(x, a) \in \mathsf{G}$ and $\mathbb{P}$-almost all $\omega \in \Omega$; and

Assumption 3.3.1 implies that each state-action pair will be visited infinitely many times by temporal-difference learning, and $\alpha_t(x, a) \to 0$ for each $(x, a) \in \mathsf{G}$. Next, we outline the assumptions associated with state-action-dependent discounting. Given a $|\mathsf{X}| \times |\mathsf{X}|$ matrix L let $\rho(L) := \max\{|\lambda| \colon \lambda \text{ is an eigenvalue of } L\}$ be the spectral radius.

Assumption 3.3.2 (eventual-discounting). For any $\sigma \in \Sigma$, $\rho(L_\sigma) < 1$, where $L_\sigma(x, x') := \beta(x, \sigma(x), x')P(x, \sigma(x), x')$ for all $(x, x') \in \mathsf{X} \times \mathsf{X}$.

Assumption 3.3.3 (eventual-discounting). There is a non-negative $|\mathsf{X}| \times |\mathsf{X}|$ matrix L such that $\beta(x, a, x')P(x, a, x') \leqslant L(x, x')$ for all $(x, a, x') \in \mathsf{G} \times \mathsf{X}$ and $\rho(L) < 1$.

Assumption 3.3.2 or 3.3.3 implies that the MDP is eventually discounting. Compared to a constant discount factor, Assumption 3.3.2 and 3.3.3 are more general in the sense that discount factors are state-action-dependent and can exceed one under some states and actions. Moreover, Assumption 3.3.3 is sufficient to Assumption 3.3.2. We can show that the policy operator or the Bellman operator is globally stable if either Assumption 3.3.2 or 3.3.3 holds. As discussed in Section 3.2.2, the assumptions imply that there is $n \in \mathbb{N}$ such that

$$\sup_{\sigma \in \Sigma} \sup_{x \in \mathsf{X}} \mathbb{E}_x^\sigma \prod_{t=0}^{n-1} \beta_t < 1$$

where $\beta_t = \beta(x_t, \sigma(x_t), x_{t+1})$ and $\mathbb{E}_x^\sigma$ denote the expectation conditioning on $x_0 = x$ and the transition probability follows $P(x, \sigma(x), x')$ for all $(x, x') \in \mathsf{X} \times \mathsf{X}$.

The convergence results of Q-learning, SARSA, and double Q-learning are presented below with the above assumptions of eventual discounting.

PROPOSITION 3.3.1. *If Assumption 3.3.1 holds, and either Assumption 3.3.2 or 3.3.3 holds, then* $\{Q_t\}_{t \geqslant 0}$ *generated by Q-learning algorithm (3.5) converges to* Q^* *w.p.*1..

PROPOSITION 3.3.2. *If MDP is recurrent, Assumption 3.3.1 is satisfied, and either Assumption 3.3.2 or 3.3.3 holds, then the SARSA iterate* $\{Q_t\}_{t \geqslant 0}$ *, generated by (3.6), with a GLIE learning policy* π *converges to* Q^* *w.p.*1. *and the corresponding learning policy* π_t *converges to the optimal policy* σ^* *w.p.*1..

PROPOSITION 3.3.3. *If MDP is recurrent, Assumption 3.3.1 is satisfied, either Assumption 3.3.2 or 3.3.3 holds, and* $\Pr(a_{t+1} = a | Q_t, x_{t+1}) = P^R(\rho(Q_t, x_{t+1}, a_{t+1}))$*, then the SARSA iterate* $\{Q_t\}_{t \geqslant 0}$ *, generated by (3.6), with an RRR learning policy converges to* $\bar{Q}$ *w.p.*1. *Moreover, the greedy restricted policy is the optimal restricted policy.*

PROPOSITION 3.3.4. *If MDP is recurrent, Assumption 3.3.1 holds, and either Assumption 3.3.2 or 3.3.3 holds, and both* Q^A *and* Q^B *update infinitely often, then both* $\{Q_t^A\}_{t \geqslant 0}$ *and* $\{Q_t^B\}_{t \geqslant 0}$ *converge to optimal Q-value* Q^* *w.p.*1.

3.4. Proofs for Main Results

In this section, we provide the proofs for the propositions in Section 3.3.1.

3.4.1. Preliminaries. Fix $n \in \mathbb{N}$. For any positive vector $w \in \mathbb{R}^n$, define the weighted maximum norm $\| \cdot \|_w$ by

$$\|v\|_w := \max_x \frac{|v(x)|}{w(x)} \qquad (v \in \mathbb{R}^n).$$

Denote $\| \cdot \| := \| \cdot \|_{\mathbb{1}}$, where $\mathbb{1} \in \mathbb{R}^n$ is a vector of ones. A self-map F on $U \subset \mathbb{R}^n$ is *globally stable* if F has a unique fixed point u^* and $F^k u \to u^*$ for all $u \in U$.

3.4.2. Proofs of Proposition 3.3.1. The proof is completed by the following lemmas. To apply the Stochastic Approximation, we first demonstrate that the Q-factor Bellman operator is contractive under some weighted norm. Define the Q-factor Bellman operator $H \colon \mathbb{R}^{\mathsf{G}} \to \mathbb{R}^{\mathsf{G}}$ by

$$\begin{aligned} HQ(x,a) &:= \mathbb{E}_{(x,a)} r(x,a,X') + \mathbb{E}_{(x,a)} \left[\beta(x,a,X') \max_{b \in \Gamma(X')} Q(X',b) \right] \\ &= B\left(x, a, \max_b Q(\cdot, b)\right) \qquad ((x,a) \in \mathsf{G}, Q \in \mathbb{R}^{\mathsf{G}}) \end{aligned}$$

We can verify that Q^* is the fixed point of H. To simplify the notation, we let $\mathcal{M}q(x) := \max_a q(x,a)$ for all $x \in \mathsf{X}$ and all $q \in \mathbb{R}^{\mathsf{G}}$ in the following lemmas.

Lemma 3.4.1. *If Assumption 3.3.2 holds, then there exist a positive vector $\varphi \in \mathbb{R}^{\mathsf{G}}$ and $\gamma < 1$ such that $\|HQ - HQ'\|_\varphi \leqslant \gamma \|Q - Q'\|_\varphi$ for all $Q, Q' \in \mathbb{R}^{\mathsf{G}}$.*

Proof. Let Assumption 3.3.2 hold. Proposition 3.2.1 implies that v^* is the unique fixed point of T. Define $Q^* \in \mathbb{R}^{\mathsf{G}}$ by

$$Q^*(x,a) = r(x,a) + \mathbb{E}_{x,a} \beta(x,a,X') v^*(X') \quad ((x,a) \in \mathsf{G}.)$$

Since $v^* = Tv^*$, we have

$$v^*(x) = \max_{a \in \Gamma(x')} \{ r(x,a) + \mathbb{E}_{x,a} \beta(x,a,X') v^*(X') \} = \max_{a \in \Gamma(x)} Q^*(x,a),$$

for all $x \in \mathsf{X}$. It implies

$$\begin{aligned} HQ^*(x,a) &= r(x,a) + \mathbb{E}_{x,a} \left[\beta(x,a,X') \max_{a' \in \Gamma(X')} Q^*(X',a') \right] \\ &= r(x,a) + \mathbb{E}_{x,a} \beta(x,a,X') v^*(X') = Q^*(x,a) \end{aligned}$$

for all $(x, a) \in \mathsf{G}$. Hence, Q^* is the fixed point of H. Next, we construct the weighted vector φ by considering the constant reward $\hat{r}(x, a, x') \equiv -1$ for all $(x, a, x') \in \mathsf{G} \times \mathsf{X}$. Let $\hat{v}$ be the fixed point of T with reward $\hat{r}$ and $\hat{Q}(x, a) = \mathbb{E}_{x,a}[\hat{r}(x, a, X') + \beta(x, a, X')\hat{v}(X')]$ for $(x, a) \in \mathsf{G}$. Then, $\hat{Q}$ is the fixed point of H when $\hat{r} = -\mathbb{1}$, so we have[3]

$$\begin{aligned}
\hat{Q}(x, a) = H\hat{Q}(x, a) &= \sum_{x'} P(x, a, x')[\hat{r}(x, a, x') + \beta(x, a, x') \max_{a'} \hat{Q}(x', a')] \\
&= -1 + \sum_{x'} P(x, a, x')\beta(x, a, x')\mathcal{M}\hat{Q}(x') \\
&\leqslant -1 + \max_{a \in \Gamma(x)} \sum_{x'} P(x, a, x')\beta(x, a, x')\mathcal{M}\hat{Q}(x') \qquad ((x, a) \in \mathsf{G})
\end{aligned}$$

Taking the maximum over A on the left, we obtain

$$\mathcal{M}\hat{Q}(x) \leqslant -1 + \max_{a \in \Gamma(x)} \sum_{x'} P(x, a, x')\beta(x, a, x')\mathcal{M}\hat{Q}(x') \qquad (x \in \mathsf{X}.) \tag{3.8}$$

Observe that $\hat{v} = \mathcal{M}\hat{Q}$ is the optimal value, and there exists an optimal policy $\hat{\sigma} \in \Sigma$ such that $\mathcal{M}\hat{Q} = v_{\hat{\sigma}}$. Since $\rho(L_{\hat{\sigma}}) < 1$, we have $L_{\hat{\sigma}}^n \mathbb{1} \to 0$ as $n \to \infty$, and $T_{\hat{\sigma}}$ is globally stable. Then, the iteration of $v_{\hat{\sigma}} = T_{\hat{\sigma}}^n v_{\hat{\sigma}} = (I + L_{\hat{\sigma}} + \cdots + L_{\hat{\sigma}}^n)(-\mathbb{1})$ for all $n \in \mathbb{N}$ converges and yields

$$\mathcal{M}\hat{Q}(x) = v_{\hat{\sigma}}(x) = \lim_{n \to \infty} \mathbb{E}_x^{\hat{\sigma}} \sum_{t=0}^{n} \prod_{i=0}^{t-1} \beta(X_i, \hat{\sigma}(X_i), X_{i+1})(-1) \leqslant -1$$

for all $x \in \mathsf{X}$. Hence, we have $\hat{v} \leqslant -\mathbb{1}$. Let $\varphi = -\mathcal{M}\hat{Q} \geqslant \mathbb{1}$. Hence, (3.8) implies

$$\max_{a \in \Gamma(x)} \sum_{x'} P(x, a, x')\beta(x, a, x')\varphi(x') \leqslant \varphi(x) - 1 \leqslant \varphi(x) \max_{y} \frac{\varphi(y) - 1}{\varphi(y)} = \gamma\varphi(x),$$

---

[3] Denote $\mathbb{1} \in \mathbb{R}^{\mathsf{X}}$ as the vector of ones.

where $\gamma := \max_x\{(\varphi(x)-1)/\varphi(x)\} < 1$. Now, for all $Q, R \in \mathbb{R}^{\mathsf{G}}$, we obtain

$$\begin{aligned}
|HQ(x,a) - HR(x,a)| &= \left|\sum_{x'} \beta(x,a,x')P(x,a,x')(\max_{a'} Q(x',a') - \max_{a'} R(x',a'))\right| \\
&\leqslant \sum_{x'} \beta(x,a,x')P(x,a,x') \left|\max_{a'} Q(x',a') - \max_{a'} R(x',a')\right| \\
&\leqslant \sum_{x'} \beta(x,a,x')P(x,a,x') \max_{a'} |Q(x',a') - R(x',a')| \\
&\leqslant \max_{a\in\Gamma(x)} \sum_{x'} \beta(x,a,x')P(x,a,x') \max_{a'} |Q(x',a') - R(x',a')| \qquad (3.9)\\
&\leqslant \max_{a\in\Gamma(x)} \sum_{x'} \beta(x,a,x')P(x,a,x')\varphi(x') \max_{y} \frac{\mathcal{M}(|Q-R|)(y)}{\varphi(y)} \\
&= \|Q-R\|_{\varphi} \max_{a\in\Gamma(x)} \sum_{x'} \beta(x,a,x')P(x,a,x')\varphi(x') \\
&\leqslant \|Q-R\|_{\varphi}\gamma\varphi(x)
\end{aligned}$$

for all $(x,a) \in \mathsf{G}$. Dividing $\varphi(x)$ and taking supremum over G, we obtain the contraction $\|HQ - HR\|_{\varphi} \leqslant \gamma\|Q-R\|_{\varphi}$ for all $Q, R \in \mathbb{R}^{\mathsf{G}}$. □

To this end, Lemma 3.4.1 implies that H is globally stable and has a unique fixed point if Assumption 3.3.2 holds. Since Assumption 3.3.3 implies Assumption 3.3.2, H is also contractive on some weighted supremum norm under Assumption 3.3.3. Alternatively, we can show that H is eventually contracting by leveraging the properties of L, which bounds the expected discount factors.

Lemma 3.4.2. *If Assumption 3.3.3 holds, then operator H has a unique fixed point and is globally stable.*

Proof. Suppose that Assumption 3.3.3 holds. We show that there exists $k \in \mathbb{N}$ such that H^k is a contraction map. Let $Q, R \in \mathbb{R}^{\mathsf{G}}$. Since Assumption 3.3.3 holds, we obtain

$$\begin{aligned}|HQ(x,a) - HR(x,a)| &= |B(x,a,\max_{a'} Q(\cdot,a')) - B(x,a,\max_{a'} R(\cdot,a'))| \\ &\leqslant \sum_{x' \in \mathsf{X}} \beta(x,a,x')P(x,a,x') \left|\max_{a'} Q(x',a') - \max_{a'} R(x',a')\right| \\ &\leqslant \sum_{x' \in \mathsf{X}} L(x,x') \left|\max_{a'} Q(x',a') - \max_{a'} R(x',a')\right| \\ &\leqslant \sum_{x' \in \mathsf{X}} L(x,x') \max_{a'} |Q(x',a') - R(x',a')|\end{aligned}$$

for all $(x,a) \in \mathsf{G}$. Taking the maximum over A on the left, we obtain

$$\max_a |HQ(x,a) - HR(x,a)| \leqslant \sum_{x' \in \mathsf{X}} L(x,x') \max_{a'} |Q(x',a') - R(x',a')|$$

for all $(x,a) \in \mathsf{G}$. Hence, we have $\mathcal{M}(|HQ - HR|) \leqslant L\mathcal{M}(|Q - R|)$. Observe that $\mathcal{M}|H^2Q - H^2R| \leqslant L\mathcal{M}(|HQ - HR|) \leqslant LL\mathcal{M}(|Q - R|) = L^2\mathcal{M}(|Q - R|)$. Hence, induction yields

$$\begin{aligned}\mathcal{M}(|H^jQ - H^jR|) &\leqslant L^j\mathcal{M}(|Q - R|) \\ &\leqslant \|L^j\|\|Q - R\|.\end{aligned}$$

for $j \in \mathbb{N}$. Taking the maximum over X on the left, we obtain $\|H^jQ - H^jR\| \leqslant \|L^j\|\|Q - R\|$ for $j \in \mathbb{N}$. Since $\rho(L) = \lim_{n\to\infty} \|L^n\|^{1/n} < 1$, the Gelfand's formula implies that there exists $k \in \mathbb{N}$ such that $\|L^k\| < 1$. To this end, H^k is a contraction map. The Banach Contraction Mapping Theorem implies that H^k has a unique fixed point Q^* in $\mathbb{R}^{\mathsf{G}}$ and $Q^* = \lim_{j\to\infty} H^{jk}Q$ for any $Q \in \mathbb{R}^{\mathsf{G}}$. To see that Q^* is the unique fixed point of H, since H^k is globally stable, we can fix $\varepsilon > 0$ and choose $j > 0$ such that

$$\|H^{jk}(HQ^*) - Q^*\| < \varepsilon.$$

This implies that $\|HH^{jk}Q^* - Q^*\| = \|HQ^* - Q^*\| < \varepsilon$. Since this holds for all $\varepsilon > 0$, we obtain $\|HQ^* - Q^*\| = 0$ so that Q^* is also a fixed point of H. Finally, by the argument

$$\lim_{m\to\infty} H^mQ = \lim_{\substack{s\to\infty \\ m=sk+t;s,t\in\mathbb{N}}} H^{sk+t}Q = \lim_{\substack{s\to\infty \\ m=sk+t;s,t\in\mathbb{N}}} H^{sk}(H^tQ) = Q^*,$$

H is globally stable, and Q^* is the unique fixed point to H. □

Lemma 3.4.3. *If Assumption 3.3.3 holds, then there exist a positive vector* $\varphi \in \mathbb{R}^{\mathsf{G}}$ *and* $\gamma < 1$ *such that* $\|HQ - HR\|_\varphi \leqslant \gamma \|Q - R\|_\varphi$ *for all* $Q, R \in \mathbb{R}^{\mathsf{G}}$.

Proof. Suppose Assumption 3.3.3 holds. Then, Lemma 3.4.2 implies that H has a unique fixed point. We construct the weighted vector φ by considering the reward $\hat{r}(x, a, x') \equiv -1$ for all $(x, a, x') \in \mathsf{G} \times \mathsf{X}$. Since there is a unique fixed point $\hat{Q} \in \mathbb{R}^{\mathsf{G}}$ when $\hat{r} = -\mathbb{1}$, we have

$$\begin{aligned}
\hat{Q}(x, a) = H\hat{Q}(x, a) &= \sum_{x'} P(x, a, x')[\hat{r}(x, a, x') + \beta(x, a, x') \max_{a'} \hat{Q}(x', a')] \\
&= -1 + \sum_{x'} P(x, a, x')\beta(x, a, x') \max_{a'} \hat{Q}(x', a') \qquad (3.10) \\
&\leqslant -1 + \sum_{x'} L(x, x')\mathcal{M}\hat{Q}(x') \qquad ((x, a) \in \mathsf{G})
\end{aligned}$$

Taking the maximum over A on the left, we obtain

$$\mathcal{M}\hat{Q}(x) \leqslant -\mathbb{1} + \sum_{x'} L(x, x')\mathcal{M}\hat{Q}(x') \qquad (x \in \mathsf{X}.)$$

Then, since $\rho(L) < 1$ and L is non-negative, we have

$$\mathcal{M}\hat{Q} \leqslant (I - L)^{-1}(-\mathbb{1}) = (I + L + L^2 + \dots)(-\mathbb{1}) \leqslant -\mathbb{1}.$$

Let $\hat{\varphi} = -\mathcal{M}\hat{Q} \geqslant \mathbb{1}$. Hence, (3.10) implies

$$\sum_{x'} L(x, x')\varphi(x') \leqslant \varphi(x) - 1 \leqslant \varphi(x) \max_y \frac{\varphi(y) - 1}{\varphi(y)} = \gamma\varphi(x),$$

where $\gamma := \max_x\{(\varphi(x)-1)/\varphi(x)\} < 1$. Now, for all $Q, R \in \mathbb{R}^{\mathsf{G}}$, we obtain

$$\begin{aligned}|HQ(x,a) - HR(x,a)| &= \left|\sum_{x'} \beta(x,a,x')P(x,a,x'))(\max_{a'} Q(x',a') - \max_{a'} R(x',a'))\right| \\ &\leqslant \sum_{x'} L(x,x')\mathcal{M}|Q-R|(x') \\ &\leqslant \left(\sum_{x'} L(x,x')\varphi(x')\right)\left(\max_y \frac{\mathcal{M}|Q-R|(y)}{\varphi(y)}\right) \\ &\leqslant \gamma\varphi(x)\max_y \frac{\mathcal{M}|Q-R|(y)}{\varphi(y)} \\ &= \gamma\varphi(x)\|Q-R\|_\varphi \qquad ((x,a)\in\mathsf{G})\end{aligned}$$

for all $(x,a)\in\mathsf{G}$. Dividing $\varphi(x)$ on both sides and taking the maximum on the left, we obtain the contraction $\|HQ - HR\|_\varphi \leqslant \gamma\|Q-R\|_\varphi$ for all $Q, R \in \mathbb{R}^{\mathsf{G}}$. □

The intuition behind contraction in the weighted supremum norm for Q-learning is discussed as follows. Rewriting Q-learning iteration (3.5), we have

$$Q_{t+1}(x_t,a_t) = [1-\alpha_t(x_t,a_t)]Q_t(x_t,a_t) + \alpha_t(x_t,a_t)\left[HQ_t(x_t,a_t) + w_t(x_t,a_t)\right],$$

where

$$w_t(x,a) = r_t - \mathbb{E}_{(x,a)}r_t + \max_{b\in\Gamma(x_{t+1})} Q_t(x_{t+1},b) - \mathbb{E}_{(x,a)}\left[\max_{b\in\Gamma(X')} Q_t(X',b)\right].$$

It is known that if H is a contraction map, then Q_t converges to the fixed point of H, i.e., Q^* (Tsitsiklis, 1994). Now, observe that the Q-learning iteration (3.5) is equivalent to

$$\begin{aligned}\frac{Q_{t+1}(x_t,a_t)}{\varphi(x_t,a_t)} &= [1-\alpha_t(x_t,a_t)]\frac{Q_t(x_t,a_t)}{\varphi(x_t,a_t)} \\ &\quad + \alpha_t(x_t,a_t)\left[\frac{H[\varphi(x_t,a_t)\frac{Q_t(x_t,a_t)}{\varphi(x_t,a_t)}]}{\varphi(x_t,a_t)} + \frac{w_t(x_t,a_t)}{\varphi(x_t,a_t)}\right],\end{aligned} \tag{3.11}$$

Therefore, it is equivalent to iterate $q_t(x,a) := Q_t(x,a)/\varphi(x,a)$ with respect to the map $\tilde{H}$ defined by

$$\tilde{H}q := \Phi^{-1}H(\Phi q) \qquad (q\in\mathbb{R}^{\mathsf{G}}).$$

where $\Phi q(x,a) := \varphi(x,a)q(x,a)$ and $\Phi^{-1}q(x,a) = q(x,a)/\varphi(x,a)$ for all $(x,a) \in \mathsf{G}$ and $q \in \mathbb{R}^{\mathsf{G}}$. In the above lemmas, we have $\varphi(x,a) \equiv \varphi(x)$ for all $x \in \mathsf{X}$. If H is a contractive in $\|\cdot\|_\varphi$ norm, then $\tilde{H}$ is a contraction in maximum norm.

We use the following lemma of stochastic approximation to prove convergence of algorithms. The methodology follows Jaakkola et al. (1993), Tsitsiklis (1994), Singh et al. (2000), Melo (2001), and Hasselt (2010).

Lemma 3.4.4 (Singh et al. (2000)). *Let X be a finite set. Consider a stochastic process $\{(\alpha_t, \Delta_t, F_t)\}_{t\geqslant 0}$ such that $\alpha_t, \Delta_t, F_t : X \to \mathbb{R}$ and*

$$\Delta_{t+1}(x) = (1-\alpha_t(x))\Delta_t(x) + \alpha_t(x)F_t(x) \qquad (x \in X)$$

Let $\mathscr{F}_t$ be a sequence of increasing σ-fields such that α_0, Δ_0 are $\mathscr{F}_0$-measurable and α_t, Δ_t and F_{t-1} are $\mathscr{F}_t$-measurable for $t \geqslant 1$. Assume that the following statements hold:

- *(a)* $0 \leqslant \alpha_t(x) \leqslant 1$, $\sum_t \alpha_t(x) = \infty$, $\sum_t \alpha_t^2(x) < \infty$ *w.p.1;*
- *(b)* $\|\mathbb{E}[F_t(\cdot)|\mathscr{F}_t]\|_w \leqslant \kappa\|\Delta_t\|_w + c_t$, *where* $\kappa \in (0,1)$, $w \in \mathbb{R}_+^{\mathsf{X}}$, *and c_t converges to zero w.p.1;*
- *(c)* $\operatorname{Var}[F_t(x)|\mathscr{F}_t] \leqslant K(1+\|\Delta_t\|_w)^2$, *where K is some constant.*

Then, $\Delta_t \to 0$ w.p.1.

Proof of Proposition 3.3.1. Suppose that all assumptions in the statements hold. Let $\Delta_t(x,a) = Q_t(x,a) - Q^*(x,a)$ for all $(x,a) \in \mathsf{G}$ and rearrange (3.5) as

$$\begin{aligned}\Delta_{t+1}(x_t,a_t) =&(1-\alpha_t(x_t,a_t))\Delta_t(x_t,a_t)\\ &\qquad + \alpha_t(x_t,a_t)\left[r_t + \beta_t \max_{b\in\Gamma(x_{t+1})} Q_t(x_{t+1},b) - Q^*(x_t,a_t)\right]\\ =&(1-\alpha_t(x_t,a_t))\Delta_t(x_t,a_t) + \alpha_t(x_t,a_t)F_t(x_t,a_t),\end{aligned}$$

where we write

$$F_t(x_t,a_t) = r_t + \beta_t \max_{b\in\Gamma(x_{t+1})} Q_t(x_{t+1},b) - Q^*(x_t,a_t).$$

We let $F_t(x,a) = 0$ if $(x,a) \neq (x_t, a_t)$. Here, x_{t+1} is a random sample generated from $P(x_t, a_t, \cdot)$ for all $t \in \mathbb{N}_0$. Let $\mathscr{F}_t = \sigma\{Q_0, x_0, a_0, \alpha_0, r_0, \beta_0, \dots, x_t, a_t, \alpha_t, r_{t-1}, \beta_{t-1}\}$ be the σ-field up to time t. Then, Δ_t, α_t and F_{t-1} are $\mathscr{F}_t$-measurable.

To apply Lemma 3.4.4, we need to show (1) $\|\mathbb{E}[F_t(\cdot)|\mathscr{F}_t\|_w \leqslant \kappa\|\Delta_t\|_w + c_t$, for some $\kappa \in (0,1)$, $w \in \mathbb{R}_+^{\mathsf{X}}$, and c_t converges to zero w.p.1, and (2) $\mathrm{Var}[F_t(x)|\mathscr{F}_t] \leqslant K(1+\|\Delta_t\|_w)^2$ for some constant K. From the setup of F_t, we have

$$\begin{aligned}
\mathbb{E}[F_t(x,a)|\mathscr{F}_t] &= \sum_{x' \in \mathsf{X}} P(x,a,x') \left[r(x,a,x') + \beta(x,a,x') \max_{b \in \Gamma(x')} Q_t(x',b) - Q^*(x,a) \right] \\
&= HQ_t(x,a) - Q^*(x,a)
\end{aligned}$$

Since $HQ^* = Q^*$, Lemma 3.4.1 or Lemma 3.4.3 implies

$$\|\mathbb{E}[F_t(\cdot,\cdot)|\mathscr{F}_t]\|_\varphi = \|HQ_t - HQ^*\|_\varphi \leqslant \gamma\|Q_t - Q^*\|_\varphi = \gamma\|\Delta_t\|_\varphi. \tag{3.12}$$

for some $0 < \gamma < 1$ and positive vector φ. Moreover, since X and A are finite, we have $\mathrm{Var}(r_t|\mathscr{F}_t) = \mathrm{Var}(r(x_t, a_t, X')) < \infty$ and $\mathrm{Var}(\beta_t|\mathscr{F}_t) = \mathrm{Var}(\beta(x_t, a_t, X')) < \infty$. It implies that there exists $K \in \mathbb{R}$ such that

$$\begin{aligned}
\mathrm{Var}[F_t(x,a)|\mathscr{F}_t] &= \mathbb{E}\left[\left(r_t + \beta_t \max_{b \in \Gamma(X')} Q_t(X',b) - Q^*(x,a) - (HQ_t(x,a) - Q^*(x,a)) \right)^2 \right] \\
&= \mathbb{E}\left[\left(r_t + \beta_t \max_{b \in \Gamma(X')} Q_t(X',b) - HQ_t(x,a) \right)^2 \right] \\
&= \mathrm{Var}\left[r_t + \beta_t \max_{b \in \Gamma(X')} Q_t(X',b) \right] \\
&\leqslant K(1 + \|\Delta_t\|_\varphi)^2.
\end{aligned}$$

Then, Lemma 3.4.4 shows that Δ_t converges to zero w.p.1, so Q_t converges to Q^* w.p.1. □

3.4.3. Remainging Proofs in Section 3.3.1. For the stability of SARSA, we can show that $\{Q_t\}_{t \geqslant 0}$ computed by SARSA iteration is bounded w.p.1 by Theorem 1 of Tsitsiklis (1994). Alternatively, note that the Q-value from Q-learning is an upper bound for Q values of SARSA. Moreover, consider a Q-learning process with min instead of max in the update rule:

$$Q_{t+1}(x_t,a_t) = (1-\alpha_t(x_t,a_t))Q_t(x_t,a_t) + \alpha_t(x_t,a_t)\left[r_t + \beta_t \min_{b\in\Gamma(x_{t+1})} Q_t(x_{t+1},b)\right]. \tag{3.13}$$

Clearly, Q-values from (3.13) iteration are the lower bounds for the Q-values of SARSA. Since update rule (3.13) is equivalent to the negative Q_t of the Q-learning (3.5) replacing r_t with $-r_t$, it also converges w.p.1.[4]

PROOF OF PROPOSITION 3.3.2. Suppose that MDP is recurrent, Assumption 3.3.1 is satisfied, and either Assumption 3.3.2 or 3.3.3 holds. Let $\Delta_t(x,a) = Q_t(x,a) - Q^*(x,a)$ for all $(x,a)\in \mathsf{G}$. The SARSA iterate becomes

$$\Delta_{t+1}(x_t,a_t) = (1-\alpha_t(x_t,a_t))\Delta_t(x_t,a_t) + \alpha_t(x_t,a_t)F_t(x_t,a_t),$$

where

$$\begin{aligned} F_t(x_t,a_t) =& r_t - Q^*(x_t,a_t) + \beta_t Q_t(x_{t+1},a_{t+1}) \\ =& r_t + \beta_t \max_{b\in\Gamma(x_{t+1})} Q_t(x_{t+1},b) - Q^*(x_t,a_t) + \beta_t\left[Q_t(x_{t+1},a_{t+1}) - \max_{b\in\Gamma(x_{t+1})} Q_t(x_{t+1},b)\right] \\ :=& F_t^Q(x_t,a_t) + \beta_t\left[Q_t(x_{t+1},a_{t+1}) - \max_{b\in\Gamma(x_{t+1})} Q_t(x_{t+1},b)\right] \\ :=& F_t^Q(x_t,a_t) + C_t(x_{t+1},a_{t+1}). \end{aligned}$$

Define $F_t(x,a) = F_t^Q(x,a) = C_t(x,a) = 0$ if $(x,a)\neq(x_t,a_t)$. Let

$$\mathscr{F}_t = \sigma\{Q_0,x_0,a_0,\alpha_0,r_0,\beta_0,\ldots,x_t,a_t,\alpha_t,r_{t-1},\beta_{t-1}\}$$

be the σ-filed up to time t. Then, Δ_t, α_t and F_{t-1} are $\mathscr{F}_t$-measurable. Since $\mathbb{E}[F_t^Q(x,a)|\mathscr{F}_t] = HQ_t(x,a) - Q^*(x,a)$, it follows from (3.12) that

$$\|\mathbb{E}[F_t^Q(\cdot,\cdot)|\mathscr{F}_t]\|_\varphi \leqslant \gamma\|\Delta_t\|_\varphi,$$

---

[4] We can also prove the convergence of the iteration (3.13) by the same arguments in the proofs of Q-learning.

where $\gamma < 1$ and φ are obtained from Lemma 3.4.1 or Lemma 3.4.3. Therefore,

$$\begin{aligned}\|\mathbb{E}[F_t(\cdot,\cdot)|\mathscr{F}_t]\|_\varphi &\leqslant \|\mathbb{E}[F_t^Q(\cdot,\cdot)|\mathscr{F}_t]\|_\varphi + \|\mathbb{E}[C_t(\cdot,\cdot)|\mathscr{F}_t]\|_\varphi \\ &\leqslant \gamma\|\Delta_t\|_\varphi + \|\mathbb{E}[C_t(\cdot,\cdot)|\mathscr{F}_t]\|_\varphi\end{aligned}$$

Now, $\|\mathbb{E}[C_t(\cdot,\cdot)|\mathscr{F}_t]\|_\varphi$ converges to zero since $Q_t(x,a)$ stays bounded as discussed at the beginning of this subsection, and a GLIE learning policy converges to the optimal policy, whence C_t converges to zero w.p.1. Finally, since $\mathrm{Var}(r_t|\mathscr{F}_t) < \infty$ and $\mathrm{Var}(\beta_t|\mathscr{F}_t) < \infty$ by the finiteness of X and A, we have $\mathrm{Var}(F_t|\mathscr{F}_t) \leqslant C(1 + \|\Delta_t\|_\varphi)^2$ for some constant C. Therefore, Lemma 3.4.4 concludes that Δ_t converges to zero w.p.1, and then Q_t converges to Q^* w.p.1. □

PROOF OF PROPOSITION 3.3.3. Let the conditions of the statements hold. Let $\Delta_t(x,a) \coloneqq Q_t(x,a) - \bar{Q}(x,a)$ for $(x,a) \in \mathsf{G}$, where $\bar{Q}$ is the Q-factor corresponding to the RRR learning policy. Rewriting (3.6), we have

$$\Delta_{t+1}(x_t,a_t) = (1 - \alpha_t(x_t,a_t))\Delta_t(x_t,a_t) + \alpha_t(x_t,a_t)F_t(x_t,a_t)$$

where

$$F_t(x_t,a_t) \coloneqq r_t + \beta_t Q_t(x_{t+1},a_{t+1}) - \bar{Q}(x_t,a_t)$$

for $(x,a) \in \mathsf{G}$. Define $F_t(x,a) = 0$ if $(x,a) \neq (x_t,a_t)$ and let

$$\mathscr{F}_t = \sigma\{Q_0, x_0, a_0, \alpha_0, r_0, \beta_0, \ldots, x_t, a_t, \alpha_t, r_{t-1}, \beta_{t-1}\}.$$

Recall that $\bar{Q}$ function is

$$\bar{Q}(x,a) = r(x,a) + \sum_{x'\in\mathsf{X}} P(x,a,x')\beta(x,a,x') \sum_{a'\in\mathsf{A}} P^R(\rho(\bar{Q},x',a'))\bar{Q}(x',a') \quad (x,a) \in \mathsf{G}$$

Define operator $S\colon \mathbb{R}^{\mathsf{G}} \to \mathbb{R}^{\mathsf{G}}$ by $SQ(x,a) = \sum_{a\in\mathsf{A}} P^R(\rho(Q,x,a))Q(x,a)$ for $(x,a) \in \mathsf{G}$ and $Q \in \mathbb{R}^{\mathsf{G}}$. Note that for any $Q, Q' \in \mathbb{R}^{\mathsf{G}}$ we have

$$|SQ(x,a) - SQ'(x,a)| \leqslant \max_a |Q(x,a) - Q'(x,a)| \qquad ((x,a) \in \mathsf{G}).^{5}$$

---

[5] See Singh et al. (2000) Appendix C.

Suppose that Assumption 3.3.3 holds. Let $\gamma < 1$ and φ be defined as Lemma 3.4.3. Then, since a_{t+1} follows the RRR learning policy and is selected by $P^R(\rho(Q, x, a))$, we have the expectation

$$\begin{aligned}
|\mathbb{E}[F_t(x_t, a_t)|\mathscr{F}_t]| &= |\mathbb{E}[r_t + \beta_t Q_t(x_{t+1}, a_{t+1}) - \bar{Q}(x_t, a_t)|\mathscr{F}_t]| \\
&= \left| \sum_{x' \in \mathsf{X}} P(x_t, a_t, x')\beta(x_t, a_t, x') S Q_t(x', a') - \sum_{x' \in \mathsf{X}} P(x_t, a_t, x')\beta(x_t, a_t, x') S\bar{Q}(x', a') \right| \\
&\leqslant \sum_{x' \in \mathsf{X}} P(x_t, a_t, x')\beta(x_t, a_t, x') \left| S Q_t(x', a') - S\bar{Q}(x', a') \right| \\
&\leqslant \sum_{x' \in \mathsf{X}} P(x_t, a_t, x')\beta(x_t, a_t, x') \max_{a'} |Q_t(x', a') - \bar{Q}(x', a')| \\
&\leqslant \sum_{x' \in \mathsf{X}} L(x_t, x') \max_{a'} |Q_t(x', a') - \bar{Q}(x', a')| \\
&\leqslant \sum_{x' \in \mathsf{X}} L(x_t, x')\varphi(x') \|Q_t - \bar{Q}\|_\varphi \leqslant \gamma\varphi(x_t)\|Q_t - \bar{Q}\|_\varphi
\end{aligned}$$

where the second equality follows from $\mathbb{E}(r_t|\mathscr{F}_t) = r(x_t, a_t)$ and $\Pr(a_{t+1} = a|Q_t, x_{t+1}) = P^R(\rho(Q_t, x_t, a))$, the third inequality follows from the assumption of eventual discounting, and the last two inequalities follow Assumption 3.3.3 and Lemma 3.4.3. On the other hand, suppose that Assumption 3.3.2 holds. Let γ and φ be defined as Lemma 3.4.1. Similar to the inequality (3.9) in Lemma 3.4.1, we have

$$\begin{aligned}
|\mathbb{E}[F_t(x_t, a_t)|\mathscr{F}_t]| &= |\mathbb{E}[r_t + \beta_t Q_t(x_{t+1}, a_{t+1}) - \bar{Q}(x_t, a_t)|\mathscr{F}_t]| \\
&\leqslant \sum_{x' \in \mathsf{X}} P(x_t, a_t, x')\beta(x_t, a_t, x') \left| S Q_t(x', a') - S\bar{Q}(x', a') \right| \\
&\leqslant \max_a \sum_{x' \in \mathsf{X}} P(x_t, a_t, x')\beta(x_t, a_t, x') \max_{a'} |Q_t(x', a') - \bar{Q}(x', a')| \\
&\leqslant \max_a \sum_{x' \in \mathsf{X}} P(x_t, a_t, x')\beta(x_t, a_t, x')\varphi(x') \|Q_t - \bar{Q}\|_\varphi \\
&\leqslant \gamma\varphi(x_t)\|Q_t - \bar{Q}\|_\varphi
\end{aligned}$$

Dividing $\varphi(x_t)$ and taking supremum to either one of the above inequalities, we have $\|\mathbb{E}[F_t(\cdot, \cdot)|\mathscr{F}_t]\|_\varphi \leqslant \gamma\|Q_t - \bar{Q}\|_\varphi$. Now, Lemma 3.4.4 shows that Δ_t converges to zero w.p.1, so Q_t converges to $\bar{Q}$ w.p.1. The statement that the greedy restricted policy is optimal under P^R ranking strategy follows from Theorem 3 of Singh et al. (2000), where we consider the discounting function $\beta(x, a, x')$ instead of a constant β. $\square$

PROOF OF PROPOSITION 3.3.4. Suppose that all conditions of the statement are satisfied. By symmetry, it suffices to show that Q^A converges to Q^*. To apply Lemma 3.4.4, let $\Delta_t := Q_t^A - Q^*$. Denote $\alpha_t = \alpha_t(x_t, a_t)$, $a^* := \arg\max_a Q^A(x_{t+1}, a)$ and $b^* := \arg\max_b Q^B(x_{t+1}, b)$. Double Q-learning iteration yields

$$\begin{aligned}\Delta_t(x_t, a_t) &= (1-\alpha_t)\Delta_t(x_t, a_t) + \alpha_t[r_t + \beta_t Q_t^B(x_{t+1}, a^*) - Q^*(x_t, a_t)] \\ &:= (1-\alpha_t)\Delta_t(x_t, a_t) + \alpha_t F_t(x_t, a_t)\end{aligned} \tag{3.14}$$

where

$$\begin{aligned}F_t(x_t, a_t) &:= r_t + \beta_t Q_t^B(x_{t+1}, a^*) - Q^*(x_t, a_t) \\ &= r_t + \beta_t Q_t^A(x_{t+1}, a^*) - Q^*(x_t, a_t) + \beta_t[Q_t^B(x_{t+1}, a^*) - Q_t^A(x_{t+1}, a^*)] \\ &:= F_t^Q(x_t, a_t) + \beta_t[Q_t^B(x_{t+1}, a^*) - Q_t^A(x_{t+1}, a^*)] \\ &:= F_t^Q(x_t, a_t) + C_t(x_t, a_t).\end{aligned}$$

Let $F_t(x, a) = F_t^Q(x, a) = C_t(x, a) = 0$ if $(x, a) \neq (x_t, a_t)$. Recall that $\mathbb{E}[F_t^Q(x_t, a_t)|\mathscr{F}_t] = HQ_t^A(x_t, a_t) - HQ^*(x_t, a_t)$, so we have

$$\|\mathbb{E}[F_t^Q(\cdot, \cdot)|\mathscr{F}_t]\|_\varphi \leqslant \gamma\|\Delta_t\|_\varphi,$$

where $\gamma < 1$ and φ are defined in Lemma 3.4.1 or Lemma 3.4.3. Since $\mathrm{Var}(r_t|\mathscr{F}_t) < \infty$ and $\mathrm{Var}(\beta_t|\mathscr{F}_t) < \infty$, the variance condition of Lemma 3.4.4 is satisfied. Therefore, it suffices to show that $C_t(x_t, a_t) = \beta_t(Q_t^B(x_{t+1}, a^*) - Q_t^A(x_{t+1}, a^*))$ converges to zero w.p.1.

Next, we show that $\Delta_t^{BA} \coloneqq Q_t^B - Q_t^A$ converges to zero w.p.1 by Lemma 3.4.4. Double Q-learning of Algorithm 1 implies that the update of Δ_t follows

if Q^B is updated:

$$\begin{aligned}\Delta_{t+1}^{BA}(x_t,a_t) =&Q_{t+1}^B(x_t,a_t) - Q_{t+1}^A(x_t,a_t)\\ =&(1-\alpha_t)Q_t^B(x_t,a_t) + \alpha_t(r_t + \beta_t Q^A(x_{t+1},b^*)) - Q_t^A(x_t,a_t)\\ =&(1-\alpha_t)\Delta_t^{BA}(x_t,a_t) + \alpha_t[r_t + \beta_t Q_t^A(x_{t+1},b^*) - Q_t^A(x_t,a_t)]\end{aligned}$$

if Q^A is updated:

$$\begin{aligned}\Delta_{t+1}^{BA}(x_t,a_t) =&Q_{t+1}^B(x_t,a_t) - Q_{t+1}^A(x_t,a_t)\\ =&Q_t^B - [(1-\alpha_t)Q_t^A(x_t,a_t) + \alpha_t(r_t + \beta_t Q^B(x_{t+1},a^*))]\\ =&(1-\alpha_t)\Delta_t^{BA}(x_t,a_t) - \alpha_t[r_t + \beta_t Q_t^B(x_{t+1},a^*) - Q_t^B(x_t,a_t)].\end{aligned}$$

Suppose that the probabilities of updating Q^A and Q^B are equal, and the selection of updating Q^A or Q^B is independent of the sample. Then, we have

$$\Delta_{t+1}^{BA}(x_t,a_t) = (1-\alpha_t(x_t,a_t))\Delta_t^{BA}(x_t,a_t) + \alpha_t(x_t,a_t)\tilde{F}_t(x,a), \tag{3.15}$$

where

$$\tilde{F}_t(x,a) = \begin{cases} r_t + \beta_t Q_t^A(x_{t+1},b^*) - Q_t^A(x_t,a_t), & \text{w.p.}1/2,\\ -r_t - \beta_t Q_t^B(x_{t+1},a^*) + Q_t^B(x_t,a_t), & \text{w.p.}1/2.\end{cases}$$

To apply Lemma 3.4.4 to $\{\Delta_t^{BA}\}$ process, we need to show $\|\mathbb{E}[\tilde{F}_t(\cdot,\cdot)|\mathscr{F}_t]\| \leqslant \lambda\|\Delta_t^{BA}\|$ for some $\lambda < 1$. From the definition of $\tilde{F}_t$, we obtain

$$\begin{aligned}\mathbb{E}[\tilde{F}_t(x,a)|\mathscr{F}_t] =& \frac{1}{2}\left[\mathbb{E}(r_t + \beta_t Q_t^A(x_{t+1},b^*)|\mathscr{F}_t) - Q_t^A(x_t,a_t)\right]\\ &+ \frac{1}{2}\left[-\mathbb{E}(r_t + \beta_t Q_t^B(x_{t+1},a^*)|\mathscr{F}_t) + Q_t^B(x_t,a_t)\right]\\ =& \frac{1}{2}\Delta_t^{BA}(x_t,a_t) + \frac{1}{2}\mathbb{E}[\beta_t(Q_t^A(x_{t+1},b^*) - Q_t^B(x_{t+1},a^*))|\mathscr{F}_t].\end{aligned}$$

Assume $\mathbb{E}[\beta_t Q_t^A(x_{t+1},b^*)|\mathscr{F}_t] \geqslant \mathbb{E}[\beta_t Q_t^B(x_{t+1},a^*)|\mathscr{F}_t]$. Suppose that Assumption 3.3.3 holds. Let $\gamma < 1$ and φ be defined as Lemma 3.4.3. Since the definition of a^* implies

$Q_t^A(x_{t+1}, a^*) \geqslant Q_t^A(x_{t+1}, b^*)$, we have

$$\begin{aligned}
|\mathbb{E}\{\beta_t[Q_t^A(x_{t+1}, b^*) - Q_t^B(x_{t+1}, a^*)]|\mathscr{F}_t\}| &= \mathbb{E}\{\beta_t[Q_t^A(x_{t+1}, b^*) - Q_t^B(x_{t+1}, a^*)]|\mathscr{F}_t\} \\
&\leqslant \mathbb{E}\{\beta_t[Q_t^A(x_{t+1}, a^*) - Q_t^B(x_{t+1}, a^*)]|\mathscr{F}_t\} \\
&= \sum_{x'} P(x_t, a_t, x')\beta(x_t, a_t, x')[Q_t^A(x', a^*) - Q_t^B(x', a^*)] \\
&\leqslant \sum_{x'} L(x_t, x') \max_{a'} |Q_t^A(x', a') - Q_t^B(x', a')| \\
&\leqslant \sum_{x'} L(x_t, x')\varphi(x')\|Q_t^A - Q_t^B\|_\varphi \\
&\leqslant \gamma\varphi(x_t)\|Q_t^A - Q_t^B\|_\varphi
\end{aligned}$$

Dividing both sides by $\varphi(x_t)$ and taking supremum, we have

$$\|\mathbb{E}[\beta_t(Q_t^B(x_{t+1}, \cdot) - Q_t^A(x_{t+1}, \cdot))|\mathscr{F}_t]\|_\varphi \leqslant \gamma\|\Delta_t^{BA}\|_\varphi. \tag{3.16}$$

On the other hand, suppose that Assumption 3.3.2 holds. Let γ and φ be defined as Lemma 3.4.1. The inequality (3.9) in Lemma 3.4.1 implies

$$\begin{aligned}
|\mathbb{E}\{\beta_t[Q_t^A(x_{t+1}, b^*) - Q_t^B(x_{t+1}, a^*)]|\mathscr{F}_t\}| &= \mathbb{E}\{\beta_t[Q_t^A(x_{t+1}, b^*) - Q_t^B(x_{t+1}, a^*)]|\mathscr{F}_t\} \\
&\leqslant \mathbb{E}\{\beta_t[Q_t^A(x_{t+1}, a^*) - Q_t^B(x_{t+1}, a^*)]|\mathscr{F}_t\} \\
&= \sum_{x'} P(x_t, a_t, x')\beta(x_t, a_t, x')[Q_t^A(x', a^*) - Q_t^B(x', a^*)] \\
&\leqslant \max_a \sum_{x'} P(x_t, a_t, x')\beta(x_t, a_t, x') \max_{a'} |Q_t^A(x', a') - Q_t^B(x', a')| \\
&\leqslant \max_a \sum_{x'} P(x_t, a_t, x')\beta(x_t, a_t, x')\varphi(x')\|Q_t^A - Q_t^B\|_\varphi \\
&\leqslant \gamma\varphi(x_t)\|Q_t^A - Q_t^B\|_\varphi.
\end{aligned}$$

To this end, (3.16) holds if Assumption 3.3.2 holds.

Alternatively, if we assume $\mathbb{E}[\beta_t Q_t^A(x_{t+1}, b^*)|\mathscr{F}_t] < \mathbb{E}[\beta_t Q_t^B(x_{t+1}, a^*)|\mathscr{F}_t]$, then by the fact that $Q_t^B(x_{t+1}, b^*) \geqslant Q_t^B(x_{t+1}, a^*)$, the above argument also implies (3.16). Therefore, we have

$$
\begin{aligned}
|\mathbb{E}[\tilde{F}_t(x,a)|\mathscr{F}_t]| &= \left|\frac{1}{2}\Delta_t^{BA}(x_t,a_t) + \frac{1}{2}\mathbb{E}[\beta_t(Q_t^A(x_{t+1},b^*) - Q_t^B(x_{t+1},a^*))|\mathscr{F}_t]\right| \\
&\leqslant \frac{1+\gamma}{2}\|\Delta_t^{BA}\|_\varphi
\end{aligned}
$$

where $(1+\gamma)/2 < 1$. Since $\mathrm{Var}(r_t|\mathscr{F}_t) < \infty$ and $\mathrm{Var}(\beta_t|\mathscr{F}_t) < \infty$, we have $\mathrm{Var}(\tilde{F}_t|\mathscr{F}_t) \leqslant K(1+\|\Delta_t^{BA}\|)^2$ for some constant K. Then, Lemma 3.4.4 and (3.15) yield $\Delta_t^{BA} \to 0$ w.p.1. Therefore, C_t also converges to zero w.p.1. Finally, Lemma 3.4.4 shows that the origin process (3.14) converges: $\Delta_t = Q^A - Q^* \to 0$ w.p.1. $\square$

3.5. Stability of Q-learning

The Bellman operator of an eventually discounting MDP can be viewed as a contraction map under a certain weighted norm. We find that the spectral radius $\rho(L)$ and its corresponding eigenvector play crucial roles in determining the convergence of the value function iteration. Utilising this fact, we establish a bound for optimal Q-value Q^*.

Define policy operator $T_\sigma \colon \mathbb{R}^{\mathsf{X}} \to \mathbb{R}^{\mathsf{X}}$ by

$$T_\sigma v(x) \coloneqq r(x,\sigma(x)) + \sum_{x'} P(x,\sigma(x),a')\beta(x,\sigma(x),x')v(x') \qquad (v \in \mathbb{R}^{\mathsf{X}}, x \in \mathsf{X}).$$

The Bellman operator can be written as $Tv(x) = \max_{\sigma\in\Sigma} T_\sigma v(x)$ for all $x \in \mathsf{X}$. Futhermore, assume that L is irreducible, which can be achieved when an MDP is recurrent. Since L is irreducible, Perron–Frobenius theorem implies that there exists $\varphi \in \mathbb{R}^n_+$ such that $L\varphi = \rho(L)\varphi$. We define the maps:

$$
\begin{aligned}
\tilde{T}_\sigma v &\coloneqq \Phi^{-1}T_\sigma(\Phi v); \quad \tilde{T}v \coloneqq \Phi^{-1}T(\Phi v); \qquad (v \in \mathbb{R}^{\mathsf{X}}) \\
\tilde{H}q &\coloneqq \hat{\Phi}^{-1}H(\hat{\Phi}q), \qquad\qquad (q \in \mathbb{R}^{\mathsf{G}})
\end{aligned}
$$

where $\Phi v(x) \coloneqq \varphi(x)v(x)$, $\Phi^{-1}v(x) \coloneqq v(x)/\varphi(x)$, $\hat{\Phi}q(x,a) \coloneqq \varphi(x)q(x,a)$, and $\hat{\Phi}^{-1}q(x,a) \coloneqq q(x,a)/\varphi(x)$ for all $(x,a) \in \mathsf{G}, v \in \mathbb{R}^{\mathsf{X}}$, and $q \in \mathbb{R}^{\mathsf{G}}$. Observe that $\tilde{v}_\sigma$, $\tilde{v}$ and $\tilde{q}$ are the fixed points of $\tilde{T}_\sigma, \tilde{T}$ and $\tilde{H}$, respectively, if and only if $v_\sigma = \Phi\tilde{v}_\sigma$, $v^* = \Phi\tilde{v}$ and $Q^* = \hat{\Phi}\tilde{q}$ are the fixed points of T_σ, T and H, respectively. We show that $\tilde{T}_\sigma, \tilde{T}$, and

$\tilde{H}$ are contraction maps with modulus $\rho(L)$, which further give the convergence rates of $\tilde{T}_\sigma, \tilde{T}$ and $\tilde{H}$ in the maximum norm, or the convergence rates of T_σ, T and H in $\|\cdot\|_\varphi$.

LEMMA 3.5.1. *If Assumption 3.3.3 holds, and L is irreducible, then T_σ, T, and H are contraction maps in $\|\cdot\|_\varphi$ with modulus $\rho(L)$, where φ is the eigenvector of L corresponding to eigenvalue $\rho(L)$.*

LEMMA 3.5.2. *If Assumption 3.3.3 holds, and L is irreducible, then $\tilde{T}_\sigma, \tilde{T}$ and $\tilde{H}$ are contraction maps with modulus $\rho(L)$ in maximum norm.*

Therefore, both T and H are globally stable, and we can analyze the convergence rates. For example, we have the following convergence rate and error bound for value function iteration (see, e.g., Bertsekas (2022) for the proof).

LEMMA 3.5.3. *If Assumption 3.3.3 holds, and L is irreducible, then*

(a) $\|T^k v - v^*\|_\varphi \leqslant \rho(L)^k \|v - v^*\|_\varphi$

(b) $\|T^{k+1} v - v^*\|_\varphi \leqslant \gamma \|T^{k+1} v - T^k v\|_\varphi$, *where* $\gamma = \rho(L)/(1-\rho(L))$.

Next, we use the contraction of H operator to find the boundedness of optimal Q^*. Moreover, the expectation $\mathbb{E}(Q_{t+1}(x,a)|\mathscr{F}_t)$ is bounded if Q_t is bounded.

PROPOSITION 3.5.1. *Suppose that Assumption 3.3.3 holds, L is irreducible, and $|r_t| \leqslant \bar{r}$ for all t w.p.1. Let $\{Q_t\}_{t\geqslant 0}$ be the Q-learning iteration. Then, the following statements are true.*

(a) $\|Q^*\|_\varphi \leqslant \|\bar{r}\|_\varphi/(1-\rho(L))$.

(b) If $\|Q_t\|_\varphi \leqslant \|\bar{r}\|_\varphi/(1-\rho(L))$, *then* $\|\mathbb{E}(Q_{t+1}(\cdot,\cdot)|\mathscr{F}_t)\|_\varphi \leqslant \|\bar{r}\|_\varphi/(1-\rho(L))$.

Proposition 3.5.1 shows that the Q-learning iteration is stable in the sense that Q_{t+1} is expected to be bounded by $\|\bar{r}\|_\varphi/(1-\rho(L))$ if Q_t is bounded by $\|\bar{r}\|_\varphi/(1-\rho(L))$. Finally, note that all the results also hold if Assumption 3.3.2 holds, by Lemma 3.4.1.

3.6. Learning with Concavity

In this section, we explore cases where the Q-factor Bellman operator exhibits concavity. We show that if the operator in Stochastic Approximation, which has a desired fixed point, is concave within a bounded interval, then the Stochastic Approximation iteration converges to the fixed point of that operator, provided that iterations remain within the same interval. Moreover, we apply this result to Q-learning by assuming that the Q-factor Bellman operator is concave.

3.6.1. Stochastic Approximation with Concavity. Let $H\colon \mathbb{R}^n \to \mathbb{R}^n$ be a map on $\mathbb{R}^n$ such that $Hx = (Hx(1), \dots, Hx(n))$ for all $x \in \mathbb{R}^n$. We are interested in computing the fixed point of H. The *Stochastic Approximation* algorithm consists of updating a vector $x_t \in \mathbb{R}^n$ by observation Hx_t with noise w_t to solve the fixed point of H. Let $\mathcal{T}_i \subset \mathbb{N}$ be a set of times at which an update of $x(i)$ is performed for $i \in \{1, \dots, n\}$. The Stochastic Approximation iterates are defined as follows:

$$x_{t+1} = \begin{cases} x_t(i), & \text{if } t \notin \mathcal{T}_i \\ (1-\alpha_t(i))x_t(i) + \alpha_t(i)(Hx_t(i) + w_t(i)), & \text{if } t \in \mathcal{T}_i \end{cases} \tag{3.17}$$

for all $i \in \{1, \dots, n\}$, where $\alpha_t(i) \in [0,1]$ is a step size parameter, $w_t(i)$ is a random noise, and $x_0 \in \mathbb{R}^n$.

Let $\mathscr{F}_t$ be the σ-field of the algorithm information up to and including the point at which the step-size $\alpha_t(i)$ is selected:

$$\mathscr{F}_t = \sigma\{x_0, \dots, x_t, w_0, \dots, w_{t-1}, \alpha_0, \dots, \alpha_t\}.$$

Let Ω be the sample space of all possible trajectories of $\{(x_t, \alpha_t, w_t)\}$ and $\mathscr{F} = \bigotimes_{t\in\mathbb{N}_0} \mathscr{F}_t$. Let $\mathbb{P}$ be the probability measure on $(\Omega, \mathscr{F})$. The following two assumptions regarding step sizes and noises are standard for Stochastic Approximation.

Assumption 3.6.1. The stepsizes $\{\alpha_t\}_{t\in\mathbb{N}_0}$ are a sequence of random variables defined on $(\Omega, \mathscr{F}, \mathbb{P})$ such that $\alpha_t(i) \in [0,1]$ and $\alpha_t(i) = 0$ for $t \in \mathcal{T}_i$ for all i and t. Moreover, we have

$$\sum_{t \in \mathcal{T}_i(\omega)} \alpha_t(i) = \infty, \text{ and } \sum_{t \in \mathcal{T}_i(\omega)} \alpha_t^2(i) < \infty$$

for all $i \in \{1, \dots, n\}$ and $\mathbb{P}$-almost all $\omega \in \Omega$.

Assumption 3.6.2.

(a) $\mathbb{E}[w_t(i)|\mathscr{F}_t] = 0$ for all i and t.

(b) There exist constants A and B such that for all i and t

$$\mathbb{E}[w_t^2(i)|\mathscr{F}_t] \leqslant A + B\|x_t\|^2$$

Assumption 3.6.2 ensures that the noise has a mean of zero and is bounded in the second moment, implying stability. Furthermore, we assume that H is concave on some interval.

Assumption 3.6.3.

(a) H is increasing and concave on $[u, v] \subset \mathbb{R}^n$ with $u < v$,

(b) $Hv \leqslant v$, and

(c) there exists an $\varepsilon > 0$ such that $Hu \geqslant u + \varepsilon(v - u)$.

Assumption 3.6.3 implies that $H\colon [u, v] \to [u, v]$ is globally stable and has a unique fixed point. To this end, the iteration (3.17) converges to the fixed point of H if the above assumptions hold.

Theorem 3.6.1. *If Assumption 3.6.1, 3.6.2 and 3.6.3 hold, and $\{x_t\}$ generated is by (3.17) is in $[u, v]$ with probability 1, then x_t converges to x^* with probability 1.*

3.6.2. Q-learning with Concavity. Let $c\colon \mathsf{G} \times \mathsf{X} \to \mathbb{R}_+$ denote a cost function. Suppose that the (future) value is adjusted by some concave function $\varphi\colon \mathbb{R} \to \mathbb{R}$ before taking expectation such that the Bellman equation becomes

$$v(x) = \min_{a \in \Gamma(x)} \mathbb{E}_{x,a} \varphi\left(c(x, a, X') + \beta(x, a, X')v(X')\right). \tag{3.18}$$

Let v^* be a solution to (3.18) and define $Q^*(x,a) \coloneqq \mathbb{E}_{x,a}[\varphi(c(x,a,X')+\beta(x,a,X')v^*(X'))]$ for all $(x,a) \in \mathsf{G}$. The corresponding Q-learning iteration follows

$$\begin{aligned} Q_{t+1}(x_t,a_t) &= (1-\alpha_t(x_t,a_t))Q_t(x_t,a_t) \\ &\qquad + \alpha_t(x_t,a_t)\left[\varphi\left(c_t + \beta_t \min_{b\in\Gamma(x_{t+1})} Q_t(x_{t+1},b)\right)\right] \end{aligned} \tag{3.19}$$

where $\alpha_t(x,a) \in [0,1]$ for all $(x,a) \in \mathsf{G}$, c_t and β_t satisfy $\mathbb{E}[c_t|(x,a',x')=(x_t,a_t,x_{t+1})] = c(x,a,x')$ and $\mathbb{E}[\beta_t|(x,a,x')=(x_t,a_t,x_{t+1})] = \beta(x,a,x')$, and x_{t+1} is a random successor state generated by $P(x_t,a_t,\cdot)$, given $Q_0 \in \mathbb{R}^{\mathsf{G}}$. Define the Q-factor Bellman operator $H\colon \mathbb{R}^{\mathsf{G}} \to \mathbb{R}^{\mathsf{G}}$ by

$$HQ(x,a) \coloneqq \mathbb{E}_{x,a}\varphi\left(c(x,a,X') + \beta(x,a,X')\min_{b\in\Gamma(X')} Q(X',b)\right) \tag{3.20}$$

for all $(x,a) \in \mathsf{G}$ and $Q \in \mathbb{R}^{\mathsf{G}}$.

ASSUMPTION 3.6.4.

(a) Both c and β are positive everywhere and bounded.

(b) $\varphi\colon \mathbb{R} \to \mathbb{R}$ is increasing, concave, and $\varphi(c(x,a,x')) > 0$ for all $(x,a,x') \in \mathsf{G} \times \mathsf{X}$.

(c) There is $K > 0$ such that $\varphi(\|c\| + \|\beta\|K) \leqslant K$.

Assumption 3.6.4 guarantees that $H(K\mathbb{1}) \leqslant K\mathbb{1}$. It also implies that $H0$ is everywhere positive and then there exists an $\varepsilon > 0$ such that $H0 \geqslant \varepsilon K\mathbb{1}$. To this end, H satisfies Assumption 3.6.3.

EXAMPLE 3.6.1. This example demonstrates a standard Q-learning. If $\beta \in (0,1)$, φ is an identity map, and c is positive everywhere and bounded. Suppose that H is defined by

$$HQ(x,a) = c(x,a) + \beta\mathbb{E}_{x,a}\min_{b\in\Gamma(x')} Q(x',b) \qquad ((x,a)\in\mathsf{G}, Q \in \mathbb{R}^{\mathsf{G}}).$$

Then, we have

$$\begin{aligned} H\left(\frac{\|c\|}{1-\beta}\mathbb{1}\right)(x,a) &= c(x,a) + \beta\mathbb{E}_{x,a}\min_{a'}\left\{\frac{\|c\|}{1-\beta}\mathbb{1}(x',a')\right\} \\ &\leqslant \|c\| + \beta\frac{\|c\|}{1-\beta} = \frac{\|c\|}{1-\beta}. \end{aligned}$$

Therefore, H is a self-map on $[u, v]$, where $u(x, a) \equiv 0$ and $v(x, a) \equiv \|c\|/(1 - \beta))$ for $(x, a) \in \mathsf{G}$. Also, $H0(x, a) = c(x, a) > 0$ for all $(x, a) \in \mathsf{G}$, which implies there exists an $\varepsilon > 0$ satisfying $H0 \geqslant \varepsilon(\|c\|/(1 - \beta))\mathbb{1}$.

Lemma 3.6.1. *If Assumption 3.6.4 holds, then H defined by (3.20) satisfies Assumption 3.6.3, is globally stable on $[0, K\mathbb{1}]$, and has a unique fixed point $x^* \in [0, K\mathbb{1}]$.*

Lemma 3.6.2. *Suppose that Assumption 3.3.1 and 3.6.4 hold. If $\{Q_t\}$ is a sequence generated by (3.19) with $Q_0 \in [0, K\mathbb{1}]$, then $\{Q_t\}$ is in $[0, K\mathbb{1}]$ with probability 1.*

Thus, Q-learning is stable in $[0, K\mathbb{1}]$ if Assumption 3.6.4 holds. Then, with the global stability of H, we can show the convergence of Q-learning.

Proposition 3.6.1. *Let $\{Q_t\}$ be generated by (3.19). If Assumption 3.3.1 and 3.6.4 hold, then Q_t converges to Q^* w.p.1.*

Example 3.6.2. This example modifies Assumption 3.6.4. Let $g \colon \mathbb{R} \to \mathbb{R}$ be an increasing and concave map such that $g(0) \geqslant 0$, and there is $K > 0$ satisfying $\|c\| + \|\beta\| g(K) < K$. Assume that c is positive everywhere and bounded, and β is everywhere positive. Suppose that H is defined by

$$HQ(x, a) = c(x, a) + \mathbb{E}_{x,a}\beta(x, a, x')g\left(\min_{b \in \Gamma(x')} Q(x', b)\right) \qquad ((x, a) \in \mathsf{G}, Q \in \mathbb{R}^{\mathsf{G}}).$$

Then, we have $H(K\mathbb{1})(x, a) \leqslant \|c\| + \|\beta\| g(K) < K$ for any $(x, a) \in \mathsf{G}$. Moreover, we have $H0(x, a) \geqslant c(x, a) > 0$ for any $(x, a) \in \mathsf{G}$ so as there exists an $\varepsilon > 0$ satisfying $H0 \geqslant \varepsilon K\mathbb{1}$. This example allows discount factors to be greater than one. In this case, the convergence relies on the concavity of function g. Similar to Lemma 3.6.1, H is globally stable and has a unique fixed point on $[0, K\mathbb{1}]$. A similar argument in Proposition 3.6.1 implies the convergence of Q-learning.

3.7. Appendix

Proof of Lemma 3.2.1. Suppose $\rho(L_m) < 1$. Since $\rho(L_m) = \lim_{n\to\infty} \|L_m^n \mathbb{1}\|^{1/n}$, there exists n such that $\|L_m^n \mathbb{1}\| < 1$. Since $\beta(x,\sigma(x),x')P(x,\sigma(x),x') \leqslant \max_a \beta(x,a,x')P(x,a,x')$ for all $(x,x') \in \mathsf{X}^2$ and $\sigma \in \Sigma$, we have $d_1 \leqslant \sup_x L_m\mathbb{1}(x)$. Induction yields $d_n \leqslant \|L_m^n\mathbb{1}\| < 1$. □

Proof of Lemma 3.5.1. Assume that Assumption 3.3.3 holds and L is irreducible. Since L is irreducible and $\rho(L) < 1$, the Perron-Frobenius Theorem implies that there is an everywhere positive eigenvector φ such that $L\varphi = \rho(L)\varphi$. That is, we have $\operatorname{diag}(\varphi)^{-1}L\varphi \leqslant \rho(L)\mathbb{1}$. We first show that T_σ is a contraction in $\|\cdot\|_\varphi$. Let $v, w \in \mathbb{R}^{\mathsf{X}}$ and $x \in \mathsf{X}$. By definition, we have

$$
\begin{aligned}
|T_\sigma v(x) - T_\sigma w(x)| &= \left| \sum_{x'\in\mathsf{X}} P(x,\sigma(x),x')\beta(x,\sigma(x),x')(v(x') - w(x')) \right| \\
&\leqslant \sum_{x'\in\mathsf{X}} P(x,\sigma(x),x')\beta(x,\sigma(x),x')|(v(x') - w(x')| \\
&\leqslant \sum_{x'\in\mathsf{X}} L(x,x')|(v(x') - w(x')| \\
&\leqslant \sum_{x'\in\mathsf{X}} L(x,x')\varphi(x') \max_{y\in\mathsf{X}} \frac{|(v(y) - w(y)|}{\varphi(y)} \\
&= \sum_{x'\in\mathsf{X}} L(x,x')\varphi(x')\|v - w\|_\varphi .
\end{aligned}
\tag{3.21}
$$

Dividing $\varphi(x)$ on both sides yields

$$
\frac{|T_\sigma v(x) - T_\sigma w(x)|}{\varphi(x)} \leqslant \frac{1}{\varphi(x)} \sum_{x'\in\mathsf{X}} L(x,x')\varphi(x')\|v - w\|_\varphi \leqslant \rho(L)\|v - w\|_\varphi,
$$

where the last inequality follows from $\operatorname{diag}(\varphi)^{-1}L\varphi \leqslant \rho(L)\mathbb{1}$. Taking the maximum on the left over X, we have $\|T_\sigma v - T_\sigma w\|_\varphi \leqslant \rho(L)\|v - w\|_\varphi$. Next, the similar argument in (3.21) yields

$$
T_\sigma v(x) \leqslant T_\sigma w(x) + \sum_{x'\in\mathsf{X}} L(x,x')\varphi(x')\|v - w\|_\varphi .
$$

By taking the supremum over Σ of both sides, we obtain

$$
Tv(x) \leqslant Tw(x) + \sum_{x'\in\mathsf{X}} L(x,x')\varphi(x')\|v - w\|_\varphi .
$$

By interchanging the roles of v and w and combining the two relations, we have

$$|Tv(x) - Tw(x)| \leqslant \sum_{x' \in \mathsf{X}} L(x, x')\varphi(x')\|v - w\|_\varphi.$$

Then, a similar argument shows $\|Tv - Tw\|_\varphi \leqslant \rho(L)\|v - w\|_\varphi$. Finally, we show that H is a contraction. Fix $Q, R \in \mathsf{X}^{\mathsf{G}}$. Then, we have

$$\begin{aligned}|HQ(x,a) - HR(x,a)| &= \left|\sum_{x' \in \mathsf{X}} P(x,a,x')\beta(x,a,x')(\max_{a'} Q(x',a') - \max_{a'} R(x',a'))\right| \\ &\leqslant \sum_{x' \in \mathsf{X}} L(x,x') \max_{a'} |Q(x',a') - R(x',a')| \\ &\leqslant \sum_{x' \in \mathsf{X}} L(x,x')\varphi(x')\|Q - R\|_\varphi.\end{aligned}$$

Thus, a similar argument implies that $\|HQ - HR\|_\varphi \leqslant \rho(L)\|Q - R\|_\varphi$. □

Proof of Proposition 3.5.1. Suppose that the assumptions in the statement hold. Since Q^* is the fixed point of H, $r(x,a,x') \leqslant \bar{r}$ w.p.1, and $\operatorname{diag}(\varphi)^{-1}L\varphi = \rho(L)\mathbb{1}$, we have

$$\begin{aligned}\left|\frac{Q^*(x,a)}{\varphi(x)}\right| &= \left|\sum_{x'} P(x,a,x')\left(\frac{r(x,a,x')}{\varphi(x)} + \frac{\beta(x,a,x')}{\varphi(x)}\max_b Q^*(x',b)\right)\right| \\ &\leqslant \left|\frac{\bar{r}}{\varphi(x)}\right| + \sum_{x'} P(x,a,x')\beta(x,a,x')\frac{\varphi(x')}{\varphi(x)} \max_y \max_b \left|\frac{Q^*(y,b)}{\varphi(y)}\right| \\ &\leqslant \|\bar{r}\|_\varphi + \sum_{x'} L(x,x')\frac{\varphi(x')}{\varphi(x)}\|Q^*\|_\varphi \\ &\leqslant \|\bar{r}\|_\varphi + \rho(L)\|Q^*\|_\varphi\end{aligned}$$

for all $(x, a) \in \mathsf{G}$. Taking the maximum on the left, we have $\|Q^*\|_\varphi \leqslant \|\bar{r}\|_\varphi + \rho(L)\|Q^*\|_\varphi$, which implies (a). Next, with the assumption $\|Q_t\|_\varphi \leqslant \|\bar{r}\|_\varphi/(1-\rho(L))$, we obtain

$$\begin{aligned}
\left|\frac{\mathbb{E}(Q_{t+1}(x,a)|\mathscr{F}_t)}{\varphi(x)}\right| &= \Bigg|(1-\alpha_t)\frac{Q_t(x,a)}{\varphi(x)} \\
&\quad + \alpha_t \sum_{x'} P(x,a,x')\left(\frac{r(x,a,x')}{\varphi(x)} + \frac{\beta(x,a,x')}{\varphi(x)}\max_b Q_t(x',b)\right)\Bigg| \\
&\leqslant (1-\alpha_t)\|Q_t\|_\varphi + \alpha_t\left(\|\bar{r}\|_\varphi + \sum_{x'} L(x,x')\frac{\varphi(x')}{\varphi(x)}\|Q_t\|_\varphi\right) \\
&= (1-\alpha_t)\|Q_t\|_\varphi + \alpha_t(\|\bar{r}\|_\varphi + \rho(L)\|Q_t\|_\varphi) \\
&\leqslant \frac{\|\bar{r}\|_\varphi}{1-\rho(L)}
\end{aligned}$$

for all $(x, a) \in \mathsf{G}$. Taking the maximum on the left, we conclude

$$\|\mathbb{E}(Q_{t+1}(\cdot,\cdot)|\mathscr{F}_t)\|_\varphi \leqslant \|\bar{r}\|_\varphi/(1-\rho(L)).$$

$\square$

Let E be a real Banach space where a partial ordering is defined by a cone $P \subset E$ such that $x \leqslant y$ if and only if $y - x \in P$. We write $x < y$ if $x \leqslant y$ and $x \neq y$. A cone is called *normal* if there exists a constant $N > 0$ such that $\theta \leqslant x \leqslant y$ implies $\|x\| \leqslant N\|y\|$, where θ denotes the zero element of E. An operator $A \colon E \to E$ is called *increasing* if $x, y \in E$ with $x \leqslant y$ implies $Ax \leqslant Ay$. It is called *concave* if for any $x, y \in E$ with $x \leqslant y$ and $\lambda \in [0, 1]$, we have $A(\lambda x + (1-\lambda)y) \geqslant \lambda Ax + (1-\lambda)Ay$. For any $u_0, v_0 \in E$ with $u_0 < v_0$, we define an order interval by $[u_0, v_0] := \{x \in E \colon u_0 \leqslant x \leqslant v_0\}$. Du (1990) shows the following fixed-point theorem with a concave operator (See also Zhang (2013) for the proof.)

THEOREM 3.7.1. *Suppose P is a normal cone, $u_0, v_0 \in E$, and $u_0 < v_0$. Moreover, $A \colon [u_0, v_0] \to E$ is an increasing operator. Let $h_0 = v_0 - u_0$. If A is a concave operator, $Au_0 \geqslant u_0 + \varepsilon h_0$, $Av_0 \leqslant v_0$ where $\varepsilon \in (0, 1)$, then A has a unique fixed point $x^* \in [u_0, v_0]$. Moreover, for any $x_0 \in [u_0, v_0]$, the iterative sequence $\{x_n\}$ given by $x_n = Ax_{n+1}$ for $n \in \mathbb{N}$ satisfying that*

$$\|x_n - x^*\| \leqslant M(1-\varepsilon)^n \qquad (n \in \mathbb{N})$$

where $M = N^2\|h_0\| + (N+1)N\|B\theta\|\varepsilon^{-2}$, $\varepsilon \in (0,1)$ *satisfies* $B\theta = Au_0 - u_0 \geqslant \varepsilon h_0$, *and* N *is the normal constant of* P*.*

PROOF OF THEOREM 3.6.1. Suppose that all the stated assumptions hold. Let $\{x_t\}$ be generated by (3.17). Define $U^{k+1} = HU^k$ and $L^{k+1} = HL^k$ for all $k \geqslant 0$ recursively with $U^0 = v$, $L^0 = u$. Assumption 3.6.3 implies $U^1 = Hv \leqslant v = U^0$ and $L^1 = Hu \geqslant u = L^0$. Since H is increasing, induction yields $L^k \leqslant L^{k+1}$ and $U^{k+1} \leqslant U^k$ for all k. Since the Du's Theorem 3.7.1 implies that H is globally stable on $[u,v]$, we have $U^k \to x^*$ and $L^k \to x^*$. The conclusion then follows from the proof for Theorem 2 of Tsitsiklis (1994) that for every k, there exists some $t_k \in \mathbb{N}$ such that

$$L^k \leqslant x_t \leqslant U^k \qquad \text{for all } t \geqslant t_k. \tag{3.22}$$

□

PROOF OF LEMMA 3.6.1. Let Assumption 3.6.4 hold. Then, H is a selfmap on $[0, K\mathbb{1}]$ and $H0 \geqslant \varepsilon K\mathbb{1}$. Let $\lambda \in [0,1]$ and fix $q_1, q_2 \in [0, K\mathbb{1}]$ with $q_1 \leqslant q_2$. Since φ is concave, we have

$$\begin{aligned}
&H(\lambda q_1 + (1-\lambda)q_2)(x,a) \\
&\qquad = \mathbb{E}_{x,a}\varphi\left(c(x,a,X') + \beta(x,a,X')\min_b\{\lambda q_1(X',b) + (1-\lambda)q_2(X',b)\}\right) \\
&\qquad \geqslant \mathbb{E}_{x,a}\varphi\left(c(x,a,X') + \beta(x,a,X')\left(\lambda \min_b q_1(X',b) + (1-\lambda)\min_b q_2(X',b)\right)\right) \\
&\qquad \geqslant \lambda\mathbb{E}_{x,a}\varphi\left(c(x,a,X') + \beta(x,a,X')\min_b q_1(X',b)\right) \\
&\qquad\qquad + (1-\lambda)\mathbb{E}_{x,a}\varphi\left(c(x,a,X') + \beta(x,a,X')\min_b q_2(X',b)\right) \\
&\qquad = \lambda Hq_1(x,a) + (1-\lambda)Hq_2(x,a)
\end{aligned}$$

for all $(x,a) \in \mathsf{G}$. Therefore, H is concave. Since φ is increasing, H is also increasing. It follows from the Du's theorem 3.7.1 that H has a unique fixed point Q^* in $[0, K\mathbb{1}]$ and there exists $\alpha \in (0,1)$ and $M > 0$ such that

$$\|H^m Q_0 - Q^*\| \leqslant \alpha^n M$$

for any $Q_0 \in [0, K\mathbb{1}]$ and $m \in \mathbb{N}$. $\square$

PROOF OF LEMMA 3.6.2. Let Assumption 3.3.1 and 3.6.4 hold. Fix $Q_0 \in [0, t\mathbb{1}]$. Suppose that Q_t is in $[0, K\mathbb{1}]$ for some t. Then, induction hypothesis and (3.19) imply

$$\begin{aligned} Q_{t+1}(x,a) &= (1-\alpha_t(x,a))Q_t(x,a) + \alpha_t(x,a)\varphi\left(c_t + \beta_t \min_b Q(x',b)\right) \\ &\geqslant (1-\alpha_t(x,a))0 + \alpha_t(x,a)\varphi(c_t) \geqslant 0 \end{aligned}$$

and

$$\begin{aligned} Q_{t+1}(x,a) &= (1-\alpha_t(x,a))Q_t(x,a) + \alpha_t(x,a)\varphi\left(c_t + \beta_t \min_b Q(x',b)\right) \\ &\leqslant (1-\alpha_t(x,a))K + \alpha_t(x,a)\varphi(c_t + \beta_t K) \\ &\leqslant (1-\alpha_t(x,a))K + \alpha_t(x,a)\varphi(\|c\| + \|\beta\| K) \\ &\leqslant (1-\alpha_t(x,a))K + \alpha_t(x,a)K = K \end{aligned}$$

for all $(x,a) \in \mathsf{G}$. Therefore, we conclude $Q_t \in [0, K\mathbb{1}]$ for all $t \geqslant 0$ with probability 1. $\square$

PROOF OF PROPOSITION 3.6.1. Suppose that Assumption 3.3.1 and 3.6.4 hold and let $\{Q_t\}$ be generated by (3.19). We first rewrite Q_{t+1} by

$$Q_{t+1}(x_t,a_t) = (1-\alpha_t)Q_t(x_t,a_t) + \alpha_t\left(HQ(x_t,a_t) + w_t(x_t,a_t)\right)$$

where operator H is defined by (3.20) and w_t is defined by

$$w_t(x_t,a_t) = \varphi\left(c_t + \beta_t \min_{b\in\Gamma(x_{t+1})} Q_t(x_{t+1},b)\right) - \mathbb{E}_{x_t,a_t}\varphi\left(c_t + \beta_t \min_{b\in\Gamma(x_{t+1})} Q_t(x_{t+1},b)\right).$$

Clearly, we have $\mathbb{E}[w_t | \mathscr{F}_t] = 0$, where $\mathscr{F}_t$ is defined by (3.7). Since φ is concave, there are constants $p, q > 0$ such that $\varphi(s) \leqslant p + qs$ for all s. Hence, we have $\varphi^2(s) \leqslant (p+qs)^2$. Since in addition $\mathrm{Var}(\beta_t|\mathscr{F}_t)$ and $\mathrm{Var}(c_t|\mathscr{F}_t)$ are finite, there are constants A and B such that $\mathbb{E}[w_t^2|\mathscr{F}_t] \leqslant A + B\|Q_t\|^2$. Therefore, Assumption 3.6.2 holds. The conclusion then follows from Theorem 3.6.1, Lemma 3.6.1, and Lemma 3.6.2. $\square$

CHAPTER 4

Uniqueness of Equilibria in Interactive Networks

4.1. Introduction

Network analysis enhances our comprehension of macroeconomic volatility and systemic risk by capturing agent interactions. For instance, idiosyncratic shocks propagate through the interconnections of a production network, resulting in aggregate fluctuation, where the network structure, such as hubs, sparsity, and asymmetry, significantly influences the shock propagation and the magnitude of aggregate fluctuation (Carvalho, 2008, Acemoglu et al., 2012, Carvalho, 2014, Carvalho and Tahbaz-Salehi, 2019).

In economic network analysis, a critical challenge is determining the uniqueness of equilibrium. The presence or absence of uniqueness carries various economic properties and implications, and it affects the validity of comparative statics if multiplicity exists. For example, Roukny et al. (2018) address that the multiple equilibria of a financial credit network make the probability of default indeterminate, complicating the evaluation of systemic risk. Jackson and Pernoud (2020) indicate that the multiple equilibria of a financial network can lead to a self-fulfilling cascade of defaults.[1]

To address this, we investigate a unified framework capable of efficiently and systematically verifying uniqueness across various network models, including production networks, network games, and financial networks. We clarify that, under general and widely employed assumptions, there are two types of framework: eventually contracting and non-expansive networks, referring to whether the corresponding operator is eventual contracting or non-expansive. We show the existence and (almost sure) uniqueness of equilibrium for both eventually contracting and non-expansive conditions, providing researchers with a valuable tool for promptly assessing uniqueness.

---

[1] When there are multiple equilibria, and the market has pessimistic beliefs, a bank may tend to hold cash and stop payments to others if it believes that other banks experience deterioration in credit conditions. In this case, the ex-ante fear causes the cascade of defaults, even if there is another better equilibrium that banks are solvent.

Specifically, we demonstrate that equilibrium exists and is unique when either the interaction function or the sensitivity matrix is convergent,ensuring the existence of a Banach contraction. If the spectral radius of a sensitivity matrix, weighted by the Lipschitz constants of interaction functions, is less than one, then equilibrium is not only unique but also globally stable. Conversely, when interaction functions are non-expansive but bounded, and a sensitivity matrix has a spectral radius of one, we establish that equilibrium exists and is unique almost surely, provided that the shocks are absolutely continuous. We underscore the role of absolute continuity in preventing the occurrence of multiplicity with non-zero probability. For instance, multiple equilibria may occur with strictly positive probability in some financial networks, such as Acemoglu et al. (2015a) and Acemoglu et al. (2015b).[2]

Consequently, our findings confirm that the clearing payment in the generalised Eisenberg-Noe model, studied by Liu et al. (2020), is almost surely unique under non-expansive conditions and always unique under eventually contracting conditions. Liu et al. (2020) simulate the U.S. interbank lending system to explore the contagion effect of bank failures and confirm the reduction in the network's contagion effect after the 2007-09 financial crisis.[3] While they demonstrated the existence of equilibrium, our proof of uniqueness ensures the stability and reliability of their simulations when shocks are absolutely continuous, such that the probability of multiple clearing payments is zero.

Furthermore, we study the non-bounded linear system and argue that boundedness is essential for both the existence and uniqueness of equilibrium in non-contracting networks. Leveraging this insight, we propose an algorithm to compute the generalised Eisenberg-Noe interbank lending model in Acemoglu et al. (2015a) and Acemoglu et al. (2015b), where the interaction functions are bounded identity maps. We show that the algorithm converges at most in $n2^{n-1}$ iterations.

Moreover, as a particular case of a non-expansive network, we demonstrate that the equilibrium payment in Eisenberg and Noe (2001) of credit network is always unique

---

[2] The generic uniqueness is defined in the sense that the set of shocks admitting multiplicity is measure zero.

[3] They find that banks have fewer counter-party exposures after the financial crisis.

without regularity conditions, which are also applied by Amini et al. (2016) and Staum et al. (2016). Our results correspond to the results obtained by Stachurski (2022).

In the applications involving generalised networks and the unique equilibrium, we extend our analysis to identify key players using a measure inspired by Sharkey (2017). *Key players* are identified as agents whose removal from the network results in the most significant reduction in aggregate economic states (Ballester et al., 2006, Zenou, 2016). Utilising the methodology of Sharkey (2017), we compute the measure by interpreting the equilibrium as the steady state of a continuous-time dynamic system. This measure captures the impact of shocks received from others as well as those passed on by agents, making it particularly useful for assessing systemic risk. Identified key players are characterised as either "too big to fail" or "too interconnected to fail" agents.

Related Literature. This paper studies and generalises the unified network model in Acemoglu et al. (2016b). Unlike Acemoglu et al. (2016b), we consider heterogeneous interaction functions and arbitrary sensitivity matrix. Our result is different from Acemoglu et al. (2016b) that, when a network is eventually contracting, the uniqueness of equilibrium does not merely depend on the Lipschitz contraction of the interaction functions. In our framework, the convergence of either interaction functions or the sensitivity matrix may lead to global stability. For non-expansive cases, we allow the sensitivity matrix to be either row or column stochastic.

The unified network model in this paper can be applied to determine the Nash equilibrium in network games, the equilibrium output in input-output analysis, and the clearing payments in generalised Eisenberg-Noe financial networks (Eisenberg and Noe, 2001). For example, the model can be used to describe the best response and solve Nash equilibrium in network games (Calvó-Armengol et al., 2009, Cohen-Cole et al., 2015, Blume et al., 2015, Zenou, 2016, Galeotti et al., 2020). For production networks, the model could represent the input-output relationship and determine the output equilibrium (Acemoglu et al., 2012, Bartelme and Gorodnichenko, 2015, Acemoglu et al., 2016a, 2017, Herskovic, 2018, Carvalho and Tahbaz-Salehi, 2019, Acemoglu and Azar, 2020, Herskovic et al., 2020, Pesaran and Yang, 2020). For financial networks, it calculates the clearing loan repayments, which studies the systemic risk of default cascade (Eisenberg and Noe, 2001, Cifuentes et al., 2005, Elsinger et al., 2006, Rogers and Veraart, 2013, Glasserman and Young, 2015, 2016, Acemoglu et al., 2015a, Gai and Kapadia, 2019, Veraart, 2020).

Outline. The paper is organized as follows. Section 4.2 presents the unified model and the main results of the existence and (almost sure) uniqueness of equilibrium. Section 4.3 lists out the network models with identical mathematical patterns that can be embodied in the unified framework and discusses some conditions for assumptions and extended results. Section 4.4 presents the comparative statics and discusses the tightness of assumptions and the importance of the boundedness condition. This section also discusses computation methods when the interaction functions are bounded identity maps. Section 4.5 presents a measure for identifying key players, which utilizes the property of the uniqueness of equilibrium.

4.2. General Model

Notation and Preliminary. Let $x, y \in \mathbb{R}^n$ be vectors and $f_i : \mathbb{R} \to \mathbb{R}$ be functions for all $i \in V := \{1, \dots, n\}$. In expressions involving matrix algebra, we take the convention that all vectors are row vectors, unless otherwise stated. Denote $|x|$ as $|x| := (|x_1|, \dots, |x_n|)$ and $f(x)$ as $f(x) := (f_1(x_1), \dots, f_n(x_n))$. If $f_i \equiv f$ for all $i \in V$, we write $f(x) := (f(x_1), \dots, f(x_n))$. We say that f_i is *non-expansive* if $|f_i(x) - f_i(y)| \leqslant |x - y|$ for all $x, y \in \mathbb{R}$. Also, $f = (f_i)_{i \in V}$ is *non-expansive* if f_i is non-expansive for all i. Denote $x \geqslant y$ if $x_i \geqslant y_i$ for all $i \in V$, $x > y$ if $x_i \geqslant y_i$ for all i and $x_i > y_i$ for some $i \in V$, and $x \gg y$ if $x_i > y_i$ for all $i \in V$.

Let $A = (A_{ij}), B = (B_{ij}) \in \mathbb{R}^{n \times n}$ be square matrices. Denote $A \geqslant B$ if $A_{ij} \geqslant B_{ij}$ for all $i, j \in V$ and $A \gg B$ if $A_{ij} > B_{ij}$ for all $i, j \in V$. A matrix A is *column (resp., row) stochastic* if $A \geqslant 0$ and $\sum_{i \in V} A_{ij} = 1$ for all $j \in V$ (resp., $\sum_{j \in V} A_{ij} = 1$ for all $i \in V$). A matrix A is *column (resp., row) substochastic* if $A \geqslant 0$ and $\sum_{i \in V} A_{ij} \leqslant 1$ for all $j \in V$ (resp., $\sum_{j \in V} A_{ij} \leqslant 1$ for all $i \in V$). The norm $\| \cdot \|$ refers to p-norm for vectors or matrix norm induced by p-norm. Denote $\rho(A) := \max\{|\lambda| : \lambda \text{ is an eigenvalue of } A\}$ as the *spectral radius* of A. A matrix A is *convergent* if $\lim_{k \to \infty} (A^k)_{ij} = 0$ for all $i, j \in V$, where $(A^k)_{ij}$ is the (i, j)-th entry of A^k. Let $x \in \mathbb{R}^n$ be a vector. We write $\operatorname{diag}(x)$ as the diagonal matrix with main diagonal x. We know that matrix A is convergent if and only if $\lim_{k \to \infty} \|A^k\| = 0$ or $\rho(A) < 1$. Let $T \colon \mathbb{R}^n \to \mathbb{R}^n$. We say that T is *globally stable* on $\mathbb{R}^n$ if T has a unique fixed point $x^* \in \mathbb{R}^n$ and $\lim_{k \to \infty} T^k x = x^*$ for any $x \in \mathbb{R}^n$.

4.2.1. Model. Consider an economy with $n \geqslant 2$ agents, indexed by $N = \{1, \dots, n\}$. Let $x_i \in \mathbb{R}$ denote each agent's economic state. Suppose agent j's state depends on the other agents' states such that

$$x_j = f_j \left(\sum_{i=1}^{n} x_i w_{ij} + \varepsilon_j \right) \qquad (j \in N) \tag{4.1}$$

Here, $f_j \colon \mathbb{R} \to \mathbb{R}$ is the *interaction function* describing how shocks and the other agents' states affect agent j. The term $w_{ij} \in \mathbb{R}$ represents the sensitivity extent of interaction between i and j, and $\varepsilon_j \in \mathbb{R}$ is a shock. We refer to $W := (w_{ij}) \in \mathbb{R}^{n \times n}$ as the *sensitivity*

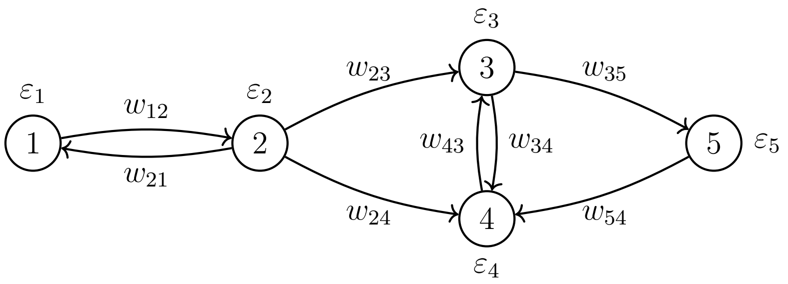


FIGURE 1

matrix, with entries w_{ij}. Let $x \coloneqq (x_1, \dots, x_n)$, $\varepsilon \coloneqq (\varepsilon_1, \dots, \varepsilon_n)$, and

$$f(xW + \varepsilon) \coloneqq \left(f_1 \left(\sum_i x_i w_{i1} + \varepsilon_1 \right), \dots, f_n \left(\sum_i x_i w_{in} + \varepsilon_n \right) \right).$$

Equation (4.1) can be rewritten in vector form as:

$$x = f(xW + \varepsilon).$$

We call (f, W, ε) a *network*. Given (f, W), a network could experience any shock realisations ε. For any shock realisation ε, we are interested in the equilibrium x satisfying (4.1) under the network structure (f, W). Let $\mathbb{P}_\varepsilon$ be the probability distribution of ε with support E_ε.

As presented in the following section, economic states can be outputs in production networks, decision choices in network games, and borrowing amounts in financial networks. The model (4.1) indicates that agent i's state influences the agent j's state if $w_{ij} \neq 0$, for $i, j \in N$. The interaction links and network structure can be presented by graph W. The degree of influence of other agents is determined by both the sensitivity w_{ij} and the interaction function f_j.

To illustrate the framework, Figure 1 demonstrates a graph example, where only the paths with non-zero sensitivity ($w_{ij} > 0$) are presented. In the graph, the equilibrium state of the agent 3 depends on the states of all other agents. However, agent 4's and 5's states also depend on agent 3's state, creating a feedback loop from agent 4 and 5 to agent 3. We observe that the network model is complex, even with only 5 agents in the network.

Unlike Acemoglu et al. (2016b), we allow heterogeneous interaction functions so that the framework includes more network models. We list the networks embodied in (4.1) in Section 4.3.

4.2.1.1. Eventually Contraction and Non-expansive Network. Observing the production networks, network games, and financial networks from literature (see Section 4.3.1), we observe generally two categories of assumptions. In the first category, the interaction function and sensitivity matrix exhibit contraction properties. For the second category, the interaction functions are non-expansive but bounded, and the sensitivity matrix is non-convergent: typically stochastic or $\rho(W) = 1$. In detail, these categories can be summarized with the following two assumptions.[4]

ASSUMPTION 4.2.1 (Eventually Contracting). $f = (f_i)$ and W satisfy

(i) f_i is Lipschitz continuous with Lipschitz constant β_i for all $i \in N$, and

(ii) $\rho(|W| \operatorname{diag}(\beta)) < 1$, where $\beta = (\beta_i)$.

ASSUMPTION 4.2.2 (Non-expansive). $f = (f_i)$ and W satisfy

(i) f_i is increasing, non-expansive, and bounded for all i, and

(ii) W is non-negative and $\rho(W) = 1$.

For instance, as illustrated in Section 4.3.1, the production networks in Carvalho (2008), Acemoglu et al. (2012), Carvalho (2014), Bartelme and Gorodnichenko (2015), Acemoglu et al. (2017), Herskovic (2018), Acemoglu and Azar (2020) and Herskovic et al. (2020), and the network games examined in Ballester et al. (2004), Ballester et al. (2006), Zenou (2012), Blume et al. (2015), Cohen-Cole et al. (2015), Zenou (2016), Galeotti et al. (2020) satisfy Assumption 4.2.1. On the other hand, the financial networks studied in Eisenberg and Noe (2001), Elsinger et al. (2005), Elsinger et al. (2006), Acemoglu et al. (2015b), Acemoglu et al. (2015a) and Liu et al. (2020) satisfy Assumption 4.2.2.

---

[4] Recall that the function $f_i \colon \mathbb{R} \to \mathbb{R}$ is bounded if there is $M > 0$ such that $|f_i(t)| \leqslant M$ for all $t \in \mathbb{R}$. The function f_i is Lipschitz continuous with Lipschitz constant β_i if $|f_i(x) - f_i(y)| \leqslant \beta_i |x - y|$ for all $x, y \in \mathbb{R}$.

We classify a network (f, W, ε) as *(eventually) contracting* if Assumption 4.2.1 holds and as *non-expansive* if Assumption 4.2.2 holds. Note that a contracting network implies that economic states tend to approach equilibrium over iterations $x \mapsto f(xW + \varepsilon)$, while the physical structure defined by graph W remains fixed. We will postpone the detailed discussion of these two assumptions to Section 4.3, where we provide more concrete examples.

Comparing these two assumptions, Assumption 4.2.1 does not require the boundedness for interaction functions, and allows for a negative sensitivity matrix. In contrast, Assumption 4.2.2 supposes that the interaction functions are bounded and increasing, and the sensitivity matrix is non-negative with a spectral radius of one. In practical applications, the spectral radius equals one when the row or column sum of the matrix is one.

As demonstrated in the main theorems, if a network satisfies the eventually contracting condition, the mapping $x \mapsto f(xW + \varepsilon)$ is a Banach contraction, ensuring the uniqueness of equilibrium. On the other hand, if a network is non-expansive, there exists an almost surely unique equilibrium.

4.2.1.2. Equilibrium. Following Acemoglu et al. (2016b), an equilibrium in this context is defined as follows.

Definition 4.2.1. Given shock realisation $(\varepsilon_1, \ldots, \varepsilon_n)$, an *equilibrium* of the economy is a collection of states $(x_1, \ldots, x_n)$ such that equation (4.1) holds simultaneously for all agents.

In other words, an equilibrium to a network (f, W, ε) is a vector of values $x = (x_i)_i$ that solves $x = f(xW + \varepsilon)$, given a fixed shock realisations ε. Since a network may have multiple equilibria, we define the concept of "almost sure uniqueness of equilibrium".

Definition 4.2.2. Let $E_\varepsilon \subset \mathbb{R}^n$ denote the set of all possible realisations of shocks ε. Define the subset M of E_ε by $M := \{\varepsilon \in E_\varepsilon : \text{Equation (4.1) has multiple equilibria}\}$. A network has an *almost surely unique* equilibrium, or an equilibrium is *almost surely unique*, if an equilibrium exists for $\varepsilon \in E_\varepsilon \setminus M$ and $\mathbb{P}_\varepsilon(M) = 0$, where $\mathbb{P}_\varepsilon$ is the distribution of ε.[5]

---

[5] Let $(\Omega, \mathscr{F}, \mathbb{P})$ be the measure space. The distribution is defined by $\mathbb{P}_\varepsilon(B) := \mathbb{P}(\varepsilon^{-1}(B))$ for all $B \in \mathscr{B}(\mathbb{R})$, where $\mathscr{B}(\mathbb{R})$ is the Borel algebra on $\mathbb{R}$.

If a network has an *almost surely unique* equilibrium, the probability that the shocks admit multiple equilibria is zero. Therefore, as we will demonstrate, the distribution of shocks determines whether there are multiple equilibria or not. In cases where the probability distribution of shocks is discrete, multiple equilibria may occur with non-zero probability. We do not adopt the terminology of generic uniqueness, as defined in Acemoglu et al. (2016b) and discussed in the subsequent remark, to avoid cases where multiple equilibria exist with strictly positive probability.

REMARK 4.2.1. In this remark, we first outline the assumptions on (f, W, ε) as presented in Acemoglu et al. (2016b) for comparison, followed by a discussion on generic uniqueness and its confusing implications. Acemoglu et al. (2016b) consider a model with homogeneous interaction functions $f_i \equiv g$. They assume that g is continuous, increasing, and either contracting or non-expansive but bounded, while W is column stochastic. The shocks ε_i are assumed to be independently and identically distributed with mean zero and constant variance.[6] According to Acemoglu et al. (2016b), they define the generic uniqueness of equilibrium as follows.[7]

DEFINITION 4.2.3. Let $E_\varepsilon \subset \mathbb{R}^n$ denote the set of all possible realisations of ε. Let M denote the set of shocks that admit multiple equilibria (i.e., $M \coloneqq \{\varepsilon \in E_\varepsilon :$ Equation (4.1) has multiple equilibria$\}$.) *Generic uniqueness* holds if the equilibrium exists for all $\varepsilon \in E_\varepsilon \setminus M$, and the Lebesgue measure of M is zero.

Under the above assumptions, Acemoglu et al. (2016b) demonstrate the existence and generic uniqueness of equilibrium for any network.[8] However, they acknowledge potential confusion since measure-zero events can also have probability one. We discuss in Example 4.2.1 that if the shock variable is discrete, there may be an arbitrarily high probability of multiplicity of equilibria. Therefore, to avoid confusion, Definition 4.2.2 is adopted, where almost sure uniqueness of equilibrium holds based on the distribution of shocks.

---

[6] Since W is column stochastic, every agent has constant total dependence on others.

[7] Acemoglu et al. (2016b) explain the generic uniqueness as "*. . . is generically unique, in the sense that the economy has multiple equilibria only for a measure zero set of realisations of agents-level shocks.*"

[8] Acemoglu et al. (2016b) show the generic uniqueness with the assumption of strong connectedness. They do not provide proof for extending the strongly connected graph to a general graph. We extend their proof to any network in Section 4.3.

This distinction is crucial as it clarifies that while generic uniqueness might hold under certain conditions, multiplicity of equilibria can still exist with a non-negligible probability depending on the shock specification.

Furthermore, the assumptions made by Acemoglu et al. (2016b) do not encompass the networks (4.4), (4.7) - (4.13) since they specifically require the sensitivity matrix to be column stochastic, and the interaction functions to be identical $f_i \equiv f$. For instance, network (4.12), demonstrated in Section 4.3.2, features a row stochastic sensitivity matrix and heterogeneous interaction functions, where $\bar{p}_i$ is not identical for all i. Consequently, we demonstrate the existence and almost sure uniqueness under broader assumptions that encompass a wider range of network models. □

4.2.2. Existence and Uniqueness. This section presents the main results. We first establish that the contraction of both interaction functions and a sensitivity matrix could lead to the existence and uniqueness of equilibrium under heterogeneous interaction functions. Next, we demonstrate that the equilibrium is almost surely unique when interaction functions are increasing, non-expansive but bounded, the spectral radius of the sensitivity matrix is one, and the shocks are absolutely continuous. The assumption of shock absolute continuity ensures that multiple equilibria do not occur with nonzero probability. Moreover, our results are applicable to any network structure.[9] Detailed proofs are provided in the Appendix.

THEOREM 4.2.1. *If Assumption 4.2.1 holds, then the equilibrium exists and is unique for any* $\varepsilon \in E_\varepsilon$.

Hence, an eventually contracting network with $\rho(|W| \operatorname{diag}(\beta)) < 1$ always has a unique equilibrium. Theorem 4.2.1 also implies that we can compute the unique equilibrium by iteration. Fix $\varepsilon \in E_\varepsilon$. Define the mapping $T : \mathbb{R}^n \to \mathbb{R}^n$ as

$$Tx := f(xW + \varepsilon). \tag{4.2}$$

---

[9] We do not assume a strong connection in the case of non-expansive conditions.

Since T is a Banach contraction, as shown in the proof ofTheorem 4.2.1, we can compute the equilibrium iteratively as $x^* = \lim_{k\to\infty} T^k x$ by starting from any initial guess $x \in \mathbb{R}^n$. We summarise the results in the following corollary.

Corollary 4.2.1. *Suppose that Assumption 4.2.1 holds. Then, we have*

(i) *T is globally stable.*

(ii) *$\|T^{k+1}x - T^k x\| \leqslant \|(|W| \operatorname{diag}(\beta))^k\| \|Tx - x\|$ for all $k \in \mathbb{N}$ and $x \in \mathbb{R}^n$.*

Corollary 4.2.1 provides the convergence rate of iteration. Since $\rho(|W| \operatorname{diag}(\beta)) < 1$, we can find $m \in \mathbb{N}$ such that $q = \|(|W| \operatorname{diag}(\beta))^m\| < 1$. Then, corollary 4.2.1 implies that the $\|T^{km+1}x - T^{km}x\| \leqslant \|(|W| \operatorname{diag}(\beta))^{km}\| \|Tx - x\| \leqslant q^k \|Tx - x\|$ for all $x \in \mathbb{R}^n$ and $k \in \mathbb{N}$.

On the other hand, in networks with non-expansive interaction functions and a sensitivity matrix having a spectral radius of one, the Banach contraction property does not apply. However, under additional assumptions of boundedness and order preservation, then equilibrium can still be determined. Specifically, any network with increasing and bounded interaction functions admits the greatest and the least equilibria.

Lemma 4.2.1. *If f_i is increasing and bounded for all $i \in N$, then the greatest and the least equilibria exist.*

We way that the shock variable ε is *absolutely continuous* if $\lambda(E) = 0$ for $E \in \mathscr{B}(E_\varepsilon)$ implies $P_\varepsilon(E) = 0$, where λ denotes the Lebesgue measure, and $E_\varepsilon \subset \mathbb{R}^n$ is the support of ε. If the shock is absolutely continuous, we can show that the equilibrium is almost surely unique.

Theorem 4.2.2. *If Assumption 4.2.2 holds, and the shock variables (ε_i) are absolutely continuous, then the equilibrium exists and is unique almost surely.*

By the Tarski's Fixed Point Theorem and the proof of Theorem 4.2.2, given that f_i is continuous for all i, we can compute the equilibrium by iteration from the bounds. In particular, let ℓ_j and u_j be the lower and upper bounds of the interaction function for all

j, if the network satisfies Assumption 4.2.2. Then, denoting $u := (u_j)$ and $\ell := (\ell_j)$, the largest equilibrium is $x^* = \lim_{m\to\infty} T^m u$, where $T : [\ell, u] \to [\ell, u]$ is defined as (4.2). If the shocks are absolutely continuous, then Theorem 4.2.2 confirms that x^* is the almost surely unique equilibrium.

Corollary 4.2.2. *If the assumptions of Theorem 4.2.2 hold, then the almost surely unique equilibrium is* $\lim_{n\to\infty} T^n u$.

Unlike the theorem in Acemoglu et al. (2016b), Theorem 4.2.2 allows sensitivity matrices to be either row or column stochastic and not necessarily strongly connected. The sensitivity matrix can even be non-stochastic, as long as its spectral radius is one (see Example 4.4.1). In section 4.3.2.3, we apply Theorem 4.2.2 to demonstrate that the financial network in Liu et al. (2020) has an almost surely unique equilibrium as a concrete example.

Notice that a non-expansive network may admit multiple equilibria with a positive probability if shocks are not absolutely continuous. This is illustrated through the following counterexample.

Example 4.2.1. This example illustrates that if the shocks are discrete, there can be multiple equilibria under Assumption 4.2.2. Suppose an economy with two agents, $n = 2$. Let $w_{12} = w_{21} = 1$ and $w_{11} = w_{22} = 0$ satisfy $\sum_h w_{hi} = 1$ for $i = 1, 2$ (see Figure 2). Assume the i.i.d. shock $\varepsilon_i \in \{1, -1\}$ for all i with probabilities $\mathbb{P}(\varepsilon_i = 1) = \mathbb{P}(\varepsilon_i = -1) = 1/2$. Then, the shocks ε_i have mean zero and constant variance, satisfying the conditions in Acemoglu et al. (2016b). Consider the interaction function as $f_i \equiv g$ for all i, where

$$g(z) = -M\mathbb{1}_{\{z<-M\}}(z) + z\mathbb{1}_{\{-M\leqslant z\leqslant M\}}(z) + M\mathbb{1}_{\{z>M\}}(z), \tag{4.3}$$

where $M \in (0, \infty)$. Then, g is a bounded and non-expansive identity mapping (i.e., $g(z) = z$ if $|z| \leqslant M$ and $|g(z)| \leqslant M$ for all $z \in \mathbb{R}$.)

Suppose that $x = (x_i) \in \mathbb{R}^n$ satisfy $x_1 = g(x_2 + \varepsilon_1)$ and $x_2 = g(x_1 + \varepsilon_1)$. Since g is bounded, we have $-M \leqslant x_1, x_2 \leqslant M$. If the realisation is $\varepsilon = (\varepsilon_1, \varepsilon_2) = (1, -1)$, then the

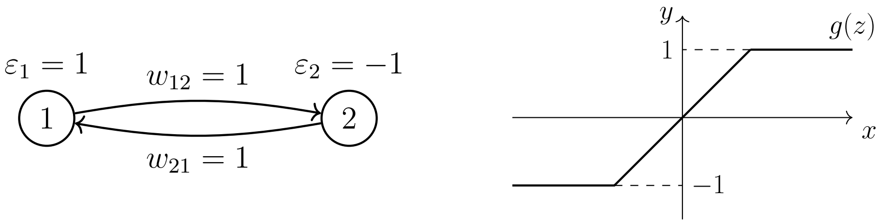


FIGURE 2

system $x = f(xW + \varepsilon)$ gives

$$x_1 = x_2 + 1$$

$$x_2 = x_1 - 1.$$

Assuming $-M \leqslant x_2 + 1, x_1 - 1 \leqslant M$, we find that $x_1 = y + 1$ and $x_2 = y$ with $-M \leqslant y \leqslant M - 1$. Therefore, there are multiple equilibria if $\varepsilon = (1, -1)$. Similarly, multiple equilibria exist if $\varepsilon = (-1, 1)$. Hence, the set of shock realisations generating multiple solutions is $M = \{\varepsilon \in \mathbb{R}^2 : \varepsilon = (1, -1) \text{ or } (-1, 1)\}$, which has Lebesgue measure zero $\lambda(M) = 0$. However, the probability that ε leads to non-uniqueness is $\mathbb{P}(\varepsilon \in M) = 1/2$, which is non-zero. □

To avoid the confusion seen in Example 4.2.1, where multiple equilibria occur with non-zero probability, we suppose that the idiosyncratic shocks are *absolutely continuous*.[10] With absolute continuity of shocks, events with measure zero also occur with zero probability. Hence, the absolute continuity precludes the case described in Example 4.2.1.

4.3. Examples and Extensions

In this section, we discuss the conditions outlined in Assumption 4.2.1 and 4.2.2. We present example models from various domains, including production networks (Section 4.3.1.1 and 4.3.1.2), network games (Section 4.3.1.3 and 4.3.1.4) and financial networks (Section 4.3.1.5, 4.3.2, 4.3.2.2 and 4.3.2.3). The goal is to demonstrate that these network models adhere to the same mathematical framework described by (4.1) and follow either Assumption 4.2.1 or Assumption 4.2.2.

---

[10] Recall that the random variable ε_i is *absolutely continuous* if there is a Lebesgue integrable function g such that $\mathbb{P}(\varepsilon_i \in A) = \int_A g(x)\lambda(\mathrm{d}x)$ for all Borel sets A and for all $i \in N$.

We begin by examining the case of eventually contracting networks, where f and W satisfy Assumption 4.2.1, ensuring $\rho(|W|\operatorname{diag}(\beta)) < 1$. A specific instance example is that the contracting condition can merely depend on the interaction functions. Taking the production network (4.6) in Section 4.3.1.2 for example, W is non-negative and its row sum is one, and $\beta_i \equiv (1-\alpha)$, the spectral radius condition is reduced to $\rho(|W|\operatorname{diag}(\beta)) = (1-\alpha) < 1$. In such case, the Lipschitz contraction property determines the uniqueness of equilibrium.

Unlike Acemoglu et al. (2016b), the contraction condition $\rho(|W|\operatorname{diag}(\beta)) < 1$ generally depends on both interaction functions and the sensitivity matrix. For instance, consider the matrices:

$$W = \begin{pmatrix} 0 & 2 \\ 4/7 & 0 \end{pmatrix}, \qquad \operatorname{diag}(\beta) = \begin{pmatrix} 5/4 & 0 \\ 0 & 2/3 \end{pmatrix}, \quad W\operatorname{diag}(\beta) = \begin{pmatrix} 0 & 4/3 \\ 5/7 & 0 \end{pmatrix}.$$

In this case, we have $\rho(W) > 1$, $\beta_1 > 1$, but $\rho(|W|\operatorname{diag}(\beta)) < 1$. Moreover, the spectral radius condition considers the sensitivity matrix by taking the absolute values of all entries $|W|$. When W is symmetric, we can reduce this condition. For example, if $W \in \mathbb{R}^{n\times n}$ is symmetric and $\beta_i \equiv \varphi$, then the condition is $\rho(|W|\operatorname{diag}(\beta)) = \varphi\rho(W) < 1$.

Furthermore, some networks may impose stricter conditions on interaction functions. For instance, except for the Lipschitz continuity, the interaction functions described by (4.4) - (4.10) in Section 4.3.1 are linear such that $f_i(a) - f_i(b) = \beta_i(a-b)$ for some $\beta_i > 0$, for all $a, b \in \mathbb{R}$ and for all $i \in N$. In this case, the contraction condition can be reduced to $\rho(W\operatorname{diag}(\beta)) < 1$.[11]

Corollary 4.3.1. *If for all $i \in N$ there is $\beta_i > 0$ such that $f_i(a) - f_i(b) = \beta_i(a-b)$ for all $a, b \in \mathbb{R}$, then the condition $\rho(W\operatorname{diag}(\beta)) < 1$ for $W \in \mathbb{R}^{n\times n}$ implies the uniqueness of equilibrium for any $\varepsilon \in \mathbb{R}^n$.*

4.3.1. Networks with Linear Interaction Functions.

In this section, we consider a simple framework where interaction functions are linear. In this case, the uniqueness of equilibrium follows Corollary 4.3.1.

---

[11] Note that the entry of W could be negative as Assumption 4.2.1.

4.3.1.1. Input-Output Analysis. The input-output analysis describes the inter-industry relationship using a matrix that tracks the flow of money.[12] It studies how the shock in one sector affects the output of the other sectors[13] This analysis also helps identify which industry or region is the most significant in the optimisation of the aggregate economy. The model is briefly introduced below.

There are n industries in a closed economy with no inventories. Each industry $i \in \{1, \ldots, n\}$ requires $w_{ij} \in [0,1]$ dollar amount of intermediate input from industry $j \in \{1, \ldots, n\}$ to produce one dollar of $i's$ output. Input-output tables empirically determine the linkage weights between sectors w_{ij}.[14] For every industry $j \in \{1, \ldots, n\}$, the gross output x_j equals the total value of its use as a final good ε_j and its use as an intermediate input to other industries:

$$x_j = \varepsilon_j + \sum_i x_i w_{ij}. \tag{4.4}$$

That means the sale of industry j to other sectors is $\sum_i x_i w_{ij}$. To compute the equilibrium, it is conventional to assume that every sector has some inputs from labour or other value-added sources, so that $\sum_j w_{ij} < 1$ for all i. Let $x = (x_i)$, $\varepsilon = (\varepsilon_i) \in \mathbb{R}^n$ and $W = (w_{ij}) \in \mathbb{R}^{n \times n}$. The vector form of (4.4) is $x = \varepsilon + xW$. Since the row sum of W is less than one, the matrix $(I - W)$ is non-singular. The unique equilibrium of output is $x = \varepsilon(I - W)^{-1}$, where $(I - W)^{-1}$ is known as the Leontief inverse.

4.3.1.2. Production Networks. This subsection presents the production network outlined in Long and Plosser (1983), Carvalho (2008), Acemoglu et al. (2012), Acemoglu et al. (2016a), Acemoglu et al. (2017), Carvalho and Tahbaz-Salehi (2019) and Acemoglu and Azar (2020). These production networks investigate how heterogeneous shocks to individual sectors can generate aggregate fluctuations through supplier-customer interconnections in a production network. Moreover, the network models illustrate that such aggregate fluctuation and the cascading effects of shocks are correlated with the structure of the networks (Carvalho (2008) and Acemoglu et al. (2012)).

---

[12] See Miller and Blair (2009) and Miller and Temurshoev (2017) for example.

[13] Fletcher (1989) uses input-output analysis to study the impact of tourism.

[14] See Timmer et al. (2015).

The economy has n competitive sectors. Each sector's output x_j follows the production function:

$$y_j = z_j^{\alpha} \ell_j^{\alpha} \prod_{i=1}^{n} y_{ij}^{(1-\alpha)w_{ij}}$$

where z_j denotes the productivity shock, ℓ_j denotes the labour input, $\alpha \in (0,1)$ is the share of labour input, y_{ij} is the intermediate input from sector i used in the production of good j, and $w_{ij} \geqslant 0$ is the share of intermediate input i in the total intermediate input. It is supposed that $\sum_i w_{ij} = 1$ for all $j = 1, \ldots, n$. Let p_j be the price of good j, and h be the labour wage. Producers maximise their profits:

$$\max_{\ell_j, y_{1j}, \ldots, y_{nj}} p_j y_j - h\ell_j - \sum_{i=1}^{n} p_i y_{ij}$$

The optimal labour input is $\ell_j = \alpha p_j y_j / h$ and the intermediate input is $y_{ij} = (1 - \alpha) w_{ij} p_j y_j / p_i$. A representative household with Cobb-Douglas preferences solves the optimal problem:

$$\max_{c_1, \ldots, c_n} u(c_1, \ldots, c_n) = A \prod_{j=1}^{n} c_j^{1/n} \qquad \text{s.t. } \sum_j p_j c_j = h \sum_j \ell_j$$

where c_j is the consumption of good h, and A is a normalisation constant. By normalising the total labour supply $\sum_i \ell_i = 1$, the first-order condition for the optimal consumption yields $c_j = h/(np_j)$. The clearing condition of the commodity market, from equation (4.4), is $y_j = c_j + \sum_k y_{jk}$. Then, we have

$$p_j y_j = \frac{h}{n} + (1 - \alpha) \sum_{k=1}^{n} w_{jk} p_k y_k \tag{4.5}$$

Let $\hat{y}_j = p_j y_j$, $\hat{y} = (\hat{y}_j) \in \mathbb{R}^n$ and $W = (w_{ij}) \in \mathbb{R}^{n \times n}$. The vector form of clearing condition is $\hat{y} = (h/n)\mathbf{1} + (1 - \alpha)\hat{y}W^{\top}$. Since the row sum of $(1 - \alpha)W^{\top}$ is strictly less than one, $[I - (1 - \alpha)W^{\top}]$ is non-singular. Thus, we have is $\hat{y} = (h/n)\mathbf{1}[1 - (1 - \alpha)W^{\top}]^{-1}$. Defining $b := \mathbf{1}[I - (1 - \alpha)W^{\top}]^{-1}$, we write $p_j y_j = b_j h/n$. Hence, we have $\ell_j = \alpha b_j / n$ and $y_{ij} = (1 - \alpha) w_{ij} y_i b_j / b_i$, so the production function yields

$$\log y_j = \mu_j + \alpha \log z_j + (1 - \alpha) \sum_i (\log y_i) w_{ij}$$

where μ_j is some constant.[15] Denote $x_j = \log y_j$ and $\varepsilon_j = (\mu_j + \alpha \log z_j)/(1-\alpha)$, it delivers

$$x_j = (1-\alpha)\left(\sum_i x_i w_{ij} + \varepsilon_j\right) \tag{4.6}$$

The equilibrium output is $x = \varepsilon[I - (1-\alpha)W]^{-1}$. Overall, there are two simple network equations (4.5) and (4.6).

4.3.1.3. Simple Network Games. In a social network game, an agent's payoff or well-being depends not only on her own actions but also on the actions of her neighbours. Social networks influence decision-making behaviours such as committing a crime and making lending decisions (Ballester et al. (2004) and Cohen-Cole et al. (2015)). Consider a simple network game as described by Zenou (2012), Zenou (2016) and Galeotti et al. (2020).

There are n players in a social network, and the social connection is represented by graph W. If agent i is connected with agent j, then $w_{ij} = 1$; otherwise, $w_{ij} = 0$. Moreover, assume that $w_{ii} = 0$ for all $i = 1, \dots, n$ by convention. Thus, $W = (w_{ij})$ is an adjacency matrix with entry w_{ij}. Assume that it is a game of strategic complement with perfect information such that players know everything about the network. Agents choose the actions $x_j \in \mathbb{R}_+$ to maximise their payoffs:

$$u_j(x_1, \dots, x_n) = \alpha_j x_j - \frac{1}{2}x_j^2 + \varphi \sum_{i=1}^n w_{ij} x_i x_j$$

where $\alpha_j > 0$ is the exogenous heterogeneity capturing individual characteristics, $\alpha_j x_j - (1/2)x_j^2$ is the individual benefits, $\varphi \in \mathbb{R}$ specifies the degree of payoff from peer influence, and $\varphi \sum_{i=1}^n w_{ij} x_i x_j$ is the peer influence, which depends on the location of agents. Hence, every agent's payoff depends on her own actions and the other agents' actions. The best-reply function in equilibrium is

$$x_j = \alpha_j + \varphi \sum_{i=1}^n x_i w_{ij} = \varphi\left(\sum_{i=1}^n x_i w_{ij} + \varepsilon_j\right) \tag{4.7}$$

where we let $\varepsilon_j = \alpha_j/\varphi$. For the equilibrium, suppose that $\varphi\,\rho(W) < 1$ so that φW is non-singular. The unique Nash equilibrium is $x^* = \varepsilon(I - \varphi W)^{-1}$. Note that we refer ε_j as

---

[15] $\mu_j = \log\big(b_j(\alpha/n)^\alpha (1-\alpha)^{1-\alpha}\big) + (1-\alpha)\sum_i w_{ij} \log(w_{ij}/b_i)$.

the shock in our model, but in a network game, it captures the observable characteristics of individual j such that it is exogenous.

4.3.1.4. Network Games with Global and Local Interaction. Ballester et al. (2006) consider both the global substitutability and local influence complementarity in network games. They investigate how to identify the "key player" that, once removed, causes the maximal decrease in aggregate activity. The model is similar to the simple network game in Section 4.3.1.3. Let $G = (g_{ij})$ be the adjacency matrix such that $g_{ii} = 0$ for all $i = 1, \ldots, n$, and $g_{ij} = 1$ if i and j are connected and $g_{ij} = 0$ otherwise. Following the setup in Section 4.3.1.3, given action profile (x_i), each agent j has an alternative payoff:

$$u_j(x_1, \ldots, x_n) = \alpha_j x_j - \frac{1}{2}(\eta - \gamma)x_j^2 - \gamma \sum_{i=1}^{n} x_i x_j + \varphi \sum_{i=1}^{n} g_{ij} x_i x_j$$

where $\alpha_j > 0$ for all j, $\eta, \varphi > 0$, $\gamma \geqslant 0$, and $\gamma \sum_{i=1}^{n} x_i x_j$ denotes the global interaction of substitution effect across all agents, and the last term represents the local interaction of strategic complement. The best-reply function is:

$$x_j = \frac{\alpha_j}{\eta} - \frac{\gamma}{\eta} \sum_{i=1}^{n} x_i + \frac{\varphi}{\eta} \sum_{i=1}^{n} x_i g_{ij} = \frac{\varphi}{\eta} \left(\sum_{i=1}^{n} x_i w_{ij} + \varepsilon_j \right) \tag{4.8}$$

where $w_{ij} = g_{ij} - \gamma/\varphi$ and $\varepsilon_j = \alpha_j/\varphi$. We see that the best-reply function (4.8) has the same form as (4.1).

Denote $J \in \mathbb{R}^{n \times n}$ as the square matrix of ones. The Nash equilibrium solves $\eta x^* = \alpha - \gamma x^* J + \varphi x^* G$, where $\alpha = (\alpha_i)$. If $(\eta I + \gamma J - \varphi G)$ or $(\varphi/\eta)W$ is non-singular, then the Nash equilibrium is $\alpha(\eta I + \gamma J - \varphi G)^{-1}$.[16]

For example, Cohen-Cole et al. (2015) present an interbank lending network with similar features. They consider a network with n banks in a lending market, and the adjacency matrix $g_{ij} = 1$ if bank j makes a loan to bank i, and $g_{ij} = 0$ otherwise. Each bank j has

---

[16] We can further solve the Nash equilibrium. Suppose that W is non-singular. Then x^* is unique and $\bar{x}^* := \sum_j x_j^*$ is constant, We further assume that $\varphi\rho(G) < \eta$. Since $x^* J = \bar{x}^* \mathbf{1}$, we also have $x^* = (\alpha - \gamma \bar{x}^* \mathbf{1})(\eta I - \varphi G)^{-1}$. Then, we can find that $\bar{x}^* = \bar{b}_\alpha/(1 + \gamma \bar{b})$, where $b_\alpha = \alpha(\eta I - \varphi G)^{-1}$, $b = \mathbf{1}(\eta I - \varphi G)^{-1}$, $\bar{b}_\alpha = b_\alpha \mathbf{1}^\top$, and $\bar{b} = b\mathbf{1}^\top$. The Nash equilibrium is $x^* = b_\alpha - \gamma[\bar{b}_\alpha/(1 + \gamma \bar{b})]b$.

the profit function, given its volume of loans x_j to other banks:

$$\pi_j = px_j - c_j x_j = \left(\theta - \sum_{i=1}^{n} x_j\right) x_j - \left(c_{0,j} - \varphi_j \sum_{i=1}^{n} x_i g_{ij}\right) x_j$$

where the $p = \theta - \sum_{i=1}^{n} x_i$ is price of loans determining the interest rate, and $c_j = c_{0,j} - \varphi_j \sum_{i=1}^{n} x_i g_{ij}$ is the marginal cost, with $\theta > 0$ and $c_{0,j}, \varphi_j > 0$ for all j. The parameter φ_j specifies the cost cut induced by each link of loan due to the collaboration between banks. Hence, since the profit increases with the links between banks, it implies local strategic competitiveness. Also, since the price decreases with the aggregate quantity of loans, there is global strategic substitutability. Under competition, each bank decides the quantity of loan to maximise its profit, so the first order condition gives

$$x_j = (\theta - c_{0,j}) - \sum_{i=1}^{n} x_i + \varphi_j \sum_{i=1}^{n} x_i g_{ij} = \varphi_j \left(\sum_{i=1}^{n} x_i w_{ij} + \varepsilon_j\right) \tag{4.9}$$

where we let $w_{ij} = g_{ij} - 1$ and $\varepsilon_j = (\theta - c_{0,j})/\varphi_j$. In this example, agents have different interaction functions since the coefficients φ_j may be heterogeneous.

4.3.1.5. Network with Cross-Holdings. Elliott et al. (2014) consider a financial network with cross-holdings and study the cascade effect of financial failure, which has the same form as input-output analysis (4.4). In their framework, banks own some share of the other banks through lending or investment, so the values of banks depend on their holdings of other banks' assets. They show that the cascade effect depends on network interconnections, in the sense that integration and diversification lead to different non-monotonic effects.

They consider an economy with n financial institutions or banks, indexed $j = 1, \dots, n$. Each organisation holds a basket of primitive assets, indexed $h = 1, \dots, m$, which could be projects that create cash flows. The share of asset h that organisation j holds is denoted as $b_{hj} \geqslant 0$. The market price of the asset h is denoted as p_h.

Organisations cross-hold some shares of the other organisations in the networks. For all $i, j = 1, \dots, n$, let d_{ij} be the debt that organisation i has to repay to organisation j or the amount of funds invested in organisation i by organisation j. Define $w_{ij} := d_{ij}/x_i$. Hence, organisation j owns $w_{ij} \in [0, 1)$ fraction of the values of organisation i. Assume that $w_{ii} = 0$ for all i. Denote the book value or equity of organisation j as x_j. The book value x_j is the total asset value that j owns, including its primitive assets and the claims on other organisations:

$$x_j = \sum_{h=1}^{m} p_h b_{hj} + \sum_{i=1}^{n} d_{ij} = \varepsilon_j + \sum_{i=1}^{n} x_i w_{ij} \tag{4.10}$$

where $\varepsilon_j = \sum_{h=1}^{m} p_h b_{hj}$. Assume that external investors hold strictly positive shares of organisation i, i.e., $1 - \sum_{j \in I} w_{ij} > 0$. Then, W is non-singular, and the equilibrium equity is given by $x = \varepsilon(I - W)^{-1}$.

4.3.2. Non-linear Interaction Functions: Financial Networks. In this section, we discuss the conditions for the uniqueness of equilibrium in financial interbank lending networks such as Eisenberg and Noe (2001), Cifuentes et al. (2005) and Acemoglu et al. (2015a). Their canonical models are used to study the contagion of default under the conditions of proportional repayments of liabilities, limited liability, and the absolute priority of debt over equity. For instance, Glasserman and Young (2015) bound the probability of default due to contagion when a bank suffers a shock.

There are n risk-neutral banks in the network, as the previous section. Each bank i has the nominal liability δ_{ij} to bank j. The total liability obligation of i is $\bar{p}_i = \sum_j \delta_{ij}$. Define relative liability as $w_{ij} = \delta_{ij}/\bar{p}_i$ if $\bar{p}_i > 0$ and $w_{ij} = 0$ otherwise.[17] Assume that $\sum_j w_{ij} = 1$ for all i, in the sense that there is no payment to agents outside the network.

All banks have the exogenous cash flow $\varepsilon_j \geqslant 0$, which can be interpreted as the net asset from outside the financial network. Let x_j be the clearing repayment for all $j = 1, \dots, n$ in equilibrium. Suppose that proportional repayment of liability holds such that bank i's clearing payments to bank j is proportional to its relative liability $x_i w_{ij}$. Then, the amount of total repayment received by j from other banks is $\sum_i x_i w_{ij}$.

Suppose that, in equilibrium, all banks follow the conditions of limited liability, $x_j \leqslant \sum_i x_i w_{ij} + \varepsilon_j$, and absolute priority, either $x_j = \bar{p}_j$ or $x_j = \sum_i x_i w_{ij} + \varepsilon_j$. The clearing

---

[17] The sensitivity matrix (w_{ij}) is also called as the *relative liability matrix* in Eisenberg-Noe model.

payment x_j in equilibrium solves

$$x_j = \min\left\{\sum_i x_i w_{ij} + \varepsilon_j, \bar{p}_j\right\} \tag{4.11}$$

for all j. To this end, the interaction functions are

$$f_j(t) = t\,\mathbb{1}_{\{t<\bar{p}_j\}}(t) + \bar{p}_j\,\mathbb{1}_{\{t\geqslant\bar{p}_j\}}(t)$$

for all j. We introduce the assumptions in Glasserman and Young (2015) for the uniqueness of equilibrium, which supposes $\sum_j w_{ij} \leqslant 1$. We write $i \to j$ if j is *accessible* from i for $i, j \in N = \{1, \ldots, n\}$: $w_{ij}^k > 0$ for some $k \in \mathbb{N}$, where w_{ij}^k is the (i, j) entry of W^k. Glasserman and Young (2015) assume that for every bank i either it has some debt outside the network such that $\sum_j w_{ij} < 1$, or there is a bank t such that $i \to t$ (an obligation path from i to t), and bank t has external debt such that $\sum_j w_{tj} < 1$. Under this assumption, W is convergent, and then the clearing payment (4.11) is unique. To see why uniqueness holds under Glasserman and Young (2015), since $\beta_i \equiv 1$ in (4.11), it is sufficient to show that $\rho(W) < 1$ when the above conditions hold.

We say that W is *weakly chained substochastic* if W is row substochastic and for each $i \in N$, either $\sum_j w_{ij} < 1$, or there exists $t \in N$ such that $i \to t$ and $\sum_j w_{tj} < 1$. We can check the convergence of W by weakly chained substochasticity (see Azimzadeh (2019) and Lemma 4.7.6.)

We briefly discuss some intuition behind Lemma 4.7.6 and weekly chained convergence. The convergence can be seen by considering a simple case: if $\sum_j w_{ij} \leqslant 1$ for all i, and for all $i \in N$ there is $t \in N$ such that $i \to t$ in one step ($w_{it} > 0$) and $\sum_j w_{tj} < 1$, then we can show that $\sum_k w_{ik}^2 = \sum_k \sum_j w_{ij} w_{jk} = \sum_j w_{ij} \sum_k w_{jk} = \sum_{j\neq t} w_{ij} \sum_k w_{jk} + w_{it} \sum_k w_{tk} < 1$ for all i. Hence, the row sums of W^2 are all strictly less than one. It implies that $\|W^2\|_\infty < 1$ so that $\rho(W) < 1$ and W is convergent. We summarise the discussion using the following lemma.

Lemma 4.3.1. *If f_i is non-expansive for all $i \in N$, W is non-negative, and either W or $W^\top$ is weakly chained substochastic, then Assumption 4.2.1 holds.*

Lemma 4.3.1 implies that some network structure of graph W gives the convergence. For instance, Assumption 4.2.1 holds as long as the network structure is *acyclic* such that there exists no feedback effect: i is not accessible from j whenever j is accessible from i for all $i \neq j$.

LEMMA 4.3.2. *If f_i is Lipschitz continuous for all i, and $W \in \mathbb{R}^{n \times n}$ is such that* graph W *is acyclic, then Assumption 4.2.1 holds.*

For the following lemma, we say that W is irreducible if $\sum_{k=1}^{\infty} W^k \gg 0$.

LEMMA 4.3.3. *If f_i is non-expansive for all $i \in N$, W is non-negative, row (column) substochastic and irreducible, and there is $t \in N$ such that $\sum_j w_{tj} < 1$ ($\sum_j w_{jt} < 1$), then Assumption 4.2.1 holds.*

Eisenberg and Noe (2001), instead, show that the clearing payment is unique if the financial network is regular.[18] In the rest of this section, we apply the proof of Theorem 4.2.2 to show the uniqueness of Eisenberg and Noe (2001) model without regularity or weekly chained substochasticity, which is also shown in Stachurski (2022)[19].

The proof of Theorem 4.2.2 implies that the set of shocks that admit the multiple equilibria is Lebesgue measure zero. If we can further preclude the realisation of such shocks, we can guarantee the uniqueness of equilibrium. Eisenberg-Noe model (4.11) of clearing payment is a special case that it precludes the possibility of such shocks. It can be shown that the multiple equilibria occur only if $\varepsilon \mathbf{1}^\top = 0$ in Eisenberg-Noe network (4.11). Recall that the cash flows are non-negative $\varepsilon \geqslant 0$ in Eisenberg-Noe model. Hence, if we further assume that there exists some $i \in N$ such that $\varepsilon_i > 0$, the equilibrium is always unique.

COROLLARY 4.3.2. *Let f, W and ε follow Eisenberg-Noe model (4.11). That is, $f_j(t) = t\, \mathbb{1}_{\{t<\bar{p}_j\}}(t) + \bar{p}_j\, \mathbb{1}_{\{t \geqslant \bar{p}_j\}}(t)$ for some $\bar{p}_j > 0$ for all j, W is non-negative and row stochastic, and $\varepsilon \geqslant 0$. Then, if $\varepsilon > 0$, the equilibrium is unique.*

---

[18] The *risk orbit* of the bank i is the set that i has a directed path to all nodes in the set. The system is *regular* if any risk orbit has at least one node i with positive cash flow $\varepsilon_i > 0$.

[19] Stachurski (2022) shows the same result beautifully by Du's theorem.

Since by convention we can set the equilibrium to be zero for all agents if $\varepsilon = 0$ in Eisenberg-Noe model, Corollary 4.3.2 implies that the equilibrium is unique for any ε. Therefore, the uniqueness of clearing payment holds without regularity or weekly chained substochasticity. Staum et al. (2016) assume that every bank has strictly positive external assets for the uniqueness of clearing payment. Amini et al. (2016) also assumes that either all banks hold external assets or the total of external assets is nonzero so that the conditions of in Eisenberg and Noe (2001) are satisfied.[20] Corollary 4.3.2 implies that we only need one bank with positive external assets to have unique clearing payments.

4.3.2.1. Extended Eisenberg-Noe Network. Acemoglu et al. (2015a) and Acemoglu et al. (2015b) consider the Eisenberg-Noe network with senior liability and asset liquidation and show that the contagion of financial default depends on both the magnitude of shock and network structure. In their setting, the shock ε_j could be negative. In detail, each bank can liquidate its assets to pay the debt. Let ℓ_j be the liquidation decision for all j. Assume that banks can only recover $\zeta \in [0, 1]$ fraction of the value of a liquidated project. The repayment decision is

$$x_j = \max\left\{\min\left\{\sum_i x_i w_{ij} + c_j + z_j - \nu + \zeta \ell_j, \bar{p}_j\right\}, 0\right\}$$

where c_j is the cash, z_j is the project return, and ν is the senior liability. A bank's ability to fulfil its liability depends on its resources, including its hoarding cash, the received repayments from its debtors, the return of invested projects minus senior liability, and the liquidated asset of a project. Except for the rules of limited liability and absolute priority, banks repay nothing if the total cash flow is negative that $\sum_i x_i w_{ij} + c_j + z_j - \nu + \zeta \ell_j < 0$. Moreover, each bank decides the amount of liquidation:

$$\ell_j = \max\left\{\min\left\{\frac{1}{\zeta}\left(\bar{p}_j - \sum_i x_i w_{ij} - e_j\right), A\right\}, 0\right\}$$

where $e_j = c_j + z_j - \nu$ and A is the total value of the invested project. A Bank can liquidate a fraction of its invested project, with value A to meet the shortfall of liability

---

[20] Amini et al. (2016) studies the equilibrium of Eisenberg-Noe interbank network with asset liquidation, which affects the equilibrium price for the illiquid asset.

$\bar{p}_j - \sum_i x_i w_{ij} - e_j$. Acemoglu et al. (2015a) show that the payment in equilibrium satisfies:[21]

$$x_j = \max\left\{\min\left\{\sum_i x_i w_{ij} + \varepsilon_j,\ \bar{p}_j\right\}, 0\right\} \tag{4.12}$$

where $\varepsilon_j = c_j + \xi_j - \nu + \zeta A$ could be negative. In particular, the interaction functions are

$$f_j(t) = t\,\mathbb{1}_{\{0<t<\bar{p}_j\}}(t) + \bar{p}_j\,\mathbb{1}_{\{t\geqslant\bar{p}_j\}}(t)$$

for all j. Acemoglu et al. (2015a) show that the clearing payment is generically unique for a strongly connected network.[22] Their proof also shows that the clearing payment (4.11) of Eisenberg-Noe network is unique, given a strongly connected network. As shown in Example 4.2.1, there exists confusion that multiple equilibria occur with arbitrarily high probability. Theorem 4.2.2 shows that the equilibrium is unique almost surely if the shock is absolutely continuous. We discuss the computation method in Section 4.4.4.

As we see in Example 4.2.1, the proof of Theorem 4.2.2 implies if there exist multiple equilibria, there must be some strongly connected subgraph.[23] It also shows that all agents who have multiple equilibria must be in or accessible from some strongly connected subgraph, where all agents in this subgraph admit multiple equilibria.[24]

Corollary 4.3.3. *Let (f, W) be such that Assumption 4.2.2 holds. Given shocks $\varepsilon \in E_\varepsilon$, if there are multiple equilibria, then any agents having multiple equilibria must be accessible from some agents in a strongly connected subgraph S, which admits multiple equilibria.*

4.3.2.2. Financial Network with Bankruptcy Cost. This subsection introduces another generalised Eisenberg-Noe model with bankruptcy cost in Glasserman and Young (2015) and Glasserman and Young (2016).[25] Consider a relative liability matrix (w_{ij}) as the previous subsection. Suppose that when each bank $j = 1, \ldots, n$ defaults, its asset is

---

[21] See Lemma B2 of Acemoglu et al. (2015a)

[22] See Definition 4.2.3 for the definition of generic uniqueness.

[23] A graph $S = (V_S, E_S)$ is a *subgraph* of graph A if $V_S \subset V$ and $E_S \subset E$.

[24] To prevent confusion, although we assume absolute continuity of shock in Theorem 4.2.2, we can specify the realized shocks or relax the absolute continuity so that there exists multiplicity, if we want.

[25] The bankruptcy costs include auditing, accounting, and legal costs, and the losses associated with asset liquidation.

further reduced by

$$\alpha_j \left[\bar{p}_j - \left(\sum_i x_i w_{ij} + \varepsilon_j\right)\right]$$

up to a maximum reduction such that the assets are entirely eliminated. In other words, a large shortfall of liability generates a higher bankruptcy cost than a small shortfall. Then, the clearing payment is

$$\begin{aligned} x_j &= \min\left\{\max\left\{\sum_i x_i w_{ij} + \varepsilon_j - \alpha_j \left[\bar{p}_j - \left(\sum_i x_i w_{ij} + \varepsilon_j\right)\right], 0\right\}, \bar{p}_j\right\} \\ &= \min\left\{\max\left\{(1+\alpha_j)\left(\sum_i x_i w_{ij} + \varepsilon_j\right) - \alpha_j \bar{p}_j,\ 0\right\}, \bar{p}_j\right\} \end{aligned} \tag{4.13}$$

Hence, the interaction functions can be written as[26]

$$f_j(t) = ((1+\alpha_j)t - \alpha_j \bar{p}_j)\mathbb{1}_{\{0 \leqslant (1+\alpha_j)t - \alpha_j \bar{p}_j < \bar{p}_j\}}(t) + \bar{p}_j \mathbb{1}_{\{t \geqslant \bar{p}_j\}}(t)$$

for all j, where $\mathbb{1}_A(t)$ is the indicator function with $A \subset \mathbb{R}$.

Following Glasserman and Young (2015) and Glasserman and Young (2016), when $\alpha_i \equiv \alpha$, the clearing payment is unique if $(1+\alpha)\max_i \sum_j w_{ij} < 1$, which implies that $(1+\alpha)W$ is non-singular and Assumption 4.2.1 holds.

4.3.2.3. Financial Network with Equity Insolvency and Illiquidity. Following Eisenberg-Noe model, Liu et al. (2020) consider a financial lending network that banks are exposed to lending and borrowing with different maturities. They show that the U.S. banking network has diminished its system risk of contagion and illiquidity from 2011 to 2014.

In this model, for each bank i, the asset in its balance sheet equals the sum of overnight lending, short-term lending, long-term lending, cash and cash equivalents ε_i and other assets OA_i, while the liability consists of overnight borrowing, short-term borrowing, other liability OL_i and equity E_i. Each period t, bank i has an obligation to repay some fraction of overnight, short-term, and long-term liability δ_{ij} to bank j. The total liability obligation is $\bar{p}_i$ in the period t. The relative liability matrix (w_{ij}) is defined as before that $w_{ij} = \delta_{ij}/\bar{p}_i$ if $\bar{p}_i > 0$ and $w_{ij} = 0$ otherwise, so that (w_{ij}) is non-negative. Assume that

---

[26] For the last term, $\mathbb{1}_{\{(1+\alpha_j)t - \alpha_j \bar{p}_j \geqslant \bar{p}_j\}}(\bar{p}_j) = \mathbb{1}_{\{t \geqslant \bar{p}_j\}}(\bar{p}_j)$.

$\sum_j w_{ij} \leqslant 1$ following Liu et al. (2020). Let Q_{ij} be the remainder of all loan obligations that i has to repay j, including overnight market, short-term and long-term loans, at the end of the period. Let x_i be the realized payment made at the end of the period. Define the equity

$$E_i = \sum_h x_h w_{hi} + \varepsilon_i - \bar{p}_i + \left(\sum_h Q_{hi} - \sum_j Q_{ij} \right) + (OA_i - OL_i)$$

Denote $B_i := \sum_h Q_{hi} - \sum_j Q_{ij} + OA_i - OL_i$ as the net remaining and other assets. Each bank fails to repay in full if either it is illiquid due to insufficient cash and incoming payment or it is insolvent so its equities are negative $E_i < 0$. The payment in equilibrium satisfies:

$$x_i = \min \left\{ \left[\sum_h x_h w_{hi} + \varepsilon_i \right]^+ , \left[\sum_h x_h w_{hi} + \varepsilon_i + B_i \right]^+ , \ \bar{p}_i \right\} \tag{4.14}$$

where $[z]^+ := \max\{z, 0\}$ for $z \in \mathbb{R}$. The interaction functions are

$$f_i(t) = \min \left\{ [t]^+ , [t + B_i]^+ , \ \bar{p}_i \right\}$$

for all i. Liu et al. (2020) show that the equilibrium payment exists. We further show that it is almost surely unique when W is stochastic and unique when W is convergent. We have $0 \leqslant W \leqslant 1$ and for agent j

$$\begin{aligned} x_j &= \min \left\{ \left[\sum_i x_i w_{ij} + \varepsilon_j \right]^+ , \left[\sum_i x_i w_{ij} + \varepsilon_j + B_j \right]^+ , \ \bar{p}_j \right\} \\ &= \begin{cases} \min \left\{ \left[\sum_i x_i w_{ij} + \varepsilon_j \right]^+ , \ \bar{p}_j \right\} & \text{if } B_j \geqslant 0, \\ \min \left\{ \left[\sum_i x_i w_{ij} + \varepsilon_j + B_j \right]^+ , \ \bar{p}_j \right\} & \text{otherwise.} \end{cases} \end{aligned}$$

Then, the interaction function is

$$f_j(z) = \begin{cases} z \mathbb{1}_{\{0 \leqslant z < \bar{p}_j\}}(z) + \bar{p}_j \mathbb{1}_{\{z \geqslant \bar{p}_j\}}(z) & \text{if } B_j \geqslant 0, \\ (z + B_j) \mathbb{1}_{\{0 \leqslant z + B_j < \bar{p}_j\}}(z) + \bar{p}_j \mathbb{1}_{\{z + B_j \geqslant \bar{p}_j\}}(z) & \text{otherwise.} \end{cases}$$

Clearly, f_j is increasing and bounded for all j. By figure 2, we know that the interaction function is non-expansive for either $B_j \geqslant 0$ or $B_j < 0$. When the shock is absolutely continuous, Theorem 4.2.2 shows that the clearing payment is unique almost surely when

W is stochastic. On the other side, if W is weekly chained substochastic or $\rho(W) < 1$, the clearing payment is unique by Theorem 4.2.1.

4.4. Comparative Statics, Tightness and Boundedness

In this section, we first study the comparative statics of how the increase in interaction functions, sensitivity matrix, and shocks affect the equilibrium. We further investigate the tightness of the conditions of Assumption 4.2.1 and the requirement of boundedness in Assumption 4.2.2. We discuss an example to show that the spectral radius condition of Assumption 4.2.1 could be also a necessary condition. We then argue the boundedness condition is essential in Assumption 4.2.2. In Section 4.4.4, we consider an algorithm to compute the equilibrium when the interaction functions are bounded identity maps as financial network (4.12).

4.4.1. Comparative Statics. This section presents some simple comparative statics between two networks. We show that the equilibrium is increasing in the shock, ε, holding all other things constant. Moreover, *ceteris paribus*, when a network has a dominant interaction function, it has a greater equilibrium. We also see that the rise in strength of interactions also increases the equilibrium. These results are valid under both Assumption 4.2.1 and Assumption 4.2.2. In detail, we have the following two lemmas.

LEMMA 4.4.1. *Let (f, W, ε) and (f', W', ε') be two networks satisfying Assumption 4.2.1, and denote their corresponding equilibrium as $\hat{x}$ and $\hat{x}'$, respectively. If f_i and f_i' are increasing functions for all $i \in N$, $f_i(t) \leqslant f_i'(t)$ for all $t \in \mathbb{R}$ and all i, $W \leqslant W'$, and $\varepsilon \leqslant \varepsilon'$, then $\hat{x} \leqslant \hat{x}'$.*

LEMMA 4.4.2. *Let (f, W, ε) and (f', W', ε') be two networks satisfying Assumption 4.2.2 such that they have unique equilibrium, denoted by $\hat{x}$ and $\hat{x}'$, respectively. Suppose that for all i we have $f_i(t) \leqslant u_i$ and $f_i'(t) \leqslant u_i'$ for all $t \in \mathbb{R}$ such that $u_i \leqslant u_i'$. If $f_i(t) \leqslant f_i'(t)$ for all $t \in \mathbb{R}$ and all i, $W \leqslant W'$, and $\varepsilon \leqslant \varepsilon'$, then $\hat{x} \leqslant \hat{x}'$.*

These two lemmas imply that the equilibrium is increasing in the shock, the sensitivity matrix, or the interaction functions. When we add edges or increase the weight of edges in a network, since the interaction functions are increasing, the strengthened interconnections lead to greater equilibrium.

About Lemma 4.4.2, note that it is possible that for two sensitivity matrices $W, W' \in \mathbb{R}^{n\times n}$ we have $\rho(W) = \rho(W') = 1$ while $W \leqslant W'$. For example, the sensitivity matrices in Example 4.4.1 both have spectral radius one. However, if we restrict the sensitivity matrix to a stochastic matrix, then we should fix $W = W'$ in Lemma 4.4.2, otherwise the row or column sum is not one anymore.

EXAMPLE 4.4.1. Consider the following specifications for f, W and ε as

$$f_i^a(t) := \min\{\max\{t, 0\}, 2\}, \quad \forall i \qquad \varepsilon^a := (0.2, -0.6, -0.2, 0.2)$$

$$f_i^a(t) := \min\{\max\{t, 0.1\}, 2\}, \quad \forall i \qquad \varepsilon^b := (0.2, 0, -0.2, 0.2),$$

and

$$W^a := \begin{pmatrix} 0 & 2 & 0 & 0 \\ 0.5 & 0 & 0.5 & 0 \\ 0 & 0 & 0 & 0.8 \\ 0 & 0 & 0.8 & 0 \end{pmatrix} \qquad W^b := \begin{pmatrix} 0 & 2 & 0.1 & 0.8 \\ 0.5 & 0 & 0.8 & 0.1 \\ 0 & 0 & 0 & 0.9 \\ 0 & 0 & 0.9 & 0 \end{pmatrix}. \tag{4.15}$$

The equilibrium for $(f^a, W^a, \varepsilon^a)$ is $(0.2, 0, 0, 0.2)$, the equilibrium for $(f^b, W^a, \varepsilon^a)$ is $(0.25, 0.1, 0.1, 0.28)$, the equilibrium for $(f^a, W^b, \varepsilon^a)$ is $(0.2, 0, 0.7579, 1.0421)$, the equilibrium for $(f^a, W^a, \varepsilon^b)$ is $(1.2, 2, 2, 1.8)$, and the equilibrium for $(f^b, W^b, \varepsilon^b)$ is $(1.2, 2, 2, 2)$. Last, for the network with both W^a and W^b, the multiplicity exists if $\varepsilon \in \mathbb{R}^4$ satisfies $2\varepsilon_1 + \varepsilon_2 = 0$. This condition does not hold almost surely if the shock is absolutely continuous. □

4.4.2. Tightness of Condition. In some cases, the spectral radius condition in Assumption 4.2.1, $\rho(|W| \operatorname{diag}(\beta)) < 1$, can be the necessary condition for the existence and uniqueness of equilibrium. For example, consider the system with two agents such that the interaction function is $f_i(t) = \sqrt{t^2}$ for all $t \in \mathbb{R}$ for all $i = 1, 2$, and the realisation of shocks are positive $\varepsilon_1, \varepsilon_2 > 0$. Assume that agents' states do not affect themselves and have a positive influence on each other, such that $w_{11} = w_{22} = 0$ and $w_{12}, w_{21} > 0$:

$$W = c \begin{pmatrix} 0 & 1 \\ \lambda & 0 \end{pmatrix}$$

where $\lambda, c > 0$. Then, the interaction functions are Lipschitz continuous with a Lipschitz constant $\beta_i = 1$ for $i = 1, 2$. Further, we can see that the eigenvalues of W are $\pm c\sqrt{\lambda}$, whence the spectral radius is less than one, $\rho(W) < 1$, if and only if $c^2\lambda < 1$. Now, the equilibrium follows

$$\begin{aligned} x_1 &= \sqrt{(c\lambda x_2 + \varepsilon_1)^2} = c\lambda x_2 + \varepsilon_1 \\ x_2 &= \sqrt{(cx_1 + \varepsilon_2)^2} = cx_1 + \varepsilon_2. \end{aligned}$$

for $x_1, x_2 \in [0, \infty)$. Since we can show that this system has a unique solution if and only if $c^2\lambda < 1$, the condition $\rho(W) < 1$ is a necessary condition for the existence and uniqueness of equilibrium.

As another example, considering the linear system $x = xW + \varepsilon$, the equilibrium could exist and be unique when $\rho(W) > 1$, but $T\colon x \mapsto f(xW + \varepsilon)$ is not globally stable. In some applications, we want the equilibrium to be non-negative or positive. For instance, the equilibrium in the production networks (4.4) and (4.5) should be non-negative. Since $(I - W)^{-1} \geqslant 0$ if and only if $\rho(W) < 1$, Assumption 4.2.1 is sufficient and necessary for non-negative equilibrium.[27]

In some other cases, the spectral radius condition is not a necessary condition for the existence. The equilibrium may still exist when $\rho(|W| \operatorname{diag}(\beta)) \geqslant 1$. However, the system is not globally stable anymore when the spectral radius is greater than one. Also, the uniqueness of equilibrium could fail. For example, suppose that the equilibrium satisfies

$$(x_1, x_2) = \min\left\{ \max\left\{ \left[(x_1, x_2) \begin{pmatrix} 0 & 2 \\ 3 & 0 \end{pmatrix} + (\varepsilon_1, \varepsilon_2) \right], (0, 0) \right\}, (5, 5) \right\}.$$

The multiplicity and global stability depend on the shocks. For example, if we have $(\varepsilon_1, \varepsilon_1) = (-6, 2)$, then there are multiple equilibria $x^* = (0, 2)$ or $(5, 5)$. More, it is not

---

[27] See Lemma 4.7.7.

globally stable, since some iteration may not converge, say, the iteration starting from $(0, 5)$.

4.4.3. Linear System and Boundedness Condition. This part attempts to argue that when the interaction function is non-expansive and the sensitivity matrix is not convergent, the boundedness is essential for existence of the equilibrium. That is, the boundedness condition in Assumption 4.2.2 cannot be precluded. We use a linear system to illustrate this concept. In particular, if the interaction system is an identity mapping, and the sensitivity matrix has spectral radius one $\rho(W) = 1$, then the equilibrium does not exist with probability one.

In detail, consider a linear system:

$$x = xW + \varepsilon \tag{4.16}$$

where x and ε are vectors in $\in \mathbb{R}^n$ and $W \in \mathbb{R}^n \times \mathbb{R}^n$ is a non-negative matrix. By Theorem 4.2.2, we know that the linear system has a unique solution x if the spectral radius of W is less than one. For instance, the study of input-output analysis assumes that the producers have positive value-added and then that $\sum_j w_{ij} < 1$ for all i (Antràs et al., 2012). Note that (4.16) may not have the solution if the scalar one is the eigenvalue of W, since $(I - W)$ is not invertible. If one is not the eigenvalue of W, then the solution always exists and is unique. Therefore, when the sensitivity matrix is stochastic, the solution does not exist almost surely. Similar to almost sure uniqueness, we define almost sure non-existence as follows.

Definition 4.4.1. Let E denote the set of shocks that the solution exists (i.e., $E \coloneqq \{\varepsilon \in E_\varepsilon$: Equation (4.16) has a solution$\}$.) We say that the solution *does not exist almost surely* if $\text{Prob}(\varepsilon \in E) = 0$.

Lemma 4.4.3. *If W is non-negative and $\rho(W) = 1$, and the shocks (ε_i) are absolutely continuous, then the solution of linear system (4.16) does not exist almost surely.*

Comparing Theorem 4.2.2 and Lemma 4.4.3, we see that boundedness plays a key role in guaranteeing the existence of equilibrium. Boundedness also helps to pin down the uniqueness of equilibrium. We provide some intuition as below. Consider again the interaction function of a bounded identity map (4.3) and a row stochastic and irreducible matrix W. From the proof of Lemma 4.7.4, we see that the multiple equilibria x must satisfy $-M\mathbf{1} \leqslant x = xW + \varepsilon \leqslant M\mathbf{1}$. Since W is stochastic, $\mathbf{1}$ is the right eigenvector, whence we have $\varepsilon\mathbf{1}^\top = (x - xW)\mathbf{1}^\top = 0$ if there exists multiplicity. To notice how boundedness pins down the unique solution, suppose on the contrary that $\varepsilon\mathbf{1}^\top \neq 0$. If x is a solution in $[-M\mathbf{1}, M\mathbf{1}]$, then we have $x\mathbf{1}^\top = xW\mathbf{1}^\top + \varepsilon\mathbf{1}^\top = x\mathbf{1}^\top + \varepsilon\mathbf{1}^\top \neq x\mathbf{1}^\top$, a contradiction. Thus, we have $xW + \varepsilon \nleqslant M\mathbf{1}$ or $-M\mathbf{1} \nleqslant xW + \varepsilon$. Without loss of generality, assume that node i is such that $(xW + \varepsilon)_i > M$, whence i's equilibrium state equals M. Remove such node i from the graph and consider the shock $\varepsilon_s + x_i w_{is}$ for all remaining agents $s \neq i$. Let $\overline{W}$ be the submatrix of W by removing the column i and row i from W. Using Lemma 4.7.9, we have $\rho(\overline{W}) < 1$. Therefore, the remaining network with vertices $N \setminus \{i\}$ has a unique solution by Theorem 4.2.1. We see that when $\varepsilon\mathbf{1}^\top \neq 0$ in this example, the boundedness pins down the unique solution.

4.4.4. Algorithm for Bounded Identity Map. In this section, we provide an algorithm to compute the equilibrium under the bounded identity maps of interaction functions. We show that the algorithm converges in at most $n2^{n-1}$ iterations. From the discussion in Section 4.4.3, we know that the boundedness condition is the key to the uniqueness of equilibrium. It implies that we can compute the equilibrium by assuming that there exists an agent whose state is always equal to the upper or lower bound of her sensitivity function. In detail, consider the interaction functions

$$f_j(t) = \min\left\{\max\left\{t, \ell_j\right\}, u_j\right\} \tag{4.17}$$

for all $j \in N = \{1, \ldots, n\}$, where $u_i > \ell_j$ for all $j \in N$. The bounded identity maps are the generalised Eisenberg-Noe model (4.12). Assume that the sensitivity matrix W is row or column stochastic as the financial network. Following the discussion in Section 4.4.3, we conclude the lemma below. Denote $u = (u_i)$ and $\ell = (\ell_i)$.

Lemma 4.4.4. *Let u, ℓ be such that $u \gg \ell$, f be defined as (4.17), and $W \geqslant 0$ be row/column stochastic. Given ε, if the equilibrium x^* is unique, then there is $j \in N$ such that either $x_j^* = u_j$ or $x_j^* = \ell_j$.*

In fact, for any strongly connected subgraph $G_s \subset \text{graph}\, W$, we can show that there exists an agent j in subgraph G_s such that $x_j^* = u_j$ or $x_j^* = \ell_j$. Lemma 4.4.4 implies that every agent's equilibrium state is either at the upper bound, lower bound, or in between. Since there are n agents in the network, we have 3^n possibilities. Since we can pin down the equilibrium for those agents at the boundedness, we only need to decide the equilibrium for the remaining agents. Using this idea, we are able to design an algorithm that converges to the equilibrium in at most 3^n iterations. Algorithm 2 applies this concept to compute the equilibrium in finite iterations.

To illustrate the algorithm, given the states x, define the sets $A(x)$ and $B(x)$ as

$$\begin{aligned} A(x) &:= \{j \in N \colon \sum_i x_i w_{ij} + \varepsilon_j \geqslant u_j\}, \\ B(x) &:= \{j \in N \colon \sum_i x_i w_{ij} + \varepsilon_j \leqslant \ell_j\}. \end{aligned} \tag{4.18}$$

In words, $A(x)$ (resp., $B(x)$) is the set of agents whose equilibria are greater (resp., less) than or equal to the upper (resp., lower) bounds at states x. Also, define the diagonal matrix Λ^D for $D \subset N$ as

$$\Lambda_{ij}^D := \begin{cases} 1 & i = j \text{ and } i \in D, \\ 0 & \text{otherwise}, \end{cases}. \tag{4.19}$$

Hence, $\Lambda^{A(x)}$ and $\Lambda^{B(x)}$ indicate which agents are at the upper and lower bound, respectively.

In the financial network, $N \setminus A(x)$ is the set of banks that default under the clearing payment x, while $N \setminus B(x)$ is the set of banks that are able to make some payments under the clearing payment x. Inspired by Eisenberg and Noe (2001), Algorithm 2 returns the equilibrium $x^{(t)}$ given the interaction function (4.17). The inner for-loop of Algorithm 2 is the fictitious default iteration introduced by Eisenberg and Noe (2001). If we have $B(\ell) = \emptyset$, then Algorithm 2 is simply the fictitious default algorithm in Eisenberg and

Noe (2001), which converges in at most n iterations. The outer for-loop searches the set of agents whose equilibria are at the lower bounds, $B(x^*)$, from the potential candidates in the power set of $B(\ell)$, $\mathcal{P}(B(\ell))$. At some $t \in \mathbb{N}$, when the guess is correct: $P_i = B(x^*) = B^*$ for $P_i \in \mathcal{P}(B(\ell))$, we iterate the solutions from $\hat{u} := u(I - \Lambda^{P_i}) + \ell\Lambda^{P_i}$, and set $A_{t-1} = A(\hat{u}) \cap (N \setminus B(x^*))$. Next, if $P_i = b^*$, step 11 is equivalent to set $x^{(t)}$ as

$$x_j^{(t)} = \begin{cases} u_j & \forall j \in A_{t-1} \\ \ell_j & \forall j \in B^* \\ \sum_{i \in A_{t-1}} u_i w_{ij} + \sum_{i \in B^*} \ell_i w_{ij} + \sum_{i \in N \setminus (A_{t-1} \cup B^*)} x_i^{(t)} w_{ij} + \varepsilon_j & \text{otherwise.} \end{cases}$$

Note that since $A_{t-1} \supset A(x^*)$, the matrix $I - (I - \Lambda^{A_{t-1}} - \Lambda^{P_i})W(I - \Lambda^{A_{t-1}} - \Lambda^{P_i})$ in step 11 is non-singular so that $x^{(t)}$ is unique:[28]

$$\begin{aligned} x^{(t)} = [((u\Lambda^{A_{t-1}} + \ell\Lambda^{P_i})W + \varepsilon)(I - \Lambda^{A_{t-1}} - \Lambda^{P_i}) \\ + u\Lambda^{A_{t-1}} + \ell\Lambda^{P_i}][I - (I - \Lambda^{A_{t-1}} - \Lambda^{P_i})W(I - \Lambda^{A_{t-1}} - \Lambda^{P_i})]^{-1}. \end{aligned}$$

The solution $x^{(t)}$ is the equilibrium when $x_i = u_i$ for $i \in A_{t-1}$ and $x_i = \ell_i$ for $i \in B^*$. In the next step, we check whether $A_t = A(x^{(t)})$ equals A_{t-1} or not. If they are equal, the algorithm terminates and returns x^t as the equilibrium; otherwise, we repeat the step 11 of Algorithm 2 to get $x^{(t+1)}$ with the updated set A_t. Conversely, if the guess from the power set of $B(\ell)$ is not correct, then it may be the case that it raises a singular matrix error in step 11. In this case, we skip it and try another guess from $\mathcal{P}(B(\ell))$. Therefore, the convergence time depends on how many agents in $B(\ell)$. In general, the next lemma shows that Algorithm 2 converges in at most $n2^{n-1}$ iterations.[29]

Lemma 4.4.5. *Let f follow (4.17), $W \geqslant 0$ be column/row stochastic, and ε be such that the equilibrium is unique. Algorithm 2 returns the equilibrium $x^{(t)}$ in at most $n2^{n-1}$ iterations.*

---

[28] We can use the iterative method to approximate the solution.

[29] In some cases, the equilibrium x^* has the features that $x^* \gg \ell$ or $x^* \ll u$. If this is the case, then we can save time by implement the Eisenberg-Noe iteration and the reverse Eisenberg-Noe iteration firs.

Algorithm 2: Compute equilibrium given the interaction functions (4.17).

```
$t \leftarrow 0$;
if $A(u) = N$ then return $x^{(0)} \leftarrow u$ ;
else if $B(\ell) = N$ then return $x^{(0)} \leftarrow \ell$;
$\mathcal{P} \leftarrow$ the power set of $B(\ell)$;
for $i = 0;\ i < |\mathcal{P}|;\ i = i + 1$ do
    $\hat{u} \leftarrow \ell\Lambda^{P_i} + u(I - \Lambda^{P_i})$, where $P_i \subset B(\ell)$ is the $i$-th element of $\mathcal{P}$;
    $A_t \leftarrow A(\hat{u}) \cap (N \setminus P_i)$;
    for $j = 0;\ j < n;\ j = j + 1$ do
        $t \leftarrow t + 1$;
        try
            $x^{(t)} \leftarrow$ the fixed point of $x = u\Lambda^{A_{t-1}} + \ell\Lambda^{P_i} + \{[u\Lambda^{A_{t-1}} + \ell\Lambda^{P_i} + x(I - \Lambda^{A_{t-1}} - \Lambda^{P_i})]W + \varepsilon\}(I - \Lambda^{A_{t-1}} - \Lambda^{P_i})$;
        except singular matrix error break ;
        $A_t \leftarrow A(x^{(t)})$;
        if $A_t = A_{t-1}$ then break;
    if $f(x^{(t)}W + \varepsilon) = x^{(t)}$ then return $x^{(t)}$;
```

Example 4.4.2. We consider a numerical example to demonstrate Algorithm 2. Let the network (f, W, ε) be

$$f = \min\{\max\{xW + \varepsilon, 0\}, u\} \text{ where } u = (5, 10, 10, 8, 10, 10, 6)$$

$$W = \begin{pmatrix} 0 & 0.4 & 0.15 & 0 & 0.4 & 0.05 & 0 \\ 0.4 & 0 & 0.15 & 0.25 & 0 & 0.2 & 0 \\ 0.3 & 0.1 & 0 & 0.25 & 0.15 & 0.2 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix} \tag{4.20}$$

$$\varepsilon = 10^{-5}(2, 1, -1, 3, 2, -1, -2).$$

First, compute $B(\ell) = B(0) = \{3, 6, 7\}$ and $A(u) = \{1, 4, 5, 6, 7\}$. Then, the power set $\mathcal{P}(B(\ell))$ has 2^3 elements. When P_i in Algorithm 2 equals $\{3, 6, 7\}$, we have $\hat{u} = (5, 10, 0, 8, 10, 0, 0)$ and $A(\hat{u}) \cap (N \setminus P_i) = \{4, 5\}$. The iterations of Algorithm 2 give:

$x^{(t)}$	A_t
$\hat{u} = (5, 10, 0, 8, 10, 0, 0)$	$\{4, 5\}$
$x^{(1)} = (2.857 \times 10^{-5}, 2.143 \times 10^{-5}, 0, 8, 1, 0, 0)$	$\{4\}$
$x^{(2)} = (2.857 \times 10^{-5}, 2.143 \times 10^{-5}, 0, 8, 8.00003, 0, 0)$	$\{4\}$

In this case, since $|\mathcal{P}| = 2^3$ and $|N \setminus P_i| = 4$, the algorithm converges to the equilibrium in at most 2^5 iterations. Alternatively, if we use the operator T of (4.2) to iterate from the upper bound to compute equilibrium $\lim_{m\to\infty} T^m u$, the iteration time is more than 4.9×10^5 given the convergence tolerance 10^{-5}. If we iterate from the lower bound $\lim_{m\to\infty} T^m \ell$, the iteration time is more than 2.3×10^5. $\square$

Alternatively, we can use (nonlinear) programming to solve the problem (4.17) as Eisenberg and Noe (2001). Let $g\colon \mathbb{R}^n \to \mathbb{R}$ be a strictly increasing function. Define the programming problem as

$$\begin{aligned} &\max_{x\in[\ell,u]} g(x) \\ \text{subject to} \qquad & 0 \leqslant \max\{xW + \varepsilon - x, \ell - x\}. \end{aligned} \tag{4.21}$$

As shown in Eisenberg and Noe (2001), the solution to (4.21) is the (almost surely) unique equilibrium.[30]

Lemma 4.4.6. *Let f follow (4.17), $W \geqslant 0$ be stochastic, and $\varepsilon \in E_\varepsilon$. If g is strictly increasing, then any solution to programming problem (4.21) is an equilibrium.*

4.5. Key Players

In this section, we utilize the unique equilibrium to identify the most influential agent. There are many measures or centralities that evaluate the importance scores for agents in a network.[31] For instance, in input-output analysis, the output multiplier measures the overall output impact of a sector when it has a dollar-worth increase in final demand, so we can use the output multiplier to identify the most influential production sector, and then policymakers can decide which sector to bail out during recession (Miller and Blair, 2009). In network games, Ballester et al. (2006) define the "key player" as the agent that has the highest total impact on the aggregate activity once she is removed from the

---

[30] The solution to the following programming problem is also an equilibrium.

$$\min_{x\in[\ell,u]} g(x) \qquad \text{subject to } \min\{xW + \varepsilon - x, u - x\} \leqslant 0,$$

where $g\colon \mathbb{R}^n \to R$ is strictly increasing.

[31] See Das et al. (2018) for a survey of centralities.

network.[32] We provide a measure for identifying the key player by casting an equilibrium to the steady state of continuous-time dynamics, following and generalizing the control analysis in Sharkey (2017).

Observe that the equilibrium of the interaction system (4.1) can be interpreted as the steady state of the following continuous-time dynamics:

$$\frac{\mathrm{d}x}{\mathrm{d}t} = f(xW + \varepsilon) - x \tag{4.22}$$

where $x \in \mathbb{R}^n$ are economic states and $f(xW + \varepsilon) = (f_j(\sum_i x_i w_{ij} + \varepsilon_j))_{j=1}^n$ is defined as before. The *steady state* x^* of (4.22) is such that $f(x^*W + \varepsilon) - x^* = 0$. Suppose that the interaction functions are increasing. In this dynamics, an agent's equilibrium increases in others' equilibrium $f(xW + \varepsilon)$ and decreases in the amount of itself x. Clearly, if y is the equilibrium of model (4.1), then it is the steady state to the continuous-time dynamic system (4.22) that $\mathrm{d}x/\mathrm{d}t = f(yW + \varepsilon) - y = 0$. If Assumption 4.2.1 holds and then the equilibrium is unique by Theorem 4.2.1, then the steady state is also unique.

In this section, suppose that f_i is differentiable for all i. For some networks, including equation (4.12), the interaction function is non-differentiable. In this case, we can approximate the interaction function by a smooth function without loss of economic meaning.

Like Sharkey (2017), we know that the contraction condition in Theorem 4.2.1, $\rho(|W| \operatorname{diag}(\beta)) < 1$, also deliveries a stable continuous-time dynamics by Lemma 4.5.1.

Let $\mathrm{d}x\,/\mathrm{d}t = F(x(t))$ and $x(0) = x^0$ be an autonomous system, where $x(t) \in \mathbb{R}^n$ denotes the state vector, and $F\colon \mathbb{R}^n \to \mathbb{R}^n$ is a differentiable function of $x(t)$. Denote x^* as the steady state such that $F(x^*) = 0$. Recall that the steady states x^* of an autonomous system is *stable* if for every $\varepsilon > 0$ there exists $\delta > 0$ such that $\|x(0) - x^*\| < \delta$ implies $\|x(t) - x^*\| < \varepsilon$ for all $t \geqslant 0$. The steady state x^* is *asymptotically stable* if it is stable and there is $\delta > 0$ such that $\|x(0) - x^*\| < \delta$ implies $x(t) \to x^*$ as $t \to \infty$.

LEMMA 4.5.1. *Suppose that f_i is increasing and continuously differentiable for all i, and $\rho(|W| \operatorname{diag} \beta) < 1$. Then the dynamic system (4.22) is asymptotically stable.*

---

[32] They show that the key player has the highest intercentrality, a measure defined by Katz (Bonacich) centrality.

Again, if W is non-negative and $\beta_i \equiv \psi$, we can reduce the condition for asymptotic stability to $\psi\rho(W) < 1$. To have an asymptotically stable system, assume that $\rho(|W|\operatorname{diag}(\beta)) < 1$ in this section.

Denote the symbol $\circ$ as the Hadamard product such that $s \circ x := [s_1x_1, \ldots, s_nx_n]$ for $s, x \in \mathbb{R}^n$. Define an alternative continuous-time dynamics as

$$\frac{\mathrm{d}x}{\mathrm{d}t} = F(x, s) := f(xW + \varepsilon) - s \circ x \tag{4.23}$$

Here, the coefficients s specify the small shocks to the agents.

When $s = \mathbf{1}$, equation (4.23) is equal to the original system. Let x^* be the steady state for $s = \mathbf{1}$ that $F(x^*, \mathbf{1}) = 0$. Assume that $s = \mathbf{1}$ so that the analysis is around the equilibrium when there is fluctuation in the system. Following the definition of key players, we remove agent i from the dynamics (4.23) while the others are holding the same. The equivalent shock to the removal of i is

$$\frac{\partial s_i}{\partial x_i^*} x_i^*$$

The impact to the other agent j's steady state is then given by:

$$C_{ij} = \frac{\mathrm{d}x_j^*}{\mathrm{d}s_i} \frac{\partial s_i}{\partial x_i^*} x_i^*$$

where $\mathrm{d}x_j^*/\mathrm{d}s_i$ is the extent of change in j's steady state responding to the shocks. Hence, C_{ij} measures the impact on agent j when the shock is equivalent to the removal of i. With everything else remaining the same, the total impact of the removal of i is equal to $\sigma_i := \sum_j C_{ij}$. We evaluate the total impact around the steady state $(x, s) = (x^*, \mathbf{1})$ in the following lemma. Denote $f_i'(x^*W + \varepsilon) := (f_i'(\sum_h x_h^* w_{hi} + \varepsilon_i))_i$.

Lemma 4.5.2. *If f_i is differentiable for all i and $\rho(|W|\operatorname{diag}(\beta)) < 1$, then the total impact is $\sigma = \mathbf{1}\left[I - \operatorname{diag}(f'(x^*W + \varepsilon))W^\top\right]^{-1} \operatorname{diag}(x^*)$.*

Therefore, the total impact, when agents are removed in the continuous-time dynamics, is determined by the steady state or equilibrium, the network structure W, and the derivative of interaction function. The key player is the agent with the highest σ_i. The term $\operatorname{diag}(x^*)$ implies that the larger the agent's equilibrium state is, the higher the measure σ_i

is. Moreover, the first term, $\mathbf{1}[I-\mathrm{diag}(f'(x^*W+\varepsilon))W^\top]^{-1}$ implies that the impact measure depends on the interaction behavior $f'(x^*W+\varepsilon)$ and the network connections, $W^\top$. From the Neumann series, we can see that the more interconnected agents tend to have higher impact measure σ_i.[33] Also, $\mathbf{1}[I-\mathrm{diag}(f'(x^*W+\varepsilon))W^\top]^{-1}$ can be view as the authority-based Katz centrality adjusted by the interaction functions $\mathrm{diag}(f'(x^*W+\varepsilon))$. Hence, the measure σ captures either the too-big-to-fail or too-interconnected-to-fail agents.

Note that unlike Ballester et al. (2004) this control analysis does not change the network structure (i.e., W is the same after removing an agent.) On the other have, Ballester et al. (2004) assume that the sensitivity matrix is symmetric in the network (4.7), while the control analysis in continuous-time system allows arbitrary network structure.

One issue for the measure σ is that when $\mathrm{diag}\,(f'(x^*W+\varepsilon))W^\top$ has constant column sums, the measure is collapsed to $\sigma = x^*$, which is just the comparison among equilibria in magnitude.

In many applications, the derivative of the interaction function is constant. In this case, we can see that the measure σ evaluates both effects of the agents' impact on others and the perturbations received from others. We illustrate this property by the following example.

Example 4.5.1. Given the network $x = \alpha x W + \mathbf{1}$ for some $\alpha \in \mathbb{R}$, the equilibrium is $(I-\alpha W)^{-1}$ and $f_i' \equiv \alpha$. The measure for key player σ is reduced to

$$\sigma = \mathbf{1}\left(I-\alpha W^\top\right)^{-1}\mathrm{diag}\,(\mathbf{1}(I-\alpha W)^{-1}).$$

Define the hub-based Katz centrality as $\kappa_h := \mathbf{1}(I-\alpha W)^{-1}$, and the authority-based Katz centrality as $\kappa_a := \mathbf{1}(I-\alpha W^\top)^{-1}$. Then, the total impact is the element-wise multiplication of hub-based Katz centrality and authority-based Katz centrality.

$$\sigma = \kappa_a \circ \kappa_h.$$

---

[33] Neumann series: $[I-A]^{-1} = I + A + A^2 + A^3 + \cdots$ for a matrix A. A highly interconnected agent also tends to have high equilibrium x^*.

As the explanation in Sharkey (2017), the hub-based Katz centrality measures the "receiver" property that agents are affected by others, and the authority-based Katz centrality describes the "sender" property that agents influence others. The agent has high σ_i if either she is influenced significantly by others or she propagates shocks and affects others significantly.

Similarly, the network game (4.9) of the interbank lending market has the equilibrium $x^* = \varepsilon \ \operatorname{diag}(\varphi)[I - W \operatorname{diag}(\varphi)]^{-1}$, where $\varphi = (\varphi_i)$. The key player measure is

$$\sigma = \big(\mathbf{1}[I - \operatorname{diag}(\varphi)W^\top]^{-1}\big) \circ \big(\varepsilon \ \operatorname{diag}(\varphi)[I - W \operatorname{diag}(\varphi)]^{-1}\big)$$

It also illustrates the "receiver" and "sender" effect of shock transmission, where the receiver effect is weighted by $\varepsilon \operatorname{diag}(\varphi)$. Due to the term $\varepsilon \operatorname{diag}(\varphi) = \theta I - \operatorname{diag}(c_0)$ with $c_0 = (c_{0,i})$, if a bank i has lower marginal cost $c_{0,i}$ when it has no links to other banks, then it tends to have higher impact measure. Overall, the equation implies that the banks with significant impact measure may be either the too-big-to-fail or too-interconnected-to-fail. ☐

4.6. Conclusion

We demonstrate the (almost sure) uniqueness of equilibrium for a generalised and unified network model. The uniqueness of equilibrium holds if either the interaction functions or the sensitivity matrix has a contraction property such that the corresponding spectral radius is less than one. Alternatively, if the interaction functions are non-expansive and bounded, and the sensitivity matrix is non-convergent (spectral radius of one), then the equilibrium is unique almost surely, given absolutely continuous shocks. Moreover, we show that if the interaction functions are non-expansive, the boundedness of the interaction functions is essential to determine the existence and uniqueness of equilibrium. Using this idea, we can compute the equilibrium of the generalised Eisenberg-Noe interbank network in finite steps by checking which agents are at the bounds. Regarding systemic stability, we illustrate a measure for identifying key players in a unified network by interpreting the equilibrium as the steady state of a continuous-time dynamic system and computing the total impact of removing an agent. The measure possesses the desired properties to evaluate the impact of both receiving and broadcasting perturbations. Since both the magnitude of economic states in equilibrium and the strength of interconnection affect the measure, it helps identify either too-big-to-fail or too-interconnected-to-fail agents.

4.7. Appendix

Intuition of Multiple Equilibria. In this appendix, we explain some intuition for Example 4.2.1 in Section 4.2.2 with n agents that admits multiple equilibria. We suppose that $\sum_j w_{ij} = 1$ for all i and $\sum_i \varepsilon_i = 0$ as Example 4.2.1. The interaction function f is the bounded identity mapping (4.3) broadcasted to the vectors.

The row-sum assumption imply that $W\mathbf{1}^\top = \mathbf{1}^\top$, where $\mathbf{1}$ is a row vector of ones. We also assume that the matrix W is irreducible as Example 4.2.1. Suppose that there exists one solution x such that $-M\mathbf{1} \leqslant x = xW + \varepsilon \leqslant M\mathbf{1}$. Then, we have $x\mathbf{1}^\top = x\mathbf{1}^\top + \varepsilon\mathbf{1}^\top$ so that it has to be that $\varepsilon\mathbf{1}^\top = 0$. By Perron–Frobenius Theorem and irreducibility, the matrix W has a simple left eigenvector e that is strictly positive and satisfies $eW = e$. Let $y = x + te$ and $t \in \mathbb{R}$ such that $y \in [-M\mathbf{1}, M\mathbf{1}]$. Since $yW + \varepsilon = teW + xW + \varepsilon = te + x = y$, we create another solution y. We can see that the uniqueness may fail with non-zero probability if $\Pr(\sum_i \varepsilon_i = 0) > 0$.

Remark 4.7.1. Consider the generalised Eisenberg-Noe financial network (4.12) in Acemoglu et al. (2015a) and Acemoglu et al. (2015b) and assume that $\bar{p}_i \equiv M$ for all i. The interaction function is

$$f_i(z) = z\mathbb{1}_{\{0 \leqslant z < M\}}(z) + M\mathbb{1}_{\{z \geqslant M\}}(z). \tag{4.24}$$

for all i. In words, the interaction function is non-negative and $f_i(z) = 0$ if $z < 0$. Moreover, the sensitivity matrix satisfies $\sum_j w_{ij} = 1$ for all i but not $\sum_h w_{hi} = 1$ for all i. Therefore, we cannot apply the theorem of Acemoglu et al. (2016b) directly, although their proof can be extended by relaxing this assumption (see Theorem 4.2.2). Acemoglu et al. (2015a) show that the equilibrium is also generically unique for such interaction function and irreducible sensitivity matrix.

However, we see that Example 4.2.1 can be easily modified as an example to show that it may admit multiple equilibria with non-zero probability under the interaction function (4.24). The shock in Acemoglu et al. (2015a) is equal to the sum of holding cash, project return, and the liquidation of asset minus senior liability (see Section 4.3.2). In their setting, the cash, senior liability, and liquidation of asset are all constant, while the realizations of project returns are i.i.d, and only take two values. Hence, the shocks are also i.i.d. and $\varepsilon_i \in \{e_1, e_2\}$ with $e_1 > 0$ and $e_1 < 0$.[34] Therefore, it is possible that uniqueness fails with non-zero probability when we assign strictly positive probabilities to both e_1 and e_2. Hence, the generic uniqueness result in Acemoglu et al. (2015a) may be problematic when the shock variable is not absolutely continuous.

$\square$

REMARK 4.7.2. Hurd (2016) shows that the multiple solutions of generalised Eisenberg-Noe model (4.12) are eigenvectors of W. Clearly, Example 4.2.1 and the above discussion show that the claim is incorrect, and it should be that the differences in solutions are eigenvectors. $\square$

Proofs in Section 4.2.2.

Definitions. A matrix A is *irreducible* if $\sum_k A^k \gg 0$. Given a square matrix $A = (a_{ij}) \in \mathbb{R}^{n\times n}$, *graph* A is a tuple (V, E) consisting of the vertex set $V := \{1, \dots, n\}$ and the edge set $E \subset V \times V$ set such that $(i, j) \in E$ if and only if $a_{ij} \neq 0$. The graph A has a *directed path* $i \to j$ for $i, j \in V$ if $a^k_{ij} > 0$ for some $k \in \mathbb{N}$, where a^k_{ij} is the (i, j)th entry of A^k. A graph $S = (V_S, E_S)$ is a *subgraph* of graph A if $V_S \subset V$ and $E_S \subset E$. In the following model, $a_{ij} \neq 0$ means agent i's state affects agent j's state or j's state is sensitive to i's state. A vertex j is *accessible* from a vertex i if either $i = j$ or there is a directed path $i \to j$. Graph A is *strongly connected* if vertex j is accessible from vertex i for any $i, j \in V$. We know that graph A is strongly connected if and only if A is irreducible. We say a graph A is *acyclic* if i is not accessible from j whenever j is accessible from i for all $i \neq j \in V$.

The proofs of Theorem 4.2.1 and Theorem 4.2.2 follows the subsequent lemmas. To begin with, recall that we have the map $T\colon \mathbb{R}^n \to \mathbb{R}^n$:

$$Tx := f\,(xW + \varepsilon) \tag{4.25}$$

for $x \in \mathbb{R}^n$, where $f(xW + \varepsilon) = (f_i(\sum_h x_h w_{hi} + \varepsilon_i))_{i\in N}$. Clearly, the vector x is a fixed point of T if and only if it is an equilibrium.

---

[34] The holding cash and liquidation are both zero in section three of Acemoglu et al. (2015a).

Proof of Lemma 4.2.1. Suppose that f_i is increasing and bounded for all $i \in N$. Then, there is $E_i > 0$ such that $|f_i(t)| < E_i$ for all i. Let $E = (E_i)$. Define T by (4.25). Hence, it must that $Tx \in [-E, E]$ for all $x \in [-E, E]$. Note that if $x^* > E$ or $x^* < -E$ is a fixed point of T, then we have $Tx^* \leqslant E < x^*$ or $x^* < -E \leqslant Tx^*$, which contradicts the fact that x^* is a fixed point. Hence, we can restrict the domain of T on $[-E, E]$ without loss of generality, so that T is an increasing self-map on $[-E, E]$. Tarski's Fixed Point Theorem shows that there exists an equilibrium in $[-E, E]$, and the set of equilibria forms a complete lattice, whence both the highest and the lowest equilibria exist. □

The following lemma shows that T is a contraction mapping if $\| \left(|W| \operatorname{diag}(\beta)\right)^k \| < 1$.

Lemma 4.7.1. $\|T^k x - T^k \hat{x}\| \leqslant \| \left(|W| \operatorname{diag}(\beta)\right)^k \| \|x - \hat{x}\|$ *for all* $x, \hat{x}$ *in* $\mathbb{R}^n$ *and* $k \in \mathbb{N}$.

Proof of Lemma 4.7.1. Let $x, \hat{x} \in \mathbb{R}^n$, fix $\varepsilon \in E_\sigma$, and define T by (4.25). We first show that $|T^k x - T^k \hat{x}| \leqslant |x - \hat{x}| \left(|W| \operatorname{diag}(\beta)\right)^k$ for all $k \in \mathbb{N}$ by induction. Since $Tx = f(xW + \varepsilon)$ and $|f(xW + \varepsilon) - f(\hat{x}W + \varepsilon)| \leqslant |(xW + \varepsilon) - (\hat{x}W + \varepsilon)| \operatorname{diag}(\beta)$, we have

$$\begin{aligned} |Tx - T\hat{x}| &\leqslant |(xW + \varepsilon) - (\hat{x}W + \varepsilon)| \operatorname{diag}(\beta) \\ &= |(x - \hat{x})W| \operatorname{diag}(\beta) \leqslant |x - \hat{x}||W| \operatorname{diag}(\beta). \end{aligned}$$

Hence, the claim holds for $k = 1$. Suppose that the claim holds for some $m \in \mathbb{N}$, i.e. $|T^m x - T^m \hat{x}| \leqslant |x - \hat{x}|(|W| \operatorname{diag}(\beta))^m$. By iteration, we have

$$\begin{aligned} \left|T^{m+1}x - T^{m+1}\hat{x}\right| &= \left|f\left(T^m x W + \varepsilon\right) - f\left(T^m \hat{x} W + \varepsilon\right)\right| \\ &\leqslant |T^m x W - T^m \hat{x} W| \operatorname{diag}(\beta) \\ &\leqslant |T^m x - T^m \hat{x}| \, |W| \operatorname{diag}(\beta) \\ &\leqslant |x - \hat{x}| \left(|W| \operatorname{diag}(\beta)\right)^{m+1}. \end{aligned}$$

Thus, the induction implies that $|T^k x - T^k \hat{x}| \leqslant |x - \hat{x}|(|W| \operatorname{diag}(\beta))^k$ for all $k \in \mathbb{N}$. The definition of p-norm gives $\|T^k x - T^k \hat{x}\| \leqslant \| |x - \hat{x}|(|W| \operatorname{diag}(\beta))^k \| \leqslant \|(|W| \operatorname{diag}(\beta))^k\| \, \|x - \hat{x}\|$ for all $k \in \mathbb{N}$.[35] □

---

[35] For the max norm, we can take the maximum on both sides of inequality. For a *Riesz norm*, $|x| \leqslant |y|$ implies $\|x\| \leqslant \|y\|$.

We are ready to prove the first statement of Theorem 4.2.1 and its following corollaries.

Proof of Theorem 4.2.1. Let f and W satisfy Assumption 4.2.1 so that $\rho(|W| \operatorname{diag}(\beta)) < 1$, where $\beta = (\beta_j)$. Fix $\varepsilon \in E_\sigma$. Then, we have

$$\lim_{k\to\infty} \|(|W| \operatorname{diag}(\beta))^k\| = 0.$$

Thus, $\|(|W| \operatorname{diag}(\beta))^k\| < 1$ for some $k \in \mathbb{N}$. Lemma 4.7.1 and the Banach Contraction Theorem show that there is a unique equilibrium. ☐

Proof of Corollary 4.2.1. Let f, W and ε satisfy the conditions of Theorem 4.2.1. Let x^* be the equilibrium and $T\colon \mathbb{R}^n \to \mathbb{R}^n$ be defined as $Tx = f(xW + \varepsilon)$. Denote $A = |W| \operatorname{diag}(\beta)$. The first statement follows from Theorem 4.2.1 and Banach Fixed Point Theorem. From Lemma 4.7.1, we have $\|T^{k+1}x - T^k x\| = \|T^k(Tx) - T^k x\| \leqslant \|A^k\|\|Tx - x\|$ for all $k \in \mathbb{N}$. ☐

Next, we attempt to prove the almost surely uniqueness for the case with non-negative f_i and matrix with spectral radius one. Firstly, we consider that W is irreducible or the graph W is strongly connected. Note that Acemoglu et al. (2016b) use the continuity of f to show the existence of equilibrium. However, they argue that f has at most countably many discontinuous points when they show the uniqueness of equilibrium. Since f is non-expansive, it is Lipschitz continuous so that their argument is ambiguous. Alternatively, we consider at most countably many intervals with unit slope.

Lemma 4.7.2 (Acemoglu et al. (2016b)). *If f is increasing and $|f(a) - f(b)| = |a - b|$ for $a, b \in \mathbb{R}$ and $b > a$, then f is linear in the interval $[a, b]$ with a unit slope, i.e., $f(z) = z + f(a) - a$ for all $z \in [a, b]$.*[36]

Lemma 4.7.3. *If f is increasing and bounded, there are at most countably many intervals that f is linear with unit slopes in these intervals.*

Proof. Let f be an increasing and bounded mapping. Let I be the set of intervals in $\mathbb{R}$ such that f is linear and with unit slopes in these intervals. Let $[a_i, b_i] \in I$. Since

---

[36] See lemma B.2 of Acemoglu et al. (2016b).

$b_i > a_i$ and f is an increasing and linear mapping on $[a_i, b_i]$ with a unit slope, we have $f(a_i) < f(b_i)$. Then, there is a rational $c_i \in \mathbb{Q}$ such that $f(a_i) < c_i < f(b_i)$. The map $[a_i, b_i] \mapsto c_i$ is injective, so the set I is countable. □

Lemma 4.7.4. *Suppose that f_i is increasing, non-expansive, and bounded for all i, the shock is absolutely continuous for all i, and W is irreducible with $\rho(W) = 1$. Then, the equilibrium is almost surely unique.*

Proof of Lemma 4.7.4. Let $f = (f_i)$, W be such that the conditions of Lemma 4.7.4 hold. By Lemma 4.2.1 or Tarski's Fixed Point Theorem, the equilibrium exists and the set of equilibria is complete. Without loss of generality, assume that there are two different equilibria x and y with $x > y$. Since f is non-expansive, we have

$$\begin{aligned} e \coloneqq x - y &= |f(xW + \varepsilon) - f(yW + \varepsilon)| \\ &\leqslant |xW + \varepsilon - (yW + \varepsilon)| = |(x - y)W| = (x - y)W = eW. \end{aligned} \tag{4.26}$$

Let $q \in \mathbb{R}^n$ be the right eigenvector corresponding to $\rho(W)$ so that $Wq^\top = \rho(W)q^\top$. By the Perron-Frobenius Theorem and irreducibility, q is strictly positive, $q \gg 0$. Suppose that the inequality of equation (4.26) does not bind. It implies that

$$eq^\top < eWq^\top = eq^\top,$$

which is a contradiction. Hence, it must be that the equality of (4.26) binds so that $e = |f(xW + \varepsilon) - f(yW + \varepsilon)| = |(x - y)W| = eW$. We see that the vector e is the left eigenvector of W, so Perron-Frobenius Theorem shows that $e \gg 0$. Now, define $b \coloneqq xW + \varepsilon$ and $a \coloneqq yW + \varepsilon$. Then, we have $|f_i(b_i) - f_i(a_i)| = |b_i - a_i|$ for all $i \in N$. Since $b - a = eW = e \gg 0$, we have $b \gg a$. Therefore, the interval $[a_i, b_i]$ is well-defined for all i. Since f_i is increasing for all i, it follows from Lemma 4.7.2 that f_i is a linear mapping with unit slopes in the intervals $[a_i, b_i]$ for all i. That is, we get $f(z) = z + f(a) - a$ for all $z \in [a, b]$. In particular, since x is the fixed point, we have $x = f(b) = b + f(a) - a = xW + \varepsilon + f(a) - a$. Multiplying $q^\top$ on both sides, it implies that $xq^\top = xWq^\top + \varepsilon q^\top + (f(a) - a)q^\top = xq^\top + \varepsilon q^\top + (f(a) - a)q^\top$. Since q is the right eigenvector, if there are multiple fixed points, it has to be that $\varepsilon q^\top = (a - f(a))q^\top$.

Denote the sets M and A as

$$\begin{aligned} M &:= \{\varepsilon \in E_\varepsilon : \text{There are multiple equilibria.}\} \\ A &:= \{a \in E_\varepsilon : f_i \text{ is a linear map with unit slope on } [a_i, b_i] \\ &\qquad\qquad\qquad\qquad \text{with } b_i > a_i \text{ for } i = 1, \dots, n.\}. \end{aligned}$$

By the above discussion, we have $\varepsilon \in M$ only if $\varepsilon q^\top = (a - f(a))q^\top$, so $M \subset E_1 := \{\varepsilon \in E_\varepsilon : \varepsilon q^\top = (a - f(a))q^\top, a \in A\}$. Therefore, $\varepsilon_1, \dots, \varepsilon_n$ are linearly dependent when there exists multiplicity. Since Lemma 4.7.3 implies that there are countably many intervals $[a_i, b_i]$ for f_i, there are countably many vectors a, whence A is at most countable. Therefore, the dimension of E_1 is strictly less than n, and the Lebesgue measures of E_1 and M are zero. Since the measure of shocks is absolutely continuous, the probability of realizations in M is zero, $\mathbb{P}_\varepsilon(\varepsilon \in M) = 0$. ☐

Define in-subgraph and out-subgraph for the proof of Theorem 4.2.2. Let $A = (a_{ij}) \in \mathbb{R}^{n \times n}$ and graph $A = (V, E)$. Let graph $S = (V_S, E_S)$ be a *subgraph* of graph A such that $V_S \subset V$ and $E_S \subset E$. Subgraph S is an *in-subgraph* if there is no path from i to j for some $i \in V_S$ and $j \in V \setminus V_S$. Subgraph S is an *out-subgraph* if there is a path from i to j for some $i \in V_S$ and $j \in V \setminus V_S$.

PROOF OF THEOREM 4.2.2. Let $f = (f_i)$ and W be such that the conditions of Assumption 4.2.2 hold, and ε be absolutely continuous. Suppose that $\rho(W) = 1$. By Lemma 4.2.1, the equilibrium exists. Define the equivalence relation $\sim$ as $i \sim j$ if and only if i and j are accessible from each other for all $i, j \in N$. Denote $[i] := \{s \in N : s \sim i\}$ as the equivalence class. We see that $[i] \neq [j]$ if either i is not accessible from j, or j is not accessible from i, so the equivalence classes are mutually disjoint. Moreover, the set $\{[i] : i \in N\}$ forms a partition of N. By definition, each equivalence class $[i]$ is either an in-subgraph or out-subgraph. Since the number of vertices is finite, there are finite in-subgraphs, denoted by $S_1, \dots, S_m$. Let S_0 be the union of all out-subgraphs. By definition of in-subgraphs and out-subgraphs, we can permute the matrix W and write it as [37]

$$W = \begin{pmatrix} W_0 & P_1 & P_2 & \cdots & P_m \\ 0 & W_1 & 0 & \cdots & 0 \\ 0 & 0 & W_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & W_m \end{pmatrix}$$

where $W_t \in \mathbb{R}^{|S_t| \times |S_t|}$, for $t = 0, 1, \ldots, m$, is the weighting adjacency matrix of subgraph S_t, and $P_t \in \mathbb{R}^{|S_0| \times |S_t|}$ describes the path from S_0 to S_t.[38] Note that $W_1, \ldots, W_m$ are irreducible, since in-subgraphs are strongly connected. Let x be an equilibrium of states. Decomposing the state vector and shock vector into block forms corresponding to $S_0, S_1, \ldots, S_m$, we have $x = (x_0, x_1, \ldots, x_m)$ and $\varepsilon = (\varepsilon_0, \varepsilon_1, \ldots, \varepsilon_m)$, where $x_t, \varepsilon_t \in \mathbb{R}^{|S_t|}$ for $t = 0, 1, \ldots, m$. The equilibrium system can be written as

$$\begin{aligned} x_0 &= f(x_0 W_0 + \varepsilon_0) \\ x_1 &= f(x_1 W_1 + x_0 P_1 + \varepsilon_1) \\ x_2 &= f(x_2 W_2 + x_0 P_2 + \varepsilon_2) \\ &\vdots \\ x_m &= f(x_m W_m + x_0 P_m + \varepsilon_m) \end{aligned}$$

where $f(x_t W_t + x_0 P_t + \varepsilon_t) := (f_i(\sum_{h \in S_t} x_h w_{hi} + \sum_{h \in S_0} x_h w_{hi} + \varepsilon_i))_{i \in S_t}$. Since $\rho(W) = 1$, we have $\rho(W_0) \leqslant 1$. We first assume that $\rho(W_0) < 1$. Then, $x_0 = f(x_0 W_0 + \varepsilon_0)$ has a unique solution following Theorem 4.2.1. Fix $t = 1, \ldots, m$. Given the unique x_0, Lemma 4.7.4 shows that the equilibrium x_t is almost surely unique for $t = 1, \ldots, m$. It is not unique only if $(x_0 P_t + \varepsilon_t) q_t^\top = (a_t - f(a_t)) q_t^\top$ from Lemma 4.7.4, where q_t is the right eigenvector for W_t, f_i is a unit-sloping mapping in intervals $[a_{t,i}, b_{t,i}]$ for all $i \in S_t$, and a_t is defined as a in Lemma 4.7.4. Denote the set

$$E_t := \{\varepsilon \in E_\varepsilon : \text{the system } x_t = f(x_t W_t + x_0 P_t + \varepsilon_t) \text{ has multiple solutions}\}.$$

---

[37] See also (Berman and Plemmons, 1994, chapter 8).

[38] $|S_t|$ denotes the number of nodes in subgraph S_t.

We see that E_t has Lebesgue measure zero by the argument in Lemma 4.7.4. It implies that $\lambda(E_1 \cup \cdots \cup E_m) \leqslant \lambda(E_1) + \cdots + \lambda(E_m) = 0$, where λ denotes the Lebesgue measure. Let $E = \bigcup_t E_t$. Since the random variable of ε is absolutely continuous, we have $\mathbb{P}_\varepsilon(\varepsilon \in E) = \int_E g(x)\mathrm{d}\lambda = 0$, where g is the density function of shock. Overall, each in-subgraph has an almost surely unique equilibrium, and the equilibrium x is also almost surely unique.

Next, suppose that $\rho(W_0) = 1$, and W_0 is irreducible. Denote the equilibrium of $x_0 = f(x_0 W_0 + \varepsilon_0)$ as $x_0(\varepsilon_0)$. In this case, the solution $x_0(\varepsilon_0)$ is unique almost surely by Lemma 4.7.4. If x_0 is unique, the equilibrium x is almost surely unique following the above argument. If x_0 is not unique, then it must be $\varepsilon_0 q_0^\top = (a_0 - f(a_0))q_0^\top$, where q_0 is the right eigenvector of W_0 and a_0 is some vector in $R^{|S_0|}$ by Lemma 4.7.4. When x_0 is not unique, the solution x_t, for $t = 1, \ldots, m$, is not unique only if $(x_0(\varepsilon_0)P_t + \varepsilon_t)q_t^\top = (a_t - f(a_t))q_t^\top$. Note that $\{\varepsilon \in E_\varepsilon \colon \varepsilon_0 q_0^\top = (a_0 - f(a_0))q_0^\top$ and $(x_0(\varepsilon_0)P_t + \varepsilon_t)q_t^\top = (a_t - f(a_t))q_t^\top\}$ is a subset of $\{\varepsilon \in E_\varepsilon \colon \varepsilon_0 q_0^\top = (a_0 - f(a_0))q_0^\top\}$, which has measure zero. Therefore, x_t is also almost surely unique for all $t = 1, \ldots, m$.

Suppose that $\rho(W) = 1$ and W_0 is not irreducible. Considering an equivalence class as an entity, since S_0 is the collection of out-subgraphs of equivalence class $[i]$, these equivalence classes form an acyclic graph; otherwise, they will deviate from their definition. Assume that there are r equivalence classes $[i]$ in S_0. Since W_0 forms an acyclic network, W_0 can be written as an upper triangular matrix as

$$W_0 = \begin{pmatrix} B_{11} & B_{12} & B_{13} & \cdots & B_{1r} \\ 0 & B_{22} & B_{23} & \cdots & B_{2r} \\ 0 & 0 & B_{33} & \cdots & B_{3r} \\ \vdots & \vdots & & & \vdots \\ 0 & 0 & 0 & \cdots & B_{rr} \end{pmatrix}$$

where B_{tt} is a square matrix with respect to the corresponding equivalence class for $t = 1, \ldots r$. Then, we can decompose $x_0 = f(x_0 W_0 + \varepsilon_0)$ into r systems and pin down the equilibrium in order. Since each equivalence class is strongly connected, each subgraph B_{tt} is irreducible. Using the similar argument above, Lemma 4.7.4 and Theorem 4.2.1, we see that the equilibrium for the system with respect to graph B_{11} is almost surely a unique solution if $\rho(B_{11}) = 1$ and unique if $\rho(B_{11}) < 1$. For $t = 2, \dots, r$, if $\rho(B_{tt})$ is less than one, the system with respect to B_{tt} has a unique equilibrium. If the $\rho(B_{tt})$ is one, its equilibrium is almost surely unique. Overall, $x_0 = f(x_0 W_0 + \varepsilon_0)$ has almost surely a unique equilibrium following the above argument. A similar argument for the irreducible case concludes the result for the whole system $x = f(xW + \varepsilon)$. $\square$

Proof of Corollary 4.2.2. Suppose that (f, W, ε) follows the assumptions of Theorem 4.2.2. Define $T : [\ell, u] \to [\ell, u]$ as $Tx \coloneqq f(xW + \varepsilon)$ following the statement. Hence, since u is the upper bound, we get $Tu \leqslant u$. Since T is monotone given the monotonicity of f, we get $T^2 u \leqslant Tu$, whence the iteration implies that $T^n u \leqslant T^{n-1} u \leqslant \cdots \leqslant Tu \leqslant u$. Let $x_n \coloneqq T^n u$ for all $n \in \mathbb{N}$. Then, (x_n) is a non-increasing sequence and bounded below by ℓ, so it has a limit, denoted by $x^* \coloneqq \lim_{n\to\infty} T^n u$. Now, consider $Tx_n = T^{n+1} u$. Letting $n \to \infty$ on both sides of the equation, since T is continuous from above given the continuity of f, we have $Tx^* = T(\lim_{n\to\infty} x_n) = \lim_{n\to\infty} Tx_n = \lim_{n\to\infty} T^{n+1} u = x^*$. Therefore, x^* is the fixed point of T. Since Theorem 4.2.2 implies that T has a unique fixed point almost surely, x^* is the almost surely unique equilibrium. $\square$

4.7.1. Proofs in Section 4.3.

Proof of Corollary 4.3.2. Let f, W, and ε follow the conditions of Corollary 4.3.2. Let $\bar{p} = (\bar{p}_i)$. By Lemma 4.2.1, we know that the equilibrium exists. From the proof of Theorem 4.2.2, if there are multiple equilibria, there must be some irreducible subgraph of W that admit multiple equilibria. Without loss of generality, assume that W is irreducible. On the contrary, suppose that x and y are two fixed points with $x > y$. By equation (4.26) and Lemma 4.7.4, we know that it must follow $|x - y| = |f(xW + \varepsilon) - f(yW + \varepsilon)|$. Hence, we have $|x - y| = |(xW + \varepsilon)^+ \wedge \bar{p} - (yW + \varepsilon)^+ \wedge \bar{p}|$. Since $|(\alpha)^+ \wedge \gamma - (\beta)^+ \wedge \gamma| = |\alpha - \beta|$ for $\gamma > 0$ and $\alpha, \beta \in \mathbb{R}$ if and only if $\alpha, \beta \in [0, \gamma]$, we get $0 \leqslant xW + \varepsilon, yW + \varepsilon \leqslant \bar{p}$. Therefore, we have $x = f(xW + \varepsilon) = xW + \varepsilon$ and $y = f(yW + \varepsilon) = yW + \varepsilon$.

By the Perron-Frobenius theorem, the right eigenvector q of W is strictly positive. Since the row sum is one, q is a vector of ones or constants. Multiplying both side of $x = xW + \varepsilon$

by $q^\top$, we have $xq^\top = (xW + \varepsilon)q^\top = xq^\top + \varepsilon q^\top$. Thus, we get $\varepsilon q^\top = 0$, whence $\varepsilon \mathbf{1}^\top = 0$. Since $\varepsilon \geqslant 0$ and there is $\varepsilon_i > 0$ for some i, we obtain a contradiction. ☐

Proof of Corollary 4.3.1. Let f and W satisfy the conditions in Corollary 4.3.1 so that $\rho(W \operatorname{diag}(\beta)) < 1$, where $\beta = (\beta_j)$. The proof is similar to Lemma 4.7.1 and Theorem 4.2.1. In detail, we can show that $T^k x - T^k \hat{x} = (x - \hat{x})\,(W \operatorname{diag}(\beta))^k$ for all $k \in \mathbb{N}$ by the same induction in Lemma 4.7.1. Thus, we have $\|T^k x - T^k \hat{x}\| \leqslant \|x - \hat{x}\|\, \|(W \operatorname{diag}(\beta))^k\|$. If $\rho(W \operatorname{diag}(\beta)) < 1$, then $\|(W \operatorname{diag}(\beta))^k\| < 1$ for some k. The result follows from the Banach Contraction Theorem. ☐

Proof of Lemma 4.3.1. Let f and W be such that the conditions of Lemma 4.3.1 hold. Let $\beta = (\beta_i)$ be Lipschitz constants. If the first condition of Lemma 4.7.6 holds, it implies that $\|W^m\|_\infty < 1$ for some $m \in \mathbb{N}$ by Theorem 2.5 of Azimzadeh (2019). Similarly, the second condition of Lemma 4.7.6 implies that $\|W^m\|_1 = \|(W^\top)^m\|_\infty < 1$ for some $m \in \mathbb{N}$. Since $\operatorname{diag}(\beta) = I$ for non-expansiveness of f, we have $\rho(|W| \operatorname{diag}(\beta)) = \rho(W) < 1$. ☐

Proof of Lemma 4.3.2. Let f and W be such that the conditions of Lemma 4.3.2 hold. Let $\beta = (\beta_i)$ be Lipschitz constants. An acyclic graph contains some sink node i such that $\sum_j w_{ij} = 0$. For each node s in an acyclic graph, there is an acyclic path to some sink node t. Suppose that node s can reach t in K_t steps such that $w_{st}^{K_t} > 0$. Let $K^* = \max_t K_t$ be the length from node s to the farthest connected sink node t^*. (Since if there are multiple farthest sink nodes, they have the same length. Without loss of generality, assume that there is only one farthest sink node from s.) Then, we must have $\sum_u w_{su}^{K^*+1} = 0$.[39] Since s is arbitrary, we have $W^T = 0$ for some large enough $T \in \mathbb{N}$. It implies that $\rho(W) = 0$. Let $A = |W| \operatorname{diag}(\beta)$. The sink nodes for graph W are also the sink nodes for graph A. Hence, the same argument shows that $\rho(A) = 0$. ☐

Proof of Lemma 4.3.3. Let f and W be such that the conditions of Lemma 4.3.3 hold. Then, the conditions of Lemma 4.7.6 are satisfied. The result follows from Lemma 4.3.1. ☐

---

[39] If not zero, it contradicts with the definition of K^*.

Proof of Corollary 4.3.3. Let f and W be such that Assumption 4.2.2 holds. Using Lemma 4.3.2 and Theorem 4.2.1, if the network graph is acyclic, then the equilibrium is unique. Hence, if there exists multiplicity, there must be a strongly connected subgraph. We then decompose the graph into strongly connected subgraphs as the proof of Theorem 4.2.2. The proof of Theorem 4.2.2 implies that, given the realized shocks ε, the multiplicity occurs in and originates from some strongly connected in-subgraphs or out-subgraphs S (the shocks corresponding to S are linearly dependent). It implies that all agents in S admit multiple equilibria.[40] Moreover, the agents with multiple equilibria must be accessible from at least one subgraph with multiple equilibria. $\square$

Proofs in Section 4.4.

Proof of Lemma 4.4.1. Suppose that (f, W, ε) and (f', W', ε') are two networks such that Assumption 4.2.1 holds. Assume that $W \leqslant W'$, $\varepsilon \leqslant \varepsilon'$, and f_i and f_i' are increasing functions for all $i \in N$, and $f_i(t) \leqslant f_i'(t)$ for all $t \in \mathbb{R}$ and all $i \in N$. It then follows from Theorem 4.2.1 that (f, W, ε) and (f', W', ε') have unique equilibrium $\hat{x}$ and $\hat{x}'$, respectively. Define the maps $T\colon \mathbb{R}^n \to \mathbb{R}^n$ and $\hat{T}\colon \mathbb{R}^n \to \mathbb{R}^n$ as $Tx := f(xW + \varepsilon)$ and $\hat{T}x := f'(xW' + \varepsilon')$ for all $x \in \mathbb{R}^n$. We first show that $T^k x \leqslant \hat{T}^k x$ for all $x \in \mathbb{R}^n$ and $k \in \mathbb{N}$. Since $f_i(t) \leqslant f_i'(t)$ for all $t \in \mathbb{R}$ and all i, we see that $f(x) \leqslant f'(x)$ for all $x \in \mathbb{R}^n$. Then, since f_i and f_i' are increasing functions for all i, and $W \leqslant W'$, and $\varepsilon \leqslant \varepsilon'$, we obtain $Tx = f(xW + \varepsilon) \leqslant f(xW' + \varepsilon') \leqslant f'(xW' + \varepsilon') = \hat{T}x$ for all $x \in \mathbb{R}^n$. Suppose the induction hypothesis that $T^k x \leqslant \hat{T}^k x$ for some $k \in \mathbb{N}$ and all x. Hence, since $Tx \leqslant \hat{T}x$ for all $x \in \mathbb{R}^n$, we have $T(T^k x) \leqslant \hat{T}(T^k x) \leqslant \hat{T}(\hat{T}^k x)$. Therefore, the Mathematical Induction implies that $T^k x \leqslant \hat{T}^k x$ for all $x \in \mathbb{R}^n$ and all $k \in \mathbb{N}$. Since Corollary 4.2.1 shows that $\hat{x} = \lim_{k\to\infty} T^k x$ and $\hat{x}' = \lim_{k\to\infty} \hat{T}^k x$ for any $x \in \mathbb{R}^n$, we obtain $\hat{x} = \lim_{k\to\infty} T^k x \leqslant \lim_{k\to\infty} \hat{T}^k x = \hat{x}'$. $\square$

Proof of Lemma 4.4.2. Let (f, W, ε) and (f', W', ε') be two networks such that the assumptions of the statement hold. Suppose that $f_i(t) \leqslant u_i$ and $f_i'(t) \leqslant u_i'$ for all $t \in \mathbb{R}$

---

[40] Since S admit multiple equilibria (shocks are linearly dependent), given an equilibrium, we can create another equilibrium. In particular, let W_s be the submatrix of W corresponding to subgraph S, and x_s be the equilibrium for agents in S. From equation (4.26), $x_s + te_sW_s$ is another equilibrium, where e_s is the eigenvector of W_s, and $t \in \mathbb{R}$ is a parameter such that $x_s + te_sW_s$ is in the range of the interaction functions.

such that $u_i \leqslant u_i'$ for all i. Let $u := (u_i)$ and $u' := (u_i')$. Define the maps $T\colon \mathbb{R}^n \to \mathbb{R}^n$ and $\hat{T}\colon \mathbb{R}^n \to \mathbb{R}^n$ as $Tx := f(xW + \varepsilon)$ and $\hat{T}x := f'(xW' + \varepsilon')$ for all $x \in \mathbb{R}^n$. Following Theorem 4.2.2, we know that the equilibrium for both (f, W, ε) and (f', W', ε') are unique almost surely, denoted by $\hat{x}$ and $\hat{x}'$, respectively. Moreover, it follows from Corollary 4.2.2 that $\hat{x} = \lim_{k\to\infty} T^k u$ and $\hat{x}' = \lim_{k\to\infty} \hat{T}^k u'$. Now, we show that $T^k u \leqslant \hat{T}^k u'$ for all $k \in \mathbb{N}$. Since $f_i(t) \leqslant f_i'(t)$ for all t and all i, we see that $Tx = f(xW + \varepsilon) \leqslant f'(xW + \varepsilon) \leqslant f'(xW' + \varepsilon') = \hat{T}x$ for all x. Hence, since $u \leqslant u'$, we have $Tu \leqslant \hat{T}u \leqslant \hat{T}u'$. Suppose that $T^k u \leqslant \hat{T}^k u'$ for some $k \in \mathbb{N}$. Then, since $Tx \leqslant \hat{T}x$ for all x, and $\hat{T}$ is an increasing map, we get $TT^k u \leqslant \hat{T}T^k u \leqslant \hat{T}\hat{T}^k u'$. Therefore, the Mathematical Induction implies that $T^k u \leqslant \hat{T}^k u'$ for all $k \in \mathbb{N}$. Consequently, we have $\hat{x} = \lim_{k\to\infty} T^k u \leqslant \lim_{k\to\infty} \hat{T}^k u' = \hat{x}'$. ☐

Proofs in Section 4.4. Define the set of ε that the solution exists as

$$E_0 = \{\varepsilon \in E_\varepsilon : \text{Linear system (4.16) has a solution.}\}$$

Define the set of ε that allows multiple solutions as

$$E_1 := \{\varepsilon \in E_\varepsilon : \text{Linear system (4.16) has multiple solutions.}\}$$

LEMMA 4.7.5. *Suppose that ε is absolutely continuous If W is non-negative, and $\rho(W) = 1$, then $\lambda(E_0) = 0$ and $E_0 = E_1$.*

PROOF OF LEMMA 4.7.5. Let W be a non-negative, irreducible matrix W with $\rho(W) = 1$. Using Perron-Frobenius Theorem, $\rho(W)$ is an eigenvalue of W, and its corresponding left eigenvector and right eigenvector are non-negative, denoted by v and e, respectively. First, note that $x = xW + \varepsilon$ has a solution if and only if $\varepsilon e^\top = 0$. To see this, let $A := I - W$ and rewrite the linear system (4.16) as $xA = \varepsilon$. Hence, if x is a solution, then $xAe^\top = xe\top - xWe^\top = xe^\top - xe^\top = 0 = \varepsilon e^\top$. Conversely, if $\varepsilon e^\top = 0$, then since $Ae^\top = e^\top - e^\top = 0$, we have $rank(A) = rank([A|\varepsilon])$, where $[A|\varepsilon]$ is the augmented matrix for $xA = \varepsilon$. Since the solution exists if and only if $rank(A) = rank([A|\varepsilon])$, there must be a solution. Therefore, $E_0 = \{\varepsilon : \varepsilon e^\top = 0\}$ is measure zero. For the second statement, it is clear that $E_1 \subset E_0$. Suppose that $\varepsilon \in E_0$. Thus, there is x such that $x = xW + \varepsilon$. Letting

$t \in \mathbb{R}$, since v is the left eigenvector, we have $x + tv = xW + \varepsilon + tv = (x + tv)W + \varepsilon$, whence $x + tv$ is a solution. Then, $\varepsilon \in E_1$ and $E_0 \subset E_1$. □

Proof of Lemma 4.4.3. Let W and ε follow the conditions of Lemma 4.4.3. Lemma 4.7.5 and absolute continuity of shock imply that $\mathbb{P}_\varepsilon(\varepsilon \in E_0) = 0$. □

Proof of Lemma 4.4.4. Let $\ell, u \in \mathbb{R}^n$ be such that $u \gg \ell$, $f = (f_i)$ be defined as (4.17), and W is stochastic. Assume that x^* is the unique equilibrium. Suppose on the contrary that for all $j \in N$ we have $x_j^* = \sum_i x_i^* W_{ij} + \varepsilon_j < u_j$ and $x_j^* = \sum_i x_i^* W_{ij} + \varepsilon_j > \ell_j$, whence $\ell \ll x^* W + \varepsilon \ll u$ and $x^* = f(x^*) = x^* W + \varepsilon$. Since W is stochastic and hence non-negative, Perron-Frobenius theorem implies that W has a non-negative left eigenvector v with respect to the eigenvalue 1. That is, $vW = v$. Now, since $\ell \ll x^* \ll u$, we can find $\lambda \neq 0$ such that $\ell \leqslant x^* + \lambda v \leqslant u$. Then, since $(x^* + \lambda v)W + \varepsilon = x^* W + \lambda v + \varepsilon = x^* + \lambda v$, we have $\ell \leqslant (x^* + \lambda v)W + \varepsilon \leqslant u$. Therefore, we get $f((x^* + \lambda v)W + \varepsilon) = (x^* + \lambda v)$, so that $(x^* + \lambda v)$ is another equilibrium. Since x^* is the unique equilibrium, we have a contradiction. □

Proof of Lemma 4.4.5. Let (f, W, ε) follows the conditions of lemma. Let $u = (u_i)$ and $\ell = (\ell_i)$ be the upper bounds and lower bounds of f, respectively. Given the states $x \in [\ell, u]$, let $A(x)$ and $B(x)$ be the sets of agents with the equilibrium at the bounds as (4.18). Denote x^* as the unique equilibrium. Let $A^* := A(x^*) = \{j \in N \colon \sum_i x_i^* w_{ij} + \varepsilon_j \geqslant u_j\}$ be the sets of agents having the equilibrium at the upper bounds, and $B^* := B(x^*) = \{j \in N \colon \sum_i x_i^* w_{ij} + \varepsilon_j \leqslant \ell_j\}$ be the sets of agents with the equilibrium at the lower bounds. Following the algorithm, if $x^* = u$ or $x^* = \ell$, then step 2 or step 3 of Algorithm 2 terminates and returns the equilibrium. Without loss of generality, suppose that $x^* \neq u, \ell$.

Define $\mathcal{P} := \mathcal{P}(B(\ell))$ as the power set of $B(\ell)$. Then, since $B^* \subset B(\ell)$, there exists $P \in \mathcal{P}$ such that $P = B^*$. Define $\hat{u} := \ell \Lambda^{B^*} + u(I - \Lambda^{B^*})$. Hence, economic states $\hat{u}$ indicate that an agent i in B^* has economic state ℓ_i, and an agent i in $N \setminus B^*$ has state u_i.

Since $x^* > \ell_j$ for $j \in N \setminus B^*$, we can be reduced the system to:

$$\begin{aligned}
\forall j \in B^*, \qquad & x_j = \ell_j \\
\forall j \in N \setminus B^*, \qquad & x_j = \min\left\{\max\left\{\sum_{i\in N\setminus B^*} x_i w_{ij} + \sum_{i\in B^*} \ell_i w_{ij} + \varepsilon_j, \ell_j\right\}, u_j\right\} \\
& \quad = \min\left\{\sum_{i\in N\setminus B^*} x_i w_{ij} + \sum_{i\in B^*} \ell_i w_{ij} + \varepsilon_j, u_j\right\} \\
& \quad > \ell_j
\end{aligned} \tag{4.27}$$

We write (4.27) in vector form and define the operator $\hat{T} : [\ell, \hat{u}] \to [\ell, \hat{u}]$ as

$$\hat{T}x \coloneqq \{[x(I - \Lambda^{B^*})W + \ell\Lambda^{B^*}W + \varepsilon](I - \Lambda^{B^*})\} \wedge [u(I - \Lambda^{B^*})] + \ell\Lambda^{B^*}.$$

Then, x^* is the fixed point of $\hat{T}$. Following the algorithm, we let $x^{(0)} = \hat{u}$ and $A_0 \coloneqq A(\hat{u}) \cap (N \setminus B^*)$, so that $A^* \subset A_0 \subset N \setminus B^*$. Define $x^{(t)}$ for $t \in \mathbb{N}$ as the fixed point of Φ_t:

$$\begin{aligned}
\Phi_t x \coloneqq\ & u\Lambda^{A_{t-1}} + \ell\Lambda^{B^*} \\
& + [u\Lambda^{A_{t-1}}W + \ell\Lambda^{B^*}W + x(I - \Lambda^{A_{t-1}} - \Lambda^{B^*})W + \varepsilon](I - \Lambda^{A_{t-1}} - \Lambda^{B^*}).
\end{aligned} \tag{4.28}$$

We have $x^{(t)} = \Phi_t x^{(t)}$ for all $t \in \mathbb{N}$. For $t \in \mathbb{N}$, define $A_t = A(x^{(t)})$. We first show that $\hat{T}x^{(t)} \leqslant x^{(t)}$ and $x^{(t+1)} \leqslant x^{(t)}$ for all $t = 0, 1, 2, \ldots$ by induction.

Consider $t = 0$. Clearly, we have $\hat{T}x^{(0)} \leqslant x^{(0)}$. We need to check that Φ_1 has a unique fixed point. Since $A_0 \supset A^*$, the matrix $I - (I - \Lambda^{A_0} - \Lambda^{B^*})W(I - \Lambda^{A_0} - \Lambda^{B^*})$ is non-singular. Otherwise, there exists an irreducible submatrix W_s of $(I - \Lambda^{A_0} - \Lambda^{B^*})W(I - \Lambda^{A_0} - \Lambda^{B^*})$ such that $\rho(W_s) = 1$. But, using Lemma 4.7.4, the proof of Theorem 4.2.2, and Lemma 4.4.4, since the equilibrium is unique, every strongly connected subgraph must contain a node m such that its equilibrium is at the bound. Since the operation $(I - \Lambda^{A_0} - \Lambda^{B^*})W(I - \Lambda^{A_0} - \Lambda^{B^*})$ removes such node m from the graph W, Lemma 4.7.9 implies that $\rho(W_s) < 1$. Therefore, such irreducible submatrix W_s with $\rho(W_s) = 1$ does not exist, so the spectral radius of $(I - \Lambda^{A_0} - \Lambda^{B^*})W(I - \Lambda^{A_0} - \Lambda^{B^*})$ is less than one. Therefore, the fixed point of Φ_1 is unique, and we can compute the fixed point $x^{(1)}$ by the iteration from any initial guess. In particular, we can iterate from $x^{(0)}$: $x^{(1)} = \lim_{m\to\infty} \Phi_1^m(x^{(0)})$.

Moreover, since $x^{(0)} = u(I - \Lambda^{B^*}) + \ell\Lambda^{B^*}$ and then $x^{(0)}(I - \Lambda^{A_0} - \Lambda^{B^*}) = u(I - \Lambda^{A_0} - \Lambda^{B^*})$, we have

$$\begin{aligned}\Phi_1(x^{(0)}) &= u\Lambda^{A_0} + \ell\Lambda^{B^*} \\ &\qquad + [u\Lambda^{A_0}W + \ell\Lambda^{B^*}W + x^{(0)}(I - \Lambda^{A_0} - \Lambda^{B^*})W + \varepsilon](I - \Lambda^{A_0} - \Lambda^{B^*}) \\ &= u\Lambda^{A_0} + \ell\Lambda^{B^*} + [u(I - \Lambda^{B^*})W + \ell\Lambda^{B^*}W + \varepsilon](I - \Lambda^{A_0} - \Lambda^{B^*}) \\ &\leqslant u\Lambda^{A_0} + \ell\Lambda^{B^*} + u(I - \Lambda^{A_0} - \Lambda^{B^*}) \\ &= u(I - \Lambda^{B^*}) + \ell\Lambda^{B^*} = x^{(0)}\end{aligned}$$

where the inequality holds because the definition of A_0 implies that if $A_0 \subsetneq N \setminus B^*$, then $\sum_{i\in N\setminus B_0} u_i w_{ij} + \sum_{i\in B_0} \ell_i w_{ij} + \varepsilon_j < u_j$ for $j \notin A_0$; otherwise, if $A_0 = N \setminus B^*$, then $I - \Lambda^{A_0} - \Lambda^{B^*} = 0$ and $x^{(0)} = \hat{u}$ is the fixed point of $\hat{T}$. Since Φ_t is an increasing operator for all t, we get $\Phi_1^m(x^{(0)}) \leqslant \Phi_1^{m-1}(x^{(0)}) \leqslant \cdots \leqslant \Phi_1(x^{(0)}) \leqslant x^{(0)}$ for all $m \in \mathbb{N}$, so that $x^{(1)} = \lim_{m\to\infty} \Phi_0^m(x^{(0)}) \leqslant x^{(0)}$.

Suppose that $\hat{T}x^{(k)} \leqslant x^{(k)}$ holds for some $t = k \in \mathbb{N}$. Since $A_k = A(x^{(k)}) = \{j \in N : \sum_i x^{(k)} w_{ij} + \varepsilon_j \geqslant u_j\}$, we have $x^{(k)}(I - \Lambda^{A_k} - \Lambda^{B^*}) + u\Lambda^{A_k} + \ell\Lambda^{B^*} = x^{(k)}$. Hence, we have

$$\begin{aligned}\Phi_{k+1}(x^{(k)}) &= u\Lambda^{A_k} + \ell\Lambda^{B^*} \\ &\qquad + [u\Lambda^{A_k}W + \ell\Lambda^{B^*}W + x^{(k)}(I - \Lambda^{A_k} - \Lambda^{B^*})W + \varepsilon](I - \Lambda^{A_k} - \Lambda^{B^*}) \\ &= u\Lambda^{A_k} + \ell\Lambda^{B^*} + (x^{(k)}W + \varepsilon)(I - \Lambda^{A_k} - \Lambda^{B^*}) \\ &= \hat{T}x^{(k)} \leqslant x^{(k)}\end{aligned} \tag{4.29}$$

Similar to the above argument, $I - (I - \Lambda^{A_k} - \Lambda^{B^*})W(I - \Lambda^{A_k} - \Lambda^{B^*})$ is non-singular, so $x^{(k+1)}$ is unique. Then, since Φ_{k+1} is an increasing operator, and its iteration converges from any initial point, we have $x^{(k+1)} = \lim_{m\to\infty} \Phi_{k+1}^m(x^{(k)}) \leqslant \cdots \leqslant \Phi_{k+1}(x^{(k)}) \leqslant x^{(k)}$. Next, since $x^{(k+1)} \leqslant x^{(k)}$, we have $A_{k+1} = A(x^{(k+1)}) \subset A(x^{(k)}) = A_k$. If $A_{k+1} = A_k$, then using the definition of A_{k+1} we have

$$\begin{aligned}\hat{T}x^{(k+1)} &= \{[x^{(k+1)}(I-\Lambda^{B^*})W+\ell\Lambda^{B^*}W+\varepsilon](I-\Lambda^{B^*})\}\wedge[u(I-\Lambda^{B^*})]+\ell\Lambda^{B^*}\\ &= [x^{(k+1)}(I-\Lambda^{B^*}-\Lambda^{A_{k+1}})W+u\Lambda^{A_{k+1}}W+\ell\Lambda^{B^*}W+\varepsilon](I-\Lambda^{B^*})(I-\Lambda^{A_{k+1}})\\ &\quad+u\Lambda^{A_{k+1}}+\ell\Lambda^{B^*}\\ &= [x^{(k+1)}(I-\Lambda^{B^*}-\Lambda^{A_k})W+u\Lambda^{A_k}W+\ell\Lambda^{B^*}W+\varepsilon](I-\Lambda^{B^*}-\Lambda^{A_k})\\ &\quad+u\Lambda^{A_k}+\ell\Lambda^{B^*}\\ &= \Phi_{k+1}x^{(k+1)}=x^{(k+1)}\end{aligned}$$

Then, $x^{(k+1)}$ is the fixed point of $\hat{T}$ if $A_{k+1}=A_k$. Moreover, since $A_{k+1}=A_k$ implies $\Phi_{k+1}=\Phi_{k+2}$, we have $x^{(k+1)}=x^{(k+2)}$.

On the other hand, suppose that $A_{k+1}\subsetneq A_k$. Denote $\hat{T}x_j^{(k+1)}$ and $\Phi_{k+1}x_j^{(k+1)}$ as the j-th entry of $\hat{T}x^{(k+1)}$ and $\Phi_{k+1}x^{(k+1)}$, respectively. Then, for $j\in A_{k+1}$, we have $\hat{T}x_j^{(k+1)}=u_j=\Phi_{k+1}x_j^{(k+1)}=x_j^{(k+1)}$, where the second equality holds because $j\in A_k$. For $j\in A_k\setminus A_{k+1}$, we have $\hat{T}x_j^{(k+1)}<u_j=\Phi_{k+1}x_j^{(k+1)}=x_j^{(k+1)}$. For $j\in B^*$, we have $\hat{T}x_j^{(k+1)}=\ell_j=\Phi_{k+1}x_j^{(k+1)}=x_j^{(k+1)}$. For $j\in N\setminus(A_k\cup B^*)$, we have $j\notin(A_{k+1}\cup B^*)$ so that

$$\begin{aligned}\hat{T}x_j^{(k+1)} &= \sum_{i\in B^*}\ell_iw_{ij}+\sum_{i\in A_{k+1}}u_iw_{ij}+\sum_{i\in N\setminus(A_{k+1}\cup B^*)}x_i^{(k+1)}w_{ij}+\varepsilon_j\\ &= \sum_{i\in B^*}\ell_iw_{ij}+\sum_{i\in A_{k+1}}u_iw_{ij}+\sum_{i\in N\setminus(A_k\cup B^*)}x_i^{(k+1)}w_{ij}+\sum_{i\in A_k\setminus A_{k+1}}x_i^{(k+1)}w_{ij}+\varepsilon_j\\ &= \sum_{i\in B^*}\ell_iw_{ij}+\sum_{i\in A_{k+1}}u_iw_{ij}+\sum_{i\in N\setminus(A_k\cup B^*)}x_i^{(k+1)}w_{ij}+\sum_{i\in A_k\setminus A_{k+1}}u_iw_{ij}+\varepsilon_j\\ &= \sum_{i\in B^*}\ell_iw_{ij}+\sum_{i\in A_k}u_iw_{ij}+\sum_{i\in N\setminus(A_k\cup B^*)}x_i^{(k+1)}w_{ij}+\varepsilon_j\\ &= \Phi_{k+1}x_j^{(k+1)}=x_j^{(k+1)}\end{aligned}$$

where we use $x_i^{(k+1)}=u_i$ for $i\in A_k$ by (4.28). Therefore, we see that $\hat{T}x^{(k+1)}\leqslant x^{(k+1)}$ when $A_{k+1}\subsetneq A_k$. Then, we get $\hat{T}x^{(k+1)}\leqslant x^{(k+1)}$. The Mathematical Induction shows that $\hat{T}x^{(t)}\leqslant x^{(t)}$ and $x^{(t+1)}\leqslant x^{(t)}$ for all $t=0,1,2,\ldots$.

The above argument also shows that if $A_{k+1}=A_k$ for some $k>0$, then $x^{(k+1)}$ is the fixed point of $\hat{T}$. Also, (4.29) implies that $x^{(k+2)}=\Phi_{k+2}x^{(k+1)}=\hat{T}x^{(k+1)}=x^{(k+1)}$, so that $x^{(t)}$

for $t \geqslant k+1$ remains constant. On the other hand, if $x^{(k+1)}$ is not the fixed point of $\hat{T}$, since $x^{(k+1)}$ is the fixed point of Φ_{k+1}, there exists $j \in A_k \setminus A_{k+1}$ such that A_{k+1} decreases. Since there are only n nodes, and A_0 contains at most $n-1$ elements due to $x^* \neq u$, A_t and $x^{(t)}$ stop to change after at most n iterations. Since the sequence $x^{(t)}$ is constant only at fixed point by (4.29), we obtain the equilibrium in at most n iterations. Now, in searching of $P = B^*$ from the power set of $B(\ell)$, since $x^* \neq \ell$ and then $B(\ell)$ contains at most $n-1$ agents, $\mathcal{P}$ has at most 2^{n-1} elements. Then, we need at most 2^{n-1} searching time for $P = B^*$ and n iterations for each possible $P \in \mathcal{P}$ to find the fixed point, so the overall iteration is at most $n2^{n-1}$. ☐

Proof of Lemma 4.4.6. Let f follow (4.17), $W \geqslant 0$ be stochastic, and $\varepsilon \in E_\varepsilon$. Let $g\colon \mathbb{R}^n \to \mathbb{R}$ be a strictly increasing function. Suppose that x^* is a solution to programming problem (4.21). Then, x^* satisfies $x^* \in [\ell, u]$ and $x^* \leqslant \max\{x^*W + \varepsilon, \ell\}$. We want to show that x^*_j satisfies either $x^*_j = u_j$ or $x^*_j = (\max\{x^*W + \varepsilon, \ell\})_j$ for all $j \in N$, where $(\max\{x^*W + \varepsilon, \ell\})_j$ is the j-th entry of $\max\{x^*W + \varepsilon, \ell\}$. Suppose not. Then, there is $i \in N$ such that $x^*_i < u_i$ and $x^*_i < (\max\{x^*W + \varepsilon, \ell\})_i$. Then, we can find $\delta > 0$ such that $x^*_i + \delta \leqslant u_i$ and

$$x^*_i + \delta \leqslant (\max\{x^*W + \varepsilon, \ell\})_i \leqslant (\max\{(x^* + \delta \mathbb{1}^i)W + \varepsilon, \ell\})_i,$$

where $\mathbb{1}^i_j = 1$ if $j = i$ and $\mathbb{1}^i_j = 0$ otherwise. Therefore, $x^* + \delta\mathbb{1}^i$ satisfies the conditions of programming problem (4.21). Since g is strictly increasing, we have $g(x^*) < g(x^* + \delta\mathbb{1}^i)$, contradicting that x^* is the solution. Hence, it must be that either $x^*_j = u_j$ or $x^*_j = (\max\{x^*W + \varepsilon, \ell\})_j$ for all $j \in N$, so that $x^* = \min\{\max\{x^*W + \varepsilon, \ell\}, u\}$. ☐

Proofs in Section 4.5.

Proof of Lemma 4.5.1. Suppose that f_i is increasing and continuously differentiable for all i, and $\rho(|W| \operatorname{diag} \beta) < 1$. Let $F(x) \coloneqq f(xW + \varepsilon) - x$. Then, the system (4.22) is $\mathrm{d}x/\mathrm{d}t = F(x)$. We follow Lyapunov's linearization method. We have the partial derivatives

$$\frac{\partial F_i(x)}{\partial x_m} = \begin{cases} f_i'(\sum_h x_h w_{hi} + \varepsilon_i) w_{ii} - 1 & \text{if } m = i, \\ f_i'(\sum_h x_h w_{hi} + \varepsilon_i) w_{mi} & \text{if } m \neq i, \end{cases}$$

Denote $f' := (f_1'(\sum_h x_h w_{h1} + \varepsilon_1), \ldots, f_n'(\sum_h x_h w_{hn} + \varepsilon_n))$. The Jacobian matrix of F is $W \operatorname{diag}(f') - I$. Let λ be the eigenvalue of $W \operatorname{diag}(f') - I$. The dynamics is asymptotically stable if the real part of λ, denoted as $\operatorname{Re}(\lambda)$, is negative for all λ. From the characteristic equation, eigenvalue λ satisfies $0 = \det((W \operatorname{diag}(f') - I) - \lambda I) = \det((W \operatorname{diag}(f') - (\lambda+1)I)$, so $\lambda+1$ is the eigenvalue of $W \operatorname{diag}(f')$. We then have $\operatorname{Re}(\lambda)+1 \leqslant |\lambda+1| \leqslant \rho(W \operatorname{diag}(f'))$.

Since the interaction functions are Lipschitz continuous, the derivatives are bounded $|f_i'| \leqslant \beta_i$ for all i, so that $\operatorname{diag}(f') \leqslant \operatorname{diag}(\beta)$. Then, we can show that $|(W \operatorname{diag}(f'))^k| \leqslant (|W| \operatorname{diag}(\beta))^k$ for $k \geqslant 1$. Since $\rho(A) \leqslant \|A^k\|^{1/k}$ for a matrix $A \in \mathbb{R}^{n \times n}$ and $k \geqslant 1$, we have $\rho(W \operatorname{diag}(f')) \leqslant \|(W \operatorname{diag}(f'))^k\|^{1/k} \leqslant \|(|W| \operatorname{diag}(\beta))^k\|^{1/k}$ for all $k \geqslant 1$. Taking the limit of k, we have $\operatorname{Re}(\lambda) \leqslant \rho(W \operatorname{diag}(f')) - 1 \leqslant \rho(|W| \operatorname{diag}(\beta)) - 1 < 0$. Since the real part of any eigenvalue of $W \operatorname{diag}(f') - I$ is negative, the system is asymptotically stable. □

PROOF OF LEMMA 4.5.2. Let f_i be differentiable for all i and $\rho(|W| \operatorname{diag} \beta) < 1$. Define the system and function F as (4.23). Define C_{ij} as

$$C_{ij} := \frac{\mathrm{d}x_j^*}{\mathrm{d}s_i} \frac{\partial s_i}{\partial x_i^*} x_i^*$$

for $i, j \in N$. For the steady state, by (4.23), we have $x_i^* = f(\sum_h x_h^* w_{hi} + \varepsilon_i)/s_i$, so that

$$\partial x_i^* / \partial s_i = -f\left(\sum_h x_h^* w_{hi} + \varepsilon_i\right) / s_i^2 = -x_i^* / s_i.$$

For $s = \mathbf{1}$, we have

$$C_{ij} = \frac{\mathrm{d}x_j^*}{\mathrm{d}s_i} \frac{\partial s_i}{\partial x_i^*} x_i^* = \frac{\mathrm{d}x_j^*}{\mathrm{d}s_i} \left(\frac{-s_i}{x_i^*}\right) x_i^* = -\frac{\mathrm{d}x_j^*}{\mathrm{d}s_i}. \tag{4.30}$$

After the removal of i, the system goes to a new steady state near the original one. Since $F = 0$ at both new and original steady states, we have $\mathrm{d}F/\mathrm{d}s = \mathbf{0}$, where $\mathbf{0}$ is an n by n

zero matrix. The total derivative $\mathrm{d}F = (\partial F/\partial s)\mathrm{d}s + (\partial F/\partial x)\mathrm{d}x$ gives

$$\frac{\mathrm{d}F}{\mathrm{d}s} = \frac{\partial F}{\partial s} + \frac{\partial F}{\partial x}\frac{\mathrm{d}x}{\mathrm{d}s} = \mathbf{0}. \tag{4.31}$$

By (4.30) and (4.31), since $\partial f/\partial x = W\,\mathrm{diag}(f')$, we have

$$\begin{aligned} C = -\frac{\mathrm{d}x^*}{\mathrm{d}s} &= -\left[\frac{\partial F}{\partial s}\bigg|_{(x,s)=(x^*,\mathbf{1})}\right]\left[\frac{\partial F}{\partial x}\bigg|_{(x,s)=(x^*,\mathbf{1})}\right]^{-1} \\ &= \mathrm{diag}\,(x^*)\left[I - W\,\mathrm{diag}\,(f'(x^*W+\varepsilon))\right]^{-1} \end{aligned}$$

where $f'(x^*W+\varepsilon) = (f_1'(\sum_h x_h^* w_{h1}\varepsilon_1), \ldots, f_n'(\sum_h x_h^* w_{h1}\varepsilon_n))$. The result follows from that $\sigma = (C\mathbf{1}^\top)^\top$. □

Other Lemmas.

Lemma 4.7.6. *If $W \geqslant 0$ or $W^\top \geqslant 0$ is weakly chained substochastic, then W is convergent.*[41]

Lemma 4.7.7. *Let $A \in \mathbb{R}^{n\times n}$ be non-negative. Then, $\rho(A) < 1$ if and only if $(I-A)^{-1}$ exists and $(I-A)^{-1} \geqslant 0$.*

Let $A \in \mathbb{R}^{n\times n}$ be a matrix. Define A_{-i} as the submatrix of A deleting row i and column i of A.

Lemma 4.7.8. *Let $A \in \mathbb{R}^{n\times n}$ be a non-negative matrix. Then, $\rho(A_{-i}) \leqslant \rho(A)$.*

Lemma 4.7.9. *Let $A \in \mathbb{R}^{n\times n}$ be an irreducible stochastic matrix. Then, $\rho(A_{-i}) < 1$.*

Proof. Let $A = (a_{ij}) \in \mathbb{R}^{n\times n}$ be an irreducible stochastic matrix. It is enough to consider the strongly connected graph with the least number of links, the ring. That is, consider the graph such that for all $i \in N$ we have $a_{i,j} = 1$ if $j = (i+1) \pmod n$, otherwise $a_{i,j} = 0$. Fix $i \in N$ and delete row i and column i to get A_{-i}. Then, the row sum of row $i-1$ is less than one. Also, every other node has a path to node $i-1$. Therefore, A_{-i} is weakly chained substochastic, so it follows from Lemma 4.7.6 that $\rho(A_{-i}) < 1$. We can extend the argument to an irreducible matrix, since if we remove a node i from a

---

[41] $W^\top$ is weakly chained substochastic if and only if W is column substochastic, and for each $j \in N$, either $\sum_i w_{ij} < 1$, or there exists $t \in N$ such that there is a path from t to j ($t \to \cdots \to j$) and $\sum_i w_{it} < 1$.

strongly connected graph, then i must be in a ring, and any other nodes have some path to that ring. $\square$

CHAPTER 5

Explaining Systematic Departures from Gibrat's Law

5.1. Introduction

One of the most debated statistical regularities in industrial dynamics is Gibrat's Law, which predicts that the growth rate of a firm is independent of its size (Gibrat, 1931, Sutton, 1997). Gibrat's Law allows a firm's growth to be viewed as a stochastic process, introducing heterogeneity. It has successfully explained the skewed distribution of firm sizes. These properties make Gibrat's Law become a prevalent assumption or a stylised fact in theoretical work (Ijiri and Simon, 1964, Lucas, 1978). For instance, Gibrat's Law plays a crucial role in explaining the power-law distribution of firm sizes and the tent-shaped distribution of firm growth (Simon and Bonini, 1958, Gabaix, 2009, Fu et al., 2005). Gabaix and Landier (2008) model CEO pay based on an exogenous distribution of firm sizes from a random growth process.

However, empirical tests of Gibrat's Law have generated controversy. It has been observed that small firms often exhibit faster growth rates or greater volatility compared to larger firms, leading to the rejection of Gibrat's Law, especially for downstream firms.[1] For instance, in a study of Italian firms, Grazzi and Moschella (2018) finds a negative size-growth relationship, even when accounting for firm age. Nevertheless, it is worth noting that this breakdown appears systematic, suggesting that Gibrat's Law may still hold for large firms.

Recent studies have uncovered deviations from Gibrat's Law within production networks. Guilmi and Fujiwara (2020) and Hiromitsu and Wataru (2020), studying Japanese production networks, have found that small firms experience higher fluctuation. Hiromitsu

---

[1] Hall (1987), Evans (1987a),Evans (1987b),, Dunne and Hughes (1994), Almus (2000), Calvo (2006), Falk (2008), Daunfeldt and Elert (2013), Oke (2018) and Aydogan and Donduran (2019) find that firm growth and firm size are negatively correlated, considering small firm or aggregated data. Hall (1987) and Dunne and Hughes (1994) document that the growth of small firms is more volatile growths than that of large firms.

and Wataru (2020) find that upstream firms tend to be smaller than those in the other levels of networks. Since upstream firms tend to be small, Guilmi and Fujiwara (2020) argue that upstream firms also have higher volatility than downstream firms. Further, Hiromitsu and Wataru (2020) observe that firms following a power law are typically not positioned at the upstream levels of production chains.[2]Since Gibrat's Law explains a power-law distribution, it implies that upstream firms depart from Gibrat's Law.

Overall, their observation suggests that a small firm's size and growth may be affected by its relative position or the hierarchical structure of a production chain. Motivated by these studies, this paper seeks to establish an endogenous production network model to explain the breakdown of Gibrat's Law for upstream firms and the power-law distribution of firm sizes among downstream firms.

Our model establishes a balanced production chain and demonstrates both the existence and uniqueness of equilibrium. We discover that hub-like networks exhibit robust features, with prices increasing in transaction costs and subcontractors' sizes growing as they progress to downstream levels. Additionally, simulations reveal that the length of a production chain decreases in transaction costs but increases with the degree of decreasing return to management. These properties enable us to study the impact of cost shocks on prices and network structure.

Furthermore, our findings indicate that when assembly costs exponentially increase with the task, the firm sizes of the subcontractors follow a power law in each production chain. In addition, if home producers experience higher volatility in technology compared to subcontractors, Gibrat's Law breaks down. In other words, when assembly technology is more stable in the short run, the home producers at the most upstream level exhibit greater volatility in firm sizes than downstream subcontractors. If we further examine the economy across many production chains, it appears that the upstream or small firms experience more fluctuations than downstream or large firms.

---

[2] Guilmi and Fujiwara (2020) and Hiromitsu and Wataru (2020) also demonstrate that the distribution of large firms' sizes is Pareto, while that of small firms' sizes is not. This breakdown of Pareto distribution among small firms is a well-known observation (See e.g., (Saito et al., 2007, Fujiwara and Aoyama, 2010, Almsafir et al., 2015)).

Moreover, we show that it is possible to identify a locally optimal number of suppliers, although it may be unrealistic to be fixed. Finally, this paper extends a balanced production network to a more realistic model that allows firms to choose an optimal number of suppliers when they subcontract. We show that the extended network also has a unique equilibrium and exhibits robust comparative statics.

This paper modifies the framework incorporating multiple upstream partners from Kikuchi et al. (2018), Yu and Zhang (2019) and Kikuchi et al. (2021). These studies consider a perfectly competitive economy where all firms are ex-ante identical.[3] Facing market prices, firms aim to minimise their home production and outsourcing costs, which are subject to transaction costs and input prices. This framework highlights a trade-off between intra-firm coordinate costs and inter-firm transaction costs, as originally proposed by Coase (1937) and Williamson (1979). The key insight is that external transaction costs encourage firm growth, whereas intra-firm coordination costs discourage it. Consequently, a firm's boundary is determined at the point where the cost of managing an additional task equals the cost of acquiring a similar input or service from the market.

Kikuchi et al. (2018) develope a pricing function to capture this trade-off and compute equilibrium prices for all firms on a production chain using a recursive method. The model of Kikuchi et al. (2018) has the advantage that the equilibrium exists and is tractable by dynamic programming technique, thus determining firm sizes at each production stage endogenously. In their model, firm sizes increase as production moves downstream levels, capturing some realistic features. Moreover, their framework generate heterogeneity by the endogenous network but with ex-ante identical agents, helping us to understand how the network structure affects firm-size distribution. While they demonstrated the existence of equilibrium, they did not establish its uniqueness.

Building on Kikuchi et al. (2018), we construct a hub-like production network. Two reasons make us concentrate on hubs: they render a network more tractable, and some

---

[3] Some features of a multi-sourcing strategy are that it could mitigate the risk of suppliers' failure, allow firms to obtain broader supplier capabilities, and encourage competition among suppliers (see, e.g., Jin and Ryan (2012)). Considering the competition, our model also assumes that the intermediate market is perfectly competitive as Fally and Hillberry (2018) and firms are ex-ante identical.

literature suggests that hub-like firms are crucial to forming a power-law distribution.[4] Unlike Kikuchi et al. (2018), we let each firm make a binary decision on outsourcing, rather than decide a range of home-producing tasks, to focus on the hub-like production network. In this case, it can be shown that the equilibrium not only exists but also is unique. Moreover, it successfully explains the power-law distribution of firm sizes.

In our model, the structure of networks is more tractable than Kikuchi et al. (2018). In detail, subcontractors divide their allocated tasks into a fixed number of sub-modules and outsource each sub-module to upstream suppliers. Subcontractors have to pay assembly costs or transaction costs. Consequently, the minimum number of suppliers is two since transaction and assembly costs prevent firms from outsourcing to only one supplier. Therefore, hubs exist throughout the production network except at the upstream levels. In other words, home producers only exist in the most upstream level.

Related Literature. Guilmi and Fujiwara (2020) offer an explanation for the breakdown of Gibrat's Law among upstream firms, attributing it to amplified demand shocks along a production chain. However, both the production network and the prices are exogenous in their model, meaning the network is randomly given, and the price is integrated into stochastic shocks. Hence, their model is not able to study the impact of the shape of networks and cost shocks. Furthermore, Guilmi and Fujiwara (2020) do not explicitly explain why the sizes of downstream or large firms follow a power law, implying that Gibrat's Law holds for downstream firms in their model.

The higher volatility among small firms documented in Guilmi and Fujiwara (2020) does not necessarily imply that Gibrat's Law holds for large firms and breaks down for small firms. It could be the case that the breakdown occurs for both small and large firms. Therefore, if Gibrat's Law is rejected, it would be better to explain the power-law distribution without it. In response, this paper provides one origin of the power-law distribution of firm sizes without relying on Gibrat's Law and ex-ante heterogeneity. To simplify the

---

[4] For instance, Bernard et al. (2019) suggests that the number of suppliers for a firm is proportional to its firm size. The hub also plays an important role in transmitting shocks. Acemoglu et al. (2012) imply that super large suppliers with multiple customers can have a significant impact on aggregate shocks. Carvalho (2014) underscores that hub-like units provide shortcuts or act as influential conductors such that they help shocks propagate throughout the economy.

model, we focus on a balanced and hub-like production network, where firms have a fixed number of multiple suppliers if they choose to outsource. We find that the hub-like network can generate a power-law distribution of firm size for subcontractors and illustrate the departure of Gibrat's Law for upstream firms.

Another reason for the interest in the breakdown of Gibrat's Law is its connection to development policy. If small firms exhibit higher growth rates, policymakers may be motivated to invest in young and innovative small firms for economic development.[5]. While small firms may have to do with economic development, large firms could impact the business cycle. The power law of the firm size distribution is influential in the research of aggregate shock. For example, Gabaix (2011) shows that idiosyncratic shocks at the firm level can translate into aggregate fluctuations when the firm size distribution has a heavy tail, and the largest firms contribute disproportionately to the aggregate production.[6] Moreover, Carvalho and Grassi (2019) show that the dispersion of firm sizes affects the variance of aggregate output in a Hopenhayn economy, which implies that large firms' dynamics induce significant macro-level fluctuations.

The shape of the production network also affects aggregate shocks. Acemoglu et al. (2012) show how weighted out-degree and second-order interconnections of the inter-sectoral network affect aggregate volatility, where the aggregate output is a linear combination of sectoral shocks.[7] If a network is asymmetric, then aggregate volatility can not be averaged out.[8] Carvalho (2014) stresses that the shape of a network influences the propagation of sectoral shocks, where idiosyncratic shocks in a single sector propagate along a chain and then generate aggregate shocks.

---

[5] For example, a policymaker would support young small firms with high R&D intensity by relaxing financial constraints, since they can achieve significantly higher innovative sales (Schneider and Veugelers (2010)) Also, small firms have higher job creation rates than large firms (Schreyer (2000)).

[6] Gabaix (2011) shows that when the company size has infinite variance, the aggregate volatility decays at a slower rate than $1/\sqrt{N}$ such that the aggregate fluctuation is substantial even if N is large, where N denotes the number of firms in the economy. The fat tail breaks the central limit theorem, and the idiosyncratic shocks remain on aggregate output.

[7] The weighted out-degree is the output share of one sector in the production network which can be computed from the input-output matrix. The weighted out-degree is higher when the supplier is large. The second-order interconnection is higher when there is a clustering of significant sectors in the network.

[8] In the sense that the corresponding weighted out-degree sequence or second-order degree sequence has a power tail.

Similarly, Tahbaz-Salehi et al. (2016) illustrate that the rate at which aggregate volatility decays is determined by the structure of input-output networks as an economy becomes more disaggregated. They demonstrate how network interactions can propagate and amplify microeconomic shocks. Baqaee (2018) shows that networks influence the amplification and pattern of shocks. In the study by Bigio and La'O (2016), a network model with financial constraints reveals that shocks propagate idiosyncratic shocks and manifest as aggregate shocks through two channels: a fall in total factor productivity and an aggregate labour wedge distortion. Consequently, if our concern is economic growth and stability policy, it becomes imperative to understand the existence of a power-law distribution of firm sizes and the reason behind the breakdown of Gibrat's Law.

The real-world production network is too intricate and challenging to analyze comprehensively. However, there are two types of networks that serve as approximations to real-world complexities. Baldwin and Venables (2013) describe two highly simplified models: snakes and spiders. A snake represents a sequence of production stages, with value-added from the upstream stage to the downstream stage. Conversely, a spider is a process where numerous components (limbs) converge from the upstream stage to be assembled into a final output. The spider model results from the unbundling process, which breaks down production tasks into various sub-modules All production networks can be viewed as a combination of snakes and spiders.

Some research employs snakes as a model to investigate production behaviour, such as integration and specialisation. For example, Antràs and Chor (2013) show that the incentive to integrate suppliers depends on the relative position in a sequential chain. They predict that the average demand elasticity faced by final-good producers affects integration choices. Costinot et al. (2012) study wage inequality and vertical specialisation within countries with a sequential feature.[9]

---

[9] The other examples are: Costinot et al. (2013) study the sequential chain subject to failure and the connection between the degree of specialisation and the stage of the supply chain. They show that countries with lower probabilities of making mistakes specialise in later stages of production, with a continuum of sequential tasks produced in countries that are themselves sequentially ordered in equilibrium. Levine (2012) studies that the trade-off between specialisation and failure determines the optimal chain length, where a longer chain is more fragile.

In reality, firms often engage with multiple partners, forming complex interaction patterns. This paper specifically concentrates on spiders or hub-like networks. Acemoglu et al. (2012) and Barrot and Sauvagnat (2016) present evidence from the U.S. firms, showing that many firms and sectors adopt a "star network" structure. Moreover, the result derived from a sequential chain model may not hold if the assumption of a sequential structure is dropped. For example, Fattorini et al. (2017) find that demand elasticity is not a significant determinant for integration choices, contradicting the prediction of Antràs and Chor (2013) when the assumption of a unique linear sequence of production stages is removed. They argue that inputs participate in multiple stages of production, rather than being confined to a single stage as in a sequential chain. Therefore, this paper focuses on the more realistic spider networks.

Our model endogenously determines the equilibrium price and can analyze cost shocks under a multisourcing strategy. The model exhibits the price co-movement through the input-output network. Input-output linkages generate price co-movement and inflation synchronisation, where local cost shocks could be translated into global inflation (Antoun de Almeida, 2016, Auer et al., 2017, Bilgin and Yilmaz, 2018, Auer et al., 2019, Kamber and Wong, 2020).

This paper also explores how transaction costs impact the network. As transaction costs influence firm boundaries, the structure of a production network depends on both transaction costs and production costs. Aral et al. (2018) indicate that coordination information technology reducing coordination and communication cost, which is one of transaction cost, is correlated with the number of suppliers, while vendor-specific I.T. is associated with fewer suppliers. Boehm and Oberfield (2020) imply that the costs of contract enforcement affect firms' outsourcing decisions and then distort the economy. Recent economic shocks have significant impacts on transaction costs. For instance, the trade war between China and the U.S. (Amiti et al., 2019, Fajgelbaum et al., 2020). Another example is the COVID-19 pandemic. Barua (2020) and Das (2020) point out that the pandemic causes logistic problems and increases transportation costs. In our framework, transaction costs are completely passed through to the price of final goods, which is compatible with Amiti et al. (2019).

Other related literature investigates whether the positions of firms affect their performance. For example, Mahy et al. (2019) and Gagliardi et al. (2019) find that the productivity of Belgian firms increases on average as they go upstream levels and workers in more upstream firms obtain higher wages in the Belgian manufacturing industry, respectively. Chen (2017) find that upstream industries have more severe wage inequality than downstream industries in China manufacturing data, where downstream industries tend to do processing and assembly work. Szymczak et al. (2019) observe that workers in Central and Eastern European countries earn more when their industries are at the beginning or the end of production chains than in the middle. These papers demonstrate the importance of the position and structure of networks.

Section 5.2 establishes a model of a balanced production chain. It also presents the proofs for existence and uniqueness and discusses some robust properties, including comparative statics. Section 5.3 shows the distribution of firm sizes of subcontractors and the departure of Gibrat's Law for upstream firms. Section 5.4 provides an extended model that allows firms to choose an arbitrary number of suppliers. The code can be found at github.com/chien-y/Departure_Gibrat_law.

5.2. Model

A supply chain produces a single unit of the final good in a perfectly competitive market. Positioned at the most downstream, there exists one and only one firm that sells this final good to consumers. Suppose that all firms are ex-ante identical, indicating that they share the same cost functions for both in-house production and assembly work.

Within a production chain, firms collaborate to execute a continuum of tasks $[0, 1]$ to produce a final good. The interval $[0, 1]$ serves as a normalised measure of tasks required to produce one unit of good. To elaborate, the measure of tasks is denoted as $\ell \in [0, 1]$, representing the allocation of tasks in an outsourcing contract.

Moreover, each firm in the supply chain has $\kappa > 1$ number of upstream partners. Upon receiving an allocation of tasks ℓ, each firm makes a strategic decision between in-house production and subcontracting the tasks ℓ to κ suppliers. In the case of in-house production, a firm incurs a production cost $c(\ell)$. Conversely, if a firm ops to be a subcontractor, it divides the tasks into κ portions and outsources each portion, equivalent to ℓ/κ units, to its upstream partners. Subsequently, a subcontractor is responsible for assembly work, combining the κ pieces of components, incurring an assembly cost denoted as α.

Furthermore, when firms engage in the sell of intermediates or final goods, they are obligated to pay the transaction costs τ.[10] The transaction costs and the cost function satisfy the following assumptions.

Assumption 5.2.1. The transaction cost τ satisfies $0 < \tau < 1$.

Assumption 5.2.2. The cost function $c : [0, 1] \to \mathbb{R}$ is convex, $c(0) = 0$ and $c'(0) > 0$.

Under Assumption 5.2.2, given convexity of the cost function and the derivative $c'(0)$ as $\lim_{h \downarrow 0} c(h)/h$, the cost function is strictly increasing and non-negative.

Additionally, we assume that the assembly costs increase with the allocated tasks ℓ in Assumption 5.2.2. Two intuitions support this assumption. Firstly, if ℓ not only measures

---

[10] It matters little whether the transaction is paid by the buyer or the seller in the model. The transaction costs include search and information costs, bargaining and decision costs, and contract enforcement costs (Dahlman, 1979).

the number of tasks but also the diversity of tasks, a higher ℓ implies increased variety, leading to heightened complexity.[11] This increased complexity, in turn, results in higher assembly costs. Secondly, even though firms consistently divide tasks into κ pieces and subcontract κ/ℓ tasks to suppliers, a greater variety of tasks introduces more possibilities for task combinations. Firms need to strategically plan the task division and assembly line decision, incurring costs that rise with the task variety ℓ. Consequently, the assembly costs are proportional to the variety or quantity of tasks, making the assumption reasonable.

Specifically, we further assume that the assembly cost takes the form $b\ell^q$, where $q \geqslant 0$, b represents the efficiency for assembly work, and q captures the return to scale of management.

Assumption 5.2.3. The assembly cost is $\alpha(\ell) := b\ell^q$ for some $q \geqslant 0$ and $b > 0$.

In Assumption 5.2.3, it is permissible for the assembly cost function to be concave. The idea stems from the concept that labour can specialise in the assembly line, resulting in an increasing return to scale for assembly work. Aizcorbe (1992) illustrates this with an example that an automobile assembly plant experiences an increasing return to labour in the short run, attributed to the specialisation of tasks. Thus, the assumption allows for the assembly cost to exhibit concavity in contrast to the convex nature of the production cost.

Given these assumptions, each firm in the chain receives a contract assigning it ℓ tasks to complete. Each firm then minimises costs by deciding between home production and subcontracting. If a firm opts for home production, then its cost is $c(\ell)$. Conversely, if a firm chooses to subcontract, then its cost becomes $\kappa p(\ell/\kappa) + \alpha(\ell)$ for purchasing inputs from κ suppliers and conducting the assembly work, where $\kappa = 2, 3, \ldots$, and each supplier completes ℓ/κ fraction of tasks. Subsequently, accounting for the transaction costs $\tau \in (0, 1)$ associated with selling the intermediate inputs, the profit is given by

$$(1 - \tau)p(\ell) - \min\{\kappa p(\ell/\kappa) + \alpha(\ell),\, c(\ell)\}$$

---

[11] This concept is similar to ElMaraghy et al. (2013), Hu et al. (2008) and Hu et al. (2011). They show that the product variety complicates the design and operation of the assembly system and decreases the efficiency and quality of production.

for each $\ell \in [0,1]$ and $\kappa \in 2, 3, \ldots$. Note that if $\kappa = 1$, firms do not have the incentive to subcontract (choose the first term inside the minimum function) since all firms have identical costs, and the transaction costs are larger than zero. This implies that it must be $\kappa > 1$.

Let $\delta = 1/(1-\tau)$. The parameter δ represents the transaction cost if buyers instead of sellers bear it. Since the market is perfectly competitive, firms have zero profit in equilibrium. Hence, the pricing equation solves

$$p(\ell) \coloneqq \delta \min \left\{ \kappa p\left(\frac{\ell}{\kappa}\right) + \alpha(\ell), c(\ell) \right\}. \tag{5.1}$$

5.2.1. Equilibrium of Production Chain. Under the equilibrium price (5.1), the structure of the supply chain is as follows. At the most downstream level, indexed level 1, there exists a single firm that receives a task allocation ($\ell = 1$). If this firm chooses to subcontract, then there are κ firms in the next upstream level, indexed level 2, since the number of suppliers is fixed at κ in the model. All firms at level 2 receive $\ell = 1/\kappa$ units of task allocations. Again, if the firms at level 2 choose to subcontract, then there are κ^2 firms at level 3, and each supplier at level 3 implements $1/\kappa^2$ units of tasks. Therefore, by induction, there are κ^{n-1} number of firms at level $n \geqslant 1$, and each firm at level n implements κ^{1-n} units of tasks. See Figure 1 for the visual structure.

Moreover, if $m < \infty$ is the level index for the most upstream level of the chain, there are $1 + \kappa + \cdots + \kappa^{m-1} = (\kappa^m - 1)/(\kappa - 1)$ firms in the chain.

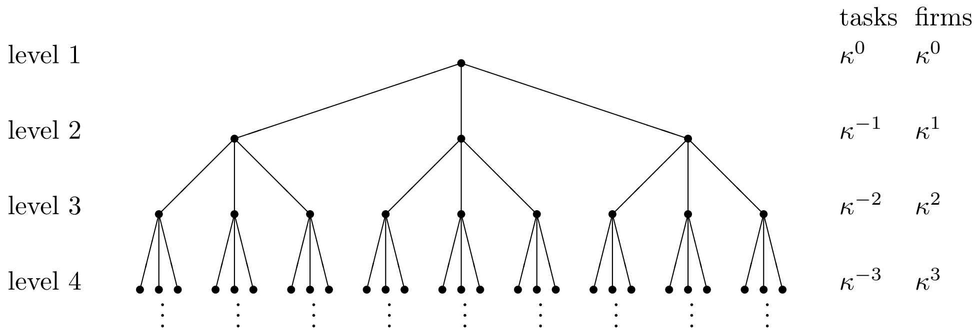


FIGURE 1. The structure of a balanced production chain. Take $\kappa = 3$ for example.

An allocation of tasks $\boldsymbol{\ell} := \{\ell_i\}$ is called *feasible* if there is $m \in \mathbb{N}$ such that $i = 1, \cdots, (\kappa^m - 1)/(\kappa - 1)$ and $\ell_i = \kappa^{1-n}$ for $(\kappa^{n-1} - 1)/(\kappa - 1) < i \leqslant (\kappa^n - 1)/(\kappa - 1)$ for some integer $1 \leqslant n \leqslant m$. Here, the integer m defines the length of the production chain. If firm i at level $n \leqslant m$ receives a feasible allocation, its allocated measure of tasks is κ^{1-n}. The length of the chain m and the length of sequence $\{\ell_i\}$ are finite for a feasible allocation.

Given a feasible ℓ with length $m \in \mathbb{N}$, the profit to firm i with the outsourcing or home production costs (5.1) is

$$\pi_i := (1 - \tau)p(\ell_i) - \min\{\kappa p(\ell_i/\kappa) + \alpha(\ell_i),\, c(\ell_i)\}, \tag{5.2}$$

for $\ell_i \in \boldsymbol{\ell}$, where the revenue is $(1 - \tau)p(\ell_i)$ with transaction cost $\tau \in (0, 1)$. Define the set of all possible tasks by $\mathcal{D} := \{\kappa^{-n} : n \in \mathbb{N} \cup \{0\}\}$.

Definition 5.2.1. Given a price function $p\colon [0, 1] \to [0, \infty)$, a feasible allocation of tasks $\boldsymbol{\ell} = \{\ell_i\}$ with length $m \in \mathbb{N}$ and the corresponding profits $\{\pi_i\}$ defined by (5.2), a tuple $(p, \boldsymbol{\ell})$ is called an *equilibrium* if

(1) $p(0) = 0$,
(2) $(1 - \tau)p(\ell) - \min\{\kappa p(\ell/\kappa) + \alpha(\ell),\, c(\ell)\} \leqslant 0$, for all $\ell \in \mathcal{D}$, and
(3) $\pi_i = 0$ for all i.

Condition (1) in definition 5.2.1 excludes the possibility of positive profits for suppliers providing initial inputs, considering $c(0) = 0$. Moreover, condition (2) ensures that no firm in the supply chain has an incentive to deviate from its decision, and inactive firms are unable to extract positive profits. Furthermore, given the assumption of perfect competition in the market, all firms in the chain are expected to have zero profits, thus establishing condition (3).

5.2.2. Existence and Uniqueness of Equilibrium. In this section, we show both the existence and uniqueness of equilibrium. A computational method is also provided. The proofs in the following sections are all presented in Appendix 5.6.

Define the set of price functions by $\mathcal{P} := \{p : \mathcal{D} \to \mathbb{R}_+ : \delta c'(0)\ell \leqslant p(\ell) \leqslant \delta c(\ell), \forall \ell \in \mathcal{D}\}$. We will show that there is a unique equilibrium price in $\mathcal{P}$. Define operator T by

$$Tp(\ell) := \delta \min \left\{ \kappa p \left(\frac{\ell}{\kappa} \right) + \alpha(\ell), c(\ell) \right\}. \tag{5.3}$$

where $\ell \in \mathcal{D}$ and $p : \mathcal{D} \to \mathbb{R}_+$. The operator T is analogous to a Bellman operator with respect to the price function (5.1). If T has a fixed point, that fixed point is also a solution to the price function (5.1) by its definition. Define $\bar{x} := \sup\{x \in (0, 1] : c'(x) \leqslant \delta c'(0)\}$ for the following proposition. The next proposition shows both the existence and uniqueness of the fixed point of T.

Proposition 5.2.1. *If Assumption 5.2.1, 5.2.2 and 5.2.3 hold, the following statements are true.*

- *(a) T is a self-map on $\mathcal{P}$.*
- *(b) T has a unique fixed point $p^* \in \mathcal{P}$.*
- *(c) p^* is a unique solution to (5.1).*
- *(d) p^* can be computed by finite iterations: $p^* = T^n p$ for all $n \geqslant 1 - \ln \bar{x} / \ln \kappa$ and any $p \in \mathcal{P}$*
- *(e) Under p^*, firms always choose to produce in-house if their task allocations are smaller or equal to $\bar{x}$, i.e. $\ell \leqslant \bar{x}$.*
- *(f) p^* is strictly increasing.*

Proposition 5.2.1 asserts the existence and uniqueness of equilibrium price p^* in $\mathcal{P}$, which can be computed by iteration with any initial guess from $\mathcal{P}$. Moreover, since c is strictly increasing, Proposition 5.2.1 shows that p^* is strictly increasing.

As indicated by Proposition 5.2.1, the bounds for $\mathcal{P}$ carry substantial economic significance, capturing all equilibrium prices. The upper bound $\delta c(\ell)$ for $\mathcal{P}$ suggests that the equilibrium price is less than or equal to the costs of home production. This upper limit aligns with the rationale that in a perfect competition market, where the price equals the total cost, outsourcing becomes economically favorable due to reduced costs, even after factoring in transaction costs.

As for the lower bound $\delta c'(0)\ell$ for $\mathcal{P}$, it represents the costs when the decreasing return to management is eliminated through outsourcing. Image a scenario where both transaction costs and assembly costs are absent; under such conditions, firms would invariably opt for outsourcing. Consequently, this leads to an infinite number of levels in the production network. The "infinite outsourcing" results in the production technology approximating a linear function when we view across the entire chain. This elimination of decreasing return of management translates to a production technology with a slope of $c'(0)$. Therefore, the lower bound not only leverages the convexity of the cost function for the proof but also reflects the cost or price when the return to management is constant.

Proposition 5.2.1 also shows that the equilibrium price can be determined through a finite number of iterations, enhancing computational feasibility. Moreover, a firm invariably becomes a home producer if its task allocation is below the critical value $\bar{x}$, which implies that the length of the production chain is always finite under equilibrium price. Note that the length of the production chain is also the number of iteration for computation $1 - \ln(\bar{x} - \kappa)$, provided in Proposition 5.2.1. As discussed in Section 5.2.1, if we ascertain the equilibrium length of the production, we can determine the task allocations. Therefore, there exists a feasible allocation $\boldsymbol{\ell}^*$ such that $(p^*, \boldsymbol{\ell}^*)$ constitutes an equilibrium for the production chain, in line with Definition 5.2.1.

Let p^* be the fixed point of operator T. Then, p^* is the solution of the price equation (5.1). Let m^* denote the number of total levels corresponding to p^*. According to Proposition 5.2.1, there exists a maximal possible length $\bar{m} \in N$ such that $m^* \leqslant \bar{m}$. Thus, m^* is finite, and the corresponding allocations $\boldsymbol{\ell}^*$ are well defined. The following proposition shows that the solution $(p^*, \boldsymbol{\ell}^*)$ constitutes a unique equilibrium.

Proposition 5.2.2. *Suppose that 5.2.1, 5.2.2 and 5.2.3 hold. If p^* is a fixed point of T, then there exists a feasible task allocations $\boldsymbol{\ell}^*$ such that $(p^*, \boldsymbol{\ell}^*)$ is a unique equilibrium for the production chain.*

Since the number of suppliers is fixed and then the network structure and firm sizes are tractable, Proposition 5.2.2 shows the uniqueness of equilibrium, although we lose some degree of freedom for network structure. Note that the proof employs the convention that firms will engage in in-house production when the costs for home production and subcontracting are equal. Proposition 5.2.2 also implies that the unique equilibrium of the production chain can be identified after we compute the equilibrium price from Proposition 5.2.1. Consequently, these results show that the upper bound and lower bound for the pricing function are deemed reasonable.

5.2.3. Properties of Equilibrium. This section attempts to characterize various properties of equilibrium $(p^*, \boldsymbol{\ell}^*)$. It includes comparative statics analysis using numerical methods and examples to reinforce the validity of the model.

5.2.3.1. Length of Production Chain, Price, and Allocations. In this section, we present some observations and intuitions about equilibrium. To begin with, we characterize the maximal attainable length and task allocations within the production chain in equilibrium. Note that the range of tasks is $\kappa^{-(m-1)}$ at level m, and the largest attainable length $\bar{m}$ is the larges integer satisfying $\kappa^{-(m-1)} \leqslant \bar{x}$ by Proposition 5.2.1. Then, the equilibrium number of levels m^* in the production chain is at most $\bar{m} := \lceil 1 - \ln(\bar{x})/\ln(\kappa) \rceil$, where $\lceil r \rceil$ denotes the smallest integer greater than $r \in \mathbb{R}$. Since $\bar{x} := \sup\{x \in (0,1] : c'(x) \leqslant \delta c'(0)\}$, it is evident that the transaction costs, the convexity of cost function and the number of suppliers influence the length of the chain.

The maximum possible number of levels $\bar{m}$ decreases with both the number of upstream partners κ and transaction costs τ, which results in an increase in $\bar{x}$.[12] This indicates that transaction costs discourage the subcontract and limit the length of the chain. Hence, we can characterize some properties of the equilibrium price, by $\kappa, \delta, \alpha(.)$ and $c(.)$ functions, before computing it explicitly.

Here, we present a method to compute the optimal length without explicitly computing the equilibrium prices. Define the function $f(\ell) := \kappa\delta c(\ell/k) + \alpha(\ell) - c(\ell)$ for $\ell \in [0,1]$. This function represents the additional costs of subcontracting compared to in-house production, assuming that the subcontracting partners are home producers. Observe that $f(0) = 0$ by our assumptions, and $f(\ell) < 0$ indicates that the outsourcing costs are less than in-house production costs. Suppose that f has a root $\hat{\ell}$ in $(0,1]$. In the appendix,

---

[12] Since $-\ln \bar{x} > 0$, $\bar{m}$ is decreasing in κ.

we demonstrate that if f is strictly concave, then a firm chooses in-house production when $\ell \leqslant \hat{\ell}$, and opts for outsourcing when $\ell > \hat{\ell}$.[13] This also implies that if $\hat{\ell} < 1$, then the equilibrium length exceeds 1.

Regarding the equilibrium prices, we are aware that p^* is strictly increasing in ℓ, as shown by Proposition 5.2.1. In addition, we can gain insight into the price function by the following iteration. If the total number of levels is $m \in \mathbb{N}$, the price of final product $p(1, m)$ follows

$$\begin{aligned} p(1,m) &= \delta\kappa p(\kappa^{-1}) + \delta b \\ &= \delta\kappa \left[\delta\kappa p(\kappa^{-2}) + \delta b \kappa^{-q}\right] + \delta b \\ &= (\delta\kappa)^2 p\left(\kappa^{-2}\right) + \left(\delta\kappa^{1-q} + 1\right)\delta b \\ &= \dots \\ &= (\delta\kappa)^{m-1} p\left(\kappa^{1-m}\right) + \left[(\delta\kappa^{1-q})^{m-2} + \dots + \delta\kappa^{1-q} + 1\right]\delta b \\ &= (\delta\kappa)^{m-1}\delta c\left(\kappa^{1-m}\right) + \left[\left(\delta\kappa^{1-q}\right)^{m-2} + \dots + \delta\kappa^{1-q} + 1\right]\delta b \\ &= (\delta\kappa)^{m-1}\delta c\left(\kappa^{1-m}\right) + \frac{\left(\delta\kappa^{1-q}\right)^{m-1} - 1}{\delta\kappa^{1-q} - 1}\delta b. \end{aligned} \tag{5.4}$$

In (5.4), the first term of the right-hand side represents the in-house production costs with tasks κ^{1-m}. The second term reflects the assembly costs, accounting for transaction costs. Clearly, if m is fixed, equation (5.4) shows that the price of finished goods increases with transaction costs δ, assembly cost parameter b, and the costs of in-house production c. However, note that the optimal length of the chain m^* will change if we alter these parameters. Consequently, the impact of these parameters on the equilibrium price is ambiguous from this perspective, as changes in these parameters may influence the optimal length of the chain, thus affecting the overall equilibrium price.

Recall that the firms become home producers if they are at level $\bar{m}$. To this end, we can also define the equilibrium price of the final good $p^*(1)$ as

$$p^*(1) := \min_{m=1,\dots,\bar{m}} \left\{ (\delta\kappa)^{m-1}\delta c\left(\kappa^{1-m}\right) + \frac{\left(\delta\kappa^{1-q}\right)^{m-1} - 1}{\delta\kappa^{1-q} - 1}\delta b \right\}. \tag{5.5}$$

---

[13] Since $f''(\ell) = \delta/\kappa c''(\ell/\kappa) - c''(\ell) + \alpha''(\ell)$, we can check that $f'' < 0$ if $\delta/\kappa < 1$, $c'' > 0$ and $\alpha'' \leqslant 0$.

This equation can be interpreted as that the organizer of the supply chain selects an optimal length, given that the number of suppliers is fixed. However, since the choice variable m is discrete, it is challenging to characterize m^* by (5.5).

5.2.3.2. Firm Sizes. Define the value-added as $c(\ell)$ for home producers and as $\alpha(\ell)$ for subcontractors. Given that the assembly cost is increasing for tasks, the value added for each subcontractor increases as it progresses downstream levels. If we consider firm size as the value-added, the model exhibits the characteristic that downstream subcontracting firms are larger than upstream subcontracting firms. This result aligns with Kikuchi et al. (2018) and Fally and Hillberry (2018), which observe a negative correlation between "upstreamness" and the value-added.[14]

On the other hand, the firm sizes for the most upstream firms or home producers are indeterminate. Home producers may be smaller than downstream firms since they receive the least amount of tasks in the production chain. However, they could also be large under some cost functions and parameters. This feature reflects the complexity of the real world, where empirical studies show that countries or industries at the upstream and downstream extremities of the chain often have higher shares of value-added than those in the middle. The most upstream firms may have higher value-added such that the curve of value-added along the chain is like a "smile" curve (Ito and Vézina, 2016, Aggarwal, 2017, Rungi and Del Prete, 2018, Stöllinger, 2019).

5.2.3.3. Comparative Statics. In a perfect competitive economy, prices reflects total costs, leading to an intuitive expectation that equilibrium prices would increase with any costs, including transaction costs, assembly costs, or production costs. The following proposition demonstrates that the shift in equilibrium price is monotonic with respect to these costs.

PROPOSITION 5.2.3. *Suppose that 5.2.1, 5.2.2 and 5.2.3 hold. The equilibrium price is increasing in transaction cost, assembly cost, and home production cost.*

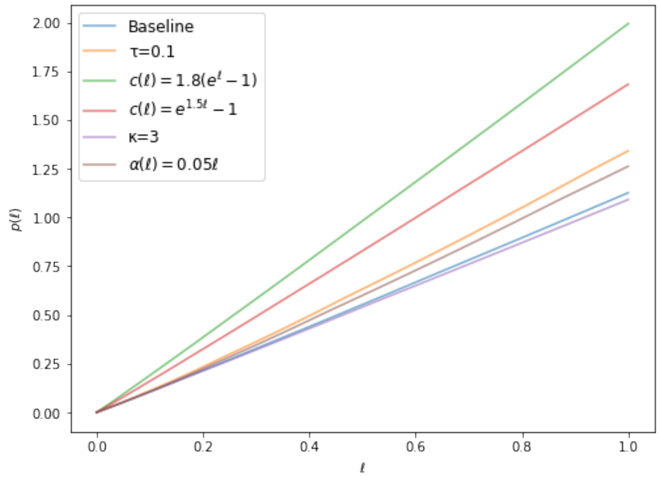
---

[14] The upstreamness is the relative position in the chain. The firm is in a more upstream position if the measure of upstreamness is higher. See Antràs et al. (2012), Antràs and Chor (2018) and Wang et al. (2017).

The proof relies on the monotonicity of operator T. It can also be shown that the price of the final good is strictly increasing in transaction costs, assembly costs, and home production costs, where the proof leverages the fact that the equilibrium price can be characterized using the optimal length, as expressed in (5.4).

EXAMPLE 5.2.1. This numerical example employs the function $c(\ell) = A(e^{\theta\ell} - 1)$ as an illustration. The change in price with respect to parameters $\tau, c(\ell), \kappa, \alpha(\ell)$ are depicted in Figure 2. The result align with Proposition 5.2.3. It also indicates that the prices are decreasing in the number of suppliers κ. The intuition for decreasing prices with κ lies in the fact that as κ increases, tasks are split into smaller pieces, reaching home production criterion $\ell \leqslant \bar{x}$ in a shorter length.

Moreover, the length is observed to increase in cost function $c(\ell)$, and decrease with transaction costs, the number of suppliers, and assembly costs. In this example, the total assembly costs decline as the length becomes shorter, but it may not be the case if $q > 1$. □



(A)

Model	Length
Baseline	5
$\tau = 0.1$	4
$c(\ell) = 1.8(e^{\ell} - 1)$	6
$c(\ell) = e^{1.5\ell} - 1$	6
$\kappa = 3$	4
$\alpha(\ell) = 0.05\ell$	4

(B)

FIGURE 2. The baseline model is with $\tau = 0.01, \alpha(\ell) = 0.01\ell, \kappa = 2, c(\ell) = e^{\ell} - 1$. We first plot the baseline price function and then plot the case for increasing $\tau = 0.1$, $c(\ell) = 1.8(e^{\ell} - 1)$, $c(\ell) = e^{1.5\ell} - 1$, $\kappa = 3$, $\alpha(\ell) = 0.05\ell$, respectively.

5.2.3.4. An Example of Assembly Cost and Cost Function. This section presents an example of a closed-form solution for the cost function and assembly cost derived from the firm's cost minimisation problem. The motivation for this example arises from the question: Does the share of intermediate inputs affect the shape of the production chain? Intuitively, if the production process involves a higher utilisation of inputs, the production entity would tend to outsource more tasks, potentially resulting in a longer chain. This section discusses how the cost share of intermediate inputs influences the structure of the production network.

Moreover, it is known that the transaction cost can impact the share of the intermediate. For example, Boehm and Oberfield (2020) point out that court congestion makes contract enforcement slow, leading plants in industries reliant on relationship-specific inputs and in states with more congested courts d to have lower shares of intermediate inputs. Given that the cost of contract enforcement is a component of transaction costs, we would expect that it influences the structure of a production network both directly through the effect of transaction costs and indirectly through the share of intermediate inputs.

Suppose that firms have the Cobb-Douglas production function as

$$q = \begin{cases} A_s[(k^a n^{1-a})^{1-\sigma} x^{\sigma}]^{\theta} & \text{if subcontract,} \\ A_h(k^a n^{1-a})^{\theta} & \text{if produce in-house,} \end{cases} \tag{5.6}$$

where q denotes output, k denotes capital, n denotes labour, $x$denotes intermediate input, and σ, a and θ are all in $(0,1)$ representing intermediate share, capital share and degree of decreasing return, respectively. Given task ℓ, wage rate w and rental rate r, if firms choose to subcontract, firms using κ pieces of intermediate ℓ/n and minimise the costs to the output ℓ

$$\min_{k,n} rk + wn + \kappa p(\ell/\kappa) \qquad \text{s.t. } \ell = A_s\left[(k^a n^{1-a})^{1-\sigma}\left(\kappa\frac{\ell}{\kappa}\right)^{\sigma}\right]^{\theta}.$$

Let the assembly cost function be the resulting minimum cost function of labour and capital as

$$\alpha(\ell) := \min_{k,n}\{rk + wn\} = rk^* + wn^* = C_a \ell^{(1-\sigma\theta)/(\theta(1-\sigma))}, \tag{5.7}$$

for a positive constant C_s.[15] Since $\theta < 1$, we have $(1-\sigma\theta)/(\theta(1-\sigma)) > 1$.

---

[15] $C_s = (1/A_s)^{1/(1-\sigma)\theta}\left[(r/w)^a\left((1-a)/a\right)^a + (w/r)^{1-a}\left(a/(1-a)\right)^{1-a}\right]$.

On the other side, if the firm is a home producer, assume that the cost is $c(\ell) = \min_{k,n}\{rk + wn\} + h\ell$, where h is a constant. Given ℓ, a home producer solves

$$\min_{k,n} rk + wn + h\ell \qquad \text{s.t. } \ell = A_h(k^a n^{1-a})^\theta. \tag{5.8}$$

Then, the resulting cost is

$$c(\ell) = C_h \ell^{1/\theta} + f\ell$$

satisfying $c'(0) > 0$, where C_h is some constant.[16] It is known that the capital share is around 1/3, and the intermediate share is around 1/2. Suppose that $\theta = 9/10$, for example, then $\alpha(\ell) = C_h \ell^{11/9}$ and $c(\ell) = C_h \ell^{10/9} + h\ell$.

This simple model also sheds light on how production efficiency affects the outsourcing decision. If assembly is significantly more efficient than home production, i.e., A_s is much higher than A_h, then C_s is much smaller than C_h. Then, outsourcing will occur if the transaction cost is also small enough.

Observe that the parameter q of assembly costs is $(1 - \sigma\theta)(\theta(1 - \sigma))$, which is increasing in σ. Since $\ell \leqslant 1$, the assembly cost $\alpha(\ell)$ is decreasing in the share of intermediate σ. Thus, if the share of intermediate is lower, the higher assembly costs make subcontracting less attractive, yielding a shorter production chain.

According to Boehm and Oberfield (2020), when courts are more congested, transaction cost τ is higher, and the share of intermediate σ is lower. Together with both changes, the model predicts that the chain is shorter, prices are higher and firms at the most upstream level will have larger vertical span of production. This aligns with the findings in Boehm and Oberfield (2020), showing that plants tend to have large vertical span of production if they are confronting higher-congested courts and relying on relationship-specific inputs.

5.3. An Economy of Balanced Production Chains

Guilmi and Fujiwara (2020) and Hiromitsu and Wataru (2020) suggest that a firm's position in a production network is related to its size, which causes the departure of Gibrat's law for upstream firms. Moreover, Gibrat's law typically breaks down for small firms due

---

[16] $C_h = (1/A_h)^{1/\theta}\left[(r/w)^a\left((1-a)/a\right)^a + (w/r)^{1-a}\left(a/(1-a)\right)^{1-a}\right]$

to their higher growth rates driven by innovation.[17] This phenomenon is particularly pronounced for small firms situated at the upstream levels of a production chain, as noted by Hiromitsu and Wataru (2020). Guilmi and Fujiwara (2020) further show that upstream firms tend to be more volatile than downstream firms. Hence, Gibrat's law breaks down for those smaller upstream firms.

This section aims to illustrate this departure within a production network. Specifically, we construct an economy based on the balanced trees of a production network. This economy has the features that the firm sizes of assemblers or subcontractors follow a power-law distribution, and home producers are more fluctuated than subcontractors, assuming the assembly technology remains unchanged in the short run. Consequently, when examining the upstream firms across all balanced chains, the upstream firms are more volatile.

5.3.1. Distribution of Firm Size. We first show that when in a balanced supply chain, where the number of suppliers is fixed for every subcontractor, the sizes of firms follow a power-law distribution. This section is motivated by the fact that the Pareto distribution of firm size holds for mid-stream and downstream firms (Hiromitsu and Wataru, 2020).

Let m^* be the optimal length of the chain in equilibrium. Recall the network structure that there are κ^{m-1} number of firms at level m, and each firm of level m implements κ^{1-m} of tasks for all $m \leqslant m^*$.

Define the firm size by the value-added v_j for firm j in the chain. If firm j is a home producer, then its value-added is defined as the cost of in-house production, $v_j = c(\ell_j)$ for $\ell_j \in \mathcal{D}$. On the other hand, if firm j is a subcontractor, then the value-added is the assembly cost, $v_j = \alpha(\ell_j)$.[18] If the subcontractor j is at level m, its value-added or firm size is $v_j = b\ell_j^q = b(\kappa^{1-m})^q = b\kappa^{(1-m)q}$.

If we also let $q = 0$ in the assembly cost function, then the distribution of firm size is flat since the value-added for all subcontractors is a constant $\alpha(\ell) \equiv b$. Alternatively, if $q > 0$, then the firm sizes, $b\kappa^{(1-m)q}$, decline exponentially as we go upstream levels, while

---

[17] e.g. Lotti et al. (2003), Calvo (2006), Daunfeldt and Elert (2013), Tang (2015) and Aydogan and Donduran (2019))

[18] A subcontractor buys intermediate inputs from suppliers and does the assembly work, so its value-added is the assembly cost.

at the same time, the number of firms κ^{1-m} at each level grow exponentially. Hence, we have the following lemma.

Let $s \geqslant 0$ be the firm size of the subcontractor, and $F(s)$ be the fraction of firms with sizes greater than or equal to s.

Proposition 5.3.1. *Suppose that Assumption 5.2.1 and 5.2.2 hold, assembly cost is* $b(\ell)^q$ *with* $b, q > 0$*, and the price function is*

$$p(\ell) = \delta \min\{kp(\ell/\kappa) + b\ell^q,\, c(\ell)\} \qquad (\ell \in \mathcal{D}).$$

Then, there exist $s_{min>0}$ *and* $\gamma > 0$ *such that* $F(s) = Cs^{-\gamma} - D$ *for all* $\bar{s} \geqslant y \geqslant s_{min}$*, where* C *and* D *are some positive constants, and* $\bar{s}$ *is the size of the largest subcontractor. In particular,* $\gamma = 1/q$*.*

Proof. Let all the stated assumptions hold. Suppose that there are m^* levels in total in equilibrium. Since there is no subcontractor if $m^* = 1$, assume that $m^* > 1$. By the discussion in section 5.3.1, there are κ^{i-1} firms at level i and the subcontractors' sizes are $b\kappa^{(1-i)q}$ at the level i of chain in equilibrium. Let $s_{min} = b\kappa^{(2-m^*)q}$, which is the sizes of the smallest subcontractors at level $m^* - 1$.[19] Let s be the firm size of one of the subcontractors. Then, there exists $m \leqslant m^*$ such that $s = b\kappa^{(1-m)q}$. Clearly, $b \geqslant s \geqslant s_{min}$.

Moreover, there are in total $1 + \kappa + \cdots + \kappa^{m^*-1}$ firms in the chain and there are $1 + \kappa + \cdots + \kappa^{m-1}$ firms at level $1, \ldots, m$.[20] In this setting, $F(s)$ is the fraction of subcontractors at level $1, \ldots, m$. Then, by definition of f and changing variables of $s = b\kappa^{(1-m)q}$, we have

$$F(s) = \frac{1 + \kappa + \cdots + \kappa^{m-1}}{1 + \kappa + \cdots + \kappa^{m^*-1}} = \frac{\kappa^m - 1}{\kappa^{m^*} - 1} = \frac{\kappa b^{1/q} s^{-1/q} - 1}{\kappa^{m^*} - 1}.$$

Let $\gamma := 1/q$, $C := \kappa b^{1/q}/(\kappa^{m^*} - 1)$ and $D := 1/(\kappa^{m^*} - 1)$. Hence, $F(s) = Cs^{-\gamma} - D$ for $b \geqslant s \geqslant s_{min}$. Generally, $F(s) \propto s^{-\gamma}$. $\square$

Proposition 5.3.1 implies that the firm sizes for subcontractors are Pareto distributed in equilibrium. In addition, when the in-house producers have the smallest size in equilibrium, then the upper tail, excluding the smallest firms, is Pareto distributed. Furthermore,

---

[19] This s_{min} captures the tail of all subcontractors. We can also choose larger sizes of subcontractors.

[20] Including the firm of size s itself.

if $q = 1$, then the above distributions follow a Zipf's law.[21] If the "span of control" parameter of assembly cost q is less than one, the assembly cost is strictly concave, and the Pareto exponent $1/q$ is greater than one by Proposition 5.3.1.

EXAMPLE 5.3.1. This part presents the computation results given the proportional assembly cost $b\ell^q$. With the parameters $\tau = 0.01, b = 0.001, \kappa = 2, q = 0.92$ and $c(\ell) = 0.01(e^{25x} - 1)$, the production network is shown in Figure 3a. There are 8 levels in equilibrium. Next, we plot the log-rank plot of the distribution of firm sizes for subcontractors in this equilibrium, see Figure 3b. Note that the ranks are the same if the firms have the same size and there are κ^{m-1} firms at level $m = 1, ..., m^*$. Define the rank as κ^{m-1} if the firm is at level m.[22] Thus, the ranks for, say, $\kappa = 2$ are $1, 2, 2, 4, 4, 4, 4, 8, \ldots$. There are 8 levels in equilibrium, so there are 7 points in Figure 3b, where κ^{m-1} points of subcontractors collapse into one point in every level. The slope of log-rank plot is -0.92 which is equal to the parameter $-q$ of assembly function $b\ell^q$ as predicted by Proposition 5.3.1.[23] In this case, the Pareto exponent is $1/q = 1.087$.

In addition, let κ increase from 2 to 5 and fix other parameters. Then, the computation indicates that the total number of levels of the chain declines to 4 in equilibrium, see Figure 4. This result is reasonable since the maximal number of levels $\bar{m}$ is decreasing in κ. On the other hand, under these particular specifications, the firms at the most upstream level have the smallest sizes. As in the previous example, the slope is $-q$ in Figure 4b, and the Pareto exponent is unchanged. The comparative statics implies that the Pareto exponent is not affected by the structure of the network if the curvature of assembly cost is identical. $\square$

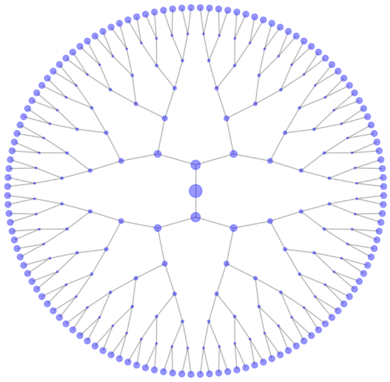
5.3.2. Breakdown of Gibrat's Law. Given a equilibrium production chain $(p^*, \boldsymbol{\ell}^*)$, Proposition 5.3.1 implies that large firms at the downstream exhibit power-law distributed sizes. In addition, it indicates that the sizes of firms at the most upstream do not necessarily follow a power-law distribution. To this end, the sizes of small firms within the

---

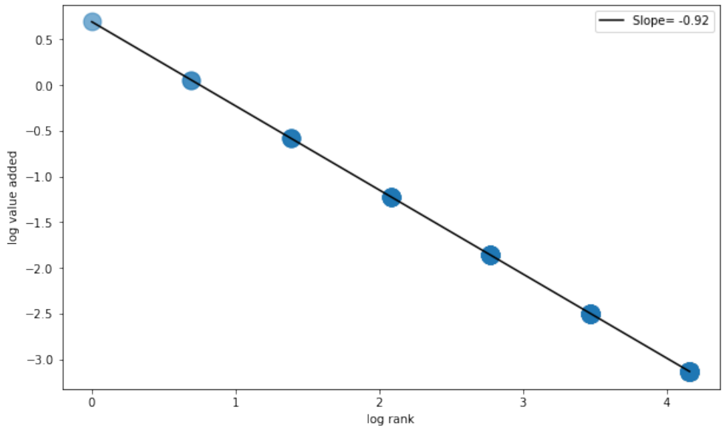
[21] For example, Axtell (2001) finds that the Pareto exponent is 1.059 , using employees as sizes, or 0.994, using receipts as sizes, for U.S. firms.

[22] Under this definition, the slope of the log-rank plot reveals the Pareto exponent for convenience.

[23] Pareto law can also be formulated as $n = Ax^{-\gamma}$, where n is the number of people having wealth $\geqslant x$. Thus, we can write $x(n) = Cn^{-1/\gamma}$ and then $\log x(n) = \log C - (1/\gamma) \log n$. The slope of log-rank is $-1/\gamma$, which is $-q$ of Proposition 5.3.1.



(A) Production network.



(B) The log-rank plot for the distribution of firm sizes.

FIGURE 3. The production chain with $\tau = 0.01, b = 0.001, \kappa = 2, q = 0.92$ and $c(\ell) = 0.01(e^{25x} - 1)$.

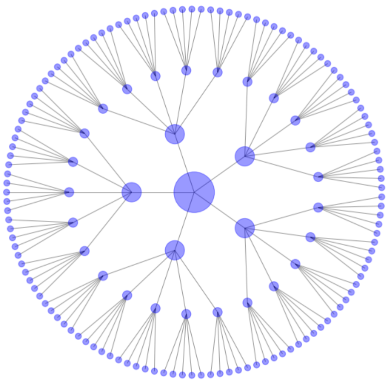
production chain generally deviate from a power law. It suggests that there exists a departure of Gibrat's Law; otherwise, the sizes of small firms would also exhibit a power-law distribution.

We restrict the time frame to the short run since we are interested in short-term volatility. This assumption is inspired by Tang (2015), who demonstrates that there is a steady state in the firm's expansion and that Gibrat's law breaks down in the short run but holds in the long run.[24]

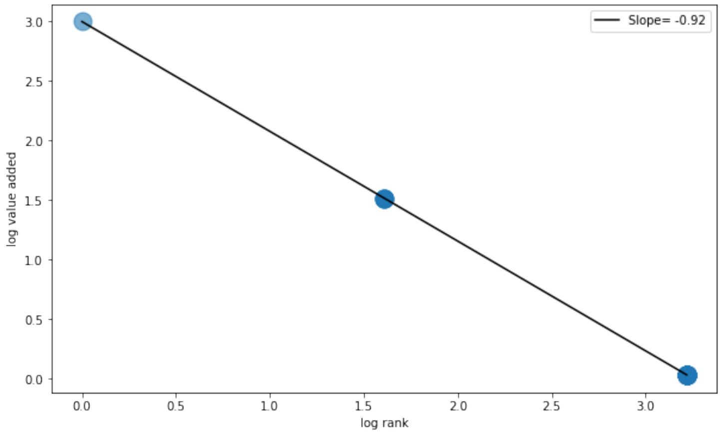
Suppose that firms solve cost minimization problems outlined in Section 5.2.3.4. Let the cost function be (5.8) and assembly cost function be (5.7). Suppose that the efficiency of in-house production grows over time in the short run, to the extent that the structure

---

[24] Tang (2015) studies firm-level data in the Swedish energy market.



(A) The network of production.



(B) The log-rank plot of the firm-size distribution for the subcontractors.

FIGURE 4. Production chain with $\tau = 0.01, b = 0.001, \kappa = 5, q = 0.92$ and $c(\ell) = 0.01(e^{25x} - 1)$.

of the network undergoes minimum change.[25] That is, let $A_{h,t+1} = \zeta_{t+1} A_{h,t}$ with growth rate $\zeta_t > 0$ and let the other parameters are fixed, indicating that C_h in (5.8) varies over time. Suppose that the growth ζ_{t+1} is a random process such that the optimal length of the production chain remains unchanged in the short run. There is no innovation in assembly technology in the short run. Consequently, the sizes of home producers change over time, while the sizes of subcontractors are stable over time. As a result, smaller firms at upstream have higher volatility, while large firms at downstream maintain more stable sizes. Furthermore, if the assembly efficiency C_s in (5.7) or assembly productivity A_s is volatile in the short run, the most upstream firms also experience greater fluctuations. This is because firms may alter their outsourcing decisions or exit the production chain.

---

[25] Hiromitsu and Wataru (2020) also shows that the structure of the production network is stable in the short run.

We expand the model from a chain to an industry, considering N heterogeneous goods in an industry. For each good $i = 1, \cdots, N$, there is a production chain dedicated to producing good i. The pricing function (5.1) determines the price for each good based on a fixed number of suppliers κ_i so that the chain is balanced.[26]

Production chain i is characterized by production cost c_i, assembly cost α_i and transaction cost τ_i. Specifically, we assume that $c_i(\ell) = C_{h,i}\ell^{\theta_i} + \varepsilon_i \ell$, where ε_i is a constant, $\alpha_i(\ell) = C_{a,i}\ell^{q_i}$ and parameter $\delta_i = 1/(1-\tau_i)$. Here, the parameters $C_{h,i}, \varepsilon_i$ and $C_{a,i}$ are selected such that home-producing firms tend to be smaller in size compared to most outsourcing firms. Suppose that Assumption 5.2.1 , 5.2.2 and 5.2.3 hold. Then, Proposition 5.2.2 confirms the existence of a unique equilibrium for each production chain. In the equilibrium, each production line has the optimal length m_i^* and price p_i^*.

Assume that each firm produces all N final goods. Each firm j is randomly at level m_j in all chains and then has a random size S_j. The cost or value added of subcontractor firm j producing good i is X_i, which follows a power law with parameter γ_i by Proposition 5.3.1. From Gabaix (2009), we know that if Y_1 and Y_2 are power-law variables, then $Y_1 + Y_2$ is also power-law variable.[27] Then, a subcontractor's size $S_j = X_1 + \cdots + X_N$ follows a power law.

Contrary to home producers, the firm sizes of home producers do not necessarily follow a power law. If further the home production technology $C_{h,i}$ is more volatile than assembly technology $C_{a,i}$, then the sizes of home producers exhibit higher volatility. Examining the entire industry, the upstream firms in the economy tend to be more volatile than downstream firms in the short run. Given that home producers typically have smaller firm sizes, smaller firms inherently possess higher volatility.

In summary, it appears that the growth rates of firms are contingent on their sizes, leading to a breakdown of Gibrat's Law for upstream firms. Simultaneously, downstream subcontractors conform to a power-law size distribution across all production chains. The model elucidates that small firms and upstream firms deviate from Gibrat's law, while large firms and downstream firms adhere to a power-law distribution of firm sizes.

---

[26] In Section 5.4, we can introduce a method to determine κ_i by the local optimality.

[27] If X and Y are power-law variables with exponent ζ_x and ζ_y satisfying $\zeta_y \geqslant \zeta_x$, then $X + Y$ is a power-law variable with exponent ζ_x.

5.4. Extension: An Endogenous Number of Suppliers

The assumption that the number of suppliers is fixed may be unrealistic. However, we can approximate the real world by choosing a reasonable fixed number of suppliers for our model. This section first shows that the local optimum for the supplier number can be found by numerical computation.

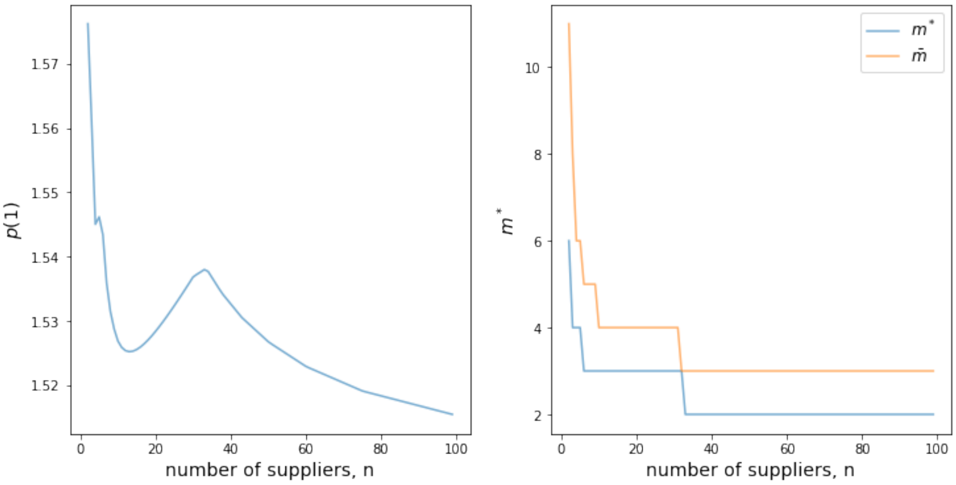
Moreover, we establish an alternative model relaxing the assumption of the fixed number of suppliers, so that subcontractors can choose an arbitrary number of suppliers. We show the existence and uniqueness of equilibrium for the extended models.

5.4.1. Local Optimality for Number of Suppliers. Throughout this section, we denote the number of suppliers as $n \in \mathbb{N} \setminus \{1\}$. To simplify the problem, suppose that the assembly cost is fixed at a constant α. As indicated by $\bar{m} = \lceil 1 - \ln \bar{x} / \ln n \rceil$ from Section 5.2, we express the maximal possible number of levels as a function of n, denoted as $\bar{m}(n)$. Then, (5.5) implies that the price of the final good is a function of n,

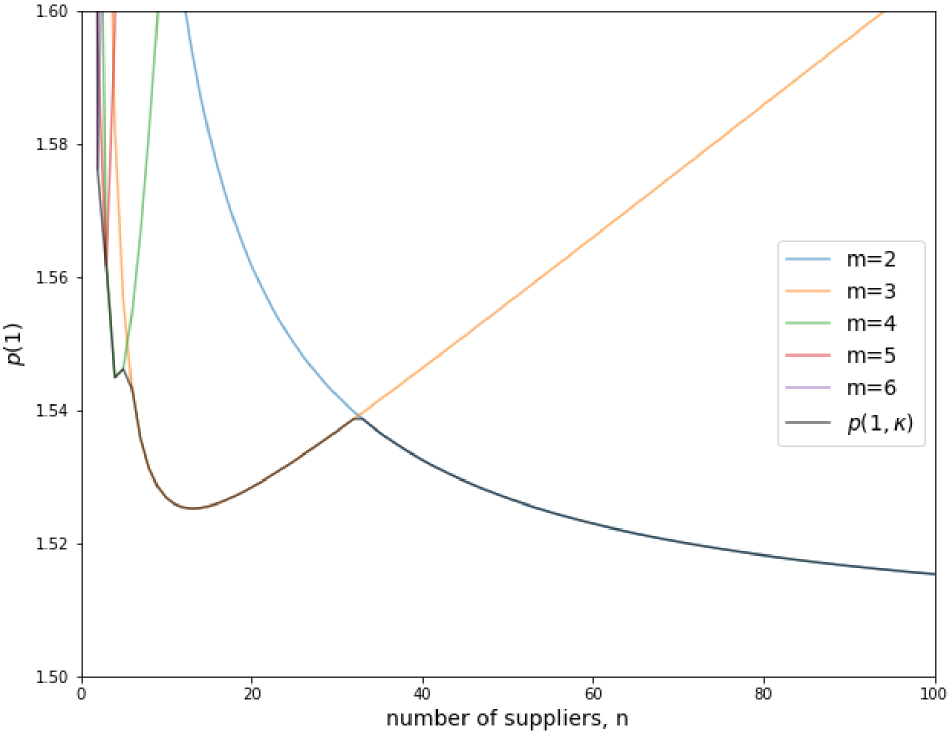
$$\begin{aligned} p^*(1,n) &:= \min_{m=1,\ldots,\bar{m}(n)} \left\{ (\delta n)^{m-1} \delta c \left(n^{1-m} \right) + \frac{(\delta n)^{m-1} - 1}{\delta n - 1} \delta \alpha \right\} \\ &= \min_{m=1,\ldots,\bar{m}(n)} \{ p(1,n,m) \}. \end{aligned} \tag{5.9}$$

In other words, the planner of the production chain strategically determines the optimal number of levels m, given the number of suppliers n in each level. This decision-making process is illustrated in Figure 5b, where the grey line represents the minimum cost, i.e. the prices of final good $p^*(1,n)$, among all possible m. Notably, for each $m = 2, \ldots, 6$, the pricing functions for the final good exhibit convexity, featuring local minimum prices. Therefore, if there are frictions to increase supplier numbers, such as contract costs and the expenses associated with searching for new suppliers, a firm may opt to remain at these locally minimised points to minimise the total cost effectively.

Moreover, Figure 5a reveals that the number of optimal levels m^* is less than $\bar{m}$ and converges to 2, and the final price generally decreases in n. This suggests that, in the absence of costs related to expanding suppliers, the most downstream firm can achieve cost reduction by increasing suppliers to an arbitrary number.



(A) Prices of final good and the corresponding number of levels



(B) Prices of final good, fixing number of levels m.

FIGURE 5. Prices of final good, given that $\tau = 0.001, \alpha(\ell) = 0.001, n = 2, c(\ell) = e^{1.5\ell} - 1$.

In general, the planner overseeing the chain has an incentive to increase the number of suppliers when there is no friction affecting the search of supplier. Since $\bar{m} = \lceil 1 - \ln \bar{x}/\ln n \rceil$, we have $\bar{m} \to 2$ as $n \to \infty$.[28] Moreover, given $m^* \leqslant \bar{m}$, we know that the optimal number of levels is $m^* \leqslant 2$ if the number of suppliers n is sufficiently large.

---

[28] $-\ln \bar{x} > 0$.

If the firm at the most downstream level has the incentive to subcontract and then there are exactly two levels in the chain $m^* = 2$, the pricing function implies $np(1/n) + \alpha = n\delta c(1/n) + \alpha < c(1)$. Since the cost function c is convex, strictly increasing, and $c(0) = 0$, the function $nc(1/n)$ is strictly decreasing in n and bounded below by zero. Hence, $nc(1/n)$ converges to some $t \geqslant 0$ as $n \to \infty$ by the monotone convergence theorem of sequence. Thus, if $2\delta c(1/2) + \alpha < c(1)$, there will be two levels for all $n \geqslant 2$, and the firm at the most downstream level will subcontract, regardless of the number of suppliers.

Note that the condition $\delta 2c(1/2) + \alpha < c(1)$ is necessary for having at least one subcontractor in equilibrium. In this case, the firm can minimise the outsourcing costs by subcontracting the tasks to as many suppliers as possible, given that the cost $\delta nc(1/n) + \alpha \downarrow \delta t + \alpha$. Therefore, if there is at least one subcontractor and no cost for expanding upstream partners, the planner has the incentive to raise the number of suppliers unlimitedly, resulting in only two levels in the chain.

5.4.2. Global Optimality of Suppliers. To curb the endogenously unbounded number of suppliers, we consider assembly costs that are intuitively proportional to the number of suppliers. The underlying rationale stems from that as subcontractors further divide tasks into smaller components, the process of assembling inputs necessitates additional effort. Precisely, let the assembly cost function be $\alpha(n)$ where n is the number of upstream suppliers and $\alpha(n)$ is strictly increasing in n.

ASSUMPTION 5.4.1. The assembly cost $\alpha : \mathbb{N}\backslash\{1\} \to \mathbb{R}$ is strictly increasing and $\alpha(2) > 0$.

Then, the profits for firms in the chain are

$$(1 - \tau)p(\ell) - \min\{np(\ell/n) + \alpha(n), c(\ell)\} \tag{5.10}$$

The corresponding Bellman equation of the pricing function is

$$p(\ell) = \delta \min\{np(\ell/n) + \alpha(n), c(\ell)\} \tag{5.11}$$

Following the proofs in section 5.2.2, we can also show the existence and uniqueness of equilibrium price and allocation under the profits (5.10). Similar to Proposition 5.2.1 and

5.2.2, the unique equilibrium can be computed by iteration, i.e. $p^* = T^k p$ for any $p \in \mathcal{P}$ and large enough k.

Lemma 5.4.1. *If Assumption 5.2.1, 5.2.2 and 5.4.1 hold, then*

(a) the solution to the Bellman equation (5.11) is an equilibrium of the production chain given the profits (5.10),

(b) the equilibrium price p^* and allocations $\boldsymbol{\ell}^*$ are unique, and

(c) $p^* = T^k p$ for any $p \in \mathcal{P}$ and sufficiently large k.

Since $\alpha(n)$ is strictly increasing by Assumption 5.4.1, there exists $\bar{n}$ such that there is only one firm or level in the chain when $n \geqslant \bar{n}$. For $n \geqslant \bar{n}$, the firm at the most downstream level does not have any incentive to subcontract, so $\bar{n}$ can be characterized by $\delta n c(1/n) + d\alpha(n) \geqslant c(1)$. The interpretation is that, if there are many pieces of modules to assemble, the sum of assembly costs and the transaction costs may be greater than the costs of production at home so that the firm does not subcontract at all.

Moreover, the global minimum exists given that $\alpha(n)$ is increasing. The intuition is that if the assembly cost $\alpha(n)$ is increasing fast enough in the number of suppliers n but still possible to subcontract, then the global minimum could be at $n = 2$. These properties are illustrated in the examples presented in Figure 6, which also depicts the upper bound for the price of final goods.[29]

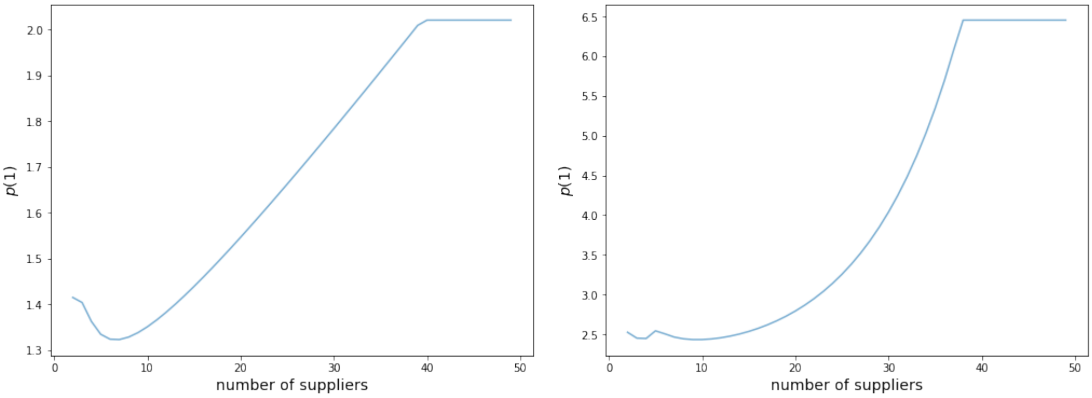
5.4.3. Choice over Number of Suppliers. This section extends the assumption that the number of suppliers is fixed along the production chain. Firms have the flexibility to select the number of suppliers when opting to subcontract their tasks. In particular, we introduce the pricing function

$$p(\ell) = \delta \min \left\{ \min_{n=2,3,4,\ldots} \{np(\ell/n) + \alpha(\ell, n)\},\, c(\ell) \right\}. \tag{5.12}$$

To provide a more comprehensive framework, suppose that the assembly cost depends on both tasks and the number of upstream partners. To prevent an unbounded partner number n^*, we have the below assumption for assembly cost.

---

[29] Since the assembly is too high for large n, firms all choose to home production and the bound is $\delta c(1)$.



(A) (B)

FIGURE 6. Examples of global minimum of prices. (a) Prices of final good are plotted given that $\tau = 0.01, \alpha(n) = 0.01(n + n^{1.1}), c(\ell) = \ell + \ell^2$. The global minimal price is at $n = 7$. (b) Prices of final good are plotted given that $\tau = 0.01, \alpha(n) = 0.1(e^{0.1n} - 1), c(\ell) = e^{2\ell} - 1$. The global minimal price is at $n = 10$.

ASSUMPTION 5.4.2. Assume that $\alpha : [0,1] \times \{2,3,\dots\} \to (0,\infty)$ is strictly increasing in both ℓ and n.

Assumption 5.4.2 makes the optimal number of suppliers n finite if there is outsourcing. Analogous to Section 5.2.2, let the set of tasks be $\mathcal{D} := \{k^{-i} : k \in \mathbb{N} \setminus \{1\}, i \in \mathbb{N} \cup \{0\}\}$, and $\mathcal{P} := \{p : \mathcal{D} \to \mathbb{R} : \delta c'(0)\ell \leqslant p(\ell) \leqslant \delta c(\ell)\}$.[30]

Under this pricing function, we say that $\boldsymbol{\ell} = \{\ell_i\}$ is a *feasible* allocation of tasks if there are $m \in \mathbb{N}$ and a finite sequence $\{n_t\}_{t=1}^m$ with $n_1 = 1$ and $n_t \geqslant 2$ for all $t = 2, \dots, m$ such that $i = 1, 2, \dots, n_1 n_2 \cdots n_m$ and $\ell_i = (n_1 n_2 \cdots n_t)^{-1}$ for $n_1 + n_1 n_2 + \cdots + n_1 n_2 \cdots n_{t-1} < i \leqslant n_1 + n_1 n_2 + \cdots + n_1 n_2 \cdots n_t$. In words, the length of the chain, m, is finite and there are $n_1 n_2 \dots n_t$ number of firms with allocation $(n_1 n_2 \dots n_t)^{-1}$ at level t. Each firm at level $t-1$ has n_t number of subcontractors if it chooses home production. At the most upstream level m, there are $n_1 \cdots n_m$ number of firms with $(n_1 \cdots n_m)^{-1}$ measure of allocated tasks.

Moreover, the profit of firm i is

$$\pi_i = (1-\tau)p(\ell_i) - \min\left\{ \min_{n=2,3,\dots} \{np(\ell_i/n) + \alpha(\ell_i, n)\},\ c(\ell_i) \right\} \tag{5.13}$$

---

[30] It can also be defined as $D = [0,1]$ and the results hold.

We can define the equilibrium under the price (5.12) as definition 5.2.1 using the above feasibility definition and profit (5.13).

DEFINITION 5.4.1. Given a price function p, a feasible allocation $\boldsymbol{\ell} = \{\ell_i\}$ and the corresponding profit defined by (5.13), $(p, \boldsymbol{\ell})$ is an equilibrium if

(1) $p(0) = 0$,

(2) for all $\ell \in \mathcal{D}$,

$$(1-\tau)p(\ell) - \min\left\{ \min_{n=2,3,\dots} \{np(\ell/n) + \alpha(\ell, n)\},\, c(\ell)\right\} \leqslant 0,$$

(3) $\pi_i = 0$ for all i.

Then, we can prove the existence of equilibrium. For the following proposition, define the operator T as

$$Tp(\ell) := \delta \min\left\{ \min_{n=2,3,4,\dots} \{np(\ell/n) + \alpha(\ell, n)\},\, c(\ell)\right\}.$$

Suppose that firms choose the minimum number of suppliers when there are multiple number of suppliers that minimise the cost. That is, suppose firms choose $\min\{\arg\min_{n\geqslant 2}\{np(\ell/n)+\alpha(\ell,n)\}\}$ as their number of suppliers if $np(\ell/n)+\alpha(\ell,n) < c(\ell)$. The proof of unique equilibrium uses this assumption to pin down the feasible task allocation.

PROPOSITION 5.4.1. *If 5.2.2, 5.2.1 and 5.4.2 hold, and the price is defined by (5.12), the following statements are true.*

(a) *The operator T is a self-map on $\mathcal{P}$ and has a unique fixed point p^* in $\mathcal{P}$*

(b) *The fixed point can be computed by $T^k p = p^*$ for all $k \geqslant 1 - \ln \bar{x}/\ln 2$ and all $p \in \mathcal{P}$.*

(c) *The production chain has a unique equilibrium $(p^*, \boldsymbol{\ell}^*)$, where p^* is the fixed point of T, and $\boldsymbol{\ell}^*$ is the corresponding task allocation under p^*.*

Hence, we can compute the unique equilibrium in a manner consistent with the basic model. The comparative statics also yield similar results as previously observed.

Proposition 5.4.2. *If 5.2.2, 5.2.1 and 5.4.2 hold, and the price is defined by (5.12), then the equilibrium price is increasing in transaction cost, assembly cost and home production cost.*

As in Section 5.2.3, we can gain insights into the equilibrium price by the properties of the cost function before computing it. The following two lemmas illustrate the equilibrium price's characteristics, demonstrating that, under certain conditions, the equilibrium price is strictly convex almost everywhere.

Suppose that the assembly function $\alpha(n)$ only depends on the number of suppliers and Assumption 5.2.3 holds. Moreover, assume that $c, \alpha \in \mathrm{C}^2$ are twice differentiable. Define $t(n, \ell) \coloneqq \delta n c(\ell/n) - c(\ell)$. Then, the second derivative $t_{22}(n, \ell) = (\delta/n)c''(\ell/n) - c''(\ell)$ implies that $t(n, .)$ is strictly concave when $0 < \tau < 0.5$ and c is strictly convex, with $n \geqslant 2$. In addition, define $f(\ell) \coloneqq \min_{n=2,3,\ldots}\{t(n, \ell) + \alpha(n)\}$. The next lemma helps us to determine the boundary of subcontracting and length of the chain.

Lemma 5.4.2. *Suppose that $t(n, .)$ is strictly concave. Then, $f(\ell)$ has a root $\hat{\ell}$ in $(0, 1]$ if and only if $p(\ell) = \delta c(\ell)$ for all $1 \leqslant \ell \leqslant \hat{\ell}$ and $p(\ell) < \delta c(\ell)$ for all $\hat{\ell} < \ell \leqslant 1$.*[31]

This lemma shows that when the transaction cost is not excessively large ($\tau < 0.5$) and the cost is strictly convex, the boundary allocation for subcontracting can be characterized by computing the root of $f(\ell)$, which is determined by the home-producing cost and assembly cost. This allows us to calculate the boundary task of subcontracting (the root $\hat{\ell}$) before knowing the equilibrium price.

Lemma 5.4.2 also shows that if the root of $f(\ell)$ is less than 1, then the length of the chain is greater than or equal to 2. Consequently, there must exist subcontractors in the chain, as the firms with $\ell \in (\hat{\ell}, 1]$ always choose to outsource. Furthermore, if the assembly is relatively economical, meaning that firms can choose numerous partners and divide the tasks in numerous modules with little cost, then there would be only two levels of length, as $1/n < \hat{\ell}$ for large n.

---

[31] This lemma also holds if the assembly cost is $\alpha(\ell, n)$ and $\alpha(., n)$ is concave.

Lemma 5.4.3. *Let p be the equilibrium price of (5.12). Suppose that the assumptions in Lemma 5.4.2 hold and further assume that c is strictly convex and $\alpha(n)$ is convex. Then p is twice differentiable with $p' > 0$ and $p'' > 0$ almost everywhere.*

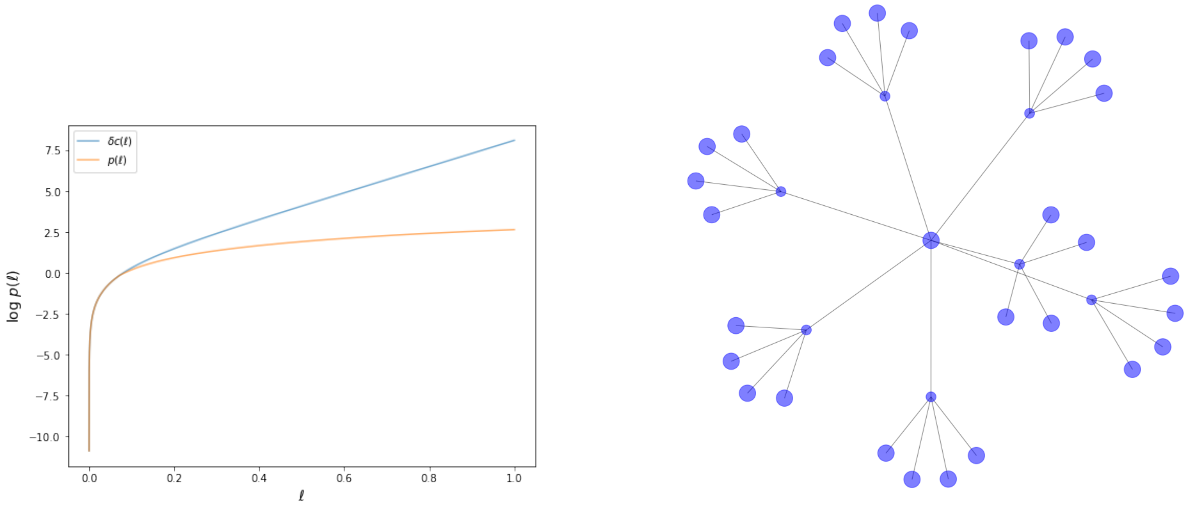
Figure 7 gives an example of a production network under pricing function (5.12), with assembly cost $\alpha(n)$ depending only on n. The scale of the pricing function is in the log. This example also demonstrates a fat-tailed distribution of the number of buyer-seller links.[32] Moreover, numerical simulations indicate that the number of suppliers may decrease as the firm goes to upstream levels, when the assembly cost is only contingent on the supplier count.

5.5. Conclusion

This paper demonstrates the influence of a hierarchical production chain. The production network with hub-like firms can generate a power-law distribution in firm size without any ex-ante heterogeneity or adhering to Gibrat's law. The production network coordinates the positions and task allocations for ex-ante identical firms in a way that a large firm tends to be a hub and be at a downstream level. The framework in this paper endeavors to capture this phenomenon through the minimization of costs. Our model provides insights into why firms following a power law often occupy downstream positions. Moreover, it explains the departure of Gibrat's law for upstream firms, which are more volatile than the downstream counterparts. The model underscores the significance of a spider network, akin to the concept introduced by Baldwin and Venables (2013).

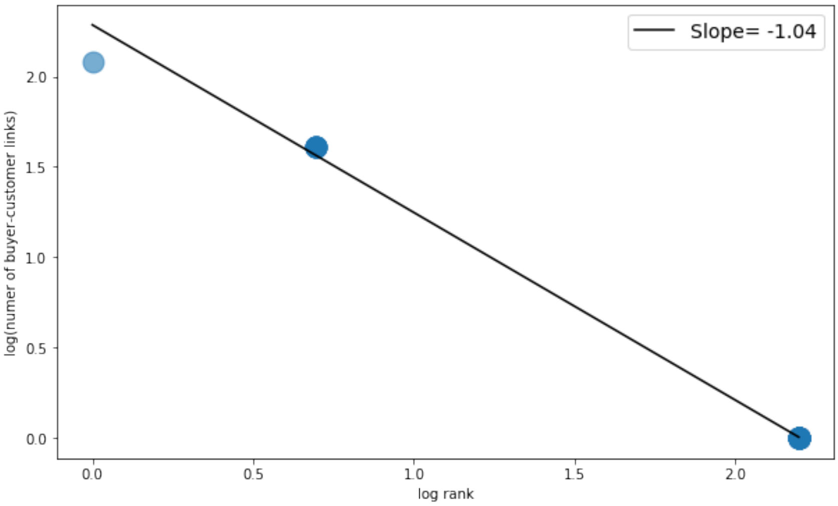
---

[32] See Bernard et al. (2019) for an empirical example.



(A) The log-price function.

(B) The network.



(C) The distribution of the number of seller-buyer links.

FIGURE 7. An example of production chain under (5.12) with $\tau = 0.1, \alpha(n) = 0.01n^{1.8}, c(\ell) = e^{8\ell}-1$. This chain has length 3 and $(n_1, n_2, n_3) = (1, 7, 4)$. Also, $\hat{\ell} = 0.0711$.

5.6. Appendix

Proofs in Section 5.2.2. Throughout this section, we assume that Assumption 5.2.1, 5.2.2 and 5.2.3 hold.

Lemma 5.6.1. *The operator T defined by (5.3) is a self-map on $\mathcal{P}$ and has a fixed point in $\mathcal{P}$. Moreover, $Tp_1 \leqslant Tp_2$ if $p_1 \leqslant p_2$ for $p_1, p_2 \in \mathcal{P}$.*

Proof. The proof is similar to Kikuchi et al. (2021). Let T be the operator of (5.3). We want to show that $T : \mathcal{P} \to \mathcal{P}$, and T preserves orders so that we can apply Knaster-Tarski fixed point theorem, with the fact that $\mathcal{P}$ is a complete lattice.[33]

Fix $p \in \mathcal{P}$. Clearly, the definition of T implies $Tp(\ell) \leqslant \delta c(\ell)$ for all ℓ. Next, we check that $\delta c'(0)\ell \leqslant Tp(\ell)$ for all ℓ. Fix $\ell \in \mathcal{D}$. Since $p \in \mathcal{P}$, $b\ell^q \geqslant 0$ and c is convex, we have

$$\begin{aligned} Tp(\ell) &= \delta \min \{\kappa p(\ell/\kappa) + b\ell^q,\, c(\ell)\} \\ &\geqslant \delta \min \{\kappa \delta c'(0)\ell/\kappa + b\ell^q,\, c(\ell)\} \\ &\geqslant \delta \min \{c'(0)\ell,\, c(\ell)\} \\ &= \delta c'(0)\ell. \end{aligned}$$

Then, $Tp \in \mathcal{P}$. In addition, suppose that $p_1, p_2 \in \mathcal{P}$ with $p_1(\ell) \leqslant p_2(\ell)$ for all $\ell \in \mathcal{D}$. Thus, we have

$$\min \{\kappa p_1(\ell/\kappa) + \alpha(\ell),\, c(\ell)\} \leqslant \min \{\kappa p_2(\ell/\kappa) + \alpha(\ell),\, c(\ell)\},$$

for all ℓ. Hence, we have $Tp_1 \leqslant Tp_2$ and T preserves orders. By Knaster-Tarski fixed point, the set of fixed points of T in $\mathcal{P}$ is also a complete lattice. Since a complete lattice is nonempty, the fixed points exist. □

Next, we characterize the price function in equilibrium and show its uniqueness. The subsequent lemma suggests that there exists a maximum number of levels. In other words, the firms will choose to home production when their task allocations are sufficiently small. Recall that $\bar{x} := \sup\{x \in (0, 1] : c'(x) \leqslant \delta c'(0)\}$.

---

[33] A partially ordered set is a *complete lattice* if every subset has both an infimum and a supremum.

Lemma 5.6.2. *Suppose that $p \in \mathcal{P}$ is a solution to price function (5.1). Then, given $\ell \in \mathcal{D}$, if $\ell \leqslant \bar{x}$, then $p(\ell) = \delta c(\ell)$. That is, firms choose to home production if $\ell \leqslant \bar{x}$.*

Proof. Suppose that $p \in \mathcal{P}$ is a solution to price function (5.1). Since $\delta = 1/(1-\tau) > 1$ with $\tau \in (0,1)$ and Assumption 5.2.2 holds, there is $x > 0$ such that $c'(x) < \delta c'(0)$. Thus, $\bar{x}$ is well-defined. Toward contradiction, suppose that there is $\ell \in \mathcal{D}$ with $\ell \leqslant \bar{x}$ such that $\kappa p(\ell/\kappa) + \alpha(\ell) < c(\ell)$. Since $p \in \mathcal{P}$, we have

$$c(\ell) > \kappa p(\ell/\kappa) + \alpha(\ell) \geqslant \kappa \delta c'(0)\frac{\ell}{\kappa} + \alpha(\ell) = \delta c'(0)\ell + \alpha(\ell) \geqslant \delta c'(0)\ell.$$

Hence, we have $c(\ell)/\ell > \delta c'(0)$. Since c is convex and $\ell \leqslant \bar{x}$, this contradicts the fact that $c(\ell)/\ell \leqslant c'(\ell) \leqslant \delta c'(0)$. Thus, it must be $\kappa p(\ell/\kappa) + \alpha(\ell) \geqslant c(\ell)$ and then $p(\ell) = \delta c(\ell)$, where we follow the convention that $\min\{\kappa p(\ell/\kappa) + \alpha(\ell), c(\ell)\} = c(\ell)$ if equality holds.[34] ☐

Lemma 5.6.3. *If $p \in \mathcal{P}$, then $Tp(\ell) = \delta c(\ell)$ for all $\ell \leqslant \bar{x}$.*

Proof. The statement follows from Lemma 5.6.2. ☐

Lemma 5.6.4. *If $p, q \in \mathcal{P}$, then $T^n p = T^n q$ for $n \geqslant 1 - \ln(\bar{x} - \kappa)$.*

Proof. Let $p, q \in \mathcal{P}$. The proof is by induction. From the proof of Lemma 5.6.2, we can show that Tp and Tq are agree on $[0, \bar{x}]$, where $\bar{x} \coloneqq \sup\{x \in (0,1] : c'(x) \leqslant \delta c'(0)\}$. Next, we prove that $T^n p = T^n q$ on $[0, \kappa^{n-1}\bar{x}]$ for all $n \in N$. Suppose that the claim is true for some $k \in \mathbb{N}$. Fix $\ell \in [0, \kappa^k \bar{x}]$. Then, since $\ell/\kappa \in [0, \kappa^{k-1}\bar{x}]$, we have

$$\begin{aligned}
T^{k+1}p(\ell) &= T(T^k p)(\ell) \\
&= \delta \min\left\{\kappa T^k p\left(\frac{\ell}{\kappa}\right) + \alpha(\ell), c(\ell)\right\} \\
&= \delta \min\left\{\kappa T^k q\left(\frac{\ell}{\kappa}\right) + \alpha(\ell), c(\ell)\right\} \\
&= T^{k+1}q(\ell)
\end{aligned}$$

---

[34] That is, firms choose to produce in-house if the costs for home production and subcontracting are identical.

Therefore, $T^n p = T^n q$ on $[0, \kappa^{n-1}\bar{x}]$ is true for all $n \in \mathbb{N}$. Since $\kappa^{n-1}\bar{x} \geqslant 1$ for $n \geqslant 1 - (\ln \bar{x})/(\ln \kappa)$, we have $T^n p = T^n q$ on $[0, 1]$ for $n \geqslant 1 - (\ln \bar{x})/(\ln \kappa)$. □

Lemma 5.6.5. *T has a unique fixed point in $\mathcal{P}$ and thus the solution to the pricing function (5.1) is also unique.*

Proof. Let p, q be the fixed points of T. By Lemma 5.6.4, $p = T^n p = T^n q = q$ for large enough n. Moreover, by the definition of T, p^* is the fixed point of T if and only if it is the solution to pricing equation (5.1). Thus, the solution exists and is unique. □

Lemma 5.6.6. *If $p \in \mathcal{P}$ is (resp. strictly) increasing, then Tp is also (resp. strictly) increasing. Moreover, if p^* is the fixed point of T, then p^* is strictly increasing.*

Proof. Since c is strictly increasing in $\ell \in [0.1]$, Tp is (strictly) increasing if p is (strictly) increasing. It follows from Lemma 5.6.4 that $p^* = T^n c$ for large enough n, which implies that p^* is strictly increasing. □

Lemma 5.6.7. *If p^* is a solution to the price equation (5.1), then there is a feasible $\boldsymbol{\ell}^*$ such that $(p^*, \boldsymbol{\ell}^*)$ is an equilibrium for the production chain.*

Proof. Let p^* be a solution to the price equation (5.1). By Lemma 5.6.2, there exists m such that $p(\ell) = \delta c(\ell)$ with $\ell = \kappa^{1-m} \leqslant \bar{x}$, where $\bar{x} = \sup\{x \in (0, 1] : c'(x) \leqslant \delta c'(0)\}$. Let $\bar{m} := \lceil 1 - \ln \bar{x} / \ln \kappa \rceil$, which is the smallest integer satisfying $\kappa^{1-\bar{m}} \leqslant \bar{x}$ and the largest possible length of chain. Therefore, the corresponding maximal level $m^* \in \mathbb{N}$ for price p^* satisfies $m^* \leqslant \bar{m}$. The corresponding allocations are $\boldsymbol{\ell}^* = \{\ell_i^*\}$ are $\ell_i^* = \kappa^{1-n}$ for $i = 1, \dots, (\kappa^m - 1)/(\kappa - 1)$ and $(\kappa^{n-1} - 1)/(\kappa - 1) < i \leqslant (\kappa^n - 1)/(\kappa - 1)$ for some $1 \leqslant n \leqslant m^*$. Therefore, $\boldsymbol{\ell}^*$ is feasible.

Since $p^* \in \mathcal{P}$, we have $\delta c'(0)\ell \leqslant p^*(\ell) \leqslant c(\ell)$ for all ℓ in $\mathcal{D}$. It implies $p^*(0) = 0$. Moreover, since the definition of price equation implies that for all $\ell \in \mathcal{D}$ we have

$$p^*(\ell) - \delta \min\{\kappa p^*(\ell/\kappa) + \alpha(\ell), c(\ell)\} = 0,$$

condition (2) and (3) of Definition 5.2.1 are clearly satisfied. Therefore, $(p^*, \boldsymbol{\ell}^*)$ is an equilibrium. □

In Lemma 5.6.7, we construct an equilibrium $(p^*, \boldsymbol{\ell}^*)$ if p^* is a unique solution to price function. We next show that $(p^*, \boldsymbol{\ell}^*)$ is the unique equilibrium. The proof in the following lemma uses the convention that the firms choose to produce in-house when the costs for home production and subcontract are indifferent.

LEMMA 5.6.8. *If the equilibrium prices and allocations exist, then they are unique.*

PROOF. Suppose that there are two arbitrary equilibria $(p_1, \boldsymbol{\ell}^1)$ and $(p_2, \boldsymbol{\ell}^2)$ following Definition 5.2.1. Since $(p_1, \boldsymbol{\ell}^1)$ and $(p_2, \boldsymbol{\ell}^2)$ are feasible, there exist $m_1, m_2 > 0$ such that the allocations $\boldsymbol{\ell}^1$ and $\boldsymbol{\ell}^2$ have m_1 and m_2 length of chain or maximal levels, respectively. Clearly, $\boldsymbol{\ell}^1 \neq \boldsymbol{\ell}^2$ if and only if $m_1 \neq m_2$, by the definition of feasibility. Without loss of generality, assume that $m_1 \geqslant m_2$. We first show that $p_1(\ell) = p_2(\ell)$ for all $\ell = 1, \kappa^{-1}, \kappa^{-2}, \ldots, \kappa^{1-m_2}$ and then extend to $\ell = 1, \kappa^{-1}, \ldots, \kappa^{1-m_2}, \ldots, \kappa^{1-m_1}$.

We claim that, for $\ell = \kappa^{-1}, \kappa^{-2}, \ldots, \kappa^{1-m_2}$, $p_1(\ell) = p_2(\ell)$ implies $p_1(\kappa\ell) = p_2(\kappa\ell)$. Let $p_1(\ell) = p_2(\ell)$ for all $\ell = \kappa^{-1}, \kappa^{-2}, \ldots, \kappa^{1-m_2}$. Since the firms are subcontractors at level $1, \ldots, m_2 - 1$ for both equilibrium, condition (3) of Definition 5.2.1 implies that

$$(1-\tau)p_1(\kappa\ell) = \kappa p_1(\ell) + \alpha(\kappa\ell) = \kappa p_2(\ell) + \alpha(\kappa\ell) = (1-\tau)p_2(\kappa\ell).$$

Therefore, we have $p_1(\kappa\ell) = p_2(\kappa\ell)$. In words, if the prices are equal at level i for $i = 2, \ldots, m_2$, then the prices are equaled at its downstream level $i-1$. To this end, if we can show that $p_1(\kappa^{1-m_2}) = p_2(\kappa^{1-m_2})$ at level m_2, then it implies that $p_1(\ell) = p_2(\ell)$ for all $\ell = 1, \kappa^{-1}, \ldots, \kappa^{1-m_2}$.

If $m_1 = m_2$, then firms are home producers at level m_1 in both equilibrium so that $p_1(\kappa^{1-m_1}) = \delta c(\kappa^{1-m_1}) = p_2(\kappa^{1-m_1})$. Then, $p_1(\ell) = p_2(\ell)$ for all $\ell = 1, \kappa^{-1}, \ldots \kappa^{1-m_1}$ and two equilibrium prices are the same.

Suppose that $m_1 > m_2$. Let $\delta = 1/(1-\tau)$ and $t = \kappa^{1-m_2}$ to simplify notations. Then, $p_1(t) \neq p_2(t)$. At level m_2, the firms in equilibrium $(p_1, \boldsymbol{\ell}^1)$ are subcontractors, while the firms in equilibrium $(p_2, \boldsymbol{\ell}^2)$ are in-house producers. Hence, condition (3) of Definition

5.2.1 implies $p_2(t) = \delta c(t)$ and the following iteration

$$\begin{aligned}
p_1(t) &= \delta\kappa p_1(t/\kappa) + \delta\alpha(t) \\
&= \delta\kappa[\delta\kappa p_1(t/\kappa^2) + \delta\alpha(t/\kappa)] + \delta\alpha(t) \\
&= (\delta\kappa)^2 p_1(t/\kappa^2) + \delta^2\kappa b(t/\kappa)^q + \delta b(t)^q \\
&= (\delta\kappa)^2 p_1(t/\kappa^2) + \delta b t^q[\delta\kappa^{1-q} + 1] \\
&= (\delta\kappa)^2[\delta\kappa p_1(t/\kappa^3) + \delta\alpha(t/\kappa^2)] + \delta b t^q[\delta\kappa^{1-q} + 1] \\
&= (\delta\kappa)^3 p_1(t/\kappa^3) + \delta b t^q[(\delta\kappa^{1-q})^2 + \delta\kappa^{1-q} + 1] \\
&= \dots \\
&= (\delta\kappa)^{m_1-m_2} p_1(t/\kappa^{m_1-m_2}) \\
&\qquad + \delta b t^q[(\delta\kappa^{1-q})^{m_1-m_2-1} + \cdots + \delta\kappa^{1-q} + 1] \\
&= (\delta\kappa)^{m_1-m_2}\delta c(\kappa^{1-m_1}) + \delta b t^q \frac{(\delta\kappa^{1-q})^{m_1-m_2} - 1}{\delta\kappa^{1-q} - 1}
\end{aligned} \tag{5.14}$$

Similarly, condition (2) of equilibrium implies $p_1(t) \leqslant \delta c(t)$ and the iteration

$$\begin{aligned}
p_2(t) &\leqslant \delta\kappa p_2(t/\kappa) + \delta\alpha(t) \\
&= (\delta\kappa)^2 p_2(t/\kappa^2) + \delta b t^q[(\delta\kappa^{1-q}) + 1] \\
&\leqslant (\delta\kappa)^2[\delta\kappa p_2(t/\kappa^3) + \delta\alpha(t/\kappa^2)] + \delta b t^q[\delta\kappa^{1-q} + 1] \\
&= (\delta\kappa)^3 p_2(t/\kappa^3) + \delta b t^q[(\delta\kappa^{1-q})^2 + \delta\kappa^{1-q} + 1] \\
&\leqslant \dots \\
&\leqslant (\delta\kappa)^{m_1-m_2} p_2(t/\kappa^{m_1-m_2}) + \delta b t^q[(\delta\kappa^{1-q})^{m_1-m_2-1} + \cdots + 1] \\
&\leqslant (\delta\kappa)^{m_1-m_2}\delta c(\kappa^{1-m_1}) + \delta b t^q \frac{(\delta\kappa^{1-q})^{m_1-m_2} - 1}{\delta\kappa^{1-q} - 1} \\
&= p_1(t)
\end{aligned} \tag{5.15}$$

where the last inequality follows from $p_2 \in \mathcal{P}$. Therefore, $\delta c(t) = p_2(t) \leqslant p_1(t) \leqslant \delta c(t)$. We have $p_1(t) = p_2(t)$ and then $p_1(\ell) = p_2(\ell)$ for all $\ell = 1, \kappa^{-1}, \dots, \kappa^{1-m_2}$ at lower levels by the previous claim. It implies that firms at level m_2 in equilibrium $(p_1, \boldsymbol{\ell}_1)$ have the same costs as the home production.

Since firms choose to produce in-house when the home production costs and subcontracting costs are the same, the firms at level m_2 under equilibrium p_1 will choose to home production since $p_1(\kappa^{1-m_2}) = \delta c(\kappa^{1-m_2}) = \delta\kappa p_1(\kappa^{-m_2}) + \delta\alpha(\kappa^{1-m_2})$. That is, the firms have no incentive to subcontract and make longer length than m_2. Then, it must be $m_1 = m_2$. This implies that $\boldsymbol{\ell}^1 = \boldsymbol{\ell}^2$ since the tasks in each level are identical for their feasibility. ☐

Proof of Proposition 5.2.1. The statements follow from Part (a) follows from Lemma 5.6.1, Part (b) and (c) follow from Lemma 5.6.5, Part (d) follows from Lemma 5.6.4, Part (e) follows from 5.6.2, and Part (f) follows from Lemma 5.6.6. ☐

Proof of Proposition 5.2.2. The statements follow from Lemma 5.6.7 and Lemma 5.6.8. ☐

Proofs in Section 5.2.3. Suppose that 5.2.1, 5.2.2 and 5.2.3 hold through out this section. Let $f(\ell) := \delta\kappa c(\ell/k) + \alpha(\ell) - c(\ell)$ for all $\ell \in [0,1]$.

Lemma 5.6.9. *Suppose that p is the equilibrium price and f is strictly concave. Then, $\hat{\ell} \in (0,1]$ is a root of f if and only if $\hat{\ell}$ satisfies $p(\ell) = \delta c(\ell)$ for all $0 \leqslant \ell \leqslant \hat{\ell}$ and $p(\ell) < \delta c(\ell)$ for all $\hat{\ell} < \ell \leqslant 1$.*

Proof of Lemma 5.6.9. Let the stated assumptions hold. Let $P = \{p : [0,1] \to \mathbb{R} : \delta c'(0)\ell \leqslant p(\ell) \leqslant \delta c(\ell), \forall \ell \in [0,1]\}$. The same arguments in Lemma 5.6.1, Lemma 5.6.2 and 5.6.4 show that $T \colon P \to P$ has a unique fixed point in P. Let p be the fixed point of P, which is also a equilibrium price.

Since $f(0) = 0$ by assumption 5.2.1, 5.2.2 and 5.2.3, and f is strictly concave, the root of f on $[0,1]$ is unique. Suppose that $\hat{\ell} \in (0,1]$ satisfies $f(\hat{\ell}) = 0$. Fix $\ell \in [0,\hat{\ell}]$. We have $f(\ell) \geqslant 0$ and $\delta\kappa c(\ell/\kappa) + \alpha(\ell) \geqslant c(\ell)$, so

$$T\delta c(\ell) = \delta \min\{\kappa\delta c(\ell/\kappa) + \alpha(\ell), c(\ell)\} = \delta c(\ell).$$

Then, $\delta c(\ell)$ is a fixed point of T if ℓ is restricted in $[0,\hat{\ell}]$. Since T only has one fixed point, we have $p(\ell) = \delta c(\ell)$.

Next, let $\ell \in (\hat{\ell}, \kappa\hat{\ell}\,]$. since $\ell/\kappa \leqslant \hat{\ell}$ and $f(\ell) < 0$, we have

$$\begin{aligned} p(\ell) &= \delta \min\{\kappa p(\ell/\kappa) + \alpha(\ell), c(\ell)\} \\ &= \delta \min\{\kappa\delta c(\ell/\kappa) + \alpha(\ell), c(\ell)\} \\ &= \delta^2 \kappa c(\ell/\kappa) + \delta\alpha(\ell) < \delta c(\ell), \end{aligned}$$

where the last inequality follows from $f(\ell) < 0$. Hence, firms subcontract when $\ell \in (\hat{\ell}, \kappa\hat{\ell}\,]$. Similarly, fixing $\ell \in (\kappa\hat{\ell}, \kappa^2\hat{\ell}\,]$, we have $p(\ell/\kappa) < \delta c(\ell/\kappa)$ and then

$$\begin{aligned} p(\ell) &= \delta \min\{\kappa p(\ell/\kappa) + \alpha(\ell), c(\ell)\} \\ &\leqslant \delta \min\{\kappa\delta c(\ell/\kappa) + \alpha(\ell), c(\ell)\} \\ &= \delta^2 \kappa c(\ell/\kappa) + \delta\alpha(\ell) < \delta c(\ell), \end{aligned}$$

where the last inequality follows from $f(\ell) < 0$. Iteration implies $p(\ell) < \delta c(\ell)$ for $\ell \in (\hat{\ell}, \kappa^n\hat{\ell}\,]$ for $n \geqslant 2$. Therefore, we have $p(\ell) < \delta c(\ell)$ for $\ell > \hat{\ell}$. For the sufficiency, suppose that $\hat{\ell}$ satisfies $p(\ell) = \delta c(\ell)$ for $\ell \leqslant \hat{\ell}$ and $p(\ell) < \delta c(\ell)$ for $\ell > \hat{\ell}$. Then, by the definition of equilibrium price p, we have $f(\ell) \geqslant 0$ for $\ell \leqslant \hat{\ell}$ and $f(\ell) < 0$ for $\ell > \hat{\ell}$. Since the root of f is unique in $(0, 1]$, $f(\hat{\ell}) = 0$. □

Lemma 5.6.10. *Following Lemma 5.6.9, if* $0 < \hat{\ell} < 1$, *then the equilibrium length has at least two levels.*

Lemma 5.6.11. *Suppose that Assumptions 5.2.2 and 5.2.3 hold, and* c *is twice differentiable. If* $f(\ell)$ *is strictly concave and has a unique root in* $(0, 1]$, *then the equilibrium price* p *is twice differentiable except for a finite number of points with* $p'(\ell) > 0$. *If further* $\alpha'' \geqslant 0$, *then* $p''(\ell) \geqslant 0$.

Proof of Lemma 5.6.11. Let all the stated assumptions hold. By Lemma 5.6.9, there is $\hat{\ell}$ such that $p(\ell) = \delta c(\ell)$ for $\ell \in [0, \hat{\ell}\,]$. Fixing $\ell \in (\hat{\ell}, \kappa\hat{\ell})$, Lemma 5.6.9 implies that $p(\ell) = \delta[\kappa\delta c(\ell/\kappa) + \alpha(\ell)] = \delta^2\kappa c(\ell/\kappa) + \delta\alpha(\ell)$. Since c and α are twice differentiable, $p'(\ell) = \delta^2 c'(\ell/\kappa) + \delta\alpha'(\ell)$ and $p''(\ell) = \delta^2/\kappa c''(\ell/\kappa) + \delta\alpha''(\ell)$. Since c and α are strictly increasing, $p'(\ell) > 0$. If a is convex, then $p''(\ell) \geqslant 0$.

Fix $\ell \in (\kappa\hat{\ell}, \kappa^2\hat{\ell})$. We have

$$\begin{aligned} p(\ell) &= \delta[\kappa p(\ell/\kappa) + \alpha(\ell)] \\ &= \delta[\kappa[\delta^2\kappa c(\ell/\kappa^2) + \delta\alpha(\ell/\kappa)] + \alpha(\ell)] \\ &= \delta^3\kappa^2 c(\ell/\kappa^2) + \delta^2\kappa\alpha(\ell/\kappa) + \delta\alpha(\ell) \end{aligned}$$

Then, we can compute $p'(\ell) = \delta^3 c'(\ell/\kappa^2) + \delta^2\alpha'(\ell/\kappa) + \delta\alpha'(\ell)$ and $p''(\ell) = \delta^3/\kappa^2 c''(\ell/\kappa^2) + \delta^2/\kappa\alpha''(\ell/\kappa) + \delta\alpha''(\ell)$ and confirm the statement for ℓ in $(\kappa\hat{\ell}, \kappa^2\hat{\ell})$.

By iteration, we conclude that p is differentiable, $p'(\ell) \geqslant 0$ and $p''(\ell) \geqslant 0$ in $[0, 1]$ except for the points $\{\hat{\ell}, \kappa\hat{\ell}, \ldots, \kappa^n\hat{\ell}\}$, where n is the greatest integer such that $\kappa^n\hat{\ell} \leqslant 1$. ☐

LEMMA 5.6.12. *If* $\delta_1 \leqslant \delta_2$ *and* $p^*_{\delta_1}, p^*_{\delta_2}$ *are the corresponding equilibrium prices, then* $p^*_{\delta_1} \leqslant p^*_{\delta_2}$.

PROOF OF LEMMA 5.6.12. Suppose that $\delta_1 \leqslant \delta_2$. Let T_1 and T_2 be the corresponding operator with respect to δ_1 and δ_2, respectively. Let $p \in \mathcal{P}$. Clearly, $T_1 p \leqslant T_2 p$. Since operator T is order preserving, we have

$$T_1(T_1 p) \leqslant T_1(T_2 p) \leqslant T_2(T_2 p).$$

By iteration, $p^*_{\delta_1} = T_1^n p \leqslant T_2^n p = p^*_{\delta_2}$ for $n \geqslant 1 - \frac{\ln \bar{x}}{\ln \kappa}$ ☐

PROOF OF PROPOSITION 5.2.3. Let all the stated assumptions hold. For the transaction cost, Lemma 5.6.12 and $\delta = 1/(1-\tau)$ shows that the equilibrium price is increasing in τ. For home production costs, suppose that $c_1, c_2 \colon \mathcal{D} \to \mathbb{R}_+$ satisfying $c_1(\ell) \leqslant c_2(\ell)$ for all $\ell \in \mathcal{D}$. Let T_1 and T_2 be the corresponding operator, respectively. Clearly, $T_1 p \leqslant T_2 p$ for all $p \in \mathcal{P}$, since we have

$$\min\{\kappa p(\ell/\kappa) + \alpha(\ell), c_1(\ell)\} \leqslant \min\{\kappa p(\ell/\kappa) + \alpha(\ell), c_2(\ell)\}.$$

Thus, $p^*_1 \leqslant p^*_2$ by the iteration as Lemma 5.6.12, where p^*_i is the equilibrium price with respect to c_i, $i = 1, 2$. Similarly, it can be shown that the price is increasing in assembly cost.

Next, we show that the price of final good is strictly increasing in transaction cost, assembly cost and home production cost. Let the price of final good $p(\alpha, c, m, \delta, \kappa)$ be defined as equation (5.4). In particular,

$$\begin{aligned} p(\alpha, c, m, \delta, \kappa) &:= (\delta\kappa)^{m-1}\delta c(\kappa^{1-m}) \\ &\quad + \delta\alpha(1) + \delta^2\kappa\alpha(\kappa^{-1}) + \cdot + \delta^{m-1}\kappa^{m-2}\alpha(\kappa^{2-m}) \end{aligned} \tag{5.16}$$

Suppose that $\alpha_1(\ell) < \alpha_2(\ell)$ for all ℓ and let p_1^* and p_2^* be the corresponding equilibrium price. Also let m_1^* and m_2^* be the corresponding optimal length in equilibrium, $m_1^* \geqslant m_2^*$. Then, by the optimality of p^*, we have $p_1^*(1) = p(\alpha_1, m_1^*; \delta, \kappa, c) \leqslant p(\alpha_1, m_2^*; \delta, \kappa, c) < p(\alpha_2, m_2^*; \delta, \kappa, c) = p_2^*(1)$. For the other task allocation $t = 1, \kappa^{-1}, \cdots, \kappa^{1-m_2^*}$, we can also characterize the price $p^*(t)$ as equation (5.14) and show that $p_1^*(t) < p_2^*(t)$.

For the production cost, suppose that $c_1(\ell) < c_2(\ell)$ for all ℓ and let p_1^* and p_2^* be the corresponding equilibrium price. Also let m_1^* and m_2^* be the corresponding optimal length in equilibrium, $m_1^* \leqslant m_2^*$. Then, by the optimality of p^*, we have $p_1^* = p(\alpha, c_1, m_1^*, \delta, \kappa) \leqslant p(\alpha, c_1, m_2^*, \delta, \kappa) < p(\alpha, c_2, m_2^*, \delta, \kappa) = p_2^*$. Similarly, we can compute $p^*(t)$ and show that $p_1^*(t) < p_2^*(t)$ by the same argument for $t = 1, \kappa^{-1}, \cdots, \kappa^{1-m_1^*}$. We can also prove the statement for transaction costs following the same procedures. $\square$

Proofs in Section 5.4. Throughout this section, suppose that 5.2.2, 5.2.1 and 5.4.2 hold, and the price is defined by (5.12).

PROOF OF LEMMA 5.4.1. Consider the operator $T : \mathcal{P} \to \mathcal{P}$ as

$$Tp(\ell) = \delta \min\{np(\ell/n) + \alpha(n), c(\ell)\}.$$

The similar proofs as in Lemma 5.6.1, Lemma 5.6.2 and Lemma 5.6.4 show that the fixed point uniquely exists. The statements then follows from Lemma 5.6.7 and 5.6.8. $\square$

LEMMA 5.6.13. *There is a unique solution to pricing function (5.12), and the equilibrium price of Definition 5.4.1 exists.*

Proof. Consider the operator $T : \mathcal{P} \to \mathcal{P}$ as

$$Tp(\ell) = \delta \min\left\{ \min_{n=2,3,4,\dots} \{np(\ell/n) + \alpha(\ell, n)\},\, c(\ell)\right\},$$

for all $\ell \in [0,1]$. We first show that T has a unique fixed point which is a equilibrium price. To the existence, following the argument of Lemma 5.6.1, it suffices to show that T is a self-map on $\mathcal{P}$ and preserves orders. Fix $\ell \in [0,1]$. Clearly, we have $Tp(\ell) \leqslant \delta c(\ell)$. Since $p \in \mathcal{P}$ and c is convex, we have

$$\begin{aligned} Tp(\ell) &\geqslant \delta \min\left\{ \min_{n=2,3,\dots} \{n\delta c'(0)\ell/n + \alpha(\ell, n)\},\, c(\ell)\right\} \\ &= \delta \min\left\{ \delta c'(0)\ell + \min_n\{\alpha(\ell, n)\},\, c(\ell)\right\} \\ &\geqslant \delta \min\{c'(0)\ell, c(\ell)\} = \delta c'(0)\ell \end{aligned}$$

Then, $Tp(\ell) \geqslant \delta c'(0)\ell$ and $T : \mathcal{P} \to \mathcal{P}$. To see that T preserves orders, assume $p_1(\ell) \leqslant p_2(\ell)$ for all $\ell \in [0,1]$. Fix $\ell \in [0,1]$ and $n \geqslant 2$, we have $np_1(\ell/n) + \alpha(\ell, n) \leqslant np_2(\ell/n) + \alpha(\ell, n)$. Hence, we have $\min_n\{np_1(\ell/n) + \alpha(\ell, n)\} \leqslant np_2(\ell/n) + \alpha(\ell, n)$ for all $n \geqslant 2$, implying $\min_n\{np_1(\ell/n) + \alpha(\ell, n)\} \leqslant \min_n\{np_2(\ell/n) + \alpha(\ell, n)\}$. This implies that $Tp_1(\ell) \leqslant Tp_2(\ell)$ for all $\ell \in [0,1]$. Therefore, Knaster-Tarski Fixed Point Theorem shows that T has at least one fixed point.

Similar to Lemma 5.6.2 and 5.6.4, we next show that $T^k p = T^k q$ for $n \geqslant 1 - \ln \bar{x}/\ln 2$ for arbitrary $p, q \in \mathcal{P}$, where $\bar{x} := \sup\{x \in [0,1] : c'(x) \leqslant \delta c'(0)\}$. Let $p, q \in \mathcal{P}$. We claim that $Tp(\ell) = Tq(\ell) = \delta c(\ell)$ for $\ell \leqslant \bar{x}$. Suppose that $Tp(\ell) \neq \delta c(\ell)$ for $\ell \leqslant \bar{x}$. Then, firms choose to subcontract, so it must be $\min_n\{np(\ell/n) + \alpha(\ell, n)\} < c(\ell)$[35]. Since $p(\ell/n) \geqslant \delta c'(0)\ell/n$ for $p \in \mathcal{P}$,

$$c(\ell) > \min_n\{np(\ell/n) + \alpha(\ell, n)\} \geqslant \delta c'(0)\ell$$

Therefore, $c(\ell)/\ell > \delta c'(0)$. This contradicts that $\ell \leqslant \bar{x}$ and convexity of c. Hence, we have $Tp(\ell) = Tq(\ell)$ for $\ell \leqslant \bar{x}$. Next, suppose that $T^k p(\ell) = T^k q(\ell)$ for $\ell \in [0, 2^{k-1}\bar{x}]$. Let $\ell \in [0, 2^k\bar{x}]$. Since $nT^k p(\ell/n) + \alpha(\ell, n) = nT^k q(\ell/n) + \alpha(\ell, n)$ for all $n \geqslant 2$, we can prove

---

[35] If " = " holds, then firms choose to home production by assumption.

that $T^{k+1}p(\ell) = T^{k+1}q(\ell)$ for $\ell \in [0, 2^k\bar{x}]$. By induction, $T^k p(\ell) = T^k q(\ell)$ for all $\ell \in [0,1]$ and $k \geqslant 1 - \ln \bar{x} / \ln 2$.

By definition of T and pricing function (5.12), p is a fixed point of T if and only if it is a solution to (5.12). Suppose that there are two fixed points or solutions of (5.12) p, q and $p \neq q$. Then, by the above induction, $p = T^k p = T^k q = q$ for large enough k. Hence, T has only one fixed point and the solution of (5.12) is also unique.

Let $\delta = 1/(1-\tau)$ and $p^* \in \mathcal{P}$ be the solution to (5.12). The above induction implies that there is a maximal possible length $\bar{m}$ and the equilibrium length m^* is finite. Therefore, we can construct a feasible $\boldsymbol{\ell}$ by iteration, starting from $\ell_1 = 1$, under the price p^*. Moreover, since $p^* \in \mathcal{P}$ and $c(0) = 0$, we have $p^*(0) = 0$. Condition (1) of Definition (5.2.1) is satisfied. By definition of (5.12) and the profit (5.13), for all $\ell \in \mathcal{D}$, we have[36]

$$p^*(\ell) - \delta \min \left\{ \min_{n=2,3,4,\ldots} \{np^*(\ell/n) + \alpha(\ell, n)\},\, c(\ell) \right\} = 0.$$

Therefore, Condition (2) and (3) are satisfied, so the solution to (5.12) is the equilibrium price. □

Lemma 5.6.14. *Given the equilibrium price, the equilibrium allocation is uniquely determined.*

Proof. Let $(p, \boldsymbol{\ell}^{\mathbf{1}})$ and $(p, \boldsymbol{\ell}^{\mathbf{2}})$ be two equilibrium with the same price p and different allocations. Since $\boldsymbol{\ell}^1 = \{\ell_i^1\}$ and $\boldsymbol{\ell}^2 = \{\ell_i^2\}$ are feasible allocations, there exist $m_1^*, m_2^* \geqslant 2$ and two integer sequence $\{n_t^1\}_{t=1}^{m_1^*}$ and $\{n_t^2\}_{t=1}^{m_2^*}$ such that $\ell_i^j = (n_1^j n_2^j \cdots n_t^j)^{-1}$ for $n_1^j + \cdots + n_1^j n_2^j \cdots n_{t-1}^j < i \leqslant n_1^j + \cdots + n_1^j n_2^j \cdots n_t^j$ for $j = 1, 2$. Suppose that $m_1^* \geqslant m_2^*$ without loss of generality. That is, $\boldsymbol{\ell}^1$ has longer chain.

If $m_2^* = 1$, then, by the assumption that firms choose home production if home production and subcontract are indifferent, it can be shown that $m_1^* = m_2^* = 1$ and $\boldsymbol{\ell}^1 = \boldsymbol{\ell}^2$ by the bellowing argument. Thus, assume that $m_2^* > 1$. At level 1, using condition (3) of

---

[36] It is also valid for $\ell \in [0, 1]$.

Definition 5.4.1, we have

$$\begin{aligned}\pi &= (1-\tau)p(1) - \min\{\min_{n=2,3,\dots}\{np(1/n)+\alpha(1,n)\}, c(1)\}\\ &= (1-\tau)p(1) - n_2^1 p(1/n_2^1) + \alpha(1,n_2^1) = 0\\ &= (1-\tau)p(1) - n_2^2 p(1/n_2^2) + \alpha(1,n_2^2) = 0\end{aligned}$$

Thus, $n_2^1 = n_2^2$ by the assumption that firms choose the minimum optimal number of suppliers. If $n_s^1 = n_s^2 = n_s$ for $s = 1,\dots,t$ and $t < m_2^*$, for $\ell = (n_1 n_2 \cdots n_t)^{-1}$, then the profit at ℓ is

$$\begin{aligned}&(1-\tau)p(\ell) - \min\{\min_{n=2,3,\dots}\{np(\ell/n)+\alpha(\ell,n)\}, c(\ell)\}\\ &\qquad\qquad = (1-\tau)p(\ell) - \min_{n=2,3,\dots}\{np(\ell/n)+\alpha(\ell,n)\}\end{aligned}$$

Again, by choosing the minimal minimiser, $n_{t+1}^1 = n_{t+1}^2$. Then, by induction, we have $n_t^1 = n_t^2 = n_t$ for all $t = 1,\dots,m_2^*$.

Suppose that $m_1^* > m_2^*$. At level m_2^*, we have tasks $\ell = (n_1 n_2 \cdots n_{m_2^*})^{-1}$. The firms of $\boldsymbol{\ell}^1$ at this level choose to outsource while the firms of $\boldsymbol{\ell}^2$ choose to home production. That is, for $t = m_2^*$,

$$\begin{aligned}&(1-\tau)p(\ell) - \min\{\min_{n=2,3,\dots}\{np(\ell/n)+\alpha(\ell,n)\}, c(\ell)\}\\ &= (1-\tau)p(\ell) - \min\{n_{t+1}^1 p(\ell/n_{t+1}^1) + \alpha(\ell,n_{t+1}^1), c(\ell)\}\\ &= (1-\tau)p(\ell) - n_{t+1}^1 p(\ell/n_{t+1}^1) + \alpha(\ell,n_{t+1}^1)\\ &= (1-\tau)p(\ell) - c(\ell) = 0\end{aligned}$$

By the assumption that firms produce at home if home production and subcontract are indifferent, it should be $m_1^* = m_2^*$. Therefore, $\boldsymbol{\ell}^1 = \boldsymbol{\ell}^2$. □

Proof of Proposition 5.4.1. The statement follows from the proof of Lemma 5.6.13 and Lemma 5.6.14 □

Proof of Proposition 5.4.2. For the first statement, let $\tau_1 \leqslant \tau_2$ and δ_1, δ_2 be the corresponding parameters of transaction costs with $\delta_1 \leqslant \delta_2$. Also, let T_i be the corresponding operator for d_i, $i = 1, 2$. Clearly, $T_1 p \leqslant T_2 p$ for all p in $\mathcal{P}$. By iteration, we have

$T_1T_1p \leqslant T_1T_2p \leqslant T_2T_2p$ or $T_1^2p \leqslant T_2^2p$. Repeat the iteration, we have the fixed points that $p_1^* = T_1^k \leqslant T_2^k = p_2^*$ for $k \to \infty$.

For the assembly cost, suppose that $\alpha_1(\ell) \leqslant \alpha_2(\ell)$ for all ℓ. Let T_i and p_i^* be the corresponding operator and equilibrium price with respect to α_i, $i = 1, 2$. Fix $p \in \mathcal{P}$. Then, $\min_n\{np(\ell/n) + \alpha(\ell, n)\} \leqslant np(\ell/n) + \alpha_2(\ell, n)$. Taking the minimum on the right hand side, we have $T_1p \leqslant T_2p$. By the similar iteration, we also have $p_1^* \leqslant p_2^*$. Moreover, the same technique proves the statement for home production cost c.

We can also show that the price of final good is strictly increasing in τ, c and α. Take the home production cost for example. Let the costs be $c_1 < c_2$. Let $(p_1^*, \boldsymbol{\ell}_1)$ and $(p_2^*, \boldsymbol{\ell}_2)$ be the corresponding equilibrium. Define the equilibrium price as $p(c, \alpha, \delta)$ under parameter c, α and δ. By the optimality of equilibrium, $p_1^* = p(c_1, \boldsymbol{\ell}_1; \delta) \leqslant p(c_1, \boldsymbol{\ell}_2; \delta) < p(c_2, \boldsymbol{\ell}_2; \delta) = p_2^*$. Thus, the price of final good is increasing in home production cost. Similarly, we can prove that the price of final good is strictly increasing in transaction cost and assembly cost. $\square$

LEMMA 5.6.15. *Suppose that $\alpha(\ell, n)$ is increasing in ℓ. If p is the solution of pricing function (5.12), then it is strictly increasing in ℓ.*

PROOF. Suppose that $p \in \mathcal{P}$ is strictly increasing in ℓ. Let $\ell_1 \leqslant \ell_2$. Then,

$$\begin{aligned} \min_n\{np(\ell_1/n) + \alpha(\ell_1, n)\} &\leqslant np(\ell_1/n) + \alpha(\ell_1, n) \\ &< np(\ell_2/n) + \alpha(\ell_2, n) \quad \text{for all n} \geqslant 2. \end{aligned}$$

Hence, $\min_n\{np(\ell_1/n) + \alpha(\ell_1, n)\} < \min_n\{np(\ell_2/n) + \alpha(\ell_2, n)\}$. Since c is also strictly increasing,

$$\begin{aligned} Tp(\ell_1) &= \delta \min\{\min_n\{np(\ell_1/n) + \alpha(\ell_1, n)\}, c(\ell_1)\} \\ &< \delta \min\{\min_n\{np(\ell_2/n) + \alpha(\ell_2, n)\}, c(\ell_2)\} \\ &= Tp(\ell_2) \end{aligned}$$

$Tp(\ell)$ is also strictly increasing. Therefore, the fixed point $T^N p(\ell)$, for large enough N, is also strictly increasing. By Lemma 5.6.13, the solution is unique and equaled to the fixed point of T, so the solution of (5.12) is strictly increasing. □

Proof of Lemma 5.4.2. (⇒) Let $\ell, \ell' \in [0,1]$ with $\ell \neq \ell'$ and $\theta \in (0,1)$. Since $t(n,.)$ is strictly concave in ℓ,

$$
\begin{aligned}
&\min_{n=2,3,\ldots} \left\{ n\delta c\left(\frac{\theta\ell + (1-\theta)\ell'}{n}\right) + \alpha(n) - c(\theta\ell + (1-\theta)\ell') \right\} \\
&\quad = \min_{n=2,3,\ldots} \{t(\theta\ell + (1-\theta)\ell') + \alpha(n)\} \\
&\quad > \min_{n=2,3,\ldots} \{\theta t(\ell/n) + (1-\theta)t(\ell'/n) + \alpha(n)\} \\
&\quad \geqslant \min_{n=2,3,\ldots} \{\theta(t(\ell/n) + \alpha(n)) + (1-\theta)(t(\ell'/n) + \alpha(n))\} \\
&\quad \geqslant \theta \min_{n=2,3,\ldots} \{t(\ell/n) + \alpha(n)\} + (1-\theta) \min_{n=2,3,\ldots} \{t(\ell'/n) + \alpha(n)\}
\end{aligned}
$$

Thus, the function $f(\ell) := \min_n\{n\delta c(\ell/n) + \alpha(n)\} - c(\ell)$ is strictly concave. Together with $f(0) > 0$, the root $\hat{\ell} \in (0,1]$ is unique. Therefore, fixing $\ell \leqslant \hat{\ell}$, we have $f(\ell) \geqslant 0$ and then

$$
c(\ell) \leqslant \min_{2,3,\ldots}\{n\delta c(\ell/\kappa) + \alpha(n)\}.
$$

So, $T\delta c(\ell) = \delta c(\ell)$. Hence, $\delta c(\ell)$ is a fixed point of T considering only $\ell \in [0, \hat{\ell}]$.

Fixing $\ell \in (\hat{\ell}, 2\hat{\ell}]$. Observe that $p(\ell/n) = \delta c(\ell/n)$ for all $n \geqslant 2$. Then, since $f(\ell) < 0$, we have

$$
\begin{aligned}
p(\ell) &= \delta \min\{\min_{n=2,3,\ldots} \{np(\ell/n) + \alpha(n)\}, c(\ell)\} \\
&= \delta \min\{\min_{n=2,3,\ldots} \{n\delta c(\ell/n) + \alpha(n)\}, c(\ell)\} \\
&< \delta c(\ell).
\end{aligned}
$$

Fixing $\ell \in (2\hat{\ell}, 2^2\hat{\ell}]$. Observe that $p(\ell/2) < \delta c(\ell)$ for $\ell/2 \in (\hat{\ell}, 2\hat{\ell})$ and $p(\ell/n) = \delta c(\ell/n)$ for $n > 2$ and $\ell/n < \hat{\ell}$. Again, since $f(\ell) < 0$, we have

$$\begin{aligned} p(\ell) &= \delta \min\{\min_{n=2,3,\dots}\{np(\ell/n) + \alpha(n)\}, c(\ell)\} \\ &\leqslant \delta \min\{\min_{n=2,3,\dots}\{n\delta c(\ell/n) + \alpha(n)\}, c(\ell)\} \\ &< \delta c(\ell). \end{aligned}$$

By iteration, $p(\ell) < \delta c(\ell)$ for all $\ell \in (\hat{\ell}, 2^N\hat{\ell}]$. Hence, we obtain that $p(\ell) < \delta c(\ell)$ for all $\hat{\ell} < \ell \leqslant 1$ when N is large enough.

($\Leftarrow$) This is done by the definition of equilibrium price p and strict concavity of function f. □

Proof. By Lemma 5.4.2, there is $\hat{\ell}$ such that $p(\ell) = \delta c(\ell)$ for all $\ell \leqslant [0, \hat{\ell}]$, so $p'(\ell) = \delta c'(\ell) > 0$ and $p''(\ell) = dc''(\ell) > 0$ for all $\ell \in [0, \hat{\ell})$. Considering $\ell \in (\hat{\ell}, 1)$. Lemma 5.4.2 implies that $p(\ell) = \delta \min_{n=2,3,\dots}\{np(\ell/n) + \alpha(n)\}$ for all $\ell > \hat{\ell}$.

Considering $\ell \in (\hat{\ell}, 2\hat{\ell})$, we have $p(\ell) = \delta \min_{n=2,3,\dots}\{n\delta c(\ell/n) + \alpha(n)\}$. Relaxing the choice set of n from $\{2, 3, \dots\}$ to $[2, \infty)$. Define the optimal choice $n^*(\ell)$ as

$$n^*(\ell) := \arg\min_{n \geqslant 2}\{n\delta c(\ell/n) + \alpha(n)\}.$$[37]

The first order condition is $r(n, \ell) := \delta c(\ell/n) - (\ell/n)\delta c'(\ell/n) + \alpha'(\ell/n) = 0$. By Implicit Function Theorem,

$$\frac{\mathrm{p}n^*}{\mathrm{p}\ell} = -\frac{\mathrm{p}r}{\mathrm{p}\ell}\Big/\frac{\mathrm{p}r}{\mathrm{p}n^*} = \frac{c''(\ell/n^*)\ell/n^{*2}}{c''(\ell/n^*)\ell^2/n^{*3} + \alpha''(n)}.$$

Hence, $\mathrm{p}n^*(\ell)/\mathrm{p}\ell > 0$ by the assumption that c is strictly convex and α is convex. Since $s(n;\ell) := n\delta c(\ell/n) + \alpha(n)$ is strictly increasing in ℓ and $s''(n;\ell) = (\ell^2/n^3)\delta c''(\ell/n) + \alpha''(n) > 0$, the curve $s(n;\ell)$ on $s(n) - n$ plane is strictly convex and shifts upward as ℓ increases. Moreover, the minimum point of $s(n;\ell)$ shifts to right as ℓ increases since $\mathrm{p}n^*(\ell)/\mathrm{p}\ell > 0$. Now, restrict n back to the grids $n \geqslant 2$. Since $s(\ell;n)$ shifts upward and the minimum point shifts to right as ℓ increases, the optimal grid $n^*(\ell)$ is increasing in

---

[37] The maximum theorem implies that $n^*(\ell)$ is single-valued and continuous if we restrict the upper bound on n.

ℓ.[38] Therefore, given n^*, there is a neighborhood $U(n^*) \subset (\hat{\ell}, 2\hat{\ell})$ such that $n^*(\ell') = n^*$ for all $\ell' \in U(n^*)$.

Suppose that $n^*(\ell) = n^*$ for a fixed $\ell \in (\hat{\ell}, 2\hat{\ell})$. Then, ℓ must be on the neighborhood $U(n^*)$. If ℓ is not in the boundary of $U(n^*)$, then $p(\ell') = \delta^2 n^* c(\ell'/n^*) + \delta\alpha(n^*)$ for $\ell' \in U(n^*)$. Given $c \in \mathrm{C}^2$, this implies that $p'(\ell') = \delta^2 c'(\ell'/n^*) > 0$ and $p''(\ell') = \delta^2 c''(\ell'/n^*)/n^* > 0$ for $\ell' \in U(n^*)$. If ℓ is on the boundary of $U(n^*)$, we exclude it. Since the optimal $n^*(\ell)$'s are discrete and countable points, the set of boundary points of such U has measure 0. Since $\ell \in (\hat{\ell}, 2\hat{\ell})$ is arbitrary, $p'(\ell) > 0$ and $p''(\ell) > 0$ for almost all ℓ in $(\hat{\ell}, 2\hat{\ell})$.

Denote $p(\ell)$ by $q(\ell)$ for $\ell \in (\hat{\ell}, 2\hat{\ell})$. Then, $q' > 0$ and $q'' > 0$ almost everywhere. Fixing $\ell \in (2\hat{\ell}, 2^2\hat{\ell})$. Again,

$$p(\ell) = \delta \min_{n=2,3,\dots} \{np(\ell/n) + \alpha(n)\} = \delta n^*(\ell) q(\ell/n^*(\ell)) + \alpha(n^*(\ell)).$$

Similarly, let $N \subset (2\hat{\ell}, 2^2\hat{\ell})$ be the neighborhood of ℓ such that $n^*(\ell') = n^*(\ell) = n^*$ for all $\ell' \in N$. Excluding the boundary points of such neighborhood, suppose that ℓ is not on the boundary of N. If $n^* = 2$, then $p(\ell') = \delta n^* q(\ell'/n^*) + \delta\alpha(n^*)$ so that $p'(\ell') = \delta q'(\ell'/n^*) > 0$ and $p''(\ell') = (\delta/n^*) q''(\ell'/n^*) > 0$ for $\ell' \in N$. If $n^* > 2$, then $p(\ell') = \delta n^* \delta c(\ell'/n^*) + \delta\alpha(n^*)$ so that $p'(\ell') = \delta^2 c'(\ell'/n^*) > 0$ and $p''(\ell') = \delta^2 c''(\ell')/n^* > 0$ for all $\ell' \in N$. Either case gives that $p'(\ell) > 0$ and $p''(\ell) > 0$ for $\ell \in (\hat{\ell}, 2\hat{\ell})$ almost everywhere.

Repeating the process, we can conclude that $p'(\ell) > 0$ and $p''(\ell) > 0$ for almost all $\ell \in [0, 2^k\hat{\ell}]$. The conclusion follows by large enough k. □

---

[38] This can be seen by drawing the plot. $n^*(\ell)$ is not "strictly" increasing since the minimum grid may be the same as ℓ increases.

Bibliography

Acemoglu, D., U. Akcigit, and W. Kerr (2016a): "Networks and the Macroeconomy: An Empirical Exploration," *Nber Macroeconomics Annual*, 30, 273–335.

Acemoglu, D. and P. D. Azar (2020): "Endogenous Production Networks," *Econometrica*, 88, 33–82.

Acemoglu, D., V. M. Carvalho, A. Ozdaglar, and A. Tahbaz-Salehi (2012): "The Network Origins of Aggregate Fluctuations," *Econometrica*, 80, 1977–2016.

Acemoglu, D., A. Ozdaglar, and A. Tahbaz-Salehi (2015a): "Systemic Risk and Stability in Financial Networks," *American Economic Review*, 105, 564–608.

——— (2015b): "Systemic Risk in Endogenous Financial Networks," *Columbia business school research paper*.

——— (2016b): "Networks, Shocks, and Systemic Risk," in *The Oxford Handbook of the Economics of Networks*, ed. by Y. Bramoullé, A. Galeotti, and B. W. Rogers, Oxford University Press.

——— (2017): "Microeconomic Origins of Macroeconomic Tail Risks," *American Economic Review*, 107, 54–108.

Aggarwal, S. (2017): "Smile Curve and Its Linkages with Global Value Chains," *Journal of Economics Bibliography*, 4, 278–286.

Aizcorbe, A. M. (1992): "Procyclical Labour Productivity, Increasing Returns to Labour and Labour Hoarding in Car Assembly Plant Employment," *The Economic Journal*, 102, 860–873.

Albuquerque, R., M. Eichenbaum, V. X. Luo, and S. Rebelo (2016): "Valuation Risk and Asset Pricing," *The Journal of Finance*, 71, 2861–2904.

Aliprantis, C. D. and K. C. Border (2006): *Infinite Dimensional Analysis*, Springer.

Almsafir, M. K., I. A. Nassar, M. H. Al-Mahrouq, and J. A. Hayajneh (2015): "The Validity of Gibrat's Law: Evidence from the Service Sector in Jordan," *Procedia*

Economics and Finance, 23, 1602–1606.

Almus, M. (2000): “Testing" Gibrat’s Law" for Young Firms-empirical Results for West Germany,” *Small Business Economics*, 15, 1–12.

Amini, H., D. Filipović, and A. Minca (2016): “Uniqueness of Equilibrium in a Payment System with Liquidation Costs,” *Operations Research Letters*, 44, 1–5.

Amiti, M., S. J. Redding, and D. E. Weinstein (2019): “The Impact of the 2018 Tariffs on Prices and Welfare,” *Journal of Economic Perspectives*, 33, 187–210.

Antoun de Almeida, L. (2016): “Globalization of Inflation and Input-Output Linkages,” *University of Frankfurt mimeograph.*

Antràs, P. and D. Chor (2013): “Organizing the Global Value Chain,” *Econometrica*, 81, 2127–2204.

Antràs, P., D. Chor, T. Fally, and R. Hillberry (2012): “Measuring the Upstreamness of Production and Trade Flows,” *American Economic Review*, 102, 412–16.

Antràs, P. and D. Chor (2018): “On the Measurement of Upstreamness and Downstreamness in Global Value Chains,” Working Paper 24185, National Bureau of Economic Research.

Aral, S., Y. Bakos, and E. Brynjolfsson (2018): “Information Technology, Repeated Contracts, and the Number of Suppliers,” *Management Science*, 64, 592–612.

Arellano, C. (2008): “Default Risk and Income Fluctuations in Emerging Economies,” *American Economic Review*, 98, 690–712.

Auer, R. A., C. Borio, and A. Filardo (2017): “The Globalisation of Inflation: The Growing Importance of Global Value Chains,” CESifo Working Paper Series 6387, CESifo.

Auer, R. A., A. A. Levchenko, and P. Sauré (2019): “International Inflation Spillovers through Input Linkages,” *Review of Economics and Statistics*, 101, 507–521.

Axtell, R. L. (2001): “Zipf Distribution of US Firm Sizes,” *Science*, 293, 1818–1820.

Aydogan, Y. and M. Donduran (2019): “Concluding Gibrat’s Law with Turkish Firm Data,” *Physica A: Statistical Mechanics and its Applications*, 533, 122066.

Azimzadeh, P. (2019): “A Fast and Stable Test to Check if a Weakly Diagonally Dominant Matrix is a Nonsingular M-matrix,” *Mathematics of Computation*, 88, 783–800.

Backus, D., A. Ferriere, and S. Zin (2015): “Risk and Ambiguity in Models of Business Cycles,” *Journal of Monetary Economics*, 69, 42–63.

Balbus, L., W. Olszewski, K. Reffett, and L. P. Wozny (2022): “Iterative Monotone Comparative Statics,” Available at SSRN: `https://ssrn.com/abstract=4039543` or `http://dx.doi.org/10.2139/ssrn.4039543`.

Baldwin, R. and A. J. Venables (2013): “Spiders and Snakes: Offshoring and Agglomeration in the Global Economy,” *Journal of International Economics*, 90, 245–254.

Ballester, C., A. Calvó-Armengol, and Y. Zenou (2004): “Who’s Who in Crime Network. Wanted: The Key Player,” Tech. rep., IUI Working Paper.

——— (2006): “Who’s Who in Networks. Wanted: The Key Player,” *Econometrica*, 74, 1403–1417.

Baqaee, D. R. (2018): “Cascading Failures in Production Networks,” *Econometrica*, 86, 1819–1838.

Barrot, J.-N. and J. Sauvagnat (2016): “Input Specificity and the Propagation of Idiosyncratic Shocks in Production Networks,” *The Quarterly Journal of Economics*, 131, 1543–1592.

Bartelme, D. and Y. Gorodnichenko (2015): “Linkages and Economic Development,” Working Paper 21251, National Bureau of Economic Research.

Barua, S. (2020): “COVID-19 Pandemic and World Trade: Some Analytical Notes,” Available at SSRN: `https://ssrn.com/abstract=3577627` or `http://dx.doi.org/10.2139/ssrn.3577627`.

Bäuerle, N. and A. Jaśkiewicz (2018): “Stochastic Optimal Growth Model with Risk Sensitive Preferences,” *Journal of Economic Theory*, 173, 181–200.

Becker, G. S. and C. B. Mulligan (1997): “The Endogenous Determination of Time Preference,” *The Quarterly Journal of Economics*, 112, 729–758.

Berman, A. and R. J. Plemmons (1994): *Nonnegative Matrices in the Mathematical Sciences*, SIAM.

Bernard, A. B., A. Moxnes, and Y. U. Saito (2019): “Production Networks, Geography, and Firm Performance,” *Journal of Political Economy*, 127, 639–688.

Bertsekas, D. (2022): *Abstract Dynamic Programming*, Athena Scientific.

BERTSEKAS, D. P. AND J. N. TSITSIKLIS (1995): "Neuro-dynamic Programming: an Overview," in *Proceedings of 1995 34th IEEE Conference on Decision and Control*, IEEE, vol. 1, 560–564.

BIGIO, S. AND J. LA'O (2016): "Distortions in Production Networks," Working Paper 22212, National Bureau of Economic Research.

BILGIN, M. AND K. YILMAZ (2018): "Producer Price Inflation Connectedness and Input-Output Networks," Available at SSRN: `https://ssrn.com/abstract=3244645` or `http://dx.doi.org/10.2139/ssrn.3244645`.

BLOISE, G., C. LE VAN, AND Y. VAILAKIS (2024): "Do Not Blame Bellman: It Is Koopmans' Fault," *Econometrica*, 92, 111–140.

BLUME, L. E., W. A. BROCK, S. N. DURLAUF, AND R. JAYARAMAN (2015): "Linear Social Interactions Models," *Journal of Political Economy*, 123, 444–496.

BODENSTEIN, M. (2011): "Closing Large Open Economy Models," *Journal of International Economics*, 84, 160–177.

——— (2013): "Equilibrium Stability in Open Economy Models," *Journal of Macroeconomics*, 35, 1–13.

BOEHM, J. AND E. OBERFIELD (2020): "Misallocation in the Market for Inputs: Enforcement and the Organization of Production," *The Quarterly Journal of Economics*, 135, 2007–2058.

BOROVIČKA, J. AND J. STACHURSKI (2020): "Necessary and Sufficient Conditions for Existence and Uniqueness of Recursive Utilities," *The Journal of Finance*, 75, 1457–1493.

BÜHLER, T. AND D. A. SALAMON (2018): *Functional Analysis*, vol. 191, American Mathematical Soc.

CALVANO, E., G. CALZOLARI, V. DENICOLO, AND S. PASTORELLO (2020): "Artificial Intelligence, Algorithmic Pricing, and Collusion," *American Economic Review*, 110, 3267–3297.

CALVO, J. L. (2006): "Testing Gibrat's Law for Small, Young and Innovating Firms," *Small Business Economics*, 26, 117–123.

Calvó-Armengol, A., E. Patacchini, and Y. Zenou (2009): "Peer Effects and Social Networks in Education," *The Review of Economic Studies*, 76, 1239–1267.

Campbell, J. Y. and J. Ammer (1993): "What Moves the Stock and Bond Markets? A Variance Decomposition for Long-term Asset Returns," *The Journal of Finance*, 48, 3–37.

Carvalho, V. M. (2008): *Aggregate Fluctuations and the Network Structure of Intersectoral Trade*, The University of Chicago.

——— (2014): "From Micro to Macro via Production Networks," *Journal of Economic Perspectives*, 28, 23–48.

Carvalho, V. M. and B. Grassi (2019): "Large Firm Dynamics and the Business Cycle," *American Economic Review*, 109, 1375–1425.

Carvalho, V. M. and A. Tahbaz-Salehi (2019): "Production Networks: A Primer," *Annual Review of Economics*, 11, 635–663.

Charpentier, A., R. Elie, and C. Remlinger (2021): "Reinforcement Learning in Economics and Finance," *Computational Economics*, 1–38.

Chen, B. (2017): "Upstreamness, Exports, and Wage Inequality: Evidence from Chinese Manufacturing Data," *Journal of Asian Economics*, 48, 66–74.

Cheney, E. W. (2001): *Analysis for Applied Mathematics*, vol. 1, Springer.

Choi, H., N. C. Mark, and D. Sul (2008): "Endogenous Discounting, the World Saving Glut and the US Current Account," *Journal of International Economics*, 75, 30–53.

Christensen, T. M. (2022): "Existence and Uniqueness of Recursive Utilities without Boundedness," *Journal of Economic Theory*, 200, 105413.

Cifuentes, R., G. Ferrucci, and H. S. Shin (2005): "Liquidity Risk and Contagion," *Journal of the European Economic Association*, 3, 556–566.

Coase, R. H. (1937): "The Nature of The Firm," *Economica*, 4, 386–405.

Cochrane, J. (2009): *Asset Pricing: Revised Edition*, Princeton University Press.

Cochrane, J. H. (2011): "Presidential Address: Discount Rates," *The Journal of Finance*, 66, 1047–1108.

COHEN, J., K. M. ERICSON, D. LAIBSON, AND J. M. WHITE (2020): "Measuring Time Preferences," *Journal of Economic Literature*, 58, 299–347.

COHEN-COLE, E., E. PATACCHINI, AND Y. ZENOU (2015): "Static and dynamic networks in interbank markets," *Network Science*, 3, 98–123.

COSTINOT, A., J. VOGEL, AND S. WANG (2012): "Global Supply Chains and Wage Inequality," *American Economic Review*, 102, 396–401.

——— (2013): "An Elementary Theory of Global Supply Chains," *Review of Economic Studies*, 80, 109–144.

DAHLMAN, C. J. (1979): "The Problem of Externality," *The Journal of Law and Economics*, 22, 141–162.

DAS, K. (2020): "Impact of COVID-19 Pandemic Into Solar Energy Generation Sector," Available at SSRN: https://ssrn.com/abstract=3580341 or http://dx.doi.org/10.2139/ssrn.3580341.

DAS, K., S. SAMANTA, AND M. PAL (2018): "Study on centrality measures in social networks: a survey," *Social network analysis and mining*, 8, 1–11.

DAUNFELDT, S.-O. AND N. ELERT (2013): "When is Gibrat's Law a Law?" *Small Business Economics*, 41, 133–147.

DU, Y. (1990): "Fixed Points of Increasing Operators in Ordered Banach Spaces and Applications," *Applicable Analysis*, 38, 1–20.

DUNNE, P. AND A. HUGHES (1994): "Age, Size, Growth and Survival: UK Companies in the 1980s," *The Journal of Industrial Economics*, 42, 115–140.

DURDU, C. B., E. G. MENDOZA, AND M. E. TERRONES (2009): "Precautionary Demand for Foreign Assets in Sudden Stop Economies: An Assessment of the New Mercantilism," *Journal of Development Economics*, 89, 194–209.

DUTTA, D. AND Y. YANG (2013): "Endogenous Time Preference: Evidence from Australian Households' Behaviour," School of Economics, The University of Sydney, Available at http://hdl.handle.net/2123/9265.

EISENBERG, L. AND T. H. NOE (2001): "Systemic Risk in Financial Systems," *Management Science*, 47, 236–249.

Elliott, M., B. Golub, and M. O. Jackson (2014): “Financial Networks and Contagion,” *American Economic Review*, 104, 3115–53.

ElMaraghy, H., G. Schuh, W. ElMaraghy, F. Piller, P. Schönsleben, M. Tseng, and A. Bernard (2013): “Product Variety Management,” *CIRP Annals*, 62, 629 – 652.

Elsinger, H., A. Lehar, and M. Summer (2005): “Using Market Information for Banking System Risk Assessment,” Available at SSRN: `https://ssrn.com/abstract=787929` or `http://dx.doi.org/10.2139/ssrn.787929`.

——— (2006): “Risk Assessment for Banking Systems,” *Management Science*, 52, 1301–1314.

Epstein, L. G. and J. A. Hynes (1983): “The Rate of Time Preference and Dynamic Economic Analysis,” *Journal of Political Economy*, 91, 611–635.

Erol, S., C. Le Van, and C. Saglam (2011): “Existence, Optimality and Dynamics of Equilibria with Endogenous Time Preference,” *Journal of Mathematical Economics*, 47, 170–179.

Evans, D. S. (1987a): “The Relationship Between Firm Growth, Size, and Age: Estimates for 100 Manufacturing Industries,” *The Journal of Industrial Economics*, 567–581.

——— (1987b): “Tests of Alternative Theories of Firm Growth,” *Journal of Political Economy*, 95, 657–674.

Fajgelbaum, P. D., P. K. Goldberg, P. J. Kennedy, and A. K. Khandelwal (2020): “The Return To Protectionism,” *The Quarterly Journal of Economics*, 135, 1–55.

Falk, M. (2008): “Testing Gibrat’s Law for European Multinational Enterprises,” FIW-Research Reports 014, FIW - Research Centre International Economics.

Fally, T. and R. Hillberry (2018): “A Coasian Model of International Production Chains,” *Journal of International Economics*, 114, 299 – 315.

Fattorini, L., A. Rungi, and Z. Zhu (2017): “The Organization of Global Supply Networks,” IMT Lucca EIC Working Paper Series No. 08/2017, Available at SSRN: `https://ssrn.com/abstract=3081744` or `http://dx.doi.org/10.2139/ssrn.3081744`.

FLETCHER, J. E. (1989): "Input-output Analysis and Tourism Impact Studies," *Annals of Tourism Research*, 16, 514–529.

FU, D., F. PAMMOLLI, S. V. BULDYREV, M. RICCABONI, K. MATIA, K. YAMASAKI, AND H. E. STANLEY (2005): "The Growth of Business Firms: Theoretical Framework and Empirical Evidence," *Proceedings of the National Academy of Sciences*, 102, 18801–18806.

FUJIWARA, Y. AND H. AOYAMA (2010): "Large-scale Structure of a Nation-wide Production Network," *The European Physical Journal B*, 77, 565–580.

GABAIX, X. (2009): "Power Laws in Economics and Finance," *Annu. Rev. Econ.*, 1, 255–294.

——— (2011): "The Granular Origins of Aggregate Fluctuations," *Econometrica*, 79, 733–772.

GABAIX, X. AND A. LANDIER (2008): "Why Has CEO Pay Increased So Much?" *The Quarterly Journal of Economics*, 123, 49–100.

GAGLIARDI, N., B. MAHY, AND F. RYCX (2019): "Upstreamness, Wages and Gender: Equal Benefits for All?" *British Journal of Industrial Relations.*

GAI, P. AND S. KAPADIA (2019): "Networks and Systemic Risk in the Financial System," *Oxford Review of Economic Policy*, 35, 586–613.

GALEOTTI, A., B. GOLUB, AND S. GOYAL (2020): "Targeting Interventions in Networks," *Econometrica*, 88, 2445–2471.

GIBRAT, R. (1931): "Les inégalits économiques," *Sirey.*

GLASSERMAN, P. AND H. P. YOUNG (2015): "How Likely is Contagion in Financial Networks?" *Journal of Banking & Finance*, 50, 383–399.

——— (2016): "Contagion in Financial Networks," *Journal of Economic Literature*, 54, 779–831.

GRAZZI, M. AND D. MOSCHELLA (2018): "Small, Young, and Exporters: New Evidence on the Determinants of Firm Growth," *Journal of Evolutionary Economics*, 28, 125–152.

GUEST, R. S. AND I. M. MCDONALD (2001): "How Uzawa Preferences Improve the Simulation Properties of the Small Open Economy Model," *Journal of Macroeconomics*, 23, 417–440.

GUILMI, C. D. AND Y. FUJIWARA (2020): “Does the Supply Network Shape the Firm Size Distribution? The Japanese Case,” CAMA Working Papers 2020-66, Centre for Applied Macroeconomic Analysis, Crawford School of Public Policy, The Australian National University.

HALL, B. H. (1987): “The Relationship Between Firm Size and Firm Growth in the US Manufacturing Sector,” *The Journal of Industrial Economics*, 35, 583–606.

HANSEN, L. P. AND E. RENAULT (2010): “Pricing Kernels,” *Encyclopedia of Quantitative Finance.*

HANSEN, L. P. AND T. J. SARGENT (1995): “Discounted Linear Exponential Quadratic Gaussian Control,” *IEEE Transactions on Automatic control*, 40, 968–971.

HARRISON, J. M. AND D. M. KREPS (1978): “Speculative Investor Behavior in a Stock Market with Heterogeneous Expectations,” *The Quarterly Journal of Economics*, 92, 323–336.

HASHIMZADE, N., O. KIRSANOV, AND T. KIRSANOVA (2023): “Distributional Effects of Endogenous Discounting,” *Mathematical Social Sciences*, 122, 1–6.

HASSELT, H. (2010): “Double Q-learning,” *Advances in Neural Information Processing Systems*, 23.

HATCHONDO, J. C., L. MARTINEZ, AND H. SAPRIZA (2009): “Heterogeneous Borrowers in Quantitative Models of Sovereign Default,” *International Economic Review*, 50, 1129–1151.

HATCHONDO, J. C., L. MARTINEZ, AND C. SOSA-PADILLA (2016): “Debt Dilution and Sovereign Default Risk,” *Journal of Political Economy*, 124, 1383–1422.

HELPMAN, E. AND A. RAZIN (1982): “Dynamics of a Floating Exchange Rate Regime,” *Journal of Political Economy*, 90, 728–754.

HERNÁNDEZ-LERMA, O. AND J. B. LASSERRE (2012a): *Discrete-time Markov Control Processes: Basic Optimality Criteria*, vol. 30, Springer Science & Business Media.

——— (2012b): *Further Topics on Discrete-time Markov Control Processes*, vol. 42, Springer Science & Business Media.

HERSKOVIC, B. (2018): “Networks in Production: Asset Pricing Implications,” *The Journal of Finance*, 73, 1785–1818.

Herskovic, B., B. Kelly, H. Lustig, and S. Van Nieuwerburgh (2020): "Firm Volatility in Granular Networks," *Journal of Political Economy*, 128, 4097–4162.

Hills, T. S. and T. Nakata (2018): "Fiscal Multipliers at the Zero Lower Bound: the Role of Policy Inertia," *Journal of Money, Credit and Banking*, 50, 155–172.

Hills, T. S., T. Nakata, and S. Schmidt (2019): "Effective Lower Bound Risk," *European Economic Review*, 120, 103321.

Hiromitsu, G. and S. Wataru (2020): "Macroscopic Structure and Evolution in the Japanese Production Network," Discussion papers 20049, Research Institute of Economy, Trade and Industry (RIETI).

Hu, S., J. Ko, L. Weyand, H. ElMaraghy, T. Lien, Y. Koren, H. Bley, G. Chryssolouris, N. Nasr, and M. Shpitalni (2011): "Assembly System Design and Operations for Product Variety," *CIRP Annals*, 60, 715 – 733.

Hu, S., X. Zhu, H. Wang, and Y. Koren (2008): "Product Variety and Manufacturing Complexity in Assembly Systems and Supply Chains," *CIRP Annals*, 57, 45 – 48.

Hubmer, J., P. Krusell, and A. A. Smith Jr (2021): "Sources of US Wealth Inequality: Past, Present, and Future," *NBER Macroeconomics Annual*, 35, 391–455.

Huffman, D., R. Maurer, and O. S. Mitchell (2019): "Time Discounting and Economic Decision-making in the Older Population," *The Journal of the Economics of Ageing*, 14, 100121.

Hurd, T. R. (2016): *Contagion!: Systemic Risk in Financial Networks*, Springer.

Ijiri, Y. and H. A. Simon (1964): "Business Firm Growth and Size," *The American Economic Review*, 54, 77–89.

Ito, T. and P.-L. Vézina (2016): "Production Fragmentation, Upstreamness, and Value Added: Evidence from Factory Asia 1990-2005," *Journal of the Japanese and International Economies*, 42, 1–9.

Izadi, H. and M. S. Lamsoo (2022): "The Role of the Frisch Elasticity on Households' Behavior Using the Endogenous Discount Factor Model," *Çankırı Karatekin Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 12, 251–267.

Jaakkola, T., M. Jordan, and S. Singh (1993): "Convergence of Stochastic Iterative Dynamic Programming Algorithms," *Advances in Neural Information Processing*

Systems, 6.

Jackson, M. O. and A. Pernoud (2020): "Credit freezes, equilibrium multiplicity, and optimal bailouts in financial networks," *ArXiv Preprint ArXiv:2012.12861*.

Jasso-Fuentes, H., R. R. López-Martínez, and J. A. Minjárez-Sosa (2022): "Some Advances on Constrained Markov Decision Processes in Borel Spaces with Random State-dependent Discount Factors," *Optimization*, 1–27.

Jin, Y. and J. K. Ryan (2012): "Price and Service Competition in an Outsourced Supply Chain," *Production and Operations Management*, 21, 331–344.

Justiniano, A. and G. E. Primiceri (2008): "The Time-varying Volatility of Macroeconomic Fluctuations," *American Economic Review*, 98, 604–641.

Kamber, G. and B. Wong (2020): "Global Factors and Trend Inflation," *Journal of International Economics*, 122, 103265.

Kikuchi, T., K. Nishimura, and J. Stachurski (2018): "Span of Control, Transaction Costs, and the Structure of Production Chains," *Theoretical Economics*, 13, 729–760.

Kikuchi, T., K. Nishimura, J. Stachurski, and J. Zhang (2021): "Coase meets Bellman: Dynamic Programming for Production Networks," *Journal of Economic Theory*, 196, 105287.

Krasnosel'skii, M. A., G. M. Vainikko, R. Zabreyko, Y. B. Ruticki, and V. V. Stet'senko (2012): *Approximate Solution of Operator Equations*, Springer Science & Business Media.

Lawrance, E. C. (1991): "Poverty and the Rate of Time Preference: Evidence from Panel Data," *Journal of Political Economy*, 99, 54–77.

Levine, D. K. (2012): "Production Chains," *Review of Economic Dynamics*, 15, 271–282.

Liu, A., M. Paddrik, S. Y. Yang, and X. Zhang (2020): "Interbank Contagion: An Agent-based Model Approach to Endogenously Formed Networks," *Journal of Banking & Finance*, 112, 105191.

Long, J. B. and C. I. Plosser (1983): "Real Business Cycles," *Journal of Political Economy*, 91, 39–69.

Lotti, F., E. Santarelli, and M. Vivarelli (2003): "Does Gibrat's Law Hold Among Young, Small Firms?" *Journal of Evolutionary Economics*, 13, 213–235.

Lucas, R. E. (1978): "On the Size Distribution of Business Firms," *The Bell Journal of Economics*, 9, 508–523.

Lucas Jr, R. E. (1978): "Asset Prices in an Exchange Economy," *Econometrica: Journal of the Econometric Society*, 1429–1445.

Ma, Q., J. Stachurski, and A. A. Toda (2022): "Unbounded Dynamic Programming via the Q-transform," *Journal of Mathematical Economics*, 100, 102652.

Maeda, A. and M. Nagaya (2023): "Exhaustible Resource Use Under Endogenous Time Preference," *International Journal of Economic Policy Studies*, 17, 223–248.

Mahy, B., F. Rycx, G. Vermeylen, and M. Volral (2019): "Productivity, Wages and Profits: Does Firms' Position in the Value Chain Matter?" IZA Discussion Paper No. 12795, Available at SSRN: `https://ssrn.com/abstract=3495772` or `http://dx.doi.org/10.2139/ssrn.3495772`.

Marinacci, M. and L. Montrucchio (2010): "Unique Solutions for Stochastic Recursive Utilities," *Journal of Economic Theory*, 145, 1776–1804.

Melo, F. S. (2001): "Convergence of Q-learning: A Simple Proof," *Institute Of Systems and Robotics, Tech. Rep*, 1–4.

Mendoza, E. G. (1991): "Real Business Cycles in a Small Open Economy," *The American Economic Review*, 797–818.

Miller, R. E. and P. D. Blair (2009): *Input-output Analysis: Foundations and Extensions*, Cambridge University Press.

Miller, R. E. and U. Temurshoev (2017): "Output Upstreamness and Input Downstreamness of Industries/countries in World Production," *International Regional Science Review*, 40, 443–475.

Minjárez-Sosa, J. A. (2015): "Markov Control Models with Unknown Random State-Action-Dependent Discount Factors," *Top*, 23, 743–772.

Nakata, T. (2016): "Optimal Fiscal and Monetary Policy with Occasionally Binding Zero Bound Constraints," *Journal of Economic Dynamics and Control*, 73, 220–240.

Neuneier, R. (1997): “Enhancing Q-learning for Optimal Asset Allocation,” *Advances in Neural Information Processing Systems*, 10.

Obstfeld, M. (1990): “Intertemporal Dependence, Impatience, and Dynamics,” *Journal of Monetary Economics*, 26, 45–75.

Oke, B. (2018): “Testing the Gibrat’s Law: A Dynamic Panel Analysis of Firm Size and Firm Growth Nexus in Nigeria,” *Nigerian Journal of Management Sciences*, 18, 78–89.

Olszewski, W. (2021): “On Sequences of Iterations of Increasing and Continuous Mappings on Complete Lattices,” *Games and Economic Behavior*, 126, 453–459.

Park, D. and D. Ryu (2022): “Supply Chain Ethics and Transparency: An Agent-Based Model Approach with Q-learning Agents,” *Managerial and Decision Economics*, 43, 3331–3337.

Pesaran, M. H. and C. F. Yang (2020): “Econometric Analysis of Production Networks with Dominant Units,” *Journal of Econometrics*, 219, 507–541.

Ren, G. and J. Stachurski (2021): “Dynamic Programming with Value Convexity,” *Automatica*, 130, 109641.

Rogers, L. C. and L. A. Veraart (2013): “Failure and Rescue in an Interbank Network,” *Management Science*, 59, 882–898.

Rosenberg, J. V. and R. F. Engle (2002): “Empirical Pricing Kernels,” *Journal of Financial Economics*, 64, 341–372.

Roukny, T., S. Battiston, and J. E. Stiglitz (2018): “Interconnectedness as a source of uncertainty in systemic risk,” *Journal of Financial Stability*, 35, 93–106.

Rungi, A. and D. Del Prete (2018): “The Smile Curve at the Firm Level: Where Value is Added Along Supply Chains,” *Economics Letters*, 164, 38–42.

Saito, Y. U., T. Watanabe, and M. Iwamura (2007): “Do Larger Firms Have More Interfirm Relationships?” *Physica A: Statistical Mechanics and its Applications*, 383, 158–163.

Samwick, A. A. (1998): “Discount Rate Heterogeneity and Social Security Reform,” *Journal of Development Economics*, 57, 117–146.

Sargent, T. J. and J. Stachurski (2023): “Dynamic Programming Volume 1,” QuantEcon, Available at https://dp.quantecon.org/ or https://github.com/

QuantEcon/book-dp1.

Schmitt-Grohé, S. and M. Uribe (2003): "Closing Small Open Economy Models," *Journal of International Economics*, 61, 163–185.

Schneider, C. and R. Veugelers (2010): "On Young Highly Innovative Companies: Why they Matter and How (Not) to Policy Support Them," *Industrial and Corporate change*, 19, 969–1007.

Schreyer, P. (2000): "High-Growth Firms and Employment," OECD Science, Technology and Industry Working Papers 2000/3, OECD Publishing.

Sharkey, K. J. (2017): "A Control Analysis Perspective on Katz Centrality," *Scientific Reports*, 7, 1–8.

Sharma, A., R. Gupta, K. Lakshmanan, and A. Gupta (2021): "Transition Based Discount Factor for Model Free Algorithms in Reinforcement Learning," *Symmetry*, 13, 1197.

Simon, H. A. and C. P. Bonini (1958): "The Size Distribution of Business Firms," *The American Economic Review*, 607–617.

Singh, S., T. Jaakkola, M. L. Littman, and C. Szepesvári (2000): "Convergence Results for Single-step On-policy Reinforcement-learning Algorithms," *Machine learning*, 38, 287–308.

Stachurski, J. (2022): "Systemic Risk in Financial Systems: Properties of Equilibria," *ArXiv Preprint ArXiv:2202.11183*.

Stachurski, J., O. Wilms, and J. Zhang (2022a): "Asset Pricing Models with Preference Shocks: Existence and Uniqueness," Available at SSRN: https://ssrn.com/abstract=4041393 or http://dx.doi.org/10.2139/ssrn.4041393.

——— (2022b): "Unique Solutions to Power-Transformed Affine Systems," *ArXiv Preprint ArXiv:2212.00275*.

Stachurski, J. and J. Zhang (2021): "Dynamic Programming with State-dependent Discounting," *Journal of Economic Theory*, 192, 105190.

Staum, J., M. Feng, and M. Liu (2016): "Systemic Risk Components in a Network Model of Contagion," *IIE Transactions*, 48, 501–510.

STERN, M. L. (2006): "Endogenous Time Preference and Optimal Growth," *Economic Theory*, 29, 49–70.

STOKEY, N. L. (1989): *Recursive Methods in Economic Dynamics*, Harvard University Press.

STÖLLINGER, R. (2019): "Testing the Smile Curve: Functional Specialisation in GVCs and Value Creation," working paper 163, Wiiw Working Paper.

SUTTON, J. (1997): "Gibrat's Legacy," *Journal of Economic Literature*, 35, 40–59.

SZYMCZAK, S., J. WOLSZCZAK-DERLACZ, AND A. PARTEKA (2019): "Position in Global Value Chains: the Impact on Wages in Central and Eastern European Countries," Tech. rep., GUT FME Working Paper Series A.

TAHBAZ-SALEHI, A., A. OZDAGLAR, AND D. ACEMOGLU (2016): *Networks, Shocks, and Systemic Risk*, Oxford University Press, chap. 21, 569–608.

TANG, A. (2015): "Does Gibrat's Law Hold for Swedish Energy Firms?" *Empirical Economics*, 49, 659–674.

TIMMER, M. P., E. DIETZENBACHER, B. LOS, R. STEHRER, AND G. J. DE VRIES (2015): "An Illustrated User Guide to the World Input-output Database: the Case of Global Automotive Production," *Review of International Economics*, 23, 575–605.

TODA, A. A. (2021): "Perov's Contraction Principle and Dynamic Programming with Stochastic Discounting," *Operations Research Letters*, 49, 815–819.

——— (2023): "Unbounded Markov Dynamic Programming with Weighted Supremum Norm Perov Contractions," *ArXiv Preprint ArXiv:2310.04593*.

TSITSIKLIS, J. N. (1994): "Asynchronous Stochastic Approximation and Q-learning," *Machine learning*, 16, 185–202.

UZAWA, H. (1968): "Time Preference, the Consumption Function, and Optimum Asset Holdings," *Value, Capital and Growth: Papers in Honor of Sir John Hicks. The University of Edinburgh Press, Edinburgh*, 485–504.

VASILEV, A. (2022a): "A Real-Business-Cycle Model with Endogenous Discounting and a Government Sector," *Notas Económicas*, 73–86.

——— (2022b): "Um Modelo de Ciclos Económicos Reais com Taxa de Desconto Endógena e Estado," *Notas Económicas*, 71–84.

VERAART, L. A. M. (2020): "Distress and Default Contagion in Financial Networks," *Mathematical Finance*, 30, 705–737.

WALTMAN, L. AND U. KAYMAK (2008): "Q-learning Agents in a Cournot Oligopoly Model," *Journal of Economic Dynamics and Control*, 32, 3275–3293.

WANG, Z., S.-J. WEI, X. YU, AND K. ZHU (2017): "Characterizing Global Value Chains: Production Length and Upstreamness," Working Paper 23261, National Bureau of Economic Research.

WATKINS, C. J. AND P. DAYAN (1992): "Q-learning," *Machine learning*, 8, 279–292.

WATKINS, C. J. C. H. (1989): "Learning From Delayed Rewards," Ph.D. thesis, King's College, Cambridge United Kingdom.

WEI, Q. AND X. GUO (2011): "Markov Decision Processes with State-dependent Discount Factors and Unbounded Rewards/Costs," *Operations Research Letters*, 39, 369–374.

WEIL, P. (1993): "Precautionary Savings and the Permanent Income Hypothesis," *The Review of Economic Studies*, 60, 367–383.

WILLIAMSON, O. E. (1979): "Transaction-cost Economics: the Governance of Contractual Relations," *The journal of Law and Economics*, 22, 233–261.

WU, X. AND J. ZHANG (2016): "Finite Approximation of the first Passage Models for Discrete-time Markov Decision Processes with Varying Discount Factors," *Discrete Event Dynamic Systems*, 26, 669–683.

WU, X., X. ZOU, AND X. GUO (2015): "First Passage Markov Decision Processes with Constraints and Varying Discount Factors," *Frontiers of Mathematics in China*, 10, 1005–1023.

YOSHIDA, N., E. UCHIBE, AND K. DOYA (2013): "Reinforcement Learning with State-dependent Discount Factor," in *2013 IEEE Third Joint International Conference on Development and Learning and Epigenetic Robotics (ICDL)*, IEEE, 1–6.

YU, M. AND J. ZHANG (2019): "Equilibrium in Production Chains with Multiple Upstream Partners," *Journal of Mathematical Economics*, 83, 1–10.

YUE, V. Z. (2010): "Sovereign Default and Debt Renegotiation," *Journal of International Economics*, 80, 176–187.

Zenou, Y. (2012): "Networks in Economics," CEPR Discussion Papers 9021, C.E.P.R. Discussion Papers.

——— (2016): "Key Players," *Oxford Handbook on the Economics of Networks*, 244–274.

Zhang, Z. (2013): *Variational, Topological, and Partial Order Methods with Their Applications*, vol. 29 of *Developments in Mathematics*, Springer Berlin Heidelberg.

Zhou, L. (1994): "The Set of Nash Equilibria of a Supermodular Game is a Complete Lattice," *Games and Economic Behavior*, 7, 295–300.