UNIVERSITÀ CA' FOSCARI DI VENEZIA

SCUOLA DOTTORALE DI ATENEO — GRADUATE SCHOOL

VENETO GRADUATE SCHOOL OF ECONOMICS AND MANAGEMENT

Dottorato di Ricerca in Economia, Ciclo XXVII

Anno di Discussione: 2015

GROUP STRUCTURE AND COLLECTIVE BEHAVIOR

SETTORE SCIENTIFICO DISCIPLINARE DI AFFERENZA: SECS-P/01, SECS-S/06

Tesi di Dottorato di **SHILEI WANG**, Matricola 955949

Coordinatore del Dottorato

Michele Bernasconi

Tutore del Dottorando

Michele Bernasconi

A dissertation submitted to the Graduate School of Ca' Foscari University of Venice, and the Veneto Graduate School of Economics and Management of the Universities of Ca' Foscari, Padua, and Verona in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Economics.

To nothing, the mother of anything

Preface

The present dissertation is the outcome of studies on group structure and its effects on collective behavior. Its underlying concern is that individual behaviors do not merely depend on self-interested motives, but group structure should be equally important, as if, in some extreme situations, individual intentions even did not take any effect at all. Although the main topic of the dissertation should be well-defined, different chapters might still seem diverse for some readers. However, I wish people would not ask for its completeness and consistency as much as a book.

The first part of the dissertation including Chapter 1 and 2 investigates group structures in a more mathematical sense. Chapter 1 studies a time-dependent group which can be formalized as a dynamical system, and in particular, discusses its dynamics, stability, chaos, and some other properties in great detail. Chapter 2 studies a preference-dependent group whose welfare states can be characterized by some aggregation rule, and develops a number of assertions on its separable and additive representations. The second part containing Chapter 3 and 4 is devoted to studies on concrete groups which emerge institutionally from our society, *viz.*, the centralized limit order market with its roots in equity markets, and the hierarchical government network with its roots in political systems and organizations.

Some reported results in this dissertation have already been published, although they entailed minor revisions to maintain a higher typographical quality. Concretely, Section 3 and 4 of Chapter 1 were once integrated into an independent article, and published as a preprint on arXiv; Chapter 3 was published as a preprint on arXiv, and later as a research paper in the journal Algorithmic Finance (Wang [**66**]).

It might be noted that these chapter headings are shortened so as to reflect the main "objects" of the studies. To be more precise, each chapter had ever longer title when being prepared initially, that is, "*On the iterative nature of economic dynamics*" for Chapter 1, "*On the axiom of separability*" for Chapter 2, "*Dynamical trading mechanism in limit order markets*" for Chapter 3, and "*A theory of local public good provision in distributed political systems*" and "*Complex interactions in large government networks*" for Chapter 4.

In the period of preparing this dissertation, there are people I am deeply indebted to, and some I even did not realize clearly at this stage. For this reason, I hesitate to write all their names here. But people who naturally own something never claim that fact, do they? It surely applies here, and would help me do this usually difficult task in a preface.

Venice, November 2014 *S. Wang*

Contents

CHAPTER 1

On Dynamics and Chaos

We begin with a treatment of dynamical system, as an abstract representation of time-dependent group.

1. Introduction

1.1. Determinism and Chaos. In economic science, studies are framed a lot by the philosophy of determinism, or more precisely, by causal determinism in terms of Laplace's demon. For instance, economists describe individuals as rational agents, whose choices and social behaviors are determined by a set of fixed economic laws; the motion of an economic system is either totally determined by a representative agent, or deterministic as well by the belief that the economic system would be essentially rational in the whole. The only indeterminacy is usually described by the vague concept randomness, and people try to add it to the economic system so as to have a more general sense of determinism.

In this consideration, if there is a cause c, there will definitely be an effect e, simply as there exists a deterministic causal law L. But D. Hume argued that this kind of determinacy is effective only up to empirical induction. Again, suppose c is a cause, and one observes that c always leads to an effect e in the past, then one normally infers that c causes e, *i.e.*, there naïvely exists a causal law L, such that c causes e if and only if L is the causal law linking c and e. Unfortunately, our observations happen always in finite times, so the causal law L will not be logically true, but approximately or even empirically true. From this perspective, the causality which is so important in the philosophy of determinism should be thought of to be a metaphysical principle.

However, even if one had the logically true laws in a system, the outcomes might still not be deterministic. Suppose one has a set of fixed causal laws denoted by L again in a system, and there is a cause or an initial condition c, then by L one predicts that there will be a result $e = L(c)$. It might be noted that the cause c can be known and realized only as a measurable cause $m(c)$, which is approximately close to the true c, but there is no reason that $m(c) \equiv c$. Then one will predict a result $L\big(m(c)\big)$, which could be totally different from the true result $e = L(c)$, as the error between $L\big(m(c)\big)$ and $L(c)$ may not be linearly proportional to the difference between $m(c)$ and c. Consequently, the deterministic nature of a system may not generate fully predictable happenings in the future, or we can say a slight piece of imprecise measurability in a system could eventually make the system unpredictable at all.

Such a somewhat peculiar fact was realized first by H. Poincaré when he studied the three-body problem in the 1890s. As he wrote in the book "*Science and Method*" (Poincaré [**52**], p. 68):

> *But it is not always so; it may happen that small differences in the initial conditions produce very great ones in the final phenomena. A small error in the former will produce an enormous error in the latter. Prediction becomes impossible, and we have the fortuitous phenomenon.*

We can notice that this kind of (dynamical) instability in a system does not come from any stochastic or random factors, but from small measurable errors enlarged by certain deterministic laws. Such an unpredictable property of a system is called *deterministic chaos*, or simply just *chaos*.

Moreover, when one says there is a system with deterministic laws, it means that one holds a belief that she has a full knowledge of that system as it exists, and consistently also the chaotic properties it might have. When such a belief is neither perfect nor absent, one should be aware that her knowledge of the system is only partial. One then introduces the randomness in order to smooth the relation between her belief and the truth, in consequence, the *pseudo-chaos* as is mixed by chaos and randomness would conceptually emerge.

Suppose noises as a special form of randomness can be introduced intentionally into a chaotic system, so that only the laws of the system would be affected, but its states remain unchanged. Under some restricted assumptions, that system with noises could then be characterized by a certain stochastic process, for instance, an iterated random function system might be equivalent to some Markov chain, as will be discussed in Section 3.2 of this chapter. On the other hand, if introduced noises only affect the states in a chaotic system, but do not modify its laws, then the basic deterministic skeleton could be thought of to be stable as if the noises had not been introduced. To a certain extent, it seems however impossible to separate noises from chaos or vice versa, because noises entering at different times would adhere to the states and also evolve according to the laws of the system.

1.2. Order in Chaotic System. The order in a system can be attached to some basic properties of existent mathematical objects in that system, where the existence of an object means that it can be defined. For instance,

(i) a function should be differentiable on its domain;

(ii) the time domain, normally $\mathbb{Z}$ or $\mathbb{R}$, should be transitive, otherwise the dynamics driven by iteration would be impossible;

(iii) the state transition function $\varphi_t(x)$, which maps the initial state x into the state $\varphi_t(x)$ at t, should satisfy

$$\varphi_s(\varphi_t(x)) = \varphi_{s+t}(x), \quad \text{or} \quad \varphi_s \circ \varphi_t = \varphi_{s+t},$$

so that the state transition operator preserves the additive structure of the time domain.

Many well-defined systems can be thought of having an appropriate order in the above sense, because the laws that govern these systems are fairly normal in mathematical terms. However, with respect to the stability of these systems or their possible states after evolving for a sufficiently long time, there could still exist some mathematical objects not having a nice order, especially when the laws are nonlinear and can not be properly linearized.

Let us consider a system which has a set of fixed laws L, a state space X, and a time domain T such that the time t can be forward to $s > t$, or backward to $r < t$. The set of states $A \subseteq X$ in the system is called an *attractor*, if A is invariant under L with forward iteration, or $L(A) = A$, and there exists a neighborhood of A, $N(A)$ such that all state in $N(A)$ will converge to some state in A under L after a sufficiently long time, and moreover, any proper subset of A has no such properties. The essential property of the attractor A and its neighboring trajectories in $N(A)$ is stability. For any two close enough states $x, y \in N(A)$, the stability of A requires that these two iterated forward motions should be uniformly close to each other. Yet such property can not be satisfied by a lot of systems, and in these hence so-called unstable systems, the attractors will very likely demonstrate bizarre geometric representations. Such attractors are called *strange attractors*, or notably *fractals* as was coined by Mandelbrot [**41, 42**]. Nevertheless, there is still no necessary relation between instability and fractals, as some stable systems could also have fractals, and meanwhile, some unstable systems could have normal attractors (*cf.*, Milnor [**45**]).

That being stated, unstable systems can yet preserve an order which is partly indicated in some nice properties of their fractals. For instance,

(i) fractals fully embedded in the n-dimensional Euclidean space can have a Hausdorff-Besicovitch dimension between $n-1$ and n for $n \geq 1$, *e.g.*, the Koch curve has a dimension $\log 4/\log 3 \approx 1.262$ in the Euclidean plane;

(ii) fractals are continuous and nowhere differentiable;

(iii) most fractals are self-similar, and their scaled parts always have similar structures to the wholes.

1.3. Chaos in Social System. Unlike mathematically-defined systems, social systems, such as social groups, social networks, and organized markets, are complex by nature. In general, the governing laws for a social system can not be directly derived by aggregation of the laws for its agents, even if one could have a perfect knowledge of the social system. What's more, repeated interactions among many agents, learning to adapt their strategies, and self-organization pattern formations in a large social system will make its laws evolve themselves, and therefore agents' knowledge may not necessarily match the laws time by time. As a social system is now investigated from a complex perspective, its properties such like self-adaptation and self-regulation would turn out to be evidently important, and its different scales may show a disordered structure or even random pattern. If the disordered structure rooted in different scales could be ideally removed, then the relation between the whole social system and its parts would be stable, and thus the complex social system could be refined to a chaotic social system. The finer chaotic social system should be more easily represented by some mathematically-defined system, in case its social attributes could be formalized to some extent[1].

Example 1.1. The cobweb theorem was first studied by Leontief [**37**], Kaldor [**31**], and Ezekiel [**17**] in the 1930s, when they investigated the interactive dynamics of price and quantity in a single market. Let (p_t, q_t) denote the pair of price and quantity in a market at time t. Suppose the market is controlled by the demand law D and the supply law S in a delayed way, that is,

$$p_t = D(q_t), \quad q_{t+1} = S(p_t),$$

then $p_{t+1} = D\big(S(p_t)\big)$, or $p_{t+1} = C(p_t)$ for $C = D \circ S$.

Assume C takes the form of Logistic map, then

$$p_{t+1} = rp_t(1 - p_t) \quad (r > 0). \tag{1.1}$$

Suppose the price is normalized so that it is in the real interval $[0, 1]$. Notice that $p_t(1 - p_t) \leq 1/4$ for all $p_t \in [0, 1]$, so we need $r \in (0, 4]$ to ensure that p_t is always in $[0, 1]$. If $r \in (0, 1]$, p_t will converge to 0 and the good in the market will be free in the long term. If $r \in (1, 3]$, p_t will converge to either 0 or some stable price in $(0, 1)$. However, if $r \in (3, 4]$, the market will have bifurcations. The cardinality of its attractor can be $2^{n(r)}$, where $n(r) \in \mathbb{N}$ is a function of r. Once $n(r)$ is sufficiently large, the attractor would be almost dense on $[0, 1]$, and hence the market would be chaotic.

Example 1.2. The price dynamics in a speculative market was first studied by Bachelier [**3**]. One now knows that the fair price process is actually a martingale,

[1]There are many literatures on complex and chaotic social system in such fields of research like systems biology, swarm intelligence, and social psychology. Since our studies in this chapter will focus on economic and financial systems, they are not cited here, though some are quite insightful. Besides, the author should acknowledge one reviewer of this dissertation for suggesting literatures on dynamic psychological system.

so that there is no further arbitrage opportunity on the market at any time. In practice, it was initially assumed that the underlying price $P(t)$ moves approximately as a Brownian motion, *i.e.*, $P(t+dt) - P(t)$ is Gaussian with the mean 0 and the volatility $\sqrt{dt}$. Later on, Osborne [**50**] showed that $P(t)$ should follow a geometric Brownian motion, *i.e.*, $\log\big(P(t)\big)$ is a Brownian motion.

As one may know, a Brownian motion is almost surely continuous and nowhere differentiable, and thus its possible realization is essentially a fractal. Suppose there is a dynamical system with a price space $\mathbb{R}$ and a time domain $[0,+\infty)$, which has an attractor $\mathbb{R}$, then its price dynamics is *a priori* chaotic. A motion of the system can then be represented by a realization $p(t)$ of the stochastic process $P(t)$. Clearly, $p(t)$ should have some scaling properties, and might maintain a thick tail and a persistent long-memory volatility, as are constantly observed in real financial markets (see for example, Mandelbrot [**40**], and Calvet and Fisher [**11**]).

Recall that a Brownian motion $B(t)$ must be invariant under any scaling, that's to say, $B(\lambda t) = \sqrt{\lambda}B(t)$ for all $\lambda > 0$. A quite natural generalization for $B(t)$ is the fractional Brownian motion $B_h(t)$ for $h \in (0,1)$, as was proposed by Mandelbrot and Van Ness [**43**], such that $B_h(\lambda t) = \lambda^h B_h(t)$, and hence its volatility is equal to t^h. Another helpful consideration is to replace the real time t by a trading time $\tau(t)$, which potentially has different values at different real times.

In case $h = 1/2$ and $\tau(t) = t$, one then has the Brownian motion $B(t)$. In general, one would have a compound stochastic process $B_h\big(\tau(t)\big)$, which can hopefully serve as a bridge connecting randomness to chaos, with a fair ability to catch the patterns in the price dynamics.

2. Dynamical System

2.1. General Definition. Throughout this chapter, we will use $\mathbb{R}^+$ and $\mathbb{R}^-$ to mean the nonnegative and nonpositive real numbers, and use $\mathbb{Z}^+$ and $\mathbb{Z}^-$ to mean the nonnegative and nonpositive integers. For any sets X and Y, $X \times Y$ denotes their Cartesian product. The closure of a set X is denoted by $\mathrm{cl}(X)$. I denotes an index set, whose cardinality is at least 2 and at most $\aleph_0$.

Let X be a metric space with the metric $d : X \times X \to \mathbb{R}^+$, or simply let X be a subset of the n-dimensional Euclidean space. We call X a *state space*. Let T be a subset of $\mathbb{R}$. If $(T,+)$ is a semigroup, *i.e.*, $(r+s)+t = r+(s+t)$ for all $r,s,t \in T$, then we say T is a *time domain*. Usually, it can be $\mathbb{R}$, $\mathbb{Z}$, $\mathbb{R}^+$, or any interval in $\mathbb{R}$.

The abstract definitions of continuous dynamical system can be found in the books by Bhatia and Szegö [**8**], Krabs and Pickl [**33**], and Teschl [**61**]. We shall make the definition of general dynamical system in a similar way.

Definition. A *dynamical system* on X is a triplet (X,T,φ), which is governed by the function $\varphi : X \times T \to X$ with the properties:

(i) $\varphi(x, t_0) = x$ for all $x \in X$ and any initial time $t_0 \in T$,
(ii) $\varphi(\varphi(x, s), t) = \varphi(x, t + s)$ for all $x \in X$ and all $s, t \in T$.

The function φ in a dynamical system (X, T, φ) is called a *flow*. For any given state $x \in X$ we define a *motion* $\varphi_x : T \to X$ passing through x as $\varphi_x(t) = \varphi(x, t)$, and for any $t \in T$ we define a *transition* $\varphi_t : X \to X$ between the states at t_0 and t as $\varphi_t(x) = \varphi(x, t)$. Note that

$$\varphi(\varphi(x, s), t) = \varphi(\varphi_s(x), t) = \varphi_t(\varphi_s(x)),$$

and also $\varphi(x, t + s) = \varphi_{t+s}(x)$, thus $\varphi_t \circ \varphi_s = \varphi_{t+s}$ by property (ii) of the function φ in the above definition, and hence φ is a homomorphism between $(T, +)$ and $(X, \circ)$.

Suppose $T = \mathbb{R}$, then the motion $\varphi_x(t)$ passing through x at the initial time t_0 is a continuous function of t. We hence call the dynamical system a *continuous* dynamical system, and replace $\varphi_x(t)$ by $x(t)$. As for the transition function $\varphi_t(x)$, since T is now dense, we adopt its differential term as a measure for its movement. We thus define $\dot{x} = \partial\varphi_t(x)/\partial t = g(x, t)$, with the initial condition $x(t_0) = \varphi(x(t_0), t_0) = x_0$. Consequently, a continuous dynamical system can also be defined by an ordinary differential equation $\dot{x} = g(x, t)$ with $x(t_0) = x_0$.

Let W be an open subset of $X \times T$, and $(x_0, t_0) \in W$. If $g(x, t)$ is continuous on W, then by the Peano existence theorem, there exists at least one solution to make such a dynamical system work at a time domain $[t_0, t_1] \subseteq T$, where $t_1 > t_0$, or there is at least one function $y : [t_0, t_1] \to X$, such that $\dot{y}(t) = g(y(t), t)$ and $y(t_0) = x_0$. Moreover, if $g(x, t)$ is Lipschitz continuous in x, *i.e.*, there is a piecewise continuous function $\ell(t)$ such that for all $t \in T$ and all $x, y \in X$,

$$d\big(g(x, t) - g(y, t)\big) \leq \ell(t) d(x, y), \tag{1.2}$$

then by the Picard existence theorem, the solution $y(t)$ is also unique.

Suppose $T = \mathbb{Z}$, then the motion $\varphi_x(t)$ will be discrete, as it has values only at dispersed times. We hence call that system a *discrete* dynamical system, and replace $\varphi_x(t)$ by x_t. As for the transition rule, we take $t_0 = 0$, and define

$$f(x) = \varphi_1(x) = \varphi(x, 1),$$

then by the fact that $\varphi_t \circ \varphi_s = \varphi_{t+s}$, we have $\varphi_t(x) = f^t(x)$ for all $t \in \mathbb{Z}^+$. Here, f^t is defined by the recursion $f^t = f \circ f^{t-1}$ with $f^0 = \mathrm{id}_X$, where $t \in \mathbb{N}$ and id_X denotes the identity function on X. If $t \in \mathbb{Z}^-$, then $-t \in \mathbb{Z}^+$. Note that

$$\varphi_{-t}\big(\varphi_t(x)\big) = \varphi_{-t}\big(\varphi(x, t)\big) = \varphi\big(\varphi(x, t), -t\big) = \varphi(x, 0) = x,$$

so φ_t is the inverse function of φ_{-t}, and hence

$$\varphi_t = (\varphi_{-t})^{-1} = (f^{-t})^{-1} = (f^{-1})^{-t} = f^t,$$

which yields that $\varphi_t = f^t$ for all $t \in \mathbb{Z}$. The dynamics driven by φ_t is called *forward* for $t \in \mathbb{Z}^+$, and *backward* for $t \in \mathbb{Z}^-$.

Notice that a discrete dynamical system can be completely determined by the transition rule $f : X \to X$ between two consecutive time points. We then have the motion $x_t = f^t(x)$, where $t_0 = 0$ and $x_0 = \varphi(x_0, 0) = x$, and usually f is assumed to be continuous.

2.2. Basic Properties. In this subsection, we will use the notation $x\langle t\rangle$ in some definitions to denote a generic motion $\varphi_x(t)$, which can be either continuous as $x(t)$, or discrete as x_t.

Definition. For all deterministic x, the *trajectory* passing through x is the set of states $\gamma(x) = \{x\langle t\rangle : t \in T\}$.

Since T can be partitioned into $T^+ = T \cap \mathbb{R}^+$ and $T^- = T \cap \mathbb{R}^-$, $\gamma(x)$ can also be partitioned into two semi-trajectories,

$$\gamma^+(x) = \{x\langle t\rangle : t \in T^+\}, \quad \gamma^-(x) = \{x\langle t\rangle : t \in T^-\},$$

such that $\gamma^+(x) \cup \gamma^-(x) = \gamma(x)$ and $\gamma^+(x) \cap \gamma^-(x) \supseteq \{x\}$.

If $\gamma(x) = \{x\}$, x is called an *equilibrium state*. In terms of motion, if x is an equilibrium state, then $\varphi_x(t) = x\langle t\rangle = x$ for all $t \in T$. If a dynamical system starts from a state not in equilibrium, say y, then its trajectory $\gamma(y)$ can not contain any equilibrium state in a finite time. We can prove this claim by contradiction. Suppose an equilibrium state x enters into $\gamma(y)$ at a time t, which means $\varphi_y(t) = x$, then $\varphi_t(y) = x$, and thus $\varphi_{-t}(x) = y = \varphi_x(-t)$. But by the definition of equilibrium state, we should have $\varphi_x(t) = x$ for all $t \in T$, a contradiction. Therefore, the motion $\varphi_x(t)$ in a dynamical system either remains at the initial state x, or evolves forever.

Definition. A set of states $S \subseteq X$ is *invariant*, if $\varphi_t(S) = S$ for all $t \in T$.

Alternatively, we say S as a subset of X is invariant if $\gamma(x) \subseteq S$ for all $x \in S$. And these two definitions are actually equivalent. To see this fact, suppose $\varphi_t(S) = S$, then for all $x \in S$ we have $\varphi_t(x) \in S$ for all $t \in T$, so $\varphi_x(t) \in S$ for all t, and hence $\gamma(x) \subseteq S$ for all $x \in S$. On the other hand, suppose $\gamma(x) \subseteq S$ for all $x \in S$, but $\varphi_t(S) \neq S$, then there is at least one pair (y, t) for $y \in S$ with $\varphi_t(y) \notin S$, which hence implies that $\gamma(y) \nsubseteq S$, a contradiction.

Suppose S is an invariant set of states, and $y \notin S$, then $\gamma(y)$ will not move into S in a finite time, otherwise y should be contained in S. This fact is very similar to the reachability of an equilibrium state, but if we notice that an invariant set itself is an equilibrium set of states, it should be not surprising. Note that any singleton $\{x\}$ with an equilibrium state x is invariant, so if S is invariant, then $S \cup \{x\}$ will be invariant again. It is thus essential to identify the kernel of an invariant set S, which should be minimal. And in general, an invariant set S is called *minimal*, if S has no proper subset that is again invariant.

It can be shown that any bounded invariant set must have a kernel. In a discrete dynamical system, any invariant set has finite or countably infinite states, thus we can always find its minimally invariant kernel by backward induction. In a continuous dynamical system with a state space $X \subseteq \mathbb{R}^n$, any invariant set should be closed and hence compact because of its boundedness, thus by Zorn's Lemma it's not hard to see that its minimally invariant kernel also exists.

Definition. A motion $\varphi_x(t)$ passing through x is called *periodic*, if there exists some $s \neq 0$ such that $\varphi_x(t+s) = \varphi_x(t)$ for all $t \in T$.

If a motion $\varphi_x(t)$ is periodic, then the state x is called periodic as well, and its trajectory $\gamma(x)$ is called a *limit cycle*. The minimal $s \neq 0$ satisfying $\varphi_x(t+s) = \varphi_x(t)$ for all $t \in T$ is called the *period* of the motion $\varphi_x(t)$. In particular, if the period of $\varphi_x(t)$ is close enough to 0, then x will be nearly an equilibrium state, and if its period tends to be infinite, then the motion will be almost nonperiodic.

Definition. A state y is called an *ω-limit state*, if there is a sequence $(t_i,\ i \in I)$ where $t_i \in T$ for all $i \in I$, such that $\lim_{t_i \uparrow +\infty} x\langle t_i \rangle = y$, and similarly, a state z is called an *α-limit state*, if $\lim_{t_i \downarrow -\infty} x\langle t_i \rangle = z$.

The set of all the ω-limit states for a state x is called an *ω-limit set*, and we use $\omega(x)$ to denote it. The set of all the α-limit states for a state x is called an *α-limit set*, and we use $\alpha(x)$ to denote it.

Any ω-limit set must be invariant. To show this assertion, consider $\omega(x)$, then for all $y \in \omega(x)$ there must be a sequence $(t_i,\ i \in I)$ such that $\lim_{t_i \uparrow +\infty} x\langle t_i \rangle = y$, and thus for all $s \in T$ we have $\lim_{t_i \uparrow +\infty} \varphi_s(\varphi_{t_i}(x)) = \varphi_s(y)$, which directly yields that $\lim_{t_i \uparrow +\infty} x\langle t_i + s \rangle = \varphi_s(y)$. Since $(t_i + s,\ i \in I)$ is again a sequence in T, we have $\varphi_s(y) \in \omega(x)$ for all $s \in T$, and hence $\gamma(y) \subseteq \omega(x)$ for all $y \in \omega(x)$, which implies $\omega(x)$ is invariant by definition.

Also, we can define these two limit sets by means of the notion semi-trajectory. In a continuous dynamical system with $T = \mathbb{R}$, an ω-limit set $\omega(x)$ can be defined to be

$$\omega(x) = \bigcap_{t \in T} \operatorname{cl}\big(\gamma^+(x(t))\big) = \bigcap_{y \in \gamma(x)} \operatorname{cl}\big(\gamma^+(y)\big),$$

and an α-limit set $\alpha(x)$ can be defined to be

$$\alpha(x) = \bigcap_{y \in \gamma(x)} \operatorname{cl}\big(\gamma^-(y)\big).$$

If the motion $\varphi_x(t)$ passing through x is periodic, then $\omega(x) = \alpha(x) = \gamma(x)$, and $\omega(x)$ and $\alpha(x)$ are thus both limit cycles. The reason that $\omega(x)$ and $\alpha(x)$ are identical when the motion $\varphi_x(t)$ is periodic is quite direct, that is, any state $y \in \gamma(x)$ can be reached with infinite times, and hence is essentially a limit state. On the other hand, if $\gamma(x) = \omega(x) = \alpha(x)$, then either $\varphi_x(t)$ is periodic, or x is an equilibrium state.

Let's now consider a continuous dynamical system (X, T, φ) on the Euclidean plane. The following theorem on the possibilities of $\omega(x)$ can be stated:

THEOREM 1.1 (Poincaré-Bendixson). *In a planar dynamical system with an open state space $X \subseteq \mathbb{R}^2$, a time domain $\mathbb{R}$ and an ODEs transition rule $\dot{x} = g(x)$, where $g : X \to X$ is of class C^1, if $\omega(x) \neq \emptyset$ is compact and contains no equilibrium state, then $\omega(x)$ is a closed limit cycle.*

PROOF. Let $y \in \omega(x)$, then $\omega(y) \neq \emptyset$ and $\omega(y) \subseteq \omega(x)$. Assume $z \in \omega(y)$, then also $z \in \omega(x)$, and hence z is not an equilibrium state. Consider a Poincaré section J passing through z, then on $\mathbb{R}^2$, J is a transversal segment to g, and it always exists as z is not an equilibrium state. $y(t)$ will cross J with infinite times. Consider a sequence $(t_i,\ i \in I)$ in $\mathbb{R}$, such that $y(t_i) \in J$ for all $i \in I$ and $\lim_{t_i \uparrow +\infty} y(t_i) = z$, where the Poincaré map $y(t_i) = \varphi(y, t_i)$ is ordered, and hence the sequence $\big(y(t_i),\ i \in I\big)$ along J must be monotonic. Note that $\omega(x)$ is invariant, and $y \in \omega(x)$, so $y(t_i) \in \omega(x)$ for all $t_i \in \mathbb{R}$. Hence $y(t_i) \in J \cap \omega(x)$.

We now show $J \cap \omega(x)$ can not have more than one state. Suppose it has two different states, say z_1 and z_2, then there must exist two different sequences $x(r_i)$ and $x(s_i)$ converging to z_1 and z_2, respectively, where $x(r_i)$ and $x(s_i)$ are ordered intersections of $x(t)$ and J. But we can only have one monotonic sequence $x(t_i)$ along J, because all the subsequences of any monotonic sequence should have a same limit. Thus z_1 must be equal to z_2, a contradiction. Consequently, $J \cap \omega(x)$ is either $\emptyset$ or a singleton.

Since $z \in J \cap \omega(x)$, it must be $J \cap \omega(x) = \{z\}$, and thus $y(t_i) = z$ for all $t_i \in \mathbb{R}$. It hence suggests that $\gamma(y)$ should be periodic, and $\gamma(y) \subseteq \omega(x)$. For all $w \in \gamma(y)$, there always exists a Poincaré section J' passing through w, and also a sequence $x(t_i)$ which converges to w. By the same arguments as above, we can directly have $J' \cap \gamma(y) = \{w\}$, and hence $\gamma(y) \supseteq \omega(x)$. Recall that $\gamma(y) \subseteq \omega(x)$, so $\omega(x) = \gamma(y)$ must be a limit cycle, which completes the proof. □

Suppose now $\omega(x)$ contains equilibrium states. If $\omega(x)$ contains only equilibrium states, then it must be a singleton with one equilibrium state, otherwise we would have a connected limit set with isolated states. If $\omega(x)$ contains both equilibrium and non-equilibrium states, then for all non-equilibrium state $y \in \omega(x)$, $\omega(y)$ must be a singleton having one of the equilibrium states in $\omega(x)$.

In general, if $\omega(x) \neq \emptyset$ is compact and contains finite states, then $\omega(x)$ can be

(i) $\omega(x) = \{y\}$, where y is an equilibrium state;
(ii) $\omega(x)$ is a limit cycle, and contains no equilibrium state;

(iii) $\omega(x) = \{y_1, y_2, \ldots, y_m\} \cup \gamma(z_1) \cup \gamma(z_2) \cup \cdots \cup \gamma(z_n)$, in which y_i is an equilibrium state for all $i \in \{1, 2, \ldots, m\}$, and $\gamma(z_j)$ is a trajectory for all $j \in \{1, 2, \ldots, n\}$, such that either $\omega(z_j)$ or $\alpha(z_j)$, or both are singletons have an equilibrium state in $\{y_1, y_2, \ldots, y_m\}$.

However, if a dynamical system has a state space $X \subseteq \mathbb{R}^n$ for $n \geq 3$, its long-term behavior of trajectory will then begin to be quite hard to characterize.

2.3. Stability and Local Behavior. In this subsection, we focus on continuous dynamical systems, and introduce a number of concepts about stability, so that we can use them to investigate local behaviors in a dynamical system.

The δ-neighborhood of a set S in the metric space (X, d) is defined by

$$N(S, \delta) = \{x \in X : d(x, S) < \delta\},$$

where $d(x, S) = \inf_{y \in S} d(x, y)$. If S is occasionally a singleton, say $S = \{s\}$, we will then have $N(s, \delta) = \{x \in X : d(x, s) < \delta\}$.

Definition. A set of states $S \subseteq X$ is called *weakly stable*, if $\gamma(x)$ is compact for all $x \in S$.

Once the state space X is weakly stable in a dynamical system, the system itself will be called weakly stable. If $X \subseteq \mathbb{R}^n$, and x belongs to some weakly stable set, then the motion $\varphi_x(t)$ should be bounded.

Definition. A closed set $A \subseteq X$ is called an *attractor*, if it has the following properties:

(i) there is a set $B(A) \subseteq X$, such that $\emptyset \neq \omega(x) \subseteq A$ for all $x \in B(A)$,
(ii) there is a $\delta > 0$ such that the δ-neighborhood of A is a subset of $B(A)$, *i.e.*, $N(A, \delta) \subseteq B(A)$,
(iii) there is no proper subset of A satisfying property (i) and (ii).

Here, the set $B(A)$ is called a *basin* of the attractor A. If we replace property (i) and (ii) by a weaker statement that there is a $\delta > 0$ such that $\omega(x) \cap A \neq \emptyset$ for all $x \in N(A, \delta)$, then A will be called a *weak attractor*.

Definition. An attractor $A \subseteq X$ is called *Lyapunov stable*, if for all $\varepsilon > 0$ there is a $\delta > 0$, such that $\gamma^+(x) \subseteq N(A, \varepsilon)$ for all $x \in N(A, \delta)$.

Lyapunov stability states that any state in a δ-neighborhood of A will converge to some state in a corresponding ε-neighborhood of A. If there is some state in A which converges to a state not in A, we then shall say A is *unstable*. More precisely, A is called unstable, if there is a state $x \notin A$ such that $\alpha(x) \cap A \neq \emptyset$. Clearly, even if A is Lyapunov stable, there still could be some $x \notin A$ such that $\alpha(x) \cap N(A, \varepsilon) \neq \emptyset$ for $\varepsilon > 0$, and hence the ε-neighborhood of A is unstable. For this reason, we may need a stronger concept of stability:

Definition. An attractor $A \subseteq X$ is called *asymptotically stable*, if A is Lyapunov stable, and there is a $\xi > 0$, such that $\alpha(x) \cap A = \emptyset$ for all $x \in N(A, \xi) \setminus A$.

In particular, suppose $\{x\}$ is an attractor in some dynamical system, then there should be a basin $B(x)$ such that $\omega(y) = \{x\}$ for all $y \in B(x)$. Lyapunov stability requires that the trajectory of any state in an ε-neighborhood of $\{x\}$ should stay close enough to x, while asymptotic stability requires that $\gamma^+(x) = \{x\}$. If x is the unique equilibrium state in the dynamical system, and $\{x\}$ is an asymptotically stable attractor, then $\omega(y) = \{x\}$ and $\alpha(y) \cap \{x\} = \emptyset$ for all $y \in B(x) \setminus \{x\}$.

Let's now consider a dynamical system (X, T, φ) with a state space $X \subseteq \mathbb{R}^n$ and an ODEs transition rule $\dot{x} = g(x)$, where $g : X \to X$ is of class C^1.

A state $x \in X$ is called *Lyapunov stable*, if for all neighborhood $N(x, \varepsilon)$, there is a neighborhood $N(x, \delta) \subseteq N(x, \varepsilon)$, such that $y(t) \in N(x, \varepsilon)$ for all $y \in N(x, \delta)$ and all $t \in T$.

A state $x \in X$ is called *asymptotically stable*, if it is Lyapunov stable, and there is a neighborhood $N(x, \xi)$, such that $\lim_{t \uparrow +\infty} d(x, y(t)) = 0$ for all $y \in N(x, \xi)$.

A state $x \in X$ is called *exponentially stable*, if there is a neighborhood $N(x, \epsilon)$ and two constants $k, \lambda > 0$, such that when t tends to be sufficiently large, for all $y \in N(x, \epsilon)$,

$$d(x, y(t)) \leq k\mathrm{e}^{-\lambda t} d(x, y). \tag{1.3}$$

If x is an equilibrium state in the dynamical system (X, T, φ), then the motion $x(t)$ around x can be approximated by its linearization. Let $x(t) = x + \mu(t)$, where $\mu(t)$ is a small perturbation, then

$$\dot{\mu} = g(x + \mu) \approx g(x) + Dg(x)\mu,$$

with $g(x) = 0$. We therefore have a linear dynamical system

$$\dot{\mu} = Dg(x)\mu, \tag{1.4}$$

which is determined by the Jacobian $Dg(x)$. If all the eigenvalues of $Dg(x)$ have negative real parts, then $d(\mu(t), 0)$ will exponentially converge to 0 when t goes to $+\infty$. It thus suggests that the equilibrium state x should be exponentially stable, and surely Lyapunov stable.

2.4. Chaos. A quite popular definition of chaos is given by Devaney [**16**], which states that a dynamical system is chaotic if

(i) its dynamics sensitively depends on the initial state,
(ii) it is topologically transitive,
(iii) its periodic states are dense in the state space.

However, these three conditions are actually not completely independent of each other, as we might notice that condition (ii) and (iii) can imply condition (i) in a discrete dynamical system (*cf.*, Banks *et al.* [**4**]). It thus seems natural to remove one likely unnecessary condition in the above definition. Notice that some dynamical systems can be in disorder, even if they have no periodic state at all. In consideration of this fact, we remove condition (iii), and give the following definition.

Definition. A dynamical system (X, T, φ) is called *chaotic* if

(i) for all $x \in X$ and $\varepsilon > 0$, there is a $\delta > 0$ such that $d\big(\varphi_t(x), \varphi_t(y)\big) > \varepsilon$ for all $y \in N(x, \delta)$ at some $t \in T$,
(ii) for all open $S_1, S_2 \subseteq X$, there is some $t \in T$ such that $\varphi_t(S_1) \cap S_2 \neq \emptyset$, where $\varphi_t(S_1) = \{\varphi_t(x) : x \in S_1\}$.

The chaotic property can be preserved for all topologically conjugate dynamical systems, that's to say, a dynamical system is chaotic if and only if its topologically conjugation is chaotic. Suppose two dynamical systems (X_1, T_1, φ) and (X_2, T_2, ψ) are topologically conjugate, then their transition functions

$$\varphi_t : X_1 \to X_1, \quad \text{and} \quad \psi_t : X_2 \to X_2$$

should satisfy

$$\varphi_t = \phi^{-1} \circ \psi_t \circ \phi, \tag{1.5}$$

where $\phi : X_1 \to X_2$. Assume X_1 and X_2 are both compact, then (X_1, T_1, φ) is chaotic should directly imply (X_2, T_2, ψ) is chaotic too, and vice versa.

The sensitivity to the initial state in a chaotic dynamical system can be measured by the Lyapunov exponent. To some extent, the Lyapunov exponent is determined by the exponential rate of deviation caused by a small perturbation to the initial state. Evidently, it does not directly rely on the transition function and the motion, but on the vector field of the transition and motion. For instance, in a dynamical system with an ODEs transition rule $\dot{x} = g(x)$, the Lyapunov exponent at a state $x \in X$ can be obtained by the Jacobian $Dg(x)$.

Generally, in a dynamical system (X, T, φ), the Lyapunov exponent at a state $x \in X$ is defined by

$$\chi(x) = \limsup_{t \uparrow +\infty} \frac{\log d(x(t), 0)}{t}, \tag{1.6}$$

where $x(t)$ is the motion $\varphi_x(t)$ mapping T into X. Moreover,

Definition. A function $\chi : X \to \mathbb{R} \cup \{-\infty\}$ is called the *Lyapunov exponent* if

(i) $\chi(\lambda x) = \chi(x)$ for all $x \in X$ and all $\lambda \in \mathbb{R} \setminus \{0\}$,
(ii) $\chi(x + y) \leq \max\{\chi(x), \chi(y)\}$ for all $x, y \in X$,
(iii) $\chi(0) = -\infty$.

3. Iterated Function System

3.1. Discrete Dynamical System. As discussed in Section 2.1 of this chapter, the transition rule in a discrete dynamical system can be completely determined by one mapping defined on its state space. Let's now suppose the state space X is a metric space with a metric d, and the time domain T is $\mathbb{Z}$.

DEFINITION. A *discrete dynamical system* on X is a pair (X, f) with $x_{n+1} = f(x_n)$ for all $x_n, x_{n+1} \in X$ and all $n \in \mathbb{Z}$, where $f : X \to X$ is of class C^0.

The trajectory passing through a state $x \in X$ is

$$\gamma(x) = \{f^n(x) : n \in \mathbb{Z}\},$$

and its positive and negative semi-trajectories are

$$\gamma^+(x) = \{f^n(x) : n \in \mathbb{Z}^+\}, \quad \gamma^-(x) = \{f^n(x) : n \in \mathbb{Z}^-\}.$$

Evidently, the positive semi-trajectory $\gamma^+(x)$ also represents the motion starting from the state x.

A state x is an equilibrium state, if $\gamma(x) = \{x\}$, or $f(x) = x$. A set of states $S \subseteq X$ is invariant if $f(S) = S$. Recall that any nonempty ω-limit set should be invariant, and thus here, we have $f\big(\omega(x)\big) = \omega(x)$ for all $\omega(x) \neq \emptyset$.

A set of states $A \subseteq X$ is an attractor, if there is a neighborhood $N(A, \varepsilon)$ such that $f(N(A, \varepsilon)) \subseteq N(A, \varepsilon)$, and

$$\omega\big(N(A, \varepsilon)\big) = \bigcap_{n \in \mathbb{Z}^+} f^n\big(N(A, \varepsilon)\big) = A,$$

but no proper subset of A has such properties.

A state x (and also the motion $\gamma^+(x)$) is periodic, if there is a $k \in \mathbb{Z}^+$ such that $f^k(x) = x$, and the minimal $k \in \mathbb{Z}^+$ satisfying $f^k(x) = x$ is the period of $\gamma^+(x)$. If the period of $\gamma^+(x)$ is 1, then $f(x) = x$, and thus x is actually an equilibrium state. If the period of $\gamma^+(x)$ is $p < +\infty$, then

$$\gamma^+(x) = \{x, f(x), \ldots, f^p(x)\}. \tag{1.7}$$

A state x is called *finally periodic*, if there is an $m \in \mathbb{Z}^+$ such that $f^n(x)$ is a periodic state for all $n \geq m$, or equivalently stating, there is some $p \in \mathbb{Z}^+$ such that $f^{n+p}(x) = f^n(x)$ for all $n \geq m$. A state x is called *asymptotically periodic*, if there is a $y \in X$ such that

$$\lim_{n \uparrow +\infty} d\big(f^n(x), f^n(y)\big) = 0. \tag{1.8}$$

If the state space $X \subseteq \mathbb{R}$ and it is compact, we will have the following theorem:

THEOREM 1.2 (Li-Yorke). *Suppose X is an interval in $\mathbb{R}$, and $f : X \to X$ is of class C^0. If there exists a motion of period 3 in (X, f), viz., there are three distinct states $x, y, z \in X$ such that $f(x) = y$, $f(y) = z$, and $f(z) = x$, then there is some motion of period n in (X, f) for all $n \in \mathbb{N}$.*

PROOF. Let $<_S$ denote Sarkovskii's order on $\mathbb{N}$, then we have

$$3 <_S 5 <_S 7 <_S \cdots <_S 2^n <_S 2^{n-1} <_S \cdots <_S 2^2 <_S 2 <_S 1.$$

By Sarkovskii's theorem, if (X, f) has a motion of period m, then it must have some motion of period m' with $m <_S m'$. Since $3 <_S n$ for all $n \neq 3$, and there is a motion of period 3 in (X, f), the statement will thus directly follow. □

As defined in Section 2.4 of this chapter, a dynamical system is chaotic if its dynamics sensitively depends on the initial state, and its states are transitive. For the moment, a discrete dynamical system (X, f) is called chaotic if it satisfies

(i) for all $x \in X$ and any $\varepsilon > 0$, there is a $\delta > 0$ such that $d\big(f^n(x), f^n(y)\big) > \varepsilon$ for all $y \in N(x, \delta)$ and some $n \in \mathbb{Z}^+$,
(ii) for all $S_1, S_2 \subseteq X$, there is an $n \in \mathbb{Z}^+$ such that $f^n(S_1) \cap S_2 \neq \emptyset$.

In particular, when $X \subseteq \mathbb{R}$ is compact, and f is of class C^0, an alternative definition of chaos can be proposed in the sense of Li and Yorke [**38**].

DEFINITION. A discrete dynamical system (X, f) is *nonperiodically chaotic*, if there is an uncountable set $S \subseteq X$ such that

(i) $\limsup_{n\uparrow+\infty} d\big(f^n(x), f^n(y)\big) > 0$ for all distinct $x, y \in S$,
(ii) $\liminf_{n\uparrow+\infty} d\big(f^n(x), f^n(y)\big) = 0$ for all distinct $x, y \in S$,
(iii) for all $z \in X$ periodic, $\limsup_{n\uparrow+\infty} d\big(f^n(x), f^n(z)\big) > 0$ for all $x \in S$.

It might be noticed that nonperiodic chaos is a slightly weaker concept than chaos itself. That's to say, if a discrete dynamical system on $X \subseteq \mathbb{R}$ is chaotic, then it must be nonperiodically chaotic as well; but if a discrete dynamical system is nonperiodically chaotic, it may not be chaotic.

The reason behind such an assertion is constructive. If (X, f) is nonperiodically chaotic, then there is at most one asymptotically periodic state in S. Now suppose a state $u \in X$ is not asymptotically periodic, then $\omega(u)$ should have infinitely many states. Let $V \subseteq \omega(u)$ be the (minimally invariant) kernel of $\omega(u)$, and suppose there is some $v \in X$ such that $V = \omega(v)$, which hence again contains infinitely many states. Let $U = X \setminus V$, then $f^n(V) \cap U = \emptyset$ for all $n \in \mathbb{Z}^+$, and therefore V and U are not transitive, which then implies (X, f) is not chaotic.

3.2. Iterated Function System. Let's now consider a collection of contractive functions defined on the state space X with a metric d. Here, a function $f : X \to X$

is called *contractive*, if there is a $\lambda \in (0,1)$ such that $d\big(f(x), f(y)\big) \leq \lambda d(x,y)$ for all $x, y \in X$.

Let I_N denote an index set with N elements for $N \geq 2$ finite. Let

$$F = \{f_i : i \in I_N\},$$

where $f_i : X \to X$ is contractive and of class C^0 for all $i \in I_N$.

Definition. The pair (X, F) is called an *iterated function system*, if (X, f_i) is a discrete dynamical system for all $i \in I_N$.

Suppose X is compact, and let $\mathcal{Q}(X)$ denote the collection of all the nonempty compact subsets of X. Then $\mathcal{Q}(X)$ with the Hausdorff metric d_H is a compact metric space, where the Hausdorff metric d_H on $\mathcal{Q}(X)$ can be defined by the metric d on X, *i.e.*, for all $U, V \in \mathcal{Q}(X)$,

$$d_H(U,V) = \sup_{u\in U,\, v\in V} \{d(u,V), d(v,U)\}, \tag{1.9}$$

in which $d(u,V) = \inf_{v\in V} d(u,v)$ and $d(v,U) = \inf_{u\in U} d(v,u)$.

Define a mapping $H : \mathcal{Q}(X) \to \mathcal{Q}(X)$, such that for all $B \in \mathcal{Q}(X)$,

$$H(B) = \bigcup_{i\in I_N} f_i(B). \tag{1.10}$$

Here, H is called the *Hutchinson operator*. Moreover, define H^n by the recursion $H^n = H \circ H^{n-1}$ with $H^0 = \mathrm{id}_{\mathcal{Q}(X)}$, where $n \in \mathbb{Z}$ and $\mathrm{id}_{\mathcal{Q}(X)}$ denotes the identity mapping on $\mathcal{Q}(X)$.

Definition. $A \in \mathcal{Q}(X)$ is called an *attractor* of (X, F), if there is a neighborhood $N(A, \varepsilon) \in \mathcal{Q}(X)$ such that

$$H\big(N(A,\varepsilon)\big) \subseteq N(A,\varepsilon), \qquad \bigcap_{n\in\mathbb{Z}^+} H^n\big(N(A,\varepsilon)\big) = A,$$

and no proper subset of A in $\mathcal{Q}(X)$ has such properties.

Theorem 1.3. *(X, F) has a unique attractor A with $H(A) = A$.*

Proof. For all $f_i \in F$, there is a $\lambda_i \in (0,1)$ such that for all $x, y \in X$,

$$d\big(f_i(x), f_i(y)\big) \leq \lambda_i d(x,y).$$

Let $\lambda = \max_{i\in I_N} \lambda_i$, then $\lambda \in (0,1)$ as well. Note that for all $U, V \in \mathcal{Q}(X)$ we have

$$d_H\big(H(U), H(V)\big) \leq \sup_{i\in I_N} d_H\big(f_i(U), f_i(V)\big) \leq \sup_{i\in I_N} \lambda_i d_H(U,V) \leq \lambda d_H(U,V),$$

thus by the Banach fixed point theorem, there is a unique $A \in \mathcal{Q}(X)$ such that $H(A) = A$, and $\lim_{n\uparrow+\infty} H^n(B) = A$ for all $B \in \mathcal{Q}(X)$. And clearly, there exists a neighborhood $N(A,\varepsilon) \in \mathcal{Q}(X)$ serving as a basin of A.

We then show that any $B \neq A$ in $\mathcal{Q}(X)$ can not be an attractor of (X, F), which would imply A is the unique attractor of (X, F), and thus completes our proof. First, any $B \supset A$ can not be an attractor of (X, F), as for all $\varepsilon > 0$,

$$\bigcap_{n \in \mathbb{Z}^+} H^n\big(N(B, \varepsilon)\big) \subseteq A \subset B.$$

Next, any $B \subset A$ also can not be an attractor of (X, F), otherwise we would have

$$\lim_{n \uparrow +\infty} H^n\big(N(B, \varepsilon)\big) = B \subset A,$$

a contradiction. $\square$

Now consider the space I_N^ω, and for all $\mu \in I_N^\omega$ write

$$\mu = (\mu_n,\ n \in \mathbb{N}) = (\mu_1, \mu_2, \ldots, \mu_\omega),$$

where $\mu_n \in I_N$ for all $n \in \mathbb{N}$. The Baire metric between all $\mu, \upsilon \in I_N^\omega$ is

$$d_B(\mu, \upsilon) = 2^{-m},$$

where $m = \min\{n \in \mathbb{N} : \mu_n \neq \upsilon_n\}$. Clearly, (I_N^ω, d_B) is a compact metric space. Let's define a mapping $C : I_N^\omega \times \mathcal{Q}(X) \to \mathcal{Q}(X)$, such that for all $\mu \in I_N^\omega$ and $S \in \mathcal{Q}(X)$,

$$C(\mu, S) = \bigcap_{n \in \mathbb{N}} f_{\mu_\omega} \circ \cdots \circ f_{\mu_{n+1}} \circ f_{\mu_n}(S). \tag{1.11}$$

Note in addition that the motion of any state $x \in S$ can be expressed as

$$\gamma^+(x) = \{f_{\mu_n} \circ \cdots f_{\mu_2} \circ f_{\mu_1}(x) : n \in \mathbb{N}\}. \tag{1.12}$$

Suppose $B(A) = N(A, \varepsilon)$ for some $\varepsilon > 0$ is a basin of the attractor A, then for all $S \subseteq B(A)$ and $\mu \in I_N^\omega$, we have $C(\mu, S) \subseteq A$, and hence we can write

$$C\big(I_N^\omega, B(A)\big) = A. \tag{1.13}$$

It therefore suggests that the attractor A of (X, F) could be practically attained by all the ω-permutations of the transition rules in F.

Suppose there is some probability measure on I_N^ω, and in particular, we shall assume it is stationary, so that it can be fully characterized by a discrete probability measure on I_N. Let $\pi : I_N \to [0, 1]$ denote such a probability measure, which satisfies $\sum_{i \in I_N} \pi(i) = 1$. Thus at any time a function f_i stands out in F with a probability $\pi(i)$ for all $i \in I_N$.

Definition. The triplet (X, F, π) is called an *iterated random function system*.

Let σ_n be a random variable, such that $\text{Prob}(\sigma_n = i) = \pi(i)$ for all $i \in I_N$. The transition function at a time $n \in \mathbb{Z}$ can thus be denoted by a randomly indexed function f_{σ_n}. Let a random variable Z_n denote the stochastic state in the system (X, F, π) at the time $n \in \mathbb{Z}$, then we have

$$Z_{n+1} = f_{\sigma_{n+1}}(Z_n). \tag{1.14}$$

Suppose the initial time is 0, and the initial state is $x \in X$, then the random motion can be written as

$$\Gamma^+(x) = \{Z_n : n \in \mathbb{Z}^+\},$$

in which $Z_0 = x$, $Z_1 = f_{\sigma_1}(x)$, and $Z_n = f_{\sigma_n}(Z_{n-1})$ for all $n \geq 2$. Note that the stochastic process $(Z_n,\ n \in \mathbb{N})$ is a Markov chain, and it is equivalent to the iterated random function system (X, F, π). Suppose $Z_n = z$, then $\mathrm{Prob}(Z_{n+1} \in S)$ for some $S \subseteq X$ takes the value

$$P(z, S) = \sum_{i \in I_N} \pi(i) \mathbf{1}_S\big(f_i(z)\big), \tag{1.15}$$

where

$$\mathbf{1}_S(x) = \begin{cases} 1 & \text{if } x \in S \\ 0 & \text{if } x \notin S \end{cases}$$

When S is a Borel subset of X, there should be an invariant probability measure ρ, such that

$$\rho(S) = \int_X P(z, S) d\rho(z) = \sum_{i \in I_N} \pi(i) \rho\big(f_i^{-1}(S)\big). \tag{1.16}$$

Here, ρ is called a π-*balanced measure* for (X, F, π), as was proposed by Barnsley and Demko [**5**].

Let $R(\rho)$ denote the support of ρ, then $R(\rho) = \{x \in X : \rho(x) \neq 0\}$ and hence

$$R(\rho) = \bigcup_{i \in I_N} f_i\big(R(\rho)\big) = H\big(R(\rho)\big). \tag{1.17}$$

By Theorem 1.3, it directly appears that $R(\rho) = A$, and therefore the support of a π-balanced measure for (X, F, π) is exactly the unique attractor A of (X, F) for all π. As a result, we can see that the attractor A of (X, F) can also be attained by a random ω-permutation of the transition rules in F.

EXAMPLE 1.3. Let's assume $X = [0, 1]$, $I_2 = \{a, b\}$, $\pi(a) = \pi(b) = 0.5$, and $F = \{f_a, f_b\}$ with

$$f_a : x \mapsto x/3, \quad f_b : x \mapsto x/3 + 2/3.$$

The iterated random function system (X, F, π) is clearly equivalent to the following autoregressive process,

$$Z_{n+1} = Z_n/3 + \varepsilon_{n+1} \quad (n \in \mathbb{Z}^+), \tag{1.18}$$

where Z_0 is deterministic, and for all $n \in \mathbb{N}$,

$$\mathrm{Prob}(\varepsilon_n = 0) = \mathrm{Prob}(\varepsilon_n = 2/3) = 0.5.$$

The iterated function system (X, F) has a unique attractor A as a Cantor ternary set, that is,

$$A = \Big\{ \sum_{n \in \mathbb{N}} x_n/3^n : (x_n,\ n \in \mathbb{N}) \in \{0, 2\}^\omega \Big\}. \tag{1.19}$$

Let

$$B_k = \Big\{ \sum_{n \geq k} x_n/3^n : (x_n,\ n \geq k) \in \{0, 2\}^\omega \Big\} \quad (k \in \mathbb{N}),$$

then $B_1 = A$ and $B_2 = A/3$, and thus

$$f_a(A) = B_2, \quad f_b(A) = B_2 + 2/3,$$

which yield $f_a(A) \cup f_b(A) = B_1$. Recall that $H = f_a \cup f_b$ is the Hutchinson operator for (X, F), we can thus write $H(A) = A$.

In practice, the attractor A can be realized by a random motion with any initial state $x \in [0, 1]$ in (X, F, π). There are two cases to consider.

If $x \in A$, then $\Gamma^+(x) = A$ almost surely. If $x \notin A$, then there should be a finite sequence $(x_1, x_2, \ldots, x_k) \in \{0, 2\}^k$, such that

$$x = \sum_{n=1}^{k} x_n/3^n + r_k(x),$$

where $r_k(x) \leq 1/3^k$. Clearly, $r_k(x)$ will tend to be 0 when k goes to infinity. Now we have $Z_1 = x/3 + \varepsilon_1 = Y_1 + r_k(x)/3$, where $Y_1 = \sum_{n=1}^{k} x_n/3^{n+1} + \varepsilon_1 \in A$, and in general, $Z_m = Y_m + r_k(x)/3^m$, where $Y_m \in A$ and $r_k(x)/3^m \leq 1/3^{m+k}$. Evidently, there should be an ℓ such that $Z_\ell \in A$, which suggests $\Gamma^+(Z_\ell) = A$ almost surely.

3.3. Strange Attractor. The notion strange attractor first appeared in a paper on turbulence by Ruelle and Takens [**56**], but it was not precisely defined. A formal definition later was made by Ruelle [**55**], according to which, however, any attractor of a chaotic dynamical system would be strange. With further restrictions, we shall say an attractor of a dynamical system is strange if the dynamical system is chaotic, and it is a fractal.

Definition. A set is called a *fractal*, if its Hausdorff-Besicovitch dimension is greater than its topological dimension.

For all set $Z \subseteq X$, where X is a metric space with some metric, its *topological dimension* $\dim_L(Z)$ should be in $\mathbb{N} \cup \{-1, 0, +\infty\}$. In particular, if $Z = \emptyset$, then $\dim_L(Z) = -1$; if Z is discrete, then $\dim_L(Z) = 0$. In general, if all $z \in Z$ can be contained in at most $n+1$ open sets in any open cover of Z, then $\dim_L(Z) = n$; if there is no such a finite n, then $\dim_L(Z) = +\infty$.

Let $\mathcal{H}^d(Z)$ denote the d-dimensional Hausdorff measure of Z for $d \geq 0$, then the *Hausdorff-Besicovitch dimension* of Z, denoted by $\dim_H(Z)$, can be defined by

$$\dim_H(Z) = \inf\{d : \mathcal{H}^d(Z) = 0\} = \sup\{d : \mathcal{H}^d(Z) = +\infty\}.$$

In practice, a more useful notion that can be used to describe the dimension of a fractal is the box-counting dimension. The lower and upper box-counting dimensions of a set $Z \subseteq X$ are defined to be

$$\underline{\dim}_B(Z) = \liminf_{\varepsilon \downarrow 0} \frac{\log n(Z,\varepsilon)}{\log(1/\varepsilon)}, \quad \overline{\dim}_B(Z) = \limsup_{\varepsilon \downarrow 0} \frac{\log n(Z,\varepsilon)}{\log(1/\varepsilon)},$$

where $n(Z,\varepsilon)$ denotes the minimal number of sets required for an open ε-cover of Z. If X is an Euclidean space, and ε is close to 0, we would have

$$n(Z,\varepsilon) \approx \lambda(1/\varepsilon)^d,$$

where $\lambda > 0$ and $d \geq 0$.

If $\underline{\dim}_B(Z) = \overline{\dim}_B(Z)$, the *box-counting dimension* of Z is said to exist, and can be defined as

$$\dim_B(Z) = \underline{\dim}_B(Z) = \overline{\dim}_B(Z).$$

In general, it holds that

$$\dim_H(Z) \leq \underline{\dim}_B(Z) \leq \overline{\dim}_B(Z),$$

and hence we have $\dim_H(Z) \leq \dim_B(Z)$ if $\dim_B(Z)$ exists.

Let's return back to the iterated function system (X,F), and assume $X \subseteq \mathbb{R}^n$. With regards its unique attractor A, we have $\dim_H(A) = \dim_B(A)$, and when A is strange, by definition we should have

$$\dim_B(A) = \dim_H(A) > \dim_L(A).$$

Suppose all $f_i \in F$ is affine, and moreover, all distinct $f_i(A)$ and $f_j(A)$ do not overlap, that's to say, there is an open set $D \subseteq A$, such that $f_i(D) \cap f_j(D) = \emptyset$ for all distinct i,j, and $\bigcup_{i\in I_N} f_i(D) \subseteq D$. Then $\dim_B(A) = \dim_H(A)$ is the unique solution of the following equation,

$$\sum_{i\in I_N} \lambda_i^d = 1, \tag{1.20}$$

where $\lambda_i \in (0,1)$ is the Lipschitz constant of f_i for all $i \in I_N$.

A quite intuitive verification of (1.20) is as follows. Note that $A = \bigcup_{i\in I_N} f_i(A)$, and all distinct $f_i(A)$ and $f_j(A)$ do not overlap, thus we have

$$n(A,\varepsilon) = \sum_{i\in I_N} n\big(f_i(A),\varepsilon\big).$$

By scaling $f_i(A)$ and ε by λ_i simultaneously, we should see $n\big(f_i(A), \varepsilon\big) = n(A, \varepsilon/\lambda_i)$, and therefore

$$n(A, \varepsilon) = \sum_{i \in I_N} n(A, \varepsilon/\lambda_i).$$

Since $n(A, \varepsilon) \approx \lambda(1/\varepsilon)^d$, we then have $(1/\varepsilon)^d = \sum_{i \in I_N} (\lambda_i/\varepsilon)^d$ for ε close to 0, which directly yields the equation (1.20).

Example 1.4. Let $X = [0, 1]$, and pick $c \in (2, +\infty)$. We remove the interval $(1/c, 1 - 1/c)$ from $[0, 1]$ to have

$$C_1 = [0, 1/c] \cup [1 - 1/c, 1].$$

We then remove some middle parts from $[0, 1/c]$ and $[1 - 1/c, 1]$ to have

$$C_2 = [0, 1/c^2] \cup [1/c - 1/c^2, 1/c] \cup [1 - 1/c, 1 - 1/c + 1/c^2] \cup [1 - 1/c^2, 1].$$

In general, $C_{n+1} = C_n/c \cup \big(C_n/c + (1 - 1/c)\big)$ for all $n \in \mathbb{N}$. When n goes to infinity, we obtain the Cantor set

$$C = \Big\{ \sum_{n \in \mathbb{N}} x_n/c^n : (x_n,\ n \in \mathbb{N}) \in \{0, c - 1\}^\omega \Big\}, \tag{1.21}$$

which is the attractor of (X, F) for $F = \{x/c, x/c + (1 - 1/c)\}$.

Let $\varepsilon = 1/c^n$ for some $n \in \mathbb{N}$, then $n(C, \varepsilon) = 2^n$, because we need at least 2^n pairwise disjoint intervals for an open ε-cover of C. It thus appears that

$$\dim_B(C) = \lim_{n \uparrow +\infty} \log 2^n / \log c^n = \log 2 / \log c.$$

Since $c \in (2, +\infty)$, we have $\dim_B(C) \in (0, 1)$. Notice that (X, F) is an iterated affine function system, in which these two Lipschitz constants are both $1/c$, so $\dim_B(C)$ is the unique solution of the equation

$$(1/c)^d + (1/c)^d = 1,$$

which also yields $\dim_B(C) = \log 2 / \log c$.

As for a Cantor ternary set that appeared in Example 1.3, its dimension is equal to $\log 2 / \log 3 \approx 0.631$ for $c = 3$.

Example 1.5 (De Rham [**13**]). Let $X \subseteq \mathbb{C}$ be a triangle with vertices 0, a, and 1, such that $|a| < 1$ and $|1 - a| < 1$. Suppose $F = \{f_1, f_2\}$ with

$$f_1(z) = a\overline{z}, \quad f_2(z) = (1 - a)\overline{z} + a,$$

where $\overline{z}$ denotes the complex conjugate of z. (X, F) is thus an iterated affine function system on the complex (Gaussian) plane.

Suppose $a = 1/2 + i\sqrt{3}/6$, then X is equilateral, and the attractor of (X, F) is the Koch curve K. Note that $|a| = |1-a| = 1/\sqrt{3}$, so these two Lipschitz constants are both $1/\sqrt{3}$. The dimension of K is the unique solution of the equation

$$(1/\sqrt{3})^d + (1/\sqrt{3})^d = 1,$$

which directly yields $\dim_B(K) = 2\log 2/\log 3 \approx 1.262$.

Suppose $|a - 1/2| = 1/2$, *e.g.*, $a = 1/2 + i/2$, then the attractor of (X, F) turns out to be the Peano curve P. Since these two Lipschitz constants are now both $1/\sqrt{2}$, the dimension of P is the unique solution of the equation $(1/\sqrt{2})^d + (1/\sqrt{2})^d = 1$, and thus $\dim_B(P) = 2$.

4. Application

4.1. Random Utility. Consider a generic agent w in a large group W, and suppose w has a preference relation by nature. Let X denote a decision state space for the group W, and let $\gtrsim$ be a weak order on X such that

(i) either $x \gtrsim y$ or $y \gtrsim x$ for all $x, y \in X$,
(ii) $x \gtrsim y$ and $y \gtrsim z$ implies $x \gtrsim z$ for all $x, y, z \in X$.

So $\gtrsim$ can serve as a rational preference relation for w. In particular, we shall assume that there is a *utility function* $u : X \to \mathbb{R}$ such that for all $x, y \in X$,

$$x \gtrsim y \iff u(x) \geq u(y).$$

Let $\mathcal{P}(X)$ be the power set of X. A mapping $C : \mathcal{P}(X) \to \mathcal{P}(X)$ is called a *choice function* if $\emptyset \neq C(Y) \subseteq Y$ for all nonempty $Y \in \mathcal{P}(X)$.

If $y \in C(Y)$ for some $y \in Y$, and meanwhile, $u(y) \geq u(x)$ for all $x \in Y$, we say the choice made by w matches to her preference relation. It should be noted that there are two implicit assumptions under this statement, *i.e.*, the choices made by w can be perfectly observed, and w can perfectly identify and also intentionally apply her preference relation. However, it seems that empirical verifications would be unable to simultaneously support these two intertwined assumptions. The reason is that observations about the choices made by w are reasonable only if w does make her choices complying with her preference relation, and on the other hand, the true preference relation of w can be thought of as identifiable only if observations about her choices are perfect.

To overcome such difficulties in empirical verifications of consistency of choice and preference, we have to set one assumption ad hoc true, so that we could verify the other one. To begin with, if the preference relation of w is supposed to be perfectly identifiable and intentionally applied by w herself, then it will become possible to infer it from observations about her choices with some confidence level. This approach appeared in a study on stochastic utility model estimation by Manski [**44**].

Let $v : X \to \mathbb{R}$ denote a utility function consistent with observations about the choices made by w. And we shall say $v(x)$ is the observed utility if a choice $x \in X$ has been observed. It thus appears to us that

$$u(x) = v(x) + \varepsilon(x), \tag{1.22}$$

where $\varepsilon(x)$ denotes a "noise" function that might be independently distributed for all $x \in X$. In particular, the choice x can be assumed to be characterized by n independently observed attributes, $J(x) \in \mathbb{R}^n$, thus $v(x)$ admits a linearly parametric model $v(x) = \beta' J(x)$ for $\beta \in \mathbb{R}^n$. In consequence, we have

$$u(x) = \beta' J(x) + \varepsilon(x), \tag{1.22$'$}$$

in which the estimation $\hat{\beta}$ is determined by the observed data $J(x)$ for $x \in S$, where $S \subseteq X$ is a certain sample.

On the other hand, if the choices made by w are supposed to be perfectly observed, then we could discuss her identification of the true preference relation. In practice, the true preference relation might be only partially identified by w, but it should not be totally vague to her, even if she had an extremely limited cognitive ability. Suppose w has a collection of observable utility functions, which can represent her identified preference relations in different situations, and all these utility functions have an identical kernel as her invariant knowledge of her true preference relation.

Let I_N be a finite index set with $|I_N| = N \geq 2$, and let $v_i : X \to \mathbb{R}$ be a utility function of w for all $i \in I_N$. Suppose $u : X \to \mathbb{R}$ is the kernel utility function of all v_i for $i \in I_N$. Let $X_u = u(X)$, then $X_u \subseteq \mathbb{R}$. And for all $i \in I_N$, there is a contractive function $f_i : X_u \to X_u$ such that $v_i = f_i \circ u$, or $v_i(x) = f_i\big(u(x)\big)$ for all $x \in X$. Clearly, $\{v_i : i \in I_N\}$ on the domain X is now equivalent to $F = \{f_i : i \in I_N\}$ on the domain X_u.

Suppose w makes her choices along the time domain $\mathbb{Z}^+$ in such a way that at each time $t \in \mathbb{N}$, she picks a function $f_i \in F$ to form her utility function

$$u_t = f_i \circ u_{t-1}, \tag{1.23}$$

where u_{t-1} is her utility function at the time $t-1$. More concretely, at the initial time 0, the utility function of w is set as her kernel utility function, *i.e.*, $u_0(x) = u(x)$, and at the time 1, her utility function is $u_1(x) = f_i\big(u(x)\big)$ for some $i \in I_N$. In general, at any time $t \in \mathbb{N}$, her utility function is $u_t(x) = f_i\big(u_{t-1}(x)\big)$ for some $i \in I_N$. Here, the sequence of utility functions $(u_t(x),\ t \in \mathbb{Z}^+)$ can be considered as a general extension of a normal discounted utility function series, and in terms of time preference, we actually generalize $(\succsim,\ t \in \mathbb{Z}^+)$ to $(\succsim_t,\ t \in \mathbb{Z}^+)$, where $\succsim_t$ varies across time.

For the moment, we should notice that $(u_t(x),\ t \in \mathbb{Z}^+)$ is completely determined by the iterated function system (X_u, F). By Theorem 1.3, we directly see (X_u, F)

has a unique attractor, say $A \subseteq X_u$, such that $A = \bigcup_{i \in I_N} f_i(A)$. It thus suggests that some kernel utilities in A could be reached by w in the long term.

Let π denote a probability measure on I_N, then an iterated random function system (X_u, F, π) will emerge. By (1.14), we obtain

$$U_{t+1}(x) = f_{\sigma_{t+1}}\big(U_t(x)\big) \quad (t \in \mathbb{Z}^+), \tag{1.24}$$

where $U_0(x) = u(x)$, and $U_t(x)$ denotes the random utility function at any $t \in \mathbb{N}$.

If $f_i(x) = \rho_i x$ for all $i \in I_N$, where $\rho_i \in (0,1)$ and $\rho_i \neq \rho_j$ for all distinct $i, j \in I_N$, then (1.24) will be

$$U_{t+1}(x) = \xi_{t+1} U_t(x) \quad (t \in \mathbb{Z}^+), \tag{1.24$'$}$$

where $\mathrm{Prob}(\xi_t = \rho_i) = \pi(i)$ for all $i \in I_n$ and all $t \in \mathbb{N}$. Thus at any time $t \in \mathbb{N}$, the random utility function of w is

$$U_t(x) = \Big(\prod_{n=1}^{t} \xi_n\Big) u(x) = \exp\Big(\sum_{n=1}^{t} \log \xi_n\Big) u(x) = \exp\Big(-\sum_{n=1}^{t} \log(1/\xi_n)\Big) u(x).$$

Let $\delta_t = \sum_{n=1}^{t} \log(1/\xi_n)$, then the random utility function of w at $t \in \mathbb{N}$ can be written as

$$U_t(x) = \mathrm{e}^{-\delta_t} u(x). \tag{1.25}$$

When t goes to infinity, δ_t will approach infinity, and thus $U_t(x)$ will approach zero almost surely for all choice x.

If $f_i(x) = \rho x + r_i$ for all $i \in I_N$, where $\rho \in (0,1)$, $r_i > 0$, and $r_i \neq r_j$ for all distinct $i, j \in I_N$, then (1.24) will be

$$U_{t+1}(x) = \rho U_t(x) + \theta_{t+1} \quad (t \in \mathbb{Z}^+), \tag{1.24$''$}$$

in which once more $\mathrm{Prob}(\theta_t = r_i) = \pi(i)$ for all $i \in I_N$ and all $t \in \mathbb{N}$. At any time $t \in \mathbb{N}$, the random utility function of w is

$$U_t(x) = \rho^t u(x) + \sum_{n=1}^{t} \rho^{t-n} \theta_n. \tag{1.26}$$

Note that $\rho^t u(x)$ will vanish when t goes to infinity, but the remaining part will not converge almost surely, as a new piece of randomness θ_t will emerge at each time t.

4.2. Stochastic Growth. Consider an economy with a production function $Y = F(K, L)$, where Y, K, and L denote the total production, the capital input, and the labor supply in the economy, respectively. Let $y = Y/L$ and $k = K/L$, and suppose $F(K, L)$ is a homogeneous function of degree 1, then $Y/L = F(K/L, 1)$.

Define $f(k) = F(K/L, 1)$, thus the production technology of a generic agent w in that economy can be represented by

$$y = f(k) \quad (k \in \mathbb{R}^+). \tag{1.27}$$

As typically assumed, $f(k)$ should satisfy that for all $k \in \mathbb{R}^+$,

$$f'(k) > 0, \quad f''(k) < 0,$$

and the following Inada conditions (which are usually named after K. Inada, but also partly attributed to H. Uzawa [**64**]),

$$\lim_{k\downarrow 0} f'(k) = +\infty, \quad \lim_{k\uparrow +\infty} f'(k) = 0.$$

Let's now introduce a stochastic factor ξ into the economy, so that the production technology of w can be expressed as

$$y = f(k, \xi) \quad (k \in \mathbb{R}^+). \tag{1.28}$$

In case k and ξ are separable, we could consider two fundamental cases, *i.e.*, ξ is an additive shock to $f(k)$, and ξ is a multiplicative shock to $f(k)$. Similar to the studies by Mitra, Montrucchio, and Privileggi [**46**], and Mitra and Privileggi [**47**], we shall focus on the latter case, and rewrite the technology (1.28) as

$$y = \xi f(k) \quad (k \in \mathbb{R}^+), \tag{1.28$'$}$$

where $\xi > 0$ is a random variable. In practice, we can assume that the support of ξ is $\{\lambda_i : i \in I_N\}$, where I_N is a finite index set with $|I_N| = N \geq 2$, and there is a probability measure π on I_N, such that $\text{Prob}(\xi = \lambda_i) = \pi(i)$ for all $i \in I_N$.

In addition, the consumption and investment which are both necessary parts of a sustainable economy, are denoted by C and E, thus we should have $Y = C + E$. Let $c = C/L$ and $e = E/L$, then the income identity for w is $y = c + e$. Suppose the economy functions on the time domain $\mathbb{Z}^+$, so that the economic variables all become time-dependent, *i.e.*, $y_t, k_t, c_t, e_t, \xi_t$ for $t \in \mathbb{Z}^+$, then the economy can be represented by the following system,

$$\begin{cases} y_t = \xi_t f(k_t) \\ y_t = c_t + e_t \\ k_{t+1} = e_t \end{cases}$$

in which $k_0 \neq 0$ is the initial capital input, and $\xi_t, \xi_{t'}$ are independent for all distinct $t, t' \in \mathbb{Z}^+$.

Suppose w has a stationary utility function in her consumption c, $u(c)$ such that $u'(c) > 0$ and $u''(c) < 0$ for all $c \in \mathbb{R}^+$, and $\lim_{c\downarrow 0} u'(c) = +\infty$, then it clearly appears that $c_t > 0$ at any time $t \in \mathbb{Z}^+$. Assume the time preference of w can be characterized by a regular discounting $\rho \in (0, 1)$, then her additive utilities from a deterministic consumption flow $(c_0, c_1, \ldots, c_t)$ for $t \in \mathbb{Z}^+$, can be written as $\sum_{n=0}^{t} \rho^n u(c_n)$.

The steady growth path of the economy is thus determined by the equilibrium of the decision-making process for w, that is, w maximizes

$$\mathrm{E}_0 \sum_{t\in\mathbb{Z}^+} \rho^t u(c_t), \tag{1.29}$$

subject to $c_t = \xi_t f(k_t) - k_{t+1}$ for all $t \in \mathbb{Z}^+$ with $k_0 > 0$ initially given. Here, E_t denotes the expectation operator at a time $t \in \mathbb{Z}^+$.

Recall that an optimal consumption flow $(c_t,\ t \in \mathbb{Z}^+)$ should satisfy the following Euler equation,

$$u'(c_t) = \rho\,\mathrm{E}_t\left(\xi_{t+1} f'(k_{t+1}) u'(c_{t+1})\right). \tag{1.30}$$

Since $k_{t+1} = y_t - c_t$, (1.30) is equivalent to

$$u'(c_t) = \rho f'(y_t - c_t)\,\mathrm{E}_t\left(\xi_{t+1} u'(c_{t+1})\right), \tag{1.30$'$}$$

there should exist a real function $\varphi(y)$ such that $c_t = \varphi(y_t)$ for all c_t in the optimal consumption flow, which yields $k_{t+1} = y_t - \varphi(y_t)$, and thus

$$y_{t+1} = \xi_{t+1} f(k_{t+1}) = \xi_{t+1} f\big(y_t - \varphi(y_t)\big).$$

Let $\psi(y) = f\big(y - \varphi(y)\big)$, then we have the following stochastic growth process,

$$y_{t+1} = \xi_{t+1}\psi(y_t) \quad (t \in \mathbb{Z}^+). \tag{1.31}$$

Let $X_Y \subseteq \mathbb{R}^+$ be an invariant support set for y_t driven by the process (1.31), so that $y_t \in X_Y$ at any $t \in \mathbb{Z}^+$. Define $g_i(y) = \lambda_i \psi(y)$ for all $y \in X_Y$. Let $G = \{g_i : i \in I_N\}$, then the stochastic growth sequence $(y_t,\ t \in \mathbb{Z}^+)$ as is determined by (1.31) should be equivalent to the iterated random function system (X_Y, G, π).

Corresponding to the optimal consumption flow, the following optimal capital flow would directly come out,

$$k_{t+1} = y_t - \varphi(y_t) = \xi_t f(k_t) - \varphi\big(\xi_t f(k_t)\big), \tag{1.32}$$

which can also be supposed to admit an invariant support set $X_K \subseteq \mathbb{R}^+$. Define

$$m_i(k) = \lambda_i f(k) - \varphi\big(\lambda_i f(k)\big),$$

and let $M = \{m_i : i \in I_N\}$, then we have another iterated random function system (X_K, M, π), which in a sense is conjugate to the above one (X_Y, G, π).

Example 1.6. Assume $I_N = \{a, b\}$, $f(k) = \sqrt[3]{k}$, and $u(c) = \log c$. Let's suppose $(\xi_t,\ t \in \mathbb{Z}^+)$ is a Bernoulli process with

$$\mathrm{Prob}(\xi_t = \lambda_a) = q, \quad \mathrm{Prob}(\xi_t = \lambda_b) = 1 - q,$$

where $q \in (0, 1)$, and

$$1/\lambda_a^2 < \lambda_b < 1 < \lambda_a < 1/\lambda_b.$$

It thus suggests that the shock is either positive or negative, while the negative shock would not make the economy vanish as $\lambda_b\lambda_a^2 > 1$, and the positive shock would not make it too expansive as $\lambda_a\lambda_b < 1$.

In the optimal consumption flow $(c_t,\ t \in \mathbb{Z}^+)$, we might see that $c_t = (1-\rho/3)y_t$, which yields $\varphi(y_t) = (1-\rho/3)y_t$, and thus the optimal capital flow is determined by the formula

$$k_{t+1} = \rho y_t/3 = \rho\xi_t\sqrt[3]{k_t}/3.$$

Let $\kappa_t = \log k_t$ for all $t \in \mathbb{Z}^+$, then we have

$$\kappa_{t+1} = \kappa_t/3 + \log\xi_t + \log(\rho/3), \tag{1.33}$$

which should have an invariant support interval $[\alpha,\beta]$.

We now have two affine functions,

$$\ell_a(\kappa) = \kappa/3 + \big(\log\lambda_a + \log(\rho/3)\big), \quad \ell_b(\kappa) = \kappa/3 + \big(\log\lambda_b + \log\rho/3)\big).$$

Let $\Lambda = \{\ell_a, \ell_b\}$, then $([\alpha,\beta],\Lambda)$ is an iterated function system. Notice that

$$\beta/3 + \big(\log\lambda_a + \log(\rho/3)\big) = \beta, \quad \alpha/3 + \big(\log\lambda_b + \log(\rho/3)\big) = \alpha,$$

so $\log\lambda_a + \log(\rho/3) = 2\beta/3$ and $\log\lambda_b + \log(\rho/3) = 2\alpha/3$, and thus $\ell_a(\kappa)$ and $\ell_b(\kappa)$ can be also written as

$$\ell_a(\kappa) = \kappa/3 + 2\beta/3, \quad \ell_b(\kappa) = \kappa/3 + 2\alpha/3,$$

where $\beta > \alpha$ because of $\lambda_a > \lambda_b$. Let $z = (\kappa-\alpha)/(\beta-\alpha)$, then Λ on $[\alpha,\beta]$ can be transformed into a pair of functions defined on $[0,1]$, *i.e.*,

$$Z = \{z/3, z/3 + 2/3\}.$$

It therefore appears that $([\alpha,\beta],\Lambda)$ is equivalent to the iterated function system $([0,1],Z)$. By Example 1.3, we know that the unique attractor of $([0,1],Z)$ is the Cantor ternary set, and therefore the attractor of $([\alpha,\beta],\Lambda)$ should be also a Cantor set, which then conveys that the dynamics of the optimal stochastic growth in the economy should be essentially chaotic.

CHAPTER 2

On Separability

In the last chapter, we studied a generic time-dependent group whose structure can be characterized by a dynamical system. However, the collective behavior in such a group was implicitly assumed to exist independently of its single agents, as if the group itself were an intentional individual. In this chapter, we shall take a different perspective to investigate collective behavior in some structured group. In particular, the structure of a group is assumed to be characterized by the preference relations of its agents, so that such a group is preference-dependent, and its collective behavior would now not naturally emerge, but be determined by some embedded aggregation rule.

Here, we will focus on the aggregation rule to investigate how collective behavior of a preference-dependent group could be derived from individual behaviors. For the reason of simplicity, we will study preference relations defined on a multi-dimensional domain (as a peculiarly simplified representation of group), but not mention at times that our study has been rooted in group.

1. Introduction

Consider a general domain Σ, on which a rational preference relation can be defined. Suppose Σ can be formalized as a product set,

$$\Sigma = \prod_{i=1}^{n} X_i \quad (n \in \mathbb{N}).$$

EXAMPLE 2.1. Let $\Sigma = X_1 \times X_2 \times \cdots \times X_n$, where $X_i \subseteq \mathbb{R}$ for all i, and $n \in \mathbb{N}$. $\sigma \in \Sigma$ can be a bundle of n different commodities, or an array of n independent stimulus variables affecting a certain attribute of an observable object.

Let $\Sigma = X \times X$, where X is a set of sure prospects, then any $(x_1, x_2) \in \Sigma$ can represent an uncertain prospect of being x_1 and x_2 with some allocated probabilities.

Let $\Sigma = X \times T$, where X is a set of events, and T is a time domain, then any $(x, t) \in \Sigma$ denotes a realization x at a specific time $t \in T$.

Let $\Sigma = X \times Q$, where $X \subseteq \mathbb{R}^+$ and $Q \subseteq [0, 1]$, then any $(x, q) \in \Sigma$ can represent such a gambling that a gambler wins x and loses $\sqrt{x}$ with probability q and $1 - q$, respectively.

We will then study rational preferences on such product sets, and in particular, be interested in such properties as separability and additivity per se in their utility (value) representations. Such investigations can be found in the books by Fishburn [**19**, Chapter 4 & 5], Krantz *et al.* [**34**, Chapter 6 & 7], Roberts [**54**, Chapter 5], and Wakker [**65**, Chapter II & III], as well as the contributions by Gorman [**26**], Debreu [**14**], Luce and Tukey [**39**], Fishburn [**20**], Karni and Safra [**32**], Bouyssou and Marchant [**10**], and many other authors.

A utility function $u : \Sigma \to \mathbb{R}$ is called an *additive conjoint representation* of a preference relation on $\Sigma = \prod_{i=1}^{n} X_i$, if for all $(x_1, x_2, \dots, x_n) \in \Sigma$,

$$u(x_1, x_2, \dots, x_n) = \sum_{i=1}^{n} u_i(x_i), \tag{2.1}$$

where $u_i : X_i \to \mathbb{R}$ for all i. A preference relation that admits an additive conjoint representation is then called *additively separable* or just *additive*.

More generally, a utility function $u : \Sigma \to \mathbb{R}$ is called a *separable representation* of a preference relation on $\Sigma = \prod_{i=1}^{n} X_i$, if there is a function $v : \mathbb{R}^n \to \mathbb{R}$, such that for all $(x_1, x_2, \dots, x_n) \in \Sigma$,

$$u(x_1, x_2, \dots, x_n) = v\big(u_1(x_1), u_2(x_2), \dots, u_n(x_n)\big), \tag{2.2}$$

where $u_i : X_i \to \mathbb{R}$ for all i. The aggregation operator v in a separable utility function is called a *separability rule*. As a special case, the summation operator in an additive utility function is an additive separability rule. We shall call a preference relation that admits a separable representation *generally separable* or just *separable*.

The structure of this chapter is arranged as follows. Section 2 aims to connect the existing studies to our proposed investigations. Four distinct axiomatic systems, which are respectively based on the solvability, the Thomsen condition, the double cancellation condition, and the stationarity, are constructed so that they can all sufficiently generate additive preference structures, with differential capabilities yet.

In Section 3, the Thomsen condition on the plane is studied from the perspective of web geometry. We first make a more intuitive definition for the Thomsen condition by employing a transformation operator in the web-covered plane. We next show a critical lemma stating that a preference relation satisfies the Thomsen condition if and only if its corresponding 3-web on the plane is hexagonal.

In Section 4, we could then show that the Thomsen condition is also a necessary condition for the additivity of a preference structure on a general domain. And finally, we state a theorem to provide an identification rule to check whether a separable preference structure is actually additive.

2. Axiomatic Systems

In this section, we will summarize some developed axiomatic systems that can sufficiently ensure separable preference structures. Before we state these results, we have to make some formal definitions.

Let $\gtrsim$ denote a binary relation on the product set $\Sigma = \prod_{i=1}^{n} X_i$. $\gtrsim$ is called

(i) *reflexive*, if $\sigma \gtrsim \sigma$ for all $\sigma \in \Sigma$,

(ii) *transitive*, if $\sigma_1 \gtrsim \sigma_2$ and $\sigma_2 \gtrsim \sigma_3$ imply $\sigma_1 \gtrsim \sigma_3$ for all $\sigma_1, \sigma_2, \sigma_3 \in \Sigma$,

(iii) *complete*, if either $\sigma_1 \gtrsim \sigma_2$ or $\sigma_2 \gtrsim \sigma_1$ for all $\sigma_1, \sigma_2 \in \Sigma$.

A *preference relation* $\gtrsim$ on Σ is essentially a binary relation on the same Σ.

Definition. A preference relation $\gtrsim$ on a domain Σ is *rational*, if it is reflexive, transitive, and complete.

If Σ is countable, or Σ is uncountable but $\gtrsim$ is dense on it, there should be an isomorphism $u : \Sigma \to \mathbb{R}$ such that $\gtrsim$ on Σ can be preserved by $\geq$ on $\mathbb{R}$, or in terms of utility theory, there is an equivalent utility function $u : \Sigma \to \mathbb{R}$, such that $\sigma_1 \gtrsim \sigma_2$ if and only if $u(\sigma_1) \geq u(\sigma_2)$ for all $\sigma_1, \sigma_2 \in \Sigma$.

Suppose $\gtrsim$ is a rational preference on $\Sigma = \prod_{i=1}^{n} X_i$. If $\sigma_1 \gtrsim \sigma_2$ as well as $\sigma_2 \gtrsim \sigma_1$, we shall write $\sigma_1 \sim \sigma_2$, where $\sim$ is called the indifference relation corresponding to $\gtrsim$. Let $\gtrsim_i$ denote the rational preference on X_i induced from $\gtrsim$ on Σ, and let $\gtrsim_{-i}$ denote the rational preference on $\prod_{j \neq i} X_j$ induced from $\gtrsim$ on Σ, where $i \in \{1, 2, \ldots, n\}$. For all $\sigma = (x_1, x_2, \ldots, x_n)$ in Σ, one can also write $\sigma = (x_i, x_{-i})$ for all $i \in \{1, 2, \ldots, n\}$, where x_{-i} denotes the tuple of all the ordered elements in σ except x_i.

Definition. A preference relation $\gtrsim$ on Σ is *independent*, if $(x_i, x_{-i}) \gtrsim (y_i, x_{-i})$ is equivalent to $x_i \gtrsim_i y_i$ for all $i \in \{1, 2, \ldots, n\}$.

Definition. A preference relation $\gtrsim$ on Σ is *solvable*, if $(x_i, x_{-i}) \gtrsim \sigma \gtrsim (y_i, x_{-i})$ implies there exists a $z_i \in X_i$ such that $(z_i, x_{-i}) \sim \sigma$ for all $i \in \{1, 2, \ldots, n\}$.

We can then construct a somewhat restricted axiomatic system for an additive preference structure,

Theorem 2.1. *If a preference relation $\gtrsim$ is rational, independent, and solvable on $\Sigma = \prod_{i=1}^{n} X_i$, then it is additive, and admits an additive conjoint representation.*

Sketch of Proof. By the independence condition, $(x_i, x_{-i}) \gtrsim (y_i, x_{-i})$ if and only if $x_i \gtrsim_i y_i$ for all i. Since $\gtrsim$ is solvable on Σ, $\gtrsim_i$ is dense on X_i. Thus there exists a utility function $u_i : X_i \to \mathbb{R}$ representing $\gtrsim_i$. By the solvability condition, each x_i is essential, so for all $\sigma \in \Sigma$, we can have $(x_i, x_{-i}) \gtrsim \sigma \gtrsim (x_i, y_{-i})$. Then we can find a z_{-i} such that $\sigma \sim (x_i, z_{-i})$ by applying the solvability condition $n-1$ times. So there also exists a function $u_{-i} : \prod_{j \neq i} X_j \to \mathbb{R}$ representing $\gtrsim_{-i}$. We hence obtain $2n$ functions, $u_1, u_2, \ldots, u_n$, and $u_{-1}, u_{-2}, \ldots, u_{-n}$. Note that u_{-i} must be composed of all the u_j's for $j \neq i$, otherwise some x_j's would not be essential. The utility function $u : \Sigma \to \mathbb{R}$ representing $\succsim$ is thus determined by n functions, $u_1, u_2, \ldots, u_n$. Notice that the role of u_i in u_{-j} and that of u_j in u_{-i} is inverse to each other for all $i \neq j$, but the function form of u is constant, thus we must have $u_i + u_j$ in u. It therefore comes to us that $u = \sum_{i=1}^{n} u_i$, which completes the proof. $\square$

Note that the condition of solvability is a rather strong axiom of separability, as any factor can be separated from the other $n - 1$ factors. Debreu [**14**] presented an axiomatic system with a weaker separability condition for preference relations on a two-dimensional domain. To follow this tradition and also to simplify our analysis, we will focus on two-dimensional domains in the remaining parts of this chapter.

Let $\Sigma = X_1 \times X_2$, where X_1 and X_2 are general topologically connected spaces, as was similarly assumed by Debreu [**14**].

Definition. A preference relation $\succsim$ on $X_1 \times X_2$ satisfies the *Thomsen condition*, if $(x_1, y_2) \sim (y_1, x_2)$ and $(y_1, z_2) \sim (z_1, y_2)$ imply $(x_1, z_2) \sim (z_1, x_2)$ for all $x_1, y_1, z_1 \in X_1$ and $x_2, y_2, z_2 \in X_2$.

If a preference relation $\succsim$ on $X_1 \times X_2$ is rational, independent, and satisfies the Thomsen condition, then it must have an additive structure, and could be represented by a certain additive utility function. Fishburn [**19**] generalized the two-dimensional domain in Debreu's axiomatic system to an n-dimensional domain for $n \geq 2$ finite, on which the general Thomsen condition can be defined by the relation between each pair of factors (see Theorem 5.5 in Fishburn [**19**], pp. 71–76). As a graphic illustration, Figure 2.1 shows a preference relation satisfying the Thomsen condition on the real plane $\mathbb{R}^2$.

Definition. A preference relation $\succsim$ on $X_1 \times X_2$ satisfies the *double cancellation condition*, if $(x_1, y_2) \succsim (y_1, x_2)$ and $(y_1, z_2) \succsim (z_1, y_2)$ imply $(x_1, z_2) \succsim (z_1, x_2)$ for all $x_1, y_1, z_1 \in X_1$ and $x_2, y_2, z_2 \in X_2$.

Luce and Tukey [**39**] proposed a similar axiomatic system to Debreu's, in which the Thomsen condition is replaced by the double cancellation condition. The double cancellation condition is slightly stronger than the Thomsen condition, but again weaker than the solvability condition. So the range of preference relations that can be captured by the axiomatic system in Theorem 2.1 is smaller than that proposed by Luce and Tukey, and even much smaller than that proposed by Debreu.

As a special case of Debreu's axiomatic system, Fishburn and Rubinstein [**21**] stated an axiomatic system on the domain $X \times T$, where $X \subseteq \mathbb{R}$, and T is the time domain, either discrete as $\mathbb{Z}$ or dense on $\mathbb{R}$. According to Debreu's theorem, if the time preference $\succsim$ on $X \times T$ is rational, independent, and satisfies the Thomsen condition, there should be an additive conjoint representation $\mu : X \times T \to \mathbb{R}$, such

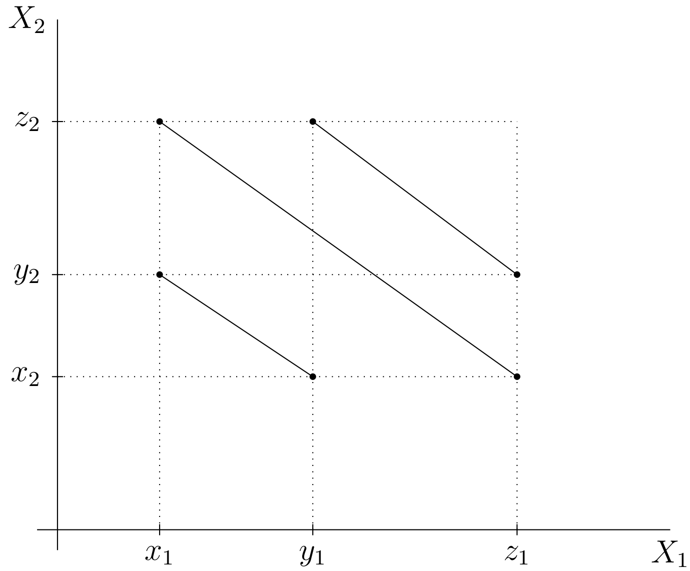


FIGURE 2.1.

that for all $(x,t) \in X \times T$,

$$\mu(x,t) = \nu(x) + \varrho(t), \tag{2.3}$$

where $\nu : X \to \mathbb{R}$ and $\varrho : T \to \mathbb{R}$.

Recall that an ordinal utility representation must be invariant to any positively monotonic transformation, thus we can have another equivalent utility representation $u(x,t) = \exp\big(\mu(x,t)\big)$, such that for all $(x,t) \in X \times T$,

$$u(x,t) = \rho(t)v(x), \tag{2.4}$$

where $\rho = \exp(\varrho)$ and $v = \exp(\nu)$.

DEFINITION. A time preference $\succsim$ on $X \times T$ is *stationary*, if $(x,t) \sim (y,s)$ implies $(x, t+\tau) \sim (y, s+\tau)$ for all $x, y \in X$ and all $t, s, t+\tau, s+\tau \in T$.

Notice that the stationarity condition means that the induced preference $\succsim_t$ on the time dimension (as a naïve time preference) is linear and independent of the dimension x, so it is much stronger than the solvability condition.

As shown by Fishburn and Rubinstein [**21**], if the time preference $\succsim$ on $X \times T$ is rational, independent, and stationary, the additive utility representation will be quasi-linear, *viz.*, $\varrho(t) = \alpha t$ for $\alpha < 0$. Thus $\mu(x,t) = \nu(x) + \alpha t$. Let $\beta = \mathrm{e}^{\alpha}$, then $\rho(t) = \exp(\alpha t) = \beta^t$, where $\beta \in (0,1)$, and hence the following exponentially scaled utility representation stands out:

$$u(x,t) = \beta^t v(x). \tag{2.4$'$}$$

Here, β is the discounting that measures the impatience of the time preference.

In sum, four axiomatic systems that are sufficient for additive representation of a preference relation on product sets have been presented, *viz.*, these ones based on the stationarity condition, the solvability condition, the double cancellation condition, and the Thomsen condition, respectively. Among them, the Thomsen condition is the weakest one, while the stationarity condition is the strongest one. Once we plan to study the necessary condition for the additivity and separability of preference relations on product sets, we should first study the Thomsen condition. In case the Thomsen condition is not necessary for an additive or separable preference structure, then all the other stronger ones could not be, either.

3. Thomsen Condition

In this section, we want to reconstruct the Thomsen condition by the theory of web geometry. We will again consider a two-dimensional domain $\Sigma = X_1 \times X_2$, where X_1 and X_2 are now general metric spaces.

Assume there always exists a diffeomorphism $f : \Sigma \to \mathbb{R}^2$ transforming Σ to an affine domain $\Sigma' = X \times Y$ of the plane $\mathbb{R}^2$. Under the diffeomorphism f, any point $(x_1, x_2) \in \Sigma$ is mapped uniquely to a corresponding point $(x, y) = f(x_1, x_2)$ in Σ', and also any smooth curve in Σ is transformed into a corresponding smooth curve in Σ'. A preference relation $\gtrsim$ on Σ is again a preference relation on Σ', as $\sigma_1 \gtrsim \sigma_2$ if and only if $f(\sigma_1) \gtrsim f(\sigma_2)$ for all $\sigma_1, \sigma_2 \in \Sigma$, where $f(\sigma_1), f(\sigma_2) \in \Sigma'$.[1] Therefore, the Thomsen condition is satisfied by $\gtrsim$ on the domain Σ, if and only if it can be satisfied by the same preference relation on the affine domain Σ'.

In $\Sigma' = X \times Y$, through a given point (x, y) there exist infinitely many smooth curves, each of which can be determined by a mapping $g : \Sigma' \to \mathbb{R}$, such that $g(x, y)$ is a constant for all (x, y) on a same curve. Thus any mapping $g : \Sigma' \to \mathbb{R}$ actually determines a family of smooth curves on Σ'. Let γ denote a regular family of smooth curves on Σ'. Suppose the collection of all the regular families of smooth curves on Σ' can be expressed as

$$\Gamma(\Sigma') = \{\gamma_i : i \in I\},$$

where I denotes an index set. The mapping that determines the family $\gamma_i \in \Gamma(\Sigma')$ is denoted by $g_i : \Sigma' \to \mathbb{R}$ for all $i \in I$. We thus have an equivalent collection of mappings $\{g_i : i \in I\}$, in other words, $\Gamma(\Sigma') \simeq \{g_i : i \in I\}$.

Definition. $\{\gamma_1, \gamma_2, \gamma_3\}$ is called a 3-*web* on Σ', if $\gamma_1, \gamma_2, \gamma_3 \in \Gamma(\Sigma')$.

[1] It means that an agent's preference relation as her subjective knowledge over Σ is shaped after her knowledge on the domain Σ and a generic transformation rule f on such a domain. So her preference relation can be kept over the domain $\Sigma' = f(\Sigma)$.

Since γ_i is solely determined by the mapping $g_i(x,y)$ for all $i \in I$, we can also use $\{g_1, g_2, g_3\}$ to denote a 3-web $\{\gamma_1, \gamma_2, \gamma_3\}$.

Example 2.2. Suppose

$$g_1(x,y) = x+y, \quad g_2(x,y) = x-y, \quad g_3(x,y) = xy,$$

then $\{x+y, x-y, xy\}$ is a 3-web on $X \times Y$.

Suppose

$$g_i(x,y) = x + a_i y \quad (i = 1,2,3),$$

where $a_1 \neq a_2 \neq a_3 \neq a_1$, then $\{x+a_1y, x+a_2y, x+a_3y\}$ is a linear 3-web on $X \times Y$, in which all the curves are lines.

A 3-web $\{\gamma_1, \gamma_2, \gamma_3\}$ is called *complete*, if

(i) there exists only one curve in the family γ_i passing any point $\sigma \in \Sigma'$,
(ii) any two distinct curves in the same family γ_i are disjoint,
(iii) any curve in a family γ_i has only one intersection with any curve in another family γ_j, for $i, j \in \{1,2,3\}$ distinct.

Note that in a complete 3-web $\{\gamma_1, \gamma_2, \gamma_3\}$, there exist exactly three different curves passing a given point in Σ'.

Now suppose there are three curves $L_i \in \gamma_i$ for $i = 1,2,3$ in a complete 3-web $\{\gamma_1, \gamma_2, \gamma_3\}$, such that $L_1 \cap L_2 \cap L_3 = \{\sigma\}$. For any point $\sigma' \neq \sigma$ on the curve L_i, we define a rule,

$$p_{ij} : L_i \to L_j, \tag{2.5}$$

such that a curve in γ_k passes both σ' and $p_{ij}(\sigma')$ for all distinct $i, j, k \in \{1,2,3\}$. Notice that $p_{ij} = p_{ji}^{-1}$, so $p_{ij} \circ p_{ji}$ is an identity function on L_i for all $j \neq i$.

Although $p_{ki} \circ p_{jk} \circ p_{ij}$ and $p_{ji} \circ p_{kj} \circ p_{ik}$ will both return back to a point in L_i, they are not necessarily identical.

For instance, if we start from a point $\sigma_1 \in L_1$, then

$$\begin{aligned} &p_{12}(\sigma_1) = \sigma_2 \in L_2, \quad p_{23}(\sigma_2) = \sigma_3 \in L_3, \quad p_{31}(\sigma_3) = \sigma_4 \in L_1, \\ &\qquad\qquad p_{12}(\sigma_4) = \sigma_5 \in L_2, \quad p_{23}(\sigma_5) = \sigma_6 \in L_3, \quad p_{31}(\sigma_6) = \sigma_7 \in L_1. \end{aligned}$$

We shall call the region shaped by the points $\sigma_1, \sigma_2, \sigma_3, \sigma_4, \sigma_5, \sigma_6, \sigma_7$ a *sequentially constructed hexagon*. If $\sigma_7 = \sigma_1$, such a hexagon is then called *closed*. In effect, we can see if $p_{31} \circ p_{23} \circ p_{12} = p_{21} \circ p_{32} \circ p_{13}$, then

$$p_{31} \circ p_{23} \circ p_{12} \circ p_{31} \circ p_{23} \circ p_{12} = p_{31} \circ p_{23} \circ p_{12} \circ (p_{21} \circ p_{32} \circ p_{13})^{-1}$$

will be an identify function, and thus any sequentially constructed hexagon will be closed. The next Figure 2.2 shows a closed sequentially constructed hexagon on the plane which is embedded with a linear 3-web.

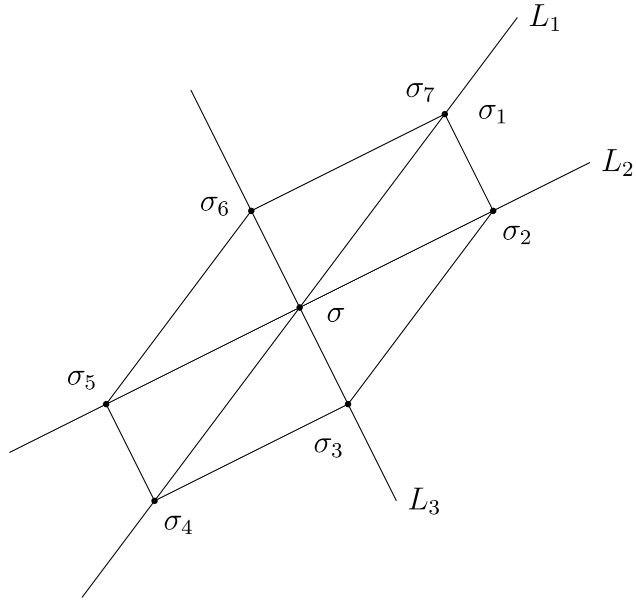


FIGURE 2.2.

DEFINITION. A 3-web on Σ' is called *hexagonal*, if any sequentially constructed hexagon in Σ' is closed.

Thomsen [**62**] proved that a planar 3-web is hexagonal if and only if it is equivalent to some linear 3-web on the plane. The Thomsen condition for a preference relation is proposed, largely because it is rather hard to geometrically identify equivalence conditions between different 3-webs before we know their algebraic representations.

LEMMA 2.2. *Assume a preference relation $\gtrsim$ on Σ' has a utility representation $u : \Sigma' \to \mathbb{R}$. $\gtrsim$ on Σ' satisfies the Thomsen condition, if and only if the* 3-web *$\{x, y, u(x, y)\}$ on Σ' is hexagonal.*

Let $\{x, y, w(x, y)\}$ be a complete 3-web on $\Sigma' = X \times Y$, where $w : \Sigma' \to \mathbb{R}$. Define a transformation operator $\bowtie$, such that for all (x_1, y_1) and (x_2, y_2) in Σ',

$$(x_1, y_1) \bowtie (x_2, y_2) = (x_1, y_2), \tag{2.6}$$

where $x_1, x_2 \in X$ and $y_1, y_2 \in Y$. Note that (x_1, y_2) and (x_1, y_1) are on a same curve $x = x_1$ in the family x, and (x_1, y_2) and (x_2, y_2) are on a same curve $y = y_2$ in the family y, then we can set rules so that (x_1, y_1) on a curve in the family y moves to (x_1, y_2) on a curve in the family $w(x, y)$, and moves to (x_2, y_2) on a curve in the family x. Thus (x_1, y_1), (x_1, y_2), and (x_2, y_2) will be consecutive in a sequentially constructed hexagon on $\{x, y, w(x, y)\}$.

Proof of Lemma 2.2. We have three regular families of smooth curves,

$$g_1(x,y)=x, \quad g_2(x,y)=y, \quad g_3(x,y)=u(x,y),$$

where x and y are coordinate bases of Σ', and $u(x,y)$ is determined by the preference relation $\gtrsim$ on Σ', thus $\{x,y,u(x,y)\}$ is a complete 3-web on Σ'.

If $\{x,y,u(x,y)\}$ on Σ' is hexagonal, then we need to show that $(x_1,y_2)\sim(x_2,y_1)$ and $(x_2,y_3)\sim(x_3,y_2)$ can sufficiently imply $(x_1,y_3)\sim(x_3,y_1)$ for all $x_1,x_2,x_3\in X$ and all $y_1,y_2,y_3\in Y$. Note that

$$(x_1,y_2)\bowtie(x_2,y_3)=(x_1,y_3), \quad (x_2,y_3)\sim(x_3,y_2),$$

and also

$$(x_3,y_2)\bowtie(x_2,y_1)=(x_3,y_1), \quad (x_2,y_1)\sim(x_1,y_2).$$

Thus the two sequences of consecutive points (x_1,y_2), (x_1,y_3), (x_2,y_3), and (x_3,y_2), as well as (x_3,y_2), (x_3,y_1), (x_2,y_1), and (x_1,y_2) are in certain sequentially constructed hexagons. Since $\{x,y,u(x,y)\}$ is hexagonal, those two sequences should form a closed hexagon in Σ'. Since $(x_2,y_3)\sim(x_3,y_2)$ and $(x_2,y_1)\sim(x_1,y_2)$, both pairs of points should be on same curves in the family $u(x,y)$. Thus the remaining two points, (x_1,y_3) and (x_3,y_1), must be also on a same curve in $u(x,y)$, as any closed hexagon is shaped by 3 curves in x, 3 curves in y, and 3 curves in $u(x,y)$. It now appears that $(x_1,y_3)\sim(x_3,y_1)$, and thus the Thomsen condition is satisfied by $\gtrsim$ on Σ', which completes the proof of the sufficient part.

If $\gtrsim$ satisfies the Thomsen condition on Σ', then

$$(x_1,y_2)\sim(x_2,y_1), \quad (x_2,y_3)\sim(x_3,y_2), \quad (x_1,y_3)\sim(x_3,y_1),$$

for all $x_1,x_2,x_3\in X$ and all $y_1,y_2,y_3\in Y$. We again have

$$(x_1,y_2)\bowtie(x_2,y_3)=(x_1,y_3), \quad (x_3,y_2)\bowtie(x_2,y_1)=(x_3,y_1).$$

So the sequence of (x_1,y_2), (x_1,y_3), and (x_2,y_3), and also that of (x_3,y_2), (x_3,y_1), and (x_2,y_1) are on sequentially constructed hexagons.

Note that (x_2,y_1) and (x_3,y_1) are on a same curve in y, and (x_2,y_1) and (x_1,y_2) are on a same curve in $u(x,y)$, so (x_3,y_1), (x_2,y_1), and (x_1,y_2) are consecutive points in a sequentially constructed hexagon. Similarly, (x_2,y_3), (x_3,y_2), and (x_3,y_1) are in a sequentially constructed hexagon as well. Thus the sequence of the seven points

$$(x_1,y_2), \quad (x_1,y_3), \quad (x_2,y_3), \quad (x_3,y_2), \quad (x_3,y_1), \quad (x_2,y_1), \quad (x_1,y_2)$$

form a closed sequentially constructed hexagon. Since x_1,x_2,x_3 and y_1,y_2,y_3 are all picked arbitrarily, the 3-web $\{x,y,u(x,y)\}$ must be hexagonal, which completes the proof of the necessary part. $\square$

By applying the transformation operator $\bowtie$ on Σ', we can have a more intuitive illustration of the Thomsen condition.

Let $\sigma_1 = (x_1, y_2)$, $\sigma_2 = (x_2, y_1)$, $\sigma_3 = (x_2, y_3)$, and $\sigma_4 = (x_3, y_2)$. So σ_1 and σ_4 are on the same curve $y = y_2$ in the family y, and σ_2 and σ_3 are on the same curve $x = x_2$ in the family x. A preference relation $\gtrsim$ on Σ' satisfies the Thomsen condition, if $\sigma_1 \sim \sigma_2$ and $\sigma_3 \sim \sigma_4$ imply

$$\sigma_1 \bowtie \sigma_3 \sim \sigma_4 \bowtie \sigma_2,$$

where $\sigma_1 \bowtie \sigma_3 = (x_1, y_3)$ and $\sigma_4 \bowtie \sigma_2 = (x_3, y_1)$. Its graphic illustration has been shown in Figure 2.3.

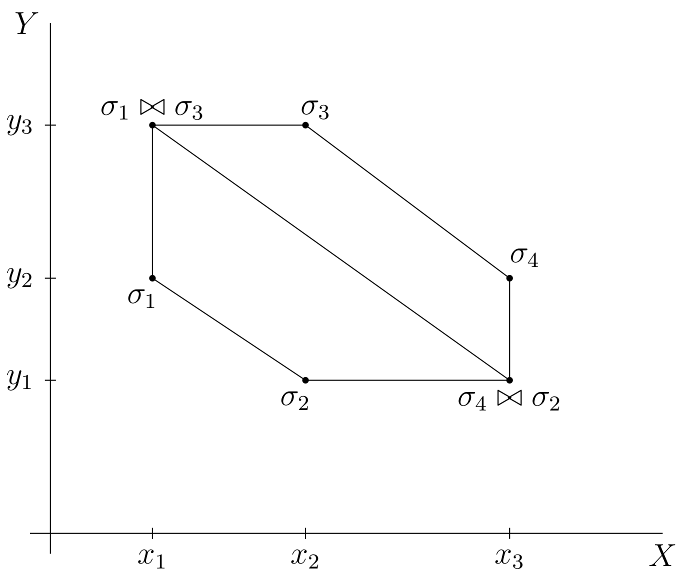


FIGURE 2.3.

Note that

$$\sigma_3 \bowtie \sigma_1 = \sigma_2 \bowtie \sigma_4 = (x_2, y_2),$$

and in addition,

$$\sigma_1 \bowtie \sigma_4 = \sigma_1, \quad \sigma_3 \bowtie \sigma_2 = \sigma_2, \quad \sigma_4 \bowtie \sigma_1 = \sigma_4, \quad \sigma_2 \bowtie \sigma_3 = \sigma_3.$$

Let $\delta \in \{\sigma_1, \sigma_2\}$ and $\delta' \in \{\sigma_1, \sigma_2\} \setminus \{\delta\}$. Let $\kappa \in \{\sigma_3, \sigma_4\}$ and $\kappa' \in \{\sigma_3, \sigma_4\} \setminus \{\kappa\}$. We therefore have a more general invariant relation between those two indifference sets $\{\sigma_1, \sigma_2\}$ and $\{\sigma_3, \sigma_4\}$: for all δ and κ,

$$\delta \bowtie \kappa \sim \kappa' \bowtie \delta', \tag{2.7}$$

by which the specific implication of the Thomsen condition can be clearly included.

4. Necessary Condition

The main result of this section is that the Thomsen condition on a general domain $\Sigma = X_1 \times X_2$, which is equivalent to $\Sigma' \subseteq \mathbb{R}^2$ through a diffeomorphism, is also necessary for a preference structure on Σ being additive. Thus a rational preference is additive on Σ can be totally captured by a pair of axioms, *viz.*, the independence condition and the Thomsen condition.

THEOREM 2.3. *If a preference relation $\succsim$ on $\Sigma' = X \times Y$ is additive, then $\succsim$ must satisfy the Thomsen condition.*

Note that any 3-web $\{\gamma_1, \gamma_2, \gamma_3\} \simeq \{x, y, u(x, y)\}$ on Σ' can represent a certain class of rational preferences on Σ', where $u : X \times Y \to \mathbb{R}$ serves as the utility function representing the family of indifference curves γ_3. A 3-web $\{x, y, u(x, y)\}$ is hexagonal, if and only if its curvature is zero (*cf.*, Akivis and Goldberg [**1**], p. 207). Recall that the curvature of the 3-web $\{x, y, u(x, y)\}$ can be defined as

$$k(u) = -\frac{1}{u_x u_y} \frac{\partial^2}{\partial x \partial y} \log(u_x/u_y), \tag{2.8}$$

where $u_x = \partial u/\partial x$ and $u_y = \partial u/\partial y$.

If $u(x, y)$ is additive, *i.e.*, $u(x, y) = u_1(x) + u_2(y)$, then

$$u_x/u_y = u_1'(x)/u_2'(y),$$

where $u_1' = \mathrm{d}u_1/\mathrm{d}x$ and $u_2' = \mathrm{d}u_2/\mathrm{d}y$, and thus its curvature $k(u) = 0$. However, if $u(x, y)$ is just in general separable, *i.e.*, $u(x, y) = v\big(u_1(x), u_2(y)\big)$, where $v : \mathbb{R}^2 \to \mathbb{R}$, $k(u)$ will not be necessarily zero. For instance, consider

$$u(x, y) = u_1(x) + \log\big(u_1(x) + u_2(y)\big), \tag{2.9}$$

in which $u_x/u_y = (u_1 + u_2 + 1)u_1'/u_2'$, then clearly, its curvature

$$k(u) = \frac{(u_1 + u_2)^2}{(u_1 + u_2 + 1)^3} \neq 0. \tag{2.10}$$

PROOF OF THEOREM 2.3. By the definition of additive preference relation, $\succsim$ on $\Sigma' = X \times Y$ admits at least one additive conjoint representation. Let $u(x, y) = u_1(x) + u_2(y)$ be such a utility representation, then $\{x, y, u_1(x) + u_2(y)\}$ is a 3-web on Σ'. Define a mapping g that maps any $(x, y) \in \Sigma'$ to $\big(u_1(x), u_2(y)\big)$, then it appears that g is bijective. Let the image of g be $\Sigma^* = Z_1 \times Z_2$. We then have an equivalent 3-web $\{z_1, z_2, z_1 + z_2\}$ on $Z_1 \times Z_2$. The linear 3-web $\{z_1, z_2, z_1 + z_2\}$ is hexagonal, so is $\{x, y, u_1(x) + u_2(y)\}$ on Σ'. By Lemma 2.2, $\succsim$ satisfies the Thomsen condition on Σ', which hence completes the proof. $\square$

Corollary 2.4. *If a preference relation $\succsim$ is additive on a domain Σ which could be transformed to an affine domain $\Sigma' \subseteq \mathbb{R}^2$ by some diffeomorphism, then $\succsim$ satisfies the Thomsen condition.*

Proof. Suppose there exists a diffeomorphism $f : \Sigma \to \mathbb{R}^2$, which transforms Σ to $\Sigma' = X \times Y$. The preference relation $\succsim$ will be again additive on Σ'. By Theorem 2.3, $\succsim$ satisfies the Thomsen condition on Σ', thus $\succsim$ also satisfies the Thomsen condition on Σ, as f is bijective between Σ and Σ'. $\square$

As a result, if a rational preference relation $\succsim$ is independent on a general domain $\Sigma = X_1 \times X_2$, the Thomsen condition is not only sufficient but also necessary for its additive structure. By the discussion above based on the curvature, we should see that the Thomsen condition on Σ' is too strong for its separability. However, there still exists some general relation between the separability and the Thomsen condition. In effect, we will next show that a separable preference structure is additive if and only if its separability rule can represent an additive preference structure on a proper subset of $\mathbb{R}^2$.

Theorem 2.5. *A separable preference relation $\succsim$ on $\Sigma = X_1 \times X_2$ with some separability rule $v : \mathbb{R}^2 \to \mathbb{R}$ is additive, if and only if v can represent a preference relation on $u_1(X_1) \times u_2(X_2)$ satisfying the Thomsen condition, where $u_i : X_i \to \mathbb{R}$ for $i = 1, 2$.*

Proof. If $v : \mathbb{R}^2 \to \mathbb{R}$ can represent a preference relation $\succsim'$ on the domain $\Sigma' = u_1(X_1) \times u_2(X_2)$ satisfying the Thomsen condition, then by Debreu's theorem [**14**], $\succsim'$ must be additive on Σ'. Thus $\succsim'$ on Σ' admits an additive conjoint representation, $v_1(x) + v_2(y)$ for all $(x, y) \in \Sigma'$, where $v_i : u_i(X_i) \to \mathbb{R}$ for $i = 1, 2$. There then must be a positively monotonic function $g : \mathbb{R} \to \mathbb{R}$, such that $v(x, y) = g\big(v_1(x) + v_2(y)\big)$ can also represent $\succsim'$ on Σ' for all $(x, y) \in \Sigma'$.

Note that for all $x = u_1(x_1)$ and $y = u_2(x_2)$, we have the utility representation for $\succsim$,

$$g\big(v_1 \circ u_1(x_1) + v_2 \circ u_2(x_2)\big) = v(u_1(x_1), u_2(x_2)),$$

where $(x_1, x_2) \in \Sigma$. Scaling $v(u_1(x_1), u_2(x_2))$ by the positively monotonic function g^{-1}, we obtain another utility representation for $\succsim$,

$$u(x_1, x_2) = v_1 \circ u_1(x_1) + v_2 \circ u_2(x_2),$$

which suggests that $\succsim$ is actually additive on Σ.

On the other hand, if $\succsim$ is separable on Σ with a separability rule v, then $\succsim$ should admit the specific utility representation $u(x_1, x_2) = v\big(u_1(x_1), u_2(x_2)\big)$, where $(x_1, x_2) \in X_1 \times X_2$. Suppose v is not additive on $\Sigma' = u_1(X_1) \times u_2(X_2)$, then $v(x, y)$, where $(x, y) \in \Sigma'$, can not be any function like $v_1(x) + v_2(y)$ up to all positively monotonic transformation. But it would imply $\succsim$ on Σ can not admit any utility function like $v_1 \circ u_1(x_1) + v_2 \circ u_2(x_2)$ up to all positively monotonic transformation, which means $\succsim$ can not be additive, a contradiction. Therefore, v must be additive on $u_1(X_1) \times u_2(X_2)$, and by Theorem 2.3, it should satisfy the Thomsen condition. $\square$

EXAMPLE 2.3. Consider the Cobb-Douglas utility function $u(x_1, x_2) = x_1^\alpha x_2^\beta$ on the domain $\Sigma = \mathbb{R}^+ \times \mathbb{R}^+$, where $\alpha, \beta \in (0, 1)$. It can be equivalently expressed as

$$u(x_1, x_2) = \exp(\alpha \log x_1 + \beta \log x_2), \tag{2.11}$$

where $(x_1, x_2) \in \Sigma$. Note that $\exp(x + y)$ is additive on $\Sigma' = \log(X_1) \times \log(X_2)$, as it is equivalent to $x + y$ on Σ' by the positively monotonic transformation "log". Thus, the Cobb-Douglas utility function represents an additive preference relation on Σ. Evidently, we know it is equivalent to the utility function $\alpha \log x_1 + \beta \log x_2$.

EXAMPLE 2.4. Consider a utility function

$$u(x_1, x_2) = \min\{\alpha u_1(x_1), \beta u_2(x_2)\} \tag{2.12}$$

on a general domain $\Sigma = X_1 \times X_2$, where $\alpha, \beta \neq 0$, and $u_i : X_i \to \mathbb{R}$ for $i = 1, 2$. It is a representation of the Leontief class of utility functions. We can observe that u is separable with a separability rule "min". However, it is not additive, as $\min\{x, y\}$ can not be additive on $u_1(X_1) \times u_2(X_2)$.

CHAPTER 3

Trading in Limit Order Market

We now move to study some concrete structured group, which usually exists as an institution in human society, either designed consciously or emerging spontaneously. For instance, a large group of traders has a structure characterized by the financial market shaped by such a group, and a group of governments has its essential structure characterized by a political network with hierarchical connections.

In this chapter, we shall study a specific equity market, the limit order market, and investigate its price formation and dynamics as the performance of its collective behavior. In the next chapter, the study will then be devoted to political network.

1. Introduction

This chapter[1] aims to investigate and understand the price dynamics in a generic limit order market, in which the best quotes, as well as the induced spread and mid-price, are evidently important information carriers for both traders and analysts. Early studies of prices formation in these financial markets mainly concentrated on their economic natures, for instance, Demsetz [**15**] considered the bid-ask spread as a markup "paid for predictable immediacy of exchange in organized markets" (Demsetz [**15**], p. 36), and regarded it as an important source in transaction costs. However, more recent studies started to explore the dynamics and evolutions appearing in these financial markets, partly because people might realize that understanding the processes and dynamics seems very likely more important and more useful than merely explaining the equilibrium states. And such a tendency seems to be necessary as well, since the real financial markets created a few catastrophes in the past century, say in particular, the crash of October 1929, and the crash of October 1987 (*cf.*, Sornette [**60**], pp. 5–7 & 12–15). The present investigations will also follow such a transformation and, as we have put at the very beginning, study trading processes in the limit order market.

Most modern stock exchanges adopt electronic order-driven platforms, in which limit order books operate to match demand with supply, and shape value for time and

[1]The author would like to acknowledge P. Z. Maymin, P. Pellizzari, M. Tolotti, and the reviewers of the journal Algorithmic Finance for their comments at different stages when writing and revising this chapter.

liquidity. They surely provide opportunities to obtain stylized empirical observations of the limit order market. There are plenty of investigations published in past few years. For instance, Lehmann and Modest [**36**] studied the trading mechanism and the liquidity in the Tokyo Stock Exchange, Biais, Hillion, and Spatt [**9**] studied the limit order book and the order flow in the Paris Bourse, Harris and Hasbrouck [**29**] measured the performance of the SuperDOT traders in the NYSE, and Al-Suhaibani and Kryzanowski [**2**] analyzed the order book and the order flow in the Saudi Stock Exchange.

As for theoretical understandings of the limit order market, the approaches of the existing studies might be roughly separated into two distinct categories, fortunately, which would also supply concrete cases in that methodological transformation as we mentioned before. Namely, one is to explain the performance of a limit order market and its stable states by modeling the behavior of traders in equilibrium. While the other is to consider the limit order market as a "super-trader" in order to either explain or predict its behavior statistically. The first category understands the strategic behavior of traders and generates testable implications by capturing some attributes of traders, for instance, being informed versus being uninformed as used by Kyle [**35**], and also Glosten and Milgrom [**25**], time preference used by Parlour [**51**], and patience versus impatience used by Foucault, Kadan, and Kandel [**22**]. The second category analyzes the limit order market by assuming there exist a few statistical laws for the dynamics or processes in the market. It is somewhat able to catch certain profiles of the market, say notably, fat tails of the price distribution, concavity of the price impact function, and the scaling law of spread to orders (see Smith *et al.* [**59**], and Farmer, Patelli, and Zovko [**18**]).

Here, we will take an intermediary position between these two quite different methodologies. That's to say, we not only agree that the performance of a limit order market and its evolution can be determined by the behavior of rational traders involved in the market, but also keep in mind that statistical mechanics could be fairly important in its dynamics and process. For this reason, we would suppose there is a large population of traders in a limit order market, and assume each trader in the population is rational, in the sense that her behavior is always strategically optimal in any game-theoretic framework. In practice, we assume that a rational trading decision in a very short time interval, as a quasi-static equilibrium, can be represented in terms of conformity with the optimal price distribution of the order book, which has a fine meaning of collective rationality. Evidently, the notions of individual rationality and collective rationality are thus supposed to be determined interchangeably.

In general, we will study how individual traders affect the price dynamics in the limit order market, how different trading blocks influence the stability of the market, how a piece of randomness in a trading system can stochastically enhance its systemic stability, and why the market would evolve more predictably, in case we can control certain factors of the market. In brief, it aims to clarify a dynamical trading mechanism in the limit order market.

The writing is organized as follows. The first section is a general introduction to the literature, the methodology, the structure, and some awaiting questions. And the last section is mixed of a summary and some additional remarks on instability and stochastic stability.

In Section 2, we first construct a basic modeling framework, and introduce atomic trading schemes as the necessary knowledge for traders in the limit order market. We then develop the switching laws for appearances of different types of traders, and show that the market capacities of accepting limit-type and market-type traders can be measured by floor functions of the log-scaled spread, and similarly the capacities of accepting buy-type and sell-type traders can be measured by floor functions of the log-scaled mid-price. These results are critical to the upcoming probabilistic setting of a random trading process.

In Section 3, we first study deterministic trading processes in a dynamical trading system from a combinatorial perspective. We recognize sufficient conditions for its general instability, and identify the necessary condition for its stability — any trading process should contain at least one reducible trading block. We next study stochastic trading processes with some certain probabilistic structures. We practically introduce two fundamental concepts — kernel region and buffering region — and show that the dynamical trading system will be stochastically stable in the kernel region, if its kernel region is moderately large and its buffering region is nonempty.

In Section 4, we would check the robustness of stochastic stability for a regular uncontrolled trading system, by setting its kernel region controlled to have some restricted properties. We show that the controlled dynamical trading system could be still stochastically stable, even if either its range of the spread or its range of the mid-price is extremely small. And thus, in a general sense, the stochastic stability of a regular trading system is robust.

2. Atomic Trading Scheme

2.1. Preliminary Framework. Consider a generic order-driven market with a large population of traders, whose attributes can be characterized by their trading directions and demands for the liquidity. As usual, the trading direction is either buying or selling initiation, while the liquidity demand determines the type of a submitted order, namely either a limit order or a market order.

The group of traders can thus be partitioned into four pairwise disjoint subgroups, each of which includes homogeneous traders, according to these two kinds of binary classification. It then appears that we have only four different types of traders:

(i) a buyer submitting a limit order,
(ii) a buyer submitting a market order,
(iii) a seller submitting a limit order,
(iv) a seller submitting a market order.

Let $\sigma(1)$, $\sigma(2)$, $\sigma(3)$, and $\sigma(4)$ denote these four types, where $\sigma : i \mapsto \sigma(i)$ is a normal permutation function defined on $\{1, 2, 3, 4\}$. Let the type space be Σ_4, then

$$\Sigma_4 = \{\sigma(1), \sigma(2), \sigma(3), \sigma(4)\}.$$

Evidently, any trader in the population must be of a unique type in Σ_4.

Assume that the depth of the limit order book is equal to the trading volume of any new order, thus a new transaction will either clear a limit order or add a more attractive limit order on the market, once there do not exist queued and hidden orders. So traders of any type in Σ_4 must affect the limit order book. We therefore call such traders *marginal traders*, in the sense that the (best) quotes will be definitely updated by their submitted orders.

Let the best bid, best ask, bid-ask spread, and mid-price in an order book be b, a, s, and m. The (best) bid-ask pair is denoted by (b, a) or $\mathbf{w}$. Note that $s = a - b$ and $m = (b + a)/2$. More generally, we define a *spread function*

$$s : \mathbf{w} \mapsto s(\mathbf{w}),$$

such that $s(\mathbf{w}) = a - b$, and define a *mid-price function*

$$m : \mathbf{w} \mapsto m(\mathbf{w}),$$

such that $m(\mathbf{w}) = (b + a)/2$, where $\mathbf{w} = (b, a)$.

Let the time domain be $\mathbb{Z}$. At each time $t \in \mathbb{Z}$, the time-dependent bid-ask pair will be denoted by (b_t, a_t) or $\mathbf{w}_t$. Once the time t is in the future, we introduce a two-dimensional random variable $\mathbf{D}_t$ to represent that unrealized bid-ask pair (B_t, A_t), where B_t and A_t are stochastic versions of b_t and a_t, respectively. Here, the normal notions of volatility and unpredictability could apply to B_t, A_t, and even $\mathbf{D}_t$.

Observe that there typically exists a tick size as the minimal change of prices in the market. Let $\tau > 0$ denote it, then we have $s > 0$, and moreover, $s \geq \tau$. Here, we shall set a stricter condition for the lower bound of the bid-ask spread, that's to say, there is a lower bound $\underline{s} > \tau$, such that $s \geq \underline{s}$ at any time.

Naturally, there also exists an upper bound of the best ask a, in virtue of the limited value of any security for all trader. Let $\overline{a}$ denote it. Besides that, note that $b \geq 0$, otherwise there would be no demand in the market, as the inverse of even the best bid would exist as a part of the ask side of the market. As a result, the bid-ask pair (b, a) should be located within a compact domain $W \subset \mathbb{R}^2$, which is defined by $b \geq 0$, $a \leq \overline{a}$, and $a - b \geq \underline{s}$. In the b-a plane, W can be geometrically represented by a triangle, whose vertices are $(0, \underline{s})$, $(0, \overline{a})$, and $(\overline{a} - \underline{s}, \overline{a})$, which has been shown in Figure 3.1.

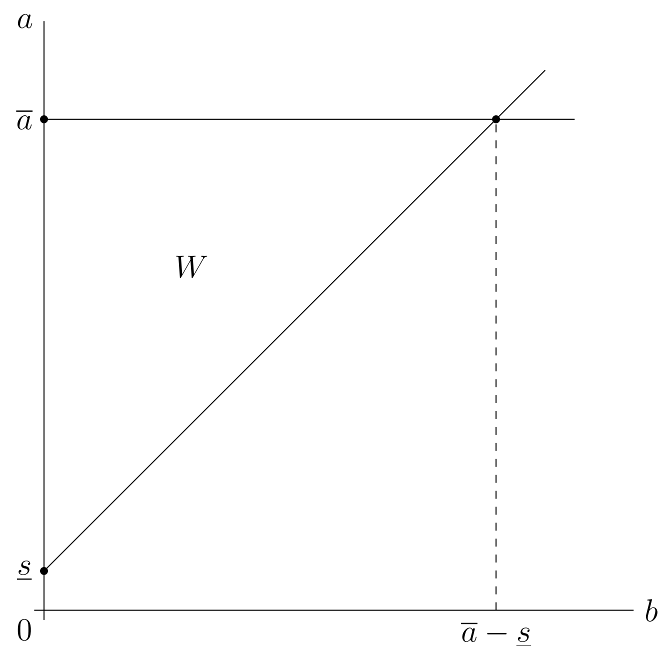


FIGURE 3.1.

Suppose that the ordered quotes on both sides of the order book are naïvely distributed over $[0, b]$ and $[a, \overline{a}]$. In addition, we assume that the difference between adjacent quotes on the same side of the order book should be proportional to the bid-ask spread. This assumption as an empirical fact was found and statistically tested by Biais, Hillion, and Spatt [**9**] using data from the Paris Bourse, and also supported by Al-Suhaibani and Kryzanowski [**2**] with evidence from the Saudi Stock Exchange. Furthermore, we set the difference of adjacent quotes equal to $\alpha(a-b)$ on the ask side of the book, and $\beta(a-b)$ on the bid side of the book, where $\alpha, \beta \in (0, 1)$.

The upper quote next to the best ask a is thus equal to $a^+ = a + \alpha(a - b)$, and the lower quote next to the best bid is equal to $b^- = b - \beta(a - b)$. Let's assume the ratio $\langle \beta, 1, \alpha \rangle$ of the differences between these four consecutive prices b^-, b, a, a^+ in the order book (see Figure 3.2) is an indicator of optimal information aggregation or market efficiency in a very short time interval. That's to say, prices, which conform to such a ratio, suggest that the market should be in a quasi-static equilibrium, so that there is no profitable perturbation that could emerge in that short time interval. In practice, the values of α and β can be roughly set as $1/2$, which was proposed by Biais, Hillion, and Spatt [**9**].

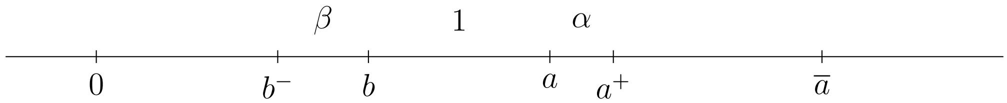


FIGURE 3.2.

2.2. Evolution of Bid-Ask Pair. Suppose that a marginal trader gets to the market at the time t, just after the bid-ask pair (b_t, a_t) forms, thus she has all the information on the market, at and before the time t. Her trading decision at t can be denoted by a price, say p_t. If she submits a limit order, then $b_t < p_t < a_t$, while if she submits a market order, we have either $p_t \geq a_t$ or $p_t \leq b_t$. If she is a buyer, then $p_t \leq a_t$, while if she is a seller, then $p_t \geq b_t$. Recall that we have assumed the order book's depth at the quotes is equal to the trading volume of any new order. So a market order will definitely clear one of the two limit orders at the best quotes, and the cleared previous limit order will be replaced by a less attractive one. And a limit order will surely lower the bid-ask spread, and improve one of the best quotes.

Type $\sigma(1)$. If the marginal trader is a buyer and submits a limit order, then the best bid b_{t+1} at the time $t+1$ will be $p_t = b_t + \rho$, where $\rho > 0$, and the best ask a_{t+1} at the time $t+1$ will remain unchanged. Suppose the marginal trader's decision process can be described by her maximizing the utility function of p_t or equivalently ρ, say $u(p_t) = u(\rho + b_t)$, where u is concave in ρ. By the requirement of rationality, we have $\rho^* \in \operatorname{argmax}_\rho u(p_t)$, subject to $\rho^* > 0$ and $a_t - b_t - \rho^* \geq \underline{s}$. According to our assumption on the optimal structure of the limit order book in a very short time interval around t, ρ^* should comply with the ratio $\langle \beta, 1 \rangle$ on the bid side, which hence implies

$$\rho^* = \beta(a_t - b_t - \rho^*), \tag{3.1}$$

so $\rho^* = \frac{\beta}{1+\beta}(a_t - b_t)$. Thus $b_{t+1} = \frac{1}{1+\beta} b_t + \frac{\beta}{1+\beta} a_t$, and $a_{t+1} = a_t$. Or $\mathbf{w}_{t+1} = \mathbf{w}_t S_1$, where

$$S_1 = \begin{pmatrix} \frac{1}{1+\beta} & 0 \\ \frac{\beta}{1+\beta} & 1 \end{pmatrix},$$

and $\mathbf{w}_{t+1}$ is the realization of the random variable $\mathbf{D}_{t+1}$.

After this trade of limit order on the bid side, the spread $s_t = a_t - b_t$ and mid-price $m_t = (b_t + a_t)/2$ will be updated respectively to

$$s_{t+1} = \frac{1}{1+\beta} s_t, \quad m_{t+1} = m_t + \frac{\beta}{2(1+\beta)} s_t.$$

Note that $s_{t+1} \geq \underline{s}$, so $s_t \geq (1+\beta)\underline{s}$. The marginal trader is of type $\sigma(1)$, only if $s_t \geq (1+\beta)\underline{s}$. And if she is of type $\sigma(1)$, she will choose an optimal improvement $\rho^* \geq \beta \underline{s}$.

Type $\sigma(2)$. If the marginal trader is again a buyer, but now she submits a market order hitting a_t, then the best ask a_{t+1} at the next period will be $a_t + \alpha(a_t - b_t)$, and the best bid b_{t+1} will remain same as b_t. Here, her decision process is overlaid with the principle of price-time priority employed in the limit order market. So her decision set is the singleton $\{a_t\}$, which then implies that $p_t = a_t$. Consequently, we have $b_{t+1} = b_t$ and $a_{t+1} = -\alpha b_t + (1+\alpha)a_t$. Or $\mathbf{w}_{t+1} = \mathbf{w}_t S_2$, where

$$S_2 = \begin{pmatrix} 1 & -\alpha \\ 0 & 1+\alpha \end{pmatrix}.$$

After this trade of market order on the bid side, the new spread and mid-price will be

$$s_{t+1} = (1+\alpha)s_t, \quad m_{t+1} = m_t + \frac{\alpha}{2}s_t.$$

Note that $s_{t+1} \le \overline{a}$, so $s_t \le \overline{a}/(1+\alpha)$. The marginal trader is of type $\sigma(2)$, only if $s_t \le \overline{a}/(1+\alpha)$.

Type $\sigma(3)$. If the marginal trader is a seller, and she submits a limit order, then the best ask a_{t+1} at the time $t+1$ will be $a_t - \theta$, where $\theta > 0$, and the best bid b_{t+1} at the time $t+1$ will remain same as b_t. Similar to the type $\sigma(1)$, this type of marginal trader's decision process can be described as maximizing her utility $v(p_t) = v(-\theta + a_t)$, where v is convex in θ. Her decision will admit the optimal choice $\theta^* \in \operatorname{argmax}_\theta v(p_t)$, subject to $\theta^* > 0$ and $a_t - \theta^* - b_t \ge \underline{s}$. By the assumption that the optimal ratio $\langle 1, \alpha\rangle$ on the ask side represents the quasi-static equilibrium around the time t in the market, we have

$$\theta^* = \alpha(a_t - \theta^* - b_t), \tag{3.2}$$

so $\theta^* = \frac{\alpha}{1+\alpha}(a_t - b_t)$. Thus $b_{t+1} = b_t$, and $a_{t+1} = \frac{\alpha}{1+\alpha}b_t + \frac{1}{1+\alpha}a_t$. Or more concisely, $\mathbf{w}_{t+1} = \mathbf{w}_t S_3$, where

$$S_3 = \begin{pmatrix} 1 & \frac{\alpha}{1+\alpha} \\ 0 & \frac{1}{1+\alpha} \end{pmatrix}.$$

After this trade of limit order on the ask side, the spread and mid-price will be updated to

$$s_{t+1} = \frac{1}{1+\alpha}s_t, \quad m_{t+1} = m_t - \frac{\alpha}{2(1+\alpha)}s_t.$$

Similar to the type $\sigma(1)$, we have $s_t \ge (1+\alpha)\underline{s}$, since s_{t+1} can not be less than $\underline{s}$. The marginal trader is of type $\sigma(3)$, only if $s_t \ge (1+\alpha)\underline{s}$. And if she is of type $\sigma(3)$, she will choose an optimal improvement $\theta^* \ge \alpha\underline{s}$.

Type $\sigma(4)$. If the marginal trader is a seller and submits a market order hitting b_t, then the best bid b_{t+1} at the time $t+1$ will be $b_t - \beta(a_t - b_t)$, and the best ask a_{t+1} will remain unchanged as a_t. Her decision is restricted to choosing p_t to maximize her utility subject to $p_t \in \{b_t\}$, so the optimal choice is $p_t = b_t$. We have $b_{t+1} = (1+\beta)b_t - \beta a_t$ and $a_{t+1} = a_t$. Or $\mathbf{w}_{t+1} = \mathbf{w}_t S_4$, where

$$S_4 = \begin{pmatrix} 1+\beta & 0 \\ -\beta & 1 \end{pmatrix}.$$

After this trade of market order on the ask side, we will have

$$s_{t+1} = (1+\beta)s_t, \quad m_{t+1} = m_t - \frac{\beta}{2}s_t.$$

Similar to the type $\sigma(2)$, we should have $s_t \le \overline{a}/(1+\beta)$. And only if $s_t \le \overline{a}/(1+\beta)$, the marginal trader could be of type $\sigma(4)$.

Type Space. Observe that any spread s can increase to $(1+\alpha)s$ or $(1+\beta)s$, and decrease to $s/(1+\alpha)$ or $s/(1+\beta)$ after a new order, so the maximal difference generated by α and β could be $|\alpha - \beta|s$. Since $\overline{a} \gg \underline{s}$ in most normal limit order markets, and both α and β are roughly close to $1/2$, $|\alpha - \beta|s$ will be extremely small in regard to $\overline{a} - \underline{s}$. Therefore, we can, without loss of generality, assume $\alpha = \beta$ to set $|\alpha - \beta|s$ exactly equal to 0. The ratio of $\langle \beta, 1, \alpha \rangle$ will then be replaced by the simpler one, $\langle \alpha, 1, \alpha \rangle$.

Consider an arbitrary initial bid-ask pair (b, a). If a marginal trader of type $\sigma(1)$ comes to the market, the bid-ask pair in the next period will be (b^+, a), where $b^+ > b$. If the marginal trader is of type $\sigma(2)$, it will be (b, a^+), where $a^+ > a$. If the marginal trader is of type $\sigma(3)$, it will be (b, a^-), where $a^- < a$. Finally, if the marginal trader is of type $\sigma(4)$, it will be (b^-, a), where $b^- < b$. Note that

$$b^+ - b = a - a^- = \frac{\alpha}{1+\alpha}s, \quad a - b^+ = a^- - b = \frac{s}{1+\alpha},$$

and also

$$a^+ - a = b - b^- = \alpha s, \quad a^+ - b = a - b^- = (1+\alpha)s.$$

The types $\sigma(2)$ and $\sigma(4)$ will cause a greater bid-ask spread than s, namely $(1+\alpha)s$, while the types $\sigma(1)$ and $\sigma(3)$ will cause a smaller bid-ask spread than s, namely $s/(1+\alpha)$. In fact, $\sigma(2)$ and $\sigma(4)$ are both market-type, while $\sigma(1)$ and $\sigma(3)$ are both limit-type (see Figure 3.3).

By similar computations as above, we can obtain

$$a + b^+ = 2m + \frac{\alpha}{1+\alpha}s, \quad a^+ + b = 2m + \alpha s,$$

and also

$$a^- + b = 2m - \frac{\alpha}{1+\alpha}s, \quad a + b^- = 2m - \alpha s.$$

The types $\sigma(1)$ and $\sigma(2)$ will generate a new mid-price greater than m, while the types $\sigma(3)$ and $\sigma(4)$ will generate a new mid-price smaller than m. We clearly see that $\sigma(1)$ and $\sigma(2)$ are both buy-type, while $\sigma(3)$ and $\sigma(4)$ are both sell-type (see again Figure 3.3).

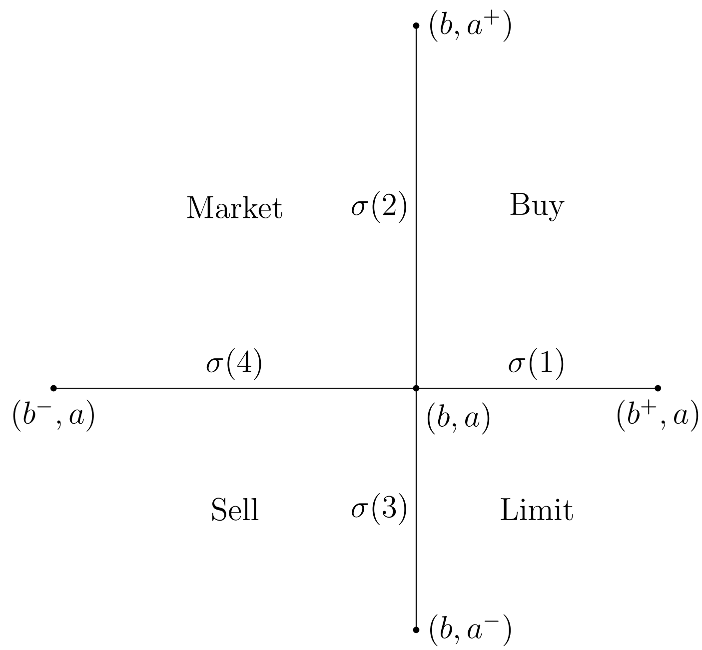


FIGURE 3.3.

2.3. Capacity and Switching Law. From the discussions on trading decisions of different marginal traders, we can see how the bid-ask spread and mid-price develop in an iterative way. Notice that a limit order will change the bid-ask spread from s to $s/(1+\alpha)$, while a market order will change it from s to $(1+\alpha)s$. It appears that any s will converge to infinity after sufficiently many market orders, and any s will converge to zero after infinitely many limit orders. Recall that s should be bounded within the interval $[\underline{s}, \overline{a}]$ (a line $a = b + s$ should pass W, see Figure 3.1), so when s is too close to its bounds, the market should have stability incentives to make s bounce away against them. The mid-price m has a similar internal stability scheme, as m should be bounded within the interval $[\underline{s}/2, \overline{a} - \underline{s}/2]$ (a line $a + b = 2m$ should pass W, see also Figure 3.1). Quite intuitively, we shall say such internal stability schemes originate in some hidden "gravitational forces" in the market, in the sense that they can control the appearances of different types of orders, and determine their switching possibilities.

We should admit that the force related to the bid-ask spread and that related to the mid-price should have different actions on the market, but similar analytical natures. In consideration of such facts, we will mainly study the force related to the bid-ask spread, and yet we will directly state similar results on the force related to the mid-price at the end of this subsection.

To investigate the gravitational force for the bid-ask spread, we first concentrate on such spreads that are very close to either $\underline{s}$ or $\overline{a}$, so that we could easily see its working principles.

If the spread is sufficiently great, the hidden force will attract limit-type traders, $\sigma(1)$ and $\sigma(3)$, and repel market-type traders, $\sigma(2)$ and $\sigma(4)$, but of course doesn't necessarily reject them until the spread is close enough to the upper bound $\overline{a}$.

On the other hand, if the bid-ask spread is extremely small, then such a hidden force will attract market-type traders and repel limit-type traders, and it will reject limit-type traders once the spread is close enough to $\underline{s}$.

Formally stating, the market with a spread s only accepts traders of type $\sigma(2)$ and type $\sigma(4)$, if

$$\underline{s} \le s < (1+\alpha)\underline{s}.$$

Suppose there exists a nonempty interval $((1-\gamma)\overline{a}, \overline{a}]$, where $0 < \gamma < 1$ and $(1-\gamma)(1+\alpha) \le 1$[2], or equivalently,

$$\alpha/(1+\alpha) \le \gamma < 1,$$

under which the market will never accept any more market-type trader. Thus the market with a spread s only accepts traders of type $\sigma(1)$ and type $\sigma(3)$, if

$$(1-\gamma)\overline{a} < s \le \overline{a}.$$

In the remaining interval,

$$(1+\alpha)\underline{s} \le s \le (1-\gamma)\overline{a},$$

the market can accept traders of any type in Σ_4. In fact, $[\underline{s}, \overline{a}]$ is now partitioned into three pairwise disjoint intervals, as shown in Figure 3.4.

FIGURE 3.4.

[2]This inequality gives a necessary condition for a zero-capacity of accepting market-type traders. If we assume $(1-\gamma)(1+\alpha) = 1$, although we could have less parameters, there would be also less room for interesting analysis.

We then establish a result stating that the capacity of accepting limit-type traders in the limit order market can be measured by a function of the bid-ask spread.

PROPOSITION 3.1. *The maximal number of limit-type traders that can be accepted continuously by a limit order market with a spread s, is determined by the function*

$$z(s) = \left\lfloor \frac{\log s - \log \underline{s}}{\log(1+\alpha)} \right\rfloor . \tag{3.3}$$

PROOF. Define a sequence of consecutive intervals,

$$[(1+\alpha)^i \underline{s}, (1+\alpha)^{i+1}\underline{s}),$$

for $i \in \{0, 1, \ldots, n\}$, and

$$n = \max\{i \in \mathbb{Z} : (1+\alpha)^i \leq \overline{a}\} - 1.$$

For all $s \in [\underline{s}, (1+\alpha)^{n+1}\underline{s})$, there is a unique $j(s) \in \{0, 1, \ldots, n\}$, such that

$$s \in [(1+\alpha)^{j(s)}\underline{s}, (1+\alpha)^{j(s)+1}\underline{s}).$$

We want to show that $z(s) = j(s)$ by induction. If $j(s) = 0$, then $s \in [\underline{s}, (1+\alpha)\underline{s})$, and the limit order market will reject limit-type orders, so $z(s) = 0$. Assume $z(s) = j(s)$ for all $j(s) \leq k$, and consider $j(s) = k+1$ such that $s \in [(1+\alpha)^{k+1}\underline{s}, (1+\alpha)^{k+2}\underline{s})$. After a marginal trader of type $\sigma(1)$ or type $\sigma(3)$ comes to the market, s will be updated to $s' = s/(1+\alpha) \in [(1+\alpha)^k\underline{s}, (1+\alpha)^{k+1}\underline{s})$. By the assumption, we know $z(s') = k$, so $z(s) = z(s') + 1 = k + 1 = j(s)$.

If $s \in [(1+\alpha)^{n+1}\underline{s}, \overline{a}]$, then $s \in [(1+\alpha)^{n+1}\underline{s}, (1+\alpha)^{n+2}\underline{s})$, as $(1+\alpha)^{n+2}\underline{s} > \overline{a}$. Here $j(s) = n + 1$, so we have $z(s) = n + 1 = j(s)$ by induction, simply as $z(s') = j(s')$ for all $j(s') = n$.

Therefore, $z(s)$ satisfies

$$(1+\alpha)^{z(s)}\underline{s} \leq s < (1+\alpha)^{z(s)+1}\underline{s},$$

and thus we have

$$z(s) \leq \frac{\log s - \log \underline{s}}{\log(1+\alpha)} < z(s) + 1,$$

which exactly defines the floor function. □

We can thus state that there is an exponential law of the bid-ask spread s to the market's capacity of accepting limit-type traders n,

$$s = (1+\alpha)^n \underline{s}, \tag{3.4}$$

where $n \in [z(s), z(s) + 1)$.

To a certain extent, the appearance probability of a new limit-type trader should be simply determined by its capacity of accepting limit-type traders. Since the capacity is a function of the spread, that probability should be also a function of the spread. Let $f : [\underline{s}, \overline{a}] \to [0,1]$ be such a function[3]. Since the log-scaled spread has been used to measure the capacity, $f(s)$ is actually a function of $\log s$, and moreover, we suppose $f(s)$ is positively linear in $\log s$. Thus we could define $f(s)$ more precisely in the following way. If $s \in [(1+\alpha)\underline{s}, (1-\gamma)\overline{a}]$, then

$$f(s) = k_1 \log s + k_2, \tag{3.5}$$

where $k_1 \geq 0$ and k_2 are constants depending on $\underline{s}, \overline{a}, \alpha, \gamma$. If $s \in [\underline{s}, (1+\alpha)\underline{s})$, then $f(s) = 0$. And if $s \in ((1-\gamma)\overline{a}, \overline{a}]$, then $f(s) = 1$.

Similar to Proposition 3.1, a result on the capacity of accepting market-type traders in the market can be proposed, and yet its proof will not be provided, as its logic is nearly the same as that of Proposition 3.1.

PROPOSITION 3.2. *The maximal number of market-type traders which can be accepted continuously by a limit order market with a spread s, is determined by the function*

$$y(s) = \left\lfloor \frac{\log((1-\gamma)\overline{a}) - \log s}{\log(1+\alpha)} + 1 \right\rfloor^{+}, \tag{3.6}$$

where $\lfloor x \rfloor^{+} = \max\{\lfloor x \rfloor, 0\}$.

Evidently, there is also an exponential law of the bid-ask spread s to the market's capacity of accepting market-type traders n,

$$s \propto (1+\alpha)^{-n}\overline{a}, \tag{3.7}$$

where $n \in [y(s), y(s)+1)$.

The appearance probability function $g : [\underline{s}, \overline{a}] \to [0,1]$, which assigns a probability of a new market-type trader appearing in the market to any spread s, can be defined as a negatively linear function of $\log s$. Concretely, if $s \in [(1+\alpha)\underline{s}, (1-\gamma)\overline{a}]$, then

$$g(s) = k_3 \log s + k_4, \tag{3.8}$$

where $k_3 \leq 0$ and k_4 are again constants determined by $\underline{s}, \overline{a}, \alpha, \gamma$. If $s \in ((1-\gamma)\overline{a}, \overline{a}]$, then $g(s) = 0$. And if $s \in [\underline{s}, (1+\alpha)\underline{s})$, then $g(s) = 1$.

Note that, in any case, a new trader appearing in the market is either limit-type or market-type, that's to say, we have $f(s) + g(s) = 1$ for all s. Intuitively, the probabilities $f(s)$ as well as $g(s)$ at any s can be thought of to be a measure of switching possibility between the limit-type and market-type traders in the market. We can thus rigorously state such a switching law between limit type and market type as follows:

[3]Note that $f(s)$ is not a probability measure over the spread domain $[\underline{s}, \overline{a}]$, as we are not considering the uncertainty in the spread, but the uncertainty in the appearance of different types of traders at any deterministic spread s.

(i) If $s \in [\underline{s}, (1+\alpha)\underline{s})$, the probability of switching from limit type to market type is 1, and that of switching from market type to limit type is 0.
(ii) If $s \in ((1-\gamma)\overline{a}, \overline{a}]$, the probability of switching from market type to limit type is 1, and that of switching from limit type to market type is 0.
(iii) If $s \in [(1+\alpha)\underline{s}, (1-\gamma)\overline{a}]$, the probability of switching from limit type to market type is decreasing according to $g(s)$, while the probability of switching from market type to limit type is increasing according to $f(s)$.

As for the mid-price as the counterpart of the spread, we can also develop, through a highly similar reasoning process, the switching law between buy type (*i.e.*, type $\sigma(1)$ and $\sigma(2)$) and sell type (*i.e.*, type $\sigma(3)$ and $\sigma(4)$). Once again, we partition the mid-price domain $[\underline{s}/2, \overline{a} - \underline{s}/2]$ into three pairwise disjoint intervals,

$$[\underline{s}/2, (1+\delta)\underline{s}/2), \quad [(1+\delta)\underline{s}/2, (1-\epsilon)(\overline{a} - \underline{s}/2)], \quad ((1-\epsilon)(\overline{a} - \underline{s}/2), \overline{a} - \underline{s}/2],$$

where $\delta > 0$ and $\delta/(1+\delta) \leq \epsilon < 1$, as we need initially assume $0 < \epsilon < 1$ and $(1+\delta)(1-\epsilon) \leq 1$. In short, we shall directly state that switching law between buy type and sell type as follows:

(i) If $m \in [\underline{s}/2, (1+\delta)\underline{s}/2)$, sell type will switch to buy type for sure.
(ii) If $m \in ((1-\epsilon)(\overline{a} - \underline{s}/2), \overline{a} - \underline{s}/2]$, buy type will switch to sell type for sure.
(iii) In the remaining domain of m, the probability of switching from buy type to sell type is increasing with $\log m$, and that of switching from sell type to buy type is decreasing with $\log m$.

3. Iterated Trading Process

3.1. Sequential Trading. Define four linear functions mapping W to itself,

$$f_i(\mathbf{w}) = \mathbf{w}S_i \quad (i \in \{1, 2, 3, 4\}),$$

where S_1, S_2, S_3, S_4 are 2×2 matrices as defined in Section 2.2 of this chapter. Recall that β in S_1 and S_4 are now replaced by α. Let F be the collection of these four functions, then $F = \{f_1, f_2, f_3, f_4\}$.

For each $i \in \{1, 2, 3, 4\}$, and all given $\mathbf{w} \in W$, define a convex set

$$L_i(\mathbf{w}) = \{\lambda \mathbf{w} + (1-\lambda)\mathbf{w}S_i : 0 \leq \lambda \leq 1\}.$$

Notice that $L_i(\mathbf{w})$ is actually the line segment between $\mathbf{w}$ and $\mathbf{w}S_i$ in the b-a plane. More precisely, we redefine $f_i(\mathbf{w})$ to be

$$f_i(\mathbf{w}) \in \begin{cases} \{\mathbf{w}S_i\} & \text{if } \mathbf{w}S_i \in W \\ L_i(\mathbf{w}) \cap \partial W & \text{if } \mathbf{w}S_i \notin W \end{cases}$$

where ∂W denotes the boundary of the closed domain W. Since both $\{\mathbf{w}S_i\}$ and $L_i(\mathbf{w}) \cap \partial W$ are singletons, $f_i(\mathbf{w})$ will take either the value of $\mathbf{w}S_i$ or the unique element in $L_i(\mathbf{w}) \cap \partial W$, hence it is essentially a well-defined function.

We should notice that the domain W will then have an absorbing barrier, in such a sense that the dynamics induced by any $f_i \in F$ will be always bounded within W. To state the sharp distinction between a state absorbed on ∂W and a state in $W \setminus \partial W$, we shall propose two useful notions to describe the states in ∂W. If the state of a limit order market is first absorbed on ∂W at a time t, we say the market is *unstable* before the time t, and it is in a *crash* or *catastrophe* at and after the time t.

DEFINITION. The pair (W, f_i) is called a *trading system* generated by the trader of type $\sigma(i)$ for all $i \in \{1, 2, 3, 4\}$.

Each trading system (W, f_i) for $i \in \{1, 2, 3, 4\}$ can create a certain dynamics of bid-ask pairs in the domain W. For any initial state $\mathbf{w} \in W$, the bid-ask pair in the trading system (W, f_i) will eventually hit the point $\mathbf{w}_i \in \partial W$, where

$$\mathbf{w}_1 = (a - \underline{s}, a), \quad \mathbf{w}_2 = (b, \overline{a}), \quad \mathbf{w}_3 = (b, b + \underline{s}), \quad \mathbf{w}_4 = (0, a).$$

These generic states are also shown geometrically in the b-a plane (see Figure 3.5).

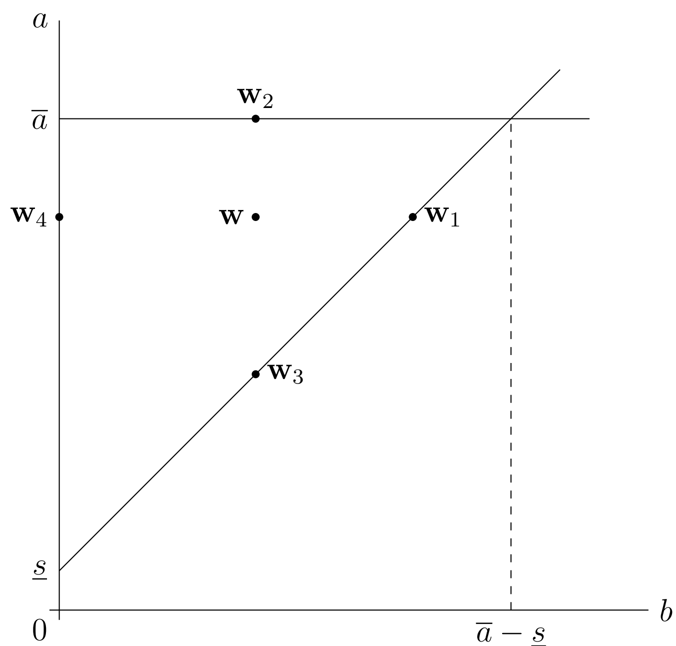


FIGURE 3.5.

PROPOSITION 3.3. *If a fixed-type marginal trader repeatedly comes to a limit order market, then the market starting from any initial state in W is unstable.*

Proof. Consider any initial state $\mathbf{w} = (b, a)$ in W. After n forward periods with the marginal trader of type $\sigma(1)$, the bid-ask pair will be $\mathbf{w}S_1^n$, which will converge to (a, a), if n goes to infinity. But in the trading system (W, f_1), the bid-ask pair should always stay in W, thus the last bid-ask pair remaining in the trading system is $(a - \underline{s}, a) \in \partial W$. It means that the trading system will monotonically move to a crash, and hence it is unstable. Similar arguments can be made for the other three types, and they will complete the proof. □

In Proposition 3.3, we actually consider a countably infinite sequence of marginal traders with a constant type, $\{q, q, \dots\}$ for all $q \in \Sigma_4$. It appears to us that the trading system involved with $\{q, q, \dots\}$ must be unstable. In general, we can consider a countably infinite sequence of marginal traders with variable types, $\{q_t,\ t \in \mathbb{Z}\}$, where $q_t \in \Sigma_4$ for all $t \in \mathbb{Z}$, and investigate the stability of the trading system involved with $\{q_t,\ t \in \mathbb{Z}\}$. Evidently, the set of all such $\{q_t,\ t \in \mathbb{Z}\}$ can be denoted by Σ_4^ω.

If the type of marginal trader at a time t is $q_t = \sigma(i)$, the atomic trading scheme functioning at that time will be f_i for all $i \in \{1, 2, 3, 4\}$. Thus the permutation function σ can relate the functioning scheme $f_i \in F$ to the marginal trader of type $\sigma(i) \in \Sigma_4$.

Definition. The triplet (W, F, σ) is called a *dynamical trading system* or an *iterated trading system*.

Since (W, f_i) is a discrete dynamical system for all $f_i \in F$, (W, F) is essentially an iterated function system. In the definition of dynamical trading system, σ is introduced additionally so as to determine the functioning schemes in (W, F) for all sequence of marginal traders in Σ_4^ω. Recall that F has a parameter $\alpha \in (0, 1)$, so the dynamical trading system (W, F, σ) depends on α as well.

First of all, we are interested in identifying certain trading blocks in a sequence of traders that will never change the state of the dynamical trading system (W, F, σ). Notice that $S_1S_4 = I$ and $S_2S_3 = I$ for all $\alpha \in (0, 1)$, where I denotes the identity matrix of order 2, thus $\{\sigma(1), \sigma(4)\}$, $\{\sigma(4), \sigma(1)\}$, $\{\sigma(2), \sigma(3)\}$, and $\{\sigma(3), \sigma(2)\}$ are all such trading blocks for all $\alpha \in (0, 1)$.

Definition. A *periodic block* is a consecutive trading block that will not change any bid-ask pair in a specific dynamical trading system.

Notice that any combination of periodic blocks is again periodic. For instance,

$$\{\sigma(1), \sigma(4), \sigma(2), \sigma(3)\}$$

is a periodic block for all $\alpha \in (0, 1)$, as $\{\sigma(1), \sigma(4)\}$ and $\{\sigma(2), \sigma(3)\}$ are both general periodic blocks. So we need to catch the invariant part of a periodic block.

Definition. A periodic block is called *minimal*, if it has no proper subtuple that is again a periodic block.

Any periodic block can be reduced into a series of minimal ones. Note that a periodic block C is either minimal or not. If C is minimal, it is equivalent to itself. If C is not minimal, we can always find a proper subtuple $C' \subset C$ such that both C' and $C \setminus C'$ are still periodic blocks. We can eventually have a series of minimal periodic blocks by applying this process recursively.

Example 3.1. If $\alpha \in (0,1)$, the periodic block

$$\{\sigma(1), \sigma(2), \sigma(3), \sigma(4)\}$$

has two minimal ones, $\{\sigma(2), \sigma(3)\}$ and $\{\sigma(1), \sigma(4)\}$, while the minimal periodic blocks of

$$\{\sigma(4), \sigma(1), \sigma(1), \sigma(4), \sigma(4), \sigma(1)\}$$

are $\{\sigma(1), \sigma(4)\}$ and double $\{\sigma(4), \sigma(1)\}$.

If $\alpha = 1/2$, the consecutive trading block

$$\{\sigma(2), \sigma(1), \sigma(2), \sigma(1), \sigma(3), \sigma(4), \sigma(3), \sigma(4), \sigma(3), \sigma(4)\}$$

is periodic, and it is also minimal.

If $\alpha = 1/3$,

$$\{\sigma(1), \sigma(2), \sigma(1), \sigma(2), \sigma(1), \sigma(2), \sigma(1), \sigma(2), \sigma(4), \sigma(3), \sigma(4), \sigma(3), \sigma(4), \sigma(3)\}$$

is a minimal periodic block.

Lemma 3.4. *The number of marginal traders in any minimal periodic block is finite and even.*

Proof. Consider a minimal periodic block C, and assume the number of traders in C is infinite. Then C must pass infinite bid-ask pairs. If not, we suppose C passes finite bid-ask pairs. Since the number of traders in C is infinite, there must exist a closed route, such that a related subset of C is a periodic block, which contradicts that C is minimal.

Assume the initial condition of C is $\mathbf{w} \in W$ with a bid-ask spread s. Note that s can be updated into either $(1+\alpha)s$ or $s/(1+\alpha)$, so any bid-ask pair on the trajectory will have a spread in the set

$$S_{\mathbf{w}} = \big\{(1+\alpha)^i s : -N_1 \leq i \leq N_2 \text{ and } i \in \mathbb{Z}\big\},$$

where $N_1, N_2 \in \mathbb{Z}^+$ are finite, as W is bounded. Let the bid-ask pair $\mathbf{w}_r$ be the first state with a spread r on the trajectory starting from $\mathbf{w}$ for all $r \in S_{\mathbf{w}}$, where $\mathbf{w}_r = \mathbf{w}$ if $r = s$. $\mathbf{w}_r$ will be updated to $\mathbf{w}_r \pm (\alpha r/(1+\alpha), \alpha r/(1+\alpha))$ by the block $\{\sigma(1), \sigma(2)\}$ or $\{\sigma(3), \sigma(4)\}$, and to $\mathbf{w}_r \pm (\alpha r, \alpha r)$ by the block $\{\sigma(2), \sigma(1)\}$ or

$\{\sigma(4), \sigma(3)\}$. So at the constant-spread line $a - b = r$, all the possible states on the trajectory have the form

$$\mathbf{w}_r + k_1(\alpha r, \alpha r) + k_2\Big(\frac{\alpha}{1+\alpha}r, \frac{\alpha}{1+\alpha}r\Big) = \mathbf{w}_r + \Big(k_1\alpha + \frac{k_2\alpha}{1+\alpha}, k_1\alpha + \frac{k_2\alpha}{1+\alpha}\Big)r,$$

where $k_1, k_2 \in \mathbb{Z}$ and they are finite, as W is bounded. Since α, k_1, k_2 are finite, all the possible states with a given spread r on the trajectory are finite, and hence all the states in W starting from $\mathbf{w}$ are finite. So C can not pass infinite bid-ask pairs, which implies the number of marginal traders in C must be finite.

Suppose C has $2n+1$ traders, where $n \in \mathbb{Z}^+$, and assume it will pass x different bid-ask pairs, the collection of which is denoted by the set P. By Proposition 3.3, any trader of type $\sigma(i)$ will definitely update the bid-ask pair in a trading system (W, f_i), so $1 < x < \infty$. Let P be the set of nodes in a graph, so any trader in C will link two different nodes in P. Since there are $2n+1$ traders, we have $2n+1$ links in this graph. But if there exists a directed circle, such that the bid-ask pair after this block will not be changed, then the number of links of any node in P should be even. So the total links in this graph should be even, which contradicts that the number of links is $2n+1$. Therefore, a block with $2n+1$ traders can not be periodic, which completes the proof. □

Note that we can have an equivalent reduced sequence of traders by identifying and then removing (minimal) periodic blocks iteratively in any sequence of traders, as we just delete some closed routes of bid-ask pairs, which will not change the dynamics in the dynamical trading system as a whole.

Definition. A sequence of marginal traders is called *irreducible*, if it contains no minimal periodic block.

Proposition 3.5. *A limit order market that accepts any irreducible sequence of traders is unstable.*

Proof. Note that a market functioning for infinite periods must contain several minimal periodic blocks, otherwise it is a minimal periodic block with infinite traders, which contradicts Lemma 3.4. Since any irreducible sequence of marginal traders contains no minimal periodic block, the number of marginal traders in any irreducible sequence must be finite, otherwise we have a market functioning for infinite periods has no minimal periodic block. If the number of marginal traders in a sequence is finite, then the market must function only for finite periods. So the bid-ask pair in the market must be absorbed on ∂W, and hence the market will be in a crash, which implies that the market must be unstable. □

It is clear that any market functioning for infinite periods will never accept an irreducible sequence of marginal traders. Or we can say the sequence of marginal traders in a stable market should be infinite and reducible, so that we can always find some minimal periodic blocks that stay in the market for finite periods.

A similar concept to periodic block is the well-known notion of hedge, as the role of risk sharing through assets diversification has a counterpart here, namely, instability sharing through orders grouping in the limit order market. Our result suggests that periodic blocks as the "hedging" units in a limit order market should be necessary for its dynamic stability.

3.2. Stochastic Trading. So far, we have studied a dynamical trading system from a combinatorial perspective. In effect, we consider all possible permutation of countably infinite marginal traders, which form the space Σ_4^ω. We find two general categories of sequences of marginal traders in Σ_4^ω, which can sufficiently cause an unstable limit order market. That's to say, the sequence of marginal traders with a constant type, as stated in Proposition 3.3, and any irreducible sequence that contains no minimal periodic block, as stated in Proposition 3.5. In addition, if we might realize that a sequence of marginal traders with a constant type is certainly irreducible, Proposition 3.3 would then become a natural corollary of Proposition 3.5 at this stage.

In this subsection, we will take a different perspective to study sequential trading processes in the limit order market. We assume there is a certain probability structure over the space Σ_4^ω, and hence the dynamics of bid-ask pairs in the dynamical trading system (W, F, σ) is random. Not surprisingly, the related limit order market would be stochastically stable, in the sense that the random trajectory in (W, F, σ) would not be absorbed on ∂W almost surely, in other words, the market would not be in a crash almost surely.

Assume the appearance probability of any type in Σ_4 at any state $\mathbf{w} \in W$ is stationary, *i.e.*, independent of the time. Define a function

$$\pi : W \to [0,1]^4,$$

such that, at any state $\mathbf{w} \in W$, $\pi(\mathbf{w})$ is the 4-tuple of the appearance probabilities of type $\sigma(1)$, type $\sigma(2)$, type $\sigma(3)$, and type $\sigma(4)$, or

$$\pi(\mathbf{w}) = \big(\pi_1(\mathbf{w}), \pi_2(\mathbf{w}), \pi_3(\mathbf{w}), \pi_4(\mathbf{w})\big),$$

with $\sum_{i=1}^4 \pi_i(\mathbf{w}) = 1$, where $\pi_i(\mathbf{w})$ denotes the appearance probability of type $\sigma(i)$ at the state $\mathbf{w}$.

We can therefore have four aggregated appearance probability functions, which can be directly induced from the original $\pi(\mathbf{w})$, and they are

(i) the market-type appearance probability function $\pi_M(\mathbf{w}) = \pi_2(\mathbf{w}) + \pi_4(\mathbf{w})$,
(ii) the limit-type appearance probability function $\pi_L(\mathbf{w}) = \pi_1(\mathbf{w}) + \pi_3(\mathbf{w})$,
(iii) the buy-type appearance probability function $\pi_B(\mathbf{w}) = \pi_1(\mathbf{w}) + \pi_2(\mathbf{w})$,

(iv) the sell-type appearance probability function $\pi_S(\mathbf{w}) = \pi_3(\mathbf{w}) + \pi_4(\mathbf{w})$.

Here, we have $\pi_M(\mathbf{w}) + \pi_L(\mathbf{w}) = 1$ and also $\pi_B(\mathbf{w}) + \pi_S(\mathbf{w}) = 1$ for all $\mathbf{w} \in W$.

Note that the value of $\pi_L(\mathbf{w})$ only depends on the spread of $\mathbf{w}$, and moreover, $\pi_L(\mathbf{w}) = f\big(s(\mathbf{w})\big)$, where $f : [\underline{s}, \overline{a}] \to [0, 1]$ was defined in Section 2.3 of this chapter. Similarly, the value of $\pi_B(\mathbf{w})$ only depends on the mid-price of $\mathbf{w}$, and specifically, we let $\pi_B(\mathbf{w}) = h\big(m(\mathbf{w})\big)$, where $h : [\underline{s}/2, \overline{a} - \underline{s}/2] \to [0, 1]$.

Recall that $f(s)$ is an increasing function of $\log s$, so $\pi_L(\mathbf{w})$ is increasing with $\log s(\mathbf{w})$, and $\pi_M(\mathbf{w})$ is decreasing with $\log s(\mathbf{w})$. Moreover, if $\mathbf{w}$ belongs to the region

$$W_M = \{\mathbf{w} : \underline{s} \leq s(\mathbf{w}) < (1 + \alpha)\underline{s}\},$$

we have $\pi_M(\mathbf{w}) = 1$ and $\pi_L(\mathbf{w}) = 0$. If $\mathbf{w}$ belongs to the region

$$W_L = \{\mathbf{w} : (1 - \gamma)\overline{a} < s(\mathbf{w}) \leq \overline{a}\},$$

we have $\pi_L(\mathbf{w}) = 1$ and $\pi_M(\mathbf{w}) = 0$. Here, $0 < \alpha < 1$ and $\alpha/(1 + \alpha) \leq \gamma < 1$.

Notice that $h(m)$ is a decreasing function of $\log m$, so $\pi_B(\mathbf{w})$ is decreasing with $\log m(\mathbf{w})$, and $\pi_S(\mathbf{w})$ is increasing with $\log m(\mathbf{w})$. Moreover, if $\mathbf{w}$ belongs to the region

$$W_B = \{\mathbf{w} : \underline{s}/2 \leq m(\mathbf{w}) < (1 + \delta)\underline{s}/2\},$$

we have $\pi_B(\mathbf{w}) = 1$ and $\pi_S(\mathbf{w}) = 0$. If $\mathbf{w}$ belongs to the region

$$W_S = \{\mathbf{w} : (1 - \epsilon)(\overline{a} - \underline{s}/2) < m(\mathbf{w}) \leq \overline{a} - \underline{s}/2\},$$

we have $\pi_S(\mathbf{w}) = 1$ and $\pi_B(\mathbf{w}) = 0$. Here, $0 < \delta < 1$ and $\delta/(1 + \delta) \leq \epsilon < 1$.

Definition. The *buffering region* of W is the largest nonclosed subset $H \subset W$ with the property that $\prod_{x \in \{L,M,B,S\}} \pi_x(\mathbf{w}) = 0$ for all $\mathbf{w} \in H$.

At any state $\mathbf{w} \in H$, there exists at least an $x \in \{L, M, B, S\}$ such that $\pi_x(\mathbf{w}) = 0$. Since $\pi_L + \pi_M = \pi_B + \pi_S = 1$, there also exists at least a $y \in \{L, M, B, S\}$ such that $\pi_y(\mathbf{w}) = 1$ at the state $\mathbf{w}$. So we can have at most two elements, say $x_1 \in \{L, M\}$ and $x_2 \in \{B, S\}$, such that $\pi_{x_1}(\mathbf{w}) = \pi_{x_2}(\mathbf{w}) = 0$, and $\pi_y(\mathbf{w}) = 1$ for $y \notin \{x_1, x_2\}$.

Definition. The *kernel region* of W is the largest closed subset $K \subseteq W$ with the property that $\pi_x(\mathbf{w}) \neq 0$ for all $x \in \{L, M, B, S\}$ and for all $\mathbf{w} \in K$.

Note that for all $\mathbf{w} \in K$ we also have $\pi_x(\mathbf{w}) \neq 1$ for all $x \in \{L, M, B, S\}$, as there exists a unique y such that $\pi_x(\mathbf{w}) = 1 - \pi_y(\mathbf{w})$, where $\pi_y(\mathbf{w}) \neq 0$.

In general, we have $K = W \setminus H$ and $K \cap H = \emptyset$. Thus we have a bipartition of the domain W, as $K \cup H = W$ and $K \cap H = \emptyset$. Since K is defined to be closed, and H is defined to be nonclosed, H may be empty, but $H \neq W$, and hence K is always nonempty. If $K = W$, then $H = \emptyset$. If $K = \{\mathbf{w}\}$, for some $\mathbf{w} \in W$ certain, then $H = W \setminus \{\mathbf{w}\}$.

In our framework, we have $\pi_L(\mathbf{w}) = 0$ for all $\mathbf{w} \in W_M$, $\pi_M(\mathbf{w}) = 0$ for all $\mathbf{w} \in W_L$, $\pi_B(\mathbf{w}) = 0$ for all $\mathbf{w} \in W_S$, and $\pi_S(\mathbf{w}) = 0$ for all $\mathbf{w} \in W_B$. Thus

$$H = W_L \cup W_M \cup W_B \cup W_S, \tag{3.9}$$

and $K = W \setminus H$, where $W_L \cap W_M = \emptyset$ and $W_B \cap W_S = \emptyset$ (see Figure 3.6).

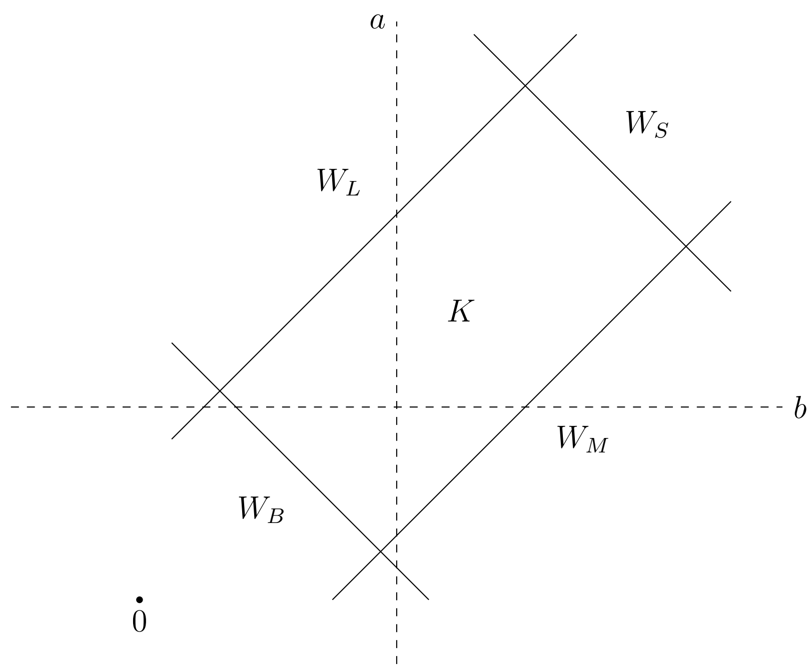


FIGURE 3.6.

Note that K is closed, while H is not closed, but $H \cup \partial K$ is also closed, where ∂K denotes the boundary of the kernel region K.

DEFINITION. The *s-range* of $R \subseteq W$ is

$$r_s(R) = \sup_{\mathbf{w}\in R} s(\mathbf{w}) - \inf_{\mathbf{w}\in R} s(\mathbf{w}). \tag{3.10}$$

DEFINITION. The *m-range* of $R \subseteq W$ is

$$r_m(R) = \sup_{\mathbf{w}\in R} m(\mathbf{w}) - \inf_{\mathbf{w}\in R} m(\mathbf{w}). \tag{3.11}$$

PROPOSITION 3.6. *If the buffering region $H \neq \emptyset$, and the kernel region K satisfies*

$$\min\{r_s(K), r_m(K)\} > \alpha(1+\alpha)(2+\alpha)\underline{s}, \tag{3.12}$$

the dynamical trading system (W, F, σ) is stochastically stable within K.

PROOF. For any trajectory starting from a bid-ask pair $\mathbf{w} \in W$, all the possible states in W are finite, as shown in the proof of Lemma 3.4. The set of all the possible states for any initial state $\mathbf{w}$ can be denoted by a corresponding lattice $\Lambda(\mathbf{w})$. Let the neighborhood of any $\mathbf{v} \in \Lambda(\mathbf{w})$ be

$$N(\mathbf{v}) = \{\mathbf{v}S_1, \mathbf{v}S_2, \mathbf{v}S_3, \mathbf{v}S_4\} \cap W.$$

Observe that $\Lambda(\mathbf{w})$ is globally defined on $W = K \cup H$, so there exist states $\mathbf{v} \in \Lambda(\mathbf{w})$ near ∂K, such that $N(\mathbf{v}) \cap H \neq \emptyset$ and $N(\mathbf{v}) \cap K \neq \emptyset$.

Note that the spread $s(\mathbf{v})$ of the state $\mathbf{v}$ can be updated to either $s(\mathbf{v})/(1+\alpha)$ or $(1+\alpha)s(\mathbf{v})$, and its mid-price $m(\mathbf{v})$ can be updated to maximally $m(\mathbf{v}) + \alpha s(\mathbf{v})/2$ and minimally $m(\mathbf{v}) - \alpha s(\mathbf{v})/2$, so

$$r_s(N(\mathbf{v})) = \frac{\alpha(2+\alpha)}{1+\alpha} s(\mathbf{v}), \quad r_m(N(\mathbf{v})) = \alpha s(\mathbf{v}).$$

Suppose $s(\mathbf{v})/(1+\alpha) \geq (1+\alpha)\underline{s}$, where $(1+\alpha)\underline{s}$ is the lower bound of the spread in K, so we have

$$r_s(N(\mathbf{v})) \geq \alpha(1+\alpha)(2+\alpha)\underline{s}, \quad r_m(N(\mathbf{v})) \geq \alpha(1+\alpha)^2\underline{s},$$

and hence

$$r_s(K) > \inf_{\mathbf{v}\in K} r_s(N(\mathbf{v})), \quad r_m(K) > \inf_{\mathbf{v}\in K} r_m(N(\mathbf{v})).$$

So for any initial state $\mathbf{w} \in W$, there exists at least a $\mathbf{v} \in \Lambda(\mathbf{w}) \cap K$ such that $N(\mathbf{v}) \subset K$.

Note that $W = K \cup H$, and both K and H are nonempty, so both H and K are proper subsets of W. Suppose $\mathbf{w} \in H$. We know $H = W_L \cup W_M \cup W_B \cup W_S$, so there exists at least an $x \in \{L, M, B, S\}$ such that $\mathbf{w} \in W_x$. Note that $\pi_x(\mathbf{v}) = 1$ for all $\mathbf{v} \in W_x$, thus $\mathbf{w}$ will move towards K along a continuous flow in $\Lambda(\mathbf{w}) \cap H$. Since the number of states in $\Lambda(\mathbf{w}) \cap H$ is finite, $\mathbf{w}$ will move into the closed kernel region K after finite periods. If the market is stochastically stable within K once $\mathbf{w} \in K$, the market will function for infinite periods, and hence it is also stochastically stable with the initial state $\mathbf{w} \in H$. Thus we only need to show the trading system is stochastically stable for any initial state $\mathbf{w} \in K$. Let $p(\mathbf{w})$ denote the probability that bid-ask pairs stay in H with an initial state $\mathbf{w}$. We need to show that $p(\mathbf{w}) = 0$ for all $\mathbf{w} \in K$.

Suppose $\mathbf{w} \in K$, then we have two possibilities, $N(\mathbf{w}) \subset K$, and $N(\mathbf{w}) \cap H \neq \emptyset$. If $N(\mathbf{w}) \cap H \neq \emptyset$, then $N(\mathbf{w}) \cap K \neq \emptyset$, otherwise $\mathbf{w} \in H$. There exist some $x \in \{L, M, B, S\}$, such that $\pi_x(\mathbf{v}) = 0$ for all $\mathbf{v} \in N(\mathbf{w}) \cap H$, and $\pi_x(\mathbf{v}) = \varepsilon_x \in [0, 1]$ for all $\mathbf{v} \in N(\mathbf{w}) \cap K$. $\pi_x(\mathbf{v}) \neq 1$ for all $\mathbf{v} \in K$ and all $x \in \{L, M, B, S\}$, otherwise there must exist a $y \in \{L, M, B, S\}$ such that $\pi_y(\mathbf{v}) = 1 - \pi_x(\mathbf{v}) = 0$, which then implies $\mathbf{v} \in H$, a contradiction. Hence $\varepsilon_x \neq 1$ for all $x \in \{L, M, B, S\}$.

If $\mathbf{w}$ moves to $\mathbf{v} \in N(\mathbf{w}) \cap H$ with a probability ε_x, it will return back to $\mathbf{v}' \in N(\mathbf{v}) \cap K$ with probability 1 in the next period, where $N(\mathbf{v}') \cap H \neq \emptyset$ as $\mathbf{v} \in N(\mathbf{v}')$. If $\mathbf{w}$ moves to $\mathbf{v} \in N(\mathbf{w}) \cap K$ with a probability $1-\varepsilon_x$, it can stay within $\Lambda(\mathbf{w}) \cap K$ with $k(\mathbf{w})$ consecutive periods, and then move into a state $\mathbf{v}'$ such that $N(\mathbf{v}') \cap H \neq \emptyset$. Recall that, for all $\mathbf{w} \in K$, we have

$$\{\mathbf{v} \in \Lambda(\mathbf{w}) \cap K : N(\mathbf{v}) \subset K\} \neq \emptyset,$$

so $k(\mathbf{w}) \geq 0$. Then we have

$$p(\mathbf{w}) = \lim_{T\uparrow\infty} \varepsilon_x p(\mathbf{w})\Big(1 - \frac{2}{T}\Big) + (1-\varepsilon_x)p(\mathbf{w})\Big(1 - \frac{k(\mathbf{w})+1}{T}\Big),$$

where $0 \leq k(\mathbf{w}) \leq T-1$. When $k(\mathbf{w}) = T-1$,

$$p(\mathbf{w}) = \lim_{T\uparrow\infty} \varepsilon_x p(\mathbf{w})\Big(1 - \frac{2}{T}\Big) = \varepsilon_x p(\mathbf{w}),$$

which generates $(1-\varepsilon_x)p(\mathbf{w}) = 0$. Since $1-\varepsilon_x \neq 0$, $p(\mathbf{w}) = 0$ for all $\mathbf{w} \in K$ with $N(\mathbf{w}) \cap H \neq \emptyset$.

If $\mathbf{w} \in K$ and $N(\mathbf{w}) \subset K$, its trajectory can either achieve a state $\mathbf{w}' \in \Lambda(\mathbf{w}) \cap K$ such that $N(\mathbf{w}') \cap H \neq \emptyset$ after $h(\mathbf{w})$ periods, where $h(\mathbf{w}) \geq 1$, or never move to such a state $\mathbf{w}'$ and thus stay within K for ever. Note that $p(\mathbf{w}') = 0$, if $\mathbf{w}' \in K$ and $N(\mathbf{w}') \cap H \neq \emptyset$. So $p(\mathbf{w}) \leq p(\mathbf{w}') = 0$, but $p(\mathbf{w}) \geq 0$, hence $p(\mathbf{w}) = 0$ for all $\mathbf{w} \in K$ with $N(\mathbf{w}) \subset K$.

As a result, $p(\mathbf{w}) = 0$ if $\mathbf{w} \in K$, and thus $p(\mathbf{w}) = 0$ for all $\mathbf{w} \in W$. So the dynamical trading system is stable within K almost surely. $\square$

Since (b_t, a_t) will stay within K almost surely, the random trajectories of b_t and a_t will also stay within bounded intervals. Note that

$$(1+\alpha)\underline{s} \leq s_t \leq r_s(K) + (1+\alpha)\underline{s}, \tag{3.13}$$

and

$$(1+\delta)\underline{s}/2 \leq m_t \leq r_m(K) + (1+\delta)\underline{s}/2, \tag{3.14}$$

almost surely for all $t \in \mathbb{Z}$, so we almost surely have

$$a_t \leq r_m(K) + r_s(K)/2 + (2+\alpha+\delta)\underline{s}/2 < r_m(K) + r_s(K)/2 + 2\underline{s}, \tag{3.15}$$

and

$$b_t \leq r_m(K) + (2+\alpha+\delta)\underline{s}/2 < r_m(K) + 2\underline{s}, \tag{3.16}$$

where $\alpha, \delta \in (0,1)$, so $(2+\alpha+\delta)/2 < 2$.

The upper bounds of b_t and a_t both have interesting implications on the roles of s-range and m-range in the limit order market. The upper bound of the best bid b_t in the market is solely determined by the m-range of the kernel region K, rather than any property of the whole domain W. The upper bound of the best ask a_t is also only related to the kernel region K, and yet it is determined by both the m-range and the s-range of K. Notice that the lower bounds of b_t and a_t are both close to $2\underline{s}$, so the *bid-range* in the market is approximately equal to $r_m(K)$, and the *ask-range* is roughly $r_m(K) + r_s(K)/2$. Evidently, the ask side of the limit order book should be more volatile than its bid side, in case that the range of a price is a meaningful indicator of its volatility.

4. Controlled Trading System

In this section, we assume again $H \neq \emptyset$, but either the s-range or the m-range of K will be less than $\alpha(1+\alpha)(2+\alpha)\underline{s}$. So the condition on the kernel region in Proposition 3.6 is no longer satisfied. We would like to check whether a random trajectory of bid-ask pairs can maintain the property of stochastic stability within certain bounded domains.

Let $U_1 = W_L \cup W_M$. Since $K \neq \emptyset$, we have $W_L \cap W_M = \emptyset$ and $W \setminus U_1 \neq \emptyset$. Define $U_2 = W \setminus U_1$, so $W = U_1 \cup U_2$ and $U_1 \cap U_2 = \emptyset$.

Similarly, let $V_1 = W_B \cup W_S$, again $W \setminus V_1 \neq \emptyset$. Define $V_2 = W \setminus V_1$, so $W = V_1 \cup V_2$ and $V_1 \cap V_2 = \emptyset$.

Note that $K = U_2 \cap V_2$ and $H = U_1 \cup V_1$. Also observe that $r_s(K) = r_s(U_2)$ and $r_m(K) = r_m(V_2)$.

4.1. Controlled Spread Dynamics. By the condition $W_L \cap W_M = \emptyset$, we have $r_s(U_2) \geq 0$, so $(1+\alpha)\underline{s} \leq (1-\gamma)\overline{a}$, or

$$\underline{s}/\overline{a} \leq \frac{1-\gamma}{1+\alpha}.$$

At the same time, we assume that the s-range of U_2 is sufficiently small, namely, $r_s(U_2) < \alpha(1+\alpha)\underline{s}$, so

$$(1-\gamma)\overline{a} - (1+\alpha)\underline{s} < \alpha(1+\alpha)\underline{s},$$

which implies that $\underline{s}/\overline{a}$ has a lower bound,

$$\underline{s}/\overline{a} > \frac{1-\gamma}{(1+\alpha)^2}.$$

Intuitively, the upper bound of the s-range of U_2 gives such a sufficient condition that any type of marginal trader will definitely update any $\mathbf{w} \in U_2$ to some $\mathbf{w}' \in U_1$.

In brief, if $0 \leq r_s(U_2) < \alpha(1+\alpha)\underline{s}$, the domain W will have the following property,

$$\frac{1-\gamma}{(1+\alpha)^2} < \underline{s}/\overline{a} \leq \frac{1-\gamma}{1+\alpha}. \tag{3.17}$$

Once the above inequality is satisfied by the domain W, we have $U_2 \neq \emptyset$, and

$$U_2 = \{\mathbf{w} : (1+\alpha)\underline{s} \leq s(\mathbf{w}) \leq (1-\gamma)\overline{a}\},$$

where $s(\mathbf{w})$ is the spread function. Let the boundary of U_2 be ∂U_2, then

$$\partial U_2 = \{\mathbf{w} : s(\mathbf{w}) = (1+\alpha)\underline{s}\} \cup \{\mathbf{w} : s(\mathbf{w}) = (1-\gamma)\overline{a}\}.$$

PROPOSITION 3.7. *If the buffering region $H = U_1 \cup V_1$ is nonempty, and the kernel region $K = U_2 \cap V_2$ satisfies*

$$0 \leq r_s(U_2) < \alpha(1+\alpha)\underline{s}, \quad r_m(V_2) > \alpha(1+\alpha)(2+\alpha)\underline{s},$$

the dynamical trading system (W, F, σ) is stochastically stable within the region

$$\{\mathbf{w} : \underline{s} \leq s(\mathbf{w}) < (1+\alpha)^3\underline{s}\} \cap V_2.$$

PROOF. Since $U_2 \neq \emptyset$ and its s-range is less than $\alpha(1+\alpha)\underline{s}$, we obtain

$$(1+\alpha)\underline{s} \leq (1-\gamma)\overline{a} < (1+\alpha)^2\underline{s}.$$

Note that

$$\min_{\mathbf{w}\in U_2} s(\mathbf{w}) = (1+\alpha)\underline{s}, \quad \max_{\mathbf{w}\in U_2} s(\mathbf{w}) = (1-\gamma)\overline{a},$$

so we have

$$\underline{s} \leq \frac{(1-\gamma)\overline{a}}{(1+\alpha)} < (1+\alpha)\underline{s} = \min_{\mathbf{w}\in U_2} s(\mathbf{w}),$$

and also

$$(1+\alpha)^2\underline{s} > (1-\gamma)\overline{a} = \max_{\mathbf{w}\in U_2} s(\mathbf{w}).$$

Since any initial state $\mathbf{w} \in H$ will definitely move into K after finite periods, we only need to show the statement is true when the initial state $\mathbf{w} \in K$. Consider any initial state $\mathbf{w} \in U_2 \cap V_2$, we have $(1+\alpha)\underline{s} \leq s(\mathbf{w}) \leq (1-\gamma)\overline{a}$. Note that $\pi_L(\mathbf{w}) > 0$ and $\pi_M(\mathbf{w}) > 0$ for all $\mathbf{w} \in K$, so $\pi_i(\mathbf{w}) \neq 0$ for all $\mathbf{w} \in K$ and all $i \in \{1, 2, 3, 4\}$. Suppose $\mathbf{w}$ is updated to $\mathbf{v} = \mathbf{w}S_1$ by a marginal trader of type $\sigma(1)$, $s(\mathbf{v}) = s(\mathbf{w})/(1+\alpha)$, which is greater than $\underline{s}$ and less than $(1-\gamma)\overline{a}/(1+\alpha)$. Since $(1-\gamma)\overline{a}/(1+\alpha) < \min_{\mathbf{w}\in U_2} s(\mathbf{w})$, we should have $\mathbf{v} \in W_M$, and hence $\pi_M(\mathbf{v}) = 1$. The trader in the next period will be market-type, namely, either type $\sigma(4)$ or type $\sigma(2)$. If she is of type $\sigma(4)$, $\mathbf{v}$ will then become $\mathbf{w}S_1S_4 = \mathbf{w}$, since $\{\sigma(1), \sigma(4)\}$ is a minimal periodic block. If she is of type $\sigma(2)$, $\mathbf{v}$ will become $\mathbf{v}' = \mathbf{w}S_1S_2$, where

$$S_1S_2 = \begin{pmatrix} \frac{1}{1+\alpha} & 0 \\ \frac{\alpha}{1+\alpha} & 1 \end{pmatrix} \begin{pmatrix} 1 & -\alpha \\ 0 & 1+\alpha \end{pmatrix} = \frac{1}{1+\alpha} \begin{pmatrix} 1 & -\alpha \\ \alpha & 1+2\alpha \end{pmatrix},$$

so $s(\mathbf{v}') = s(\mathbf{w})$. Thus after two periods, $\mathbf{w}$ will return back to itself or move to a state with the same spread as itself.

If $\mathbf{w}$ is updated to $\mathbf{v} = \mathbf{w}S_2$ by a marginal trader of type $\sigma(2)$, then $\mathbf{v} \in W_L$. So in the next period, there will come either type-$\sigma(3)$ trader or type-$\sigma(1)$ trader. The type-$\sigma(3)$ trader will update $\mathbf{v}$ to $\mathbf{w}S_2S_3 = \mathbf{w}$, as $\{\sigma(2), \sigma(3)\}$ is a minimal periodic block. The type-$\sigma(1)$ trader will update $\mathbf{v}$ to $\mathbf{w}S_2S_1$ that has the same spread as $\mathbf{w}$. If $\mathbf{w}$ is updated to $\mathbf{w}S_3$ by a marginal trader of type $\sigma(3)$, $\mathbf{w}S_3$ will then be updated either to $\mathbf{w}S_3S_2 = \mathbf{w}$, or to $\mathbf{w}S_3S_4$ that have the same spread as $\mathbf{w}$. Finally, if $\mathbf{w}$ is updated to $\mathbf{w}S_4$ by a marginal trader of type $\sigma(4)$, $\mathbf{w}S_4$ will then move either to $\mathbf{w}S_4S_1 = \mathbf{w}$, or to $\mathbf{w}S_4S_3$ such that $s(\mathbf{w}S_4S_3) = s(\mathbf{w})$.

Therefore, any initial state $\mathbf{w} \in U_2 \cap V_2$ will be updated to a state with the same spread as itself after two consecutive periods. Note that the dynamics in (W, F, σ) is repeatedly composed of those two-period dynamical blocks, thus the spread that can be achieved in such a dynamical trading system will be greater than or equal to

$$\min_{\mathbf{w}\in U_2} s(\mathbf{w})/(1+\alpha) = \underline{s},$$

and less than or equal to

$$(1+\alpha)\max_{\mathbf{w}\in U_2} s(\mathbf{w}) < (1+\alpha)^3\underline{s},$$

as $\max_{\mathbf{w}\in U_2} s(\mathbf{w}) < (1+\alpha)^2\underline{s}$.

So the trajectory of bid-ask pairs starting from any initial state $\mathbf{w} \in W$ will be bounded in the region $\{\mathbf{w} : \underline{s} \le s(\mathbf{w}) < (1+\alpha)^3\underline{s}\} \cap V_2$ almost surely. $\square$

If $0 \le r_s(U_2) < \alpha(1+\alpha)\underline{s}$, there exists a unique exponent $l \in [1, 2)$, such that

$$\underline{s}/\overline{a} = \frac{1-\gamma}{(1+\alpha)^l}. \tag{3.18}$$

Note that the upper bound of the bid-ask spread in W is $\overline{a}$, so $(1+\alpha)^3\underline{s}$ should be no greater than $\overline{a}$, or equivalently,

$$\underline{s}/\overline{a} \le \frac{1}{(1+\alpha)^3}.$$

Thus we need $(1-\gamma)(1+\alpha)^h \le 1$, where $h = 3 - l$, so $h \in (1, 2]$. It appears to be a slightly stricter requirement than $(1-\gamma)(1+\alpha) \le 1$ that we have used before, since $(1+\alpha)^h > 1+\alpha$ for $h > 1$.

Recall that $U_2 = \{\mathbf{w} : (1+\alpha)\underline{s} \le s(\mathbf{w}) \le (1-\gamma)\overline{a}\}$, we thus have two nonempty regions in U_1, which will contain buffering overflows,

$$\{\mathbf{w} : \underline{s} \le s(\mathbf{w}) < (1+\alpha)\underline{s}\}, \quad \{\mathbf{w} : (1-\gamma)\overline{a} < s(\mathbf{w}) < (1+\alpha)^3\underline{s}\}.$$

Evidently, the trajectory of bid-ask pairs will not stay exactly within the kernel region $K = U_2 \cap V_2$, but within K and parts of the buffering region $U_1 \cap V_2$ in H.

As we can see from the proof of Proposition 3.7, there are in fact eight possible two-period dynamical blocks for all $\mathbf{w} \in U_2 \cap V_2$. Half of them are minimal periodic blocks, so the bid-ask pair driven by each of them will return back to $\mathbf{w}$. While the remaining ones will update $\mathbf{w}$ to $\mathbf{w}' \neq \mathbf{w}$, with $s(\mathbf{w}') = s(\mathbf{w})$ (see Figure 3.7, where $s(\mathbf{w})$ is replaced simply by s).

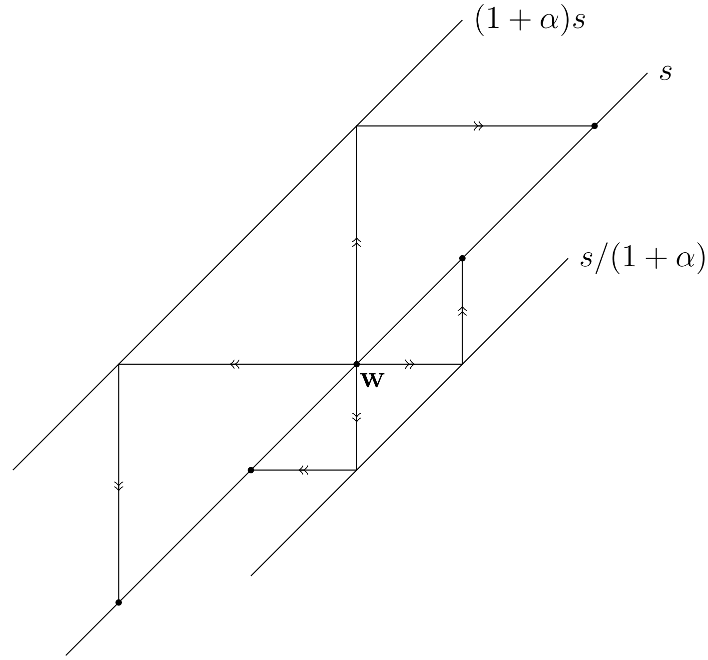


Figure 3.7.

If $\mathbf{w}' \neq \mathbf{w}$, the distance (in the b-a plane) between $\mathbf{w}$ and $\mathbf{w}'$ is

$$\frac{\sqrt{2}\alpha}{1+\alpha} s(\mathbf{w}) \quad \text{or} \quad \sqrt{2}\alpha s(\mathbf{w}),$$

and the absolute difference between $m(\mathbf{w})$ and $m(\mathbf{w}')$ is

$$\frac{\alpha}{1+\alpha} s(\mathbf{w}) \quad \text{or} \quad \alpha s(\mathbf{w}).$$

The possible trajectories in the dynamical trading system (W, F, σ) are now consecutive combinations of those eight two-period dynamical blocks. Recall that $s(\mathbf{w}) \in [(1+\alpha)\underline{s}, (1-\gamma)\overline{a}]$ for all $\mathbf{w} \in U_2 \cap V_2$. The trajectory starting from $\mathbf{w}$ will be bounded within the region

$$\{\mathbf{v} : s(\mathbf{w})/(1+\alpha) \leq s(\mathbf{v}) \leq (1+\alpha)s(\mathbf{w})\},$$

where

$$(1+\alpha)\underline{s} \in (s(\mathbf{w})/(1+\alpha), s(\mathbf{w})], \quad (1-\gamma)\overline{a} \in [s(\mathbf{w}), (1+\alpha)s(\mathbf{w})).$$

So the region containing buffering overflows are

$$\{\mathbf{v} : s(\mathbf{w})/(1+\alpha) \le s(\mathbf{v}) < (1+\alpha)\underline{s}\}, \quad \{\mathbf{v} : (1-\gamma)\overline{a} < s(\mathbf{v}) \le (1+\alpha)s(\mathbf{w})\},$$

where $s(\mathbf{w})/(1+\alpha) \ge \underline{s}$ and $(1+\alpha)s(\mathbf{w}) < (1+\alpha)^3\underline{s}$. We shall show two trajectories starting from $\mathbf{w}$ with seven periods on the b-a plane in the following Figure 3.8, where $s = s(\mathbf{w})$.

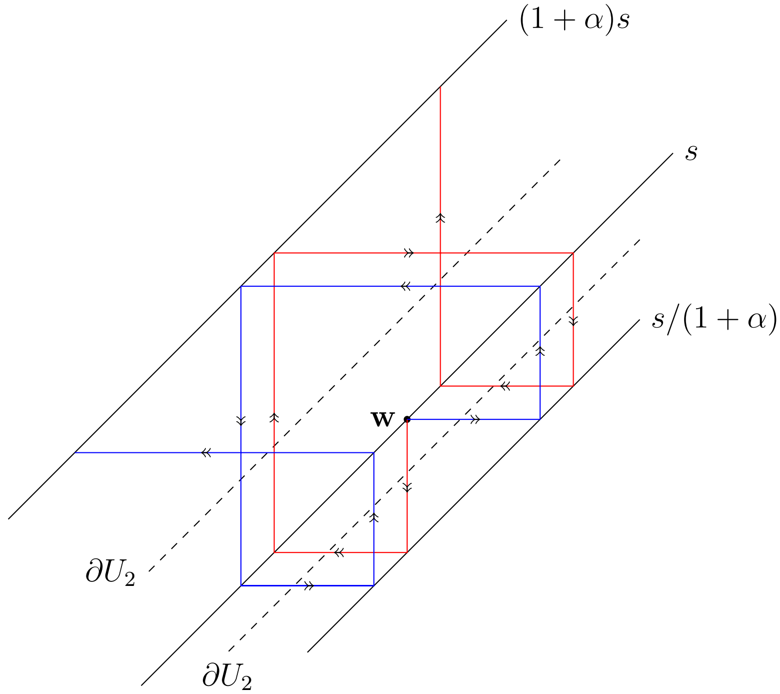


FIGURE 3.8.

4.2. Controlled Mid-Price Dynamics. If the m-range of V_2, rather than the s-range of U_2, is sufficiently small, we may establish a similar result to Proposition 3.7. Assume $0 \le r_m(V_2) < \alpha(1+\alpha)\underline{s}/2$. By the condition $r_m(V_2) \ge 0$, we have $V_2 \ne \emptyset$ and $W_B \cap W_S = \emptyset$, so $(1+\delta)\underline{s}/2 \le (1-\epsilon)(\overline{a} - \underline{s}/2)$, or

$$\frac{\underline{s}/2}{\overline{a} - \underline{s}/2} \le \frac{1-\epsilon}{1+\delta},$$

where $\delta \in (0,1)$ and $\epsilon \in [\delta/(1+\delta), 1)$. On the other hand, if $r_m(V_2) < \alpha(1+\alpha)\underline{s}/2$, then

$$(1-\epsilon)(\overline{a} - \underline{s}/2) - (1+\delta)\underline{s}/2 < \alpha(1+\alpha)\underline{s}/2,$$

which implies

$$\frac{\underline{s}/2}{\overline{a}-\underline{s}/2} > \frac{1-\epsilon}{1+\delta+\alpha(1+\alpha)}.$$

As a result, if $0 \le r_m(V_2) < \alpha(1+\alpha)\underline{s}/2$, the domain W will have the following property,

$$\frac{1-\epsilon}{1+\delta+\alpha(1+\alpha)} < \frac{\underline{s}/2}{\overline{a}-\underline{s}/2} \le \frac{1-\epsilon}{1+\delta}. \tag{3.19}$$

Under that condition, we have a nonempty V_2,

$$V_2 = \{\mathbf{w} : (1+\delta)\underline{s}/2 \le m(\mathbf{w}) \le (1-\epsilon)(\overline{a}-\underline{s}/2)\}.$$

Let the boundary of V_2 be ∂V_2, then

$$\partial V_2 = \{\mathbf{w} : m(\mathbf{w}) = (1+\delta)\underline{s}/2\} \cup \{\mathbf{w} : m(\mathbf{w}) = (1-\epsilon)(\overline{a}-\underline{s}/2)\}.$$

The lower bound in the condition (3.19) is a sufficient condition for the existence of dynamical two-period switching blocks, by which any state in $V_2 \cap U_2$ can return back into $V_2 \cap U_2$ after two periods. To confirm this claim, we only need to show that any state $\mathbf{w}$ in $\partial V_2 \cap U_2$ at a time t will be updated to some $\mathbf{w}' \in V_1$ at the time $t+1$. Note that

$$\max_{\mathbf{v}\in V_2} m(\mathbf{v}) = (1-\epsilon)(\overline{a}-\underline{s}/2), \quad \min_{\mathbf{v}\in V_2} m(\mathbf{v}) = (1+\delta)\underline{s}/2,$$

and they can be achieved by the states in ∂V_2.

Consider a generic state $\mathbf{w} \in \partial V_2 \cap U_2$ with $m(\mathbf{w}) = \max_{\mathbf{v}\in V_2} m(\mathbf{v})$. A buy-type trader will induce a greater mid-price, which means that the next state will be in V_1. A sell-type trader can generate a new mid-price,

$$m(\mathbf{w}) - \frac{\alpha}{2}s(\mathbf{w}) \quad \text{or} \quad m(\mathbf{w}) - \frac{\alpha}{2(1+\alpha)}s(\mathbf{w}).$$

A sufficient condition for the new state will be in V_1 is that the greatest mid-price should be less than $\min_{\mathbf{v}\in V_2} m(\mathbf{v})$, that's to say,

$$\max_{\mathbf{v}\in V_2} m(\mathbf{v}) - \frac{\alpha}{2(1+\alpha)} \min_{\mathbf{w}\in\partial V_2} s(\mathbf{w}) < \min_{\mathbf{v}\in V_2} m(\mathbf{v}). \tag{3.20}$$

Note that $s(\mathbf{w})/(1+\alpha) \ge (1+\alpha)\underline{s}$ for all $\mathbf{w} \in V_2 \cap U_2$, so $\min_{\mathbf{w}\in\partial V_2} s(\mathbf{w}) = (1+\alpha)^2\underline{s}$. Then we obtain

$$(1-\epsilon)(\overline{a}-\underline{s}/2) - \alpha(1+\alpha)\underline{s}/2 < (1+\delta)\underline{s}/2,$$

which is exactly equivalent to the condition $r_m(V_2) < \alpha(1+\alpha)\underline{s}/2$.

Now if $\mathbf{w} \in \partial V_2 \cap U_2$ with $m(\mathbf{w}) = \min_{\mathbf{v}\in V_2} m(\mathbf{v})$, a sell-type trader will induce a smaller mid-price, and hence the next state must be in V_1. While the new mid-price generated by a buy-type trader can be

$$m(\mathbf{w}) + \frac{\alpha}{2}s(\mathbf{w}) \quad \text{or} \quad m(\mathbf{w}) + \frac{\alpha}{2(1+\alpha)}s(\mathbf{w}).$$

A sufficient condition ensuring the new state will be in V_1 is that the smallest mid-price should be greater than $\max_{\mathbf{v}\in V_2} m(\mathbf{v})$, that's to say,

$$\min_{\mathbf{v}\in V_2} m(\mathbf{v}) + \frac{\alpha}{2(1+\alpha)} \min_{\mathbf{w}\in\partial V_2} s(\mathbf{w}) > \max_{\mathbf{v}\in V_2} m(\mathbf{v}), \tag{3.21}$$

which is again equivalent to $r_m(V_2) < \alpha(1+\alpha)\underline{s}/2$.

Proposition 3.8. *If the buffering region $H = V_1 \cup U_1$ is nonempty, and the kernel region $K = V_2 \cap U_2$ satisfies*

$$0 \le r_m(V_2) < \alpha(1+\alpha)\underline{s}/2, \quad r_s(U_2) > \alpha(1+\alpha)(2+\alpha)\underline{s},$$

the dynamical trading system (W, F, σ) is stochastically stable within the region

$$\{\mathbf{w} : \underline{m} \le m(\mathbf{w}) < \overline{m}\} \cap U_2,$$

where $\underline{m}$ and $\overline{m}$ are constants in the trading system.

The proof of Proposition 3.8 is roughly similar to that of Proposition 3.7, and thus not provided here. Once again, its trajectory of bid-ask pairs will not be bounded within the kernel region $K = V_2 \cap U_2$, but within K and parts of the buffering region $V_1 \cap U_2$, so that the trading system can contain certain buffering overflows to support its stability. In the limit, each state on the trajectory will be either in $K = V_2 \cap U_2$ or in the buffering region $V_1 \cap U_2$ with equal probability.

Since the updating process of the mid-price also depends on the spread of the states on the trajectory, the buffering region used to hold the buffering overflows is quite wide. We can see that the m-range of the region $\{\mathbf{w} : \underline{m} < m(\mathbf{w}) < \overline{m}\} \cap U_2$ is very large, by virtue of

$$\underline{m} = (1+\delta)\underline{s}/2 - \alpha(1-\gamma)\overline{a}/2, \quad \overline{m} = (1-\epsilon)(\overline{a} - \underline{s}/2) + \alpha(1-\gamma)\overline{a}/2,$$

where $(1-\gamma)\overline{a} = \max_{\mathbf{w}\in U_2} s(\mathbf{w})$. In fact, its m-range is equal to

$$\overline{m} - \underline{m} = r_m(V_2) + \alpha(1-\gamma)\overline{a}, \tag{3.22}$$

which is greater than or equal to $\alpha(1-\gamma)\overline{a}$, and less than $\alpha(1-\gamma)\overline{a} + \alpha(1+\alpha)\underline{s}/2$. So $\overline{m} - \underline{m} = O(\overline{a})$. On the other hand, in Proposition 3.7, the s-range of the region

$$\{\mathbf{w} : \underline{s} \le s(\mathbf{w}) < (1+\alpha)^3\underline{s}\} \cap V_2$$

is only $((1+\alpha)^3 - 1)\underline{s} = O(\underline{s})$, where $O(\underline{s}) \ll O(\overline{a})$.

Proposition 3.7 and 3.8 collectively produce interesting observations about the volatilities of the mid-price and the bid-ask spread on a limit order market. If the spread s in the kernel region $K = U_2 \cap V_2$ is required to be *ex ante* stable, then the random trajectory in the dynamical trading system will be bounded within a certain region with a sufficiently small s-range. However, if the mid-price m in K is required to be *ex ante* stable, the random trajectory can not remain in a certain region with a small m-range.

Suppose again the range of a time-dependent price can be thought of to be an indicator of its volatility. Let us adopt an artificial concept of "disorder" to describe the origins of price volatility, and yet we leave such an introduced concept in its obscurity. Moreover, imagine the disorder in a trading system could be classified into one intrinsic part and another external part (which are evidently similar to the terms of "self-generated disorder" and "quenched disorder" in physics), such that the intrinsic disorder has an *ex ante* controllable nature, while the external disorder has an *ex ante* uncontrollable nature, and hence carries a high volume of potential information in the trading system. In practice, more intrinsic disorder should make a price in the trading system less unpredictable, while more external disorder could make the price more unpredictable. And thus a price volatility that can be partially characterized by its unpredictability should be lower if the disorder is rather intrinsic, and higher if that is rather external.

Quite directly, we say that the disorder in the bid-ask spread is mainly intrinsic or self-generated, as we find that the bid-ask spread has an *ex ante* controllable nature. On the other hand, most of the disorder in the mid-price should be external or quenched, since we show that the mid-price has an *ex ante* uncontrollable nature. For this reason, we conclude that the volatility of the bid-ask spread should be lower than that of the mid-price in the dynamical trading system.

5. Final Remarks

In our studies, we take a dynamical perspective to investigate the microstructure of the limit order market. A limit order market is theoretically thought of to be a dynamical trading system, in which sequential trading processes are determined by either deterministic or probabilistic switches between different types of traders (represented by their optimal trading decisions in a very short time interval). In our analysis, the perfect information usually required to model traders' strategic behavior is loosened to the knowledge of the so-called atomic trading schemes. We thus set less assumptions for traders and the market, and yet still obtain a powerful ability to understand and even predict the order flow and the order book evolution in a generic limit order market. Some interesting and hopefully insightful conclusions have developed, for instance, the best ask is more volatile than the best bid, the mid-price is more volatile than the spread, and the best ask seems to have more determinants than the best bid.

To close this chapter, we would like to make some concluding remarks on these two critical notions appearing in our main results — general instability in a deterministic trading system, and stochastic stability in a stochastic trading system.

The stability of a deterministic trading system means that there should exist some bid-ask pairs such that certain trading blocks can generate convergent limit cycles attracted to them. We show that a necessary condition for general stability of the trading system is the trading block must be reducible, so that they can generate periodic bid-ask pair dynamics. Thus the notion of general instability implies that there is no stable bid-ask pair in the trading system, and it happens if the trading block is irreducible.

Once the trading block in the dynamical trading system stochastically emerges from Σ_4^ω with a certain stationary probability measure, the trading system will not be unstable for sure. One intuitive reason is that a countably infinite random trading block will be reducible almost surely. Therefore, we state that the bid-ask pair dynamics will be bounded within the domain W for sure, which gives meanings to stochastic stability of the dynamical trading system. More strictly, we show that the domain that serves for stochastic stability is only a proper subset of W, which is defined as its kernel region. However, in the meantime, the buffering region of W still has a positive role for stochastic stability, as it occasionally holds states on the bid-ask pair trajectory.

Finally, it might be worth figuring out one practical application of the notions of kernel region and buffering region. They could be used to measure the risk of systemic instability in a real market. Say, the market would have a high risk of systemic instability, if its state moves into the buffering region, while the risk of systemic instability should be acceptable with a certain confidence level, if the state stays in the kernel region.

CHAPTER 4

Collective Decision on Political Network

In this chapter, we shall study a group of governments structured as a generic political network, and investigate collective decision-making process thereof with certain economic meanings conveyed by the local public good provision.

Although the political network seems like a rather specific group structure by nature, we still formalize it in a general sense, so that we could add a third abstraction "relation-dependent group" to the former two, "time-dependent group" as studied in Chapter 1, and "preference-dependent group" as studied in Chapter 2.

1. Introduction

This chapter aims to explore and solve the optimal provision problem raised in the economic theory of public good provision. In a study of public expenditure on collective consumption by Samuelson [**57**], he claimed that there does not exist any "decentralized pricing system can serve to determine optimally these levels of collective consumption" (Samuelson [**57**], p. 388). However, that impossibility result depends on that the public expenditure can only be manipulated by the central government in an economy, and the economy relies exclusively on market mechanism. In the meanwhile, Samuelson equally admitted that the optimal decisions of collective consumption would exist, in case sufficient knowledge and ethical welfare function for the economy could be obtained. As he wrote in the article "*The pure theory of public expenditure*" (Samuelson [**57**], p. 389):

> *The failure of market catallactics in no way denies the following truth: given sufficient knowledge the optimal decisions can always be found by scanning over all the attainable states of the world and selecting the one which according to the postulated ethical welfare function is best. The solution "exists"; the problem is how to "find" it.*

Tiebout [**63**] took a different perspective to frame such a public expenditure problem. He introduced the notion of local expenditure, and then developed a simple model which can yield a solution for optimal (local) public expenditure. Since the mechanism in Tiebout's model is purely political, and also the optimality in his model is kept to local levels, the solution proposed by Tiebout in effect has a rather different nature to that of Samuelson.

In the present chapter, we shall take a perspective that combines the approaches of Samuelson and Tiebout to study the public good provision problem, so that we could propose some solutions to the optimal provision problem. With this respect, the mechanism to determine the optimal public good provisions will then have both economic and political natures, and the attained optimality will be not only local but also global. To achieve these objectives, we first introduce a political system into a typical economy, such that it has a group of economic planners (local governments) forming a government network, it has a distribution rule of political powers over the government network, and it has a collection of pairwise disjoint districts, populated by fixed citizens and administered by distinct local governments. After that, we propose a formal definition to the local public good in the economy, and give economic meanings to decision-making processes and corresponding games appearing in the political system.

A well-studied topic related to this study is comparative economic performances under different governance structures. Most studies suggest that the decentralized public good provision scheme should be more efficient than the centralized provision scheme owing to various economic reasons. For example, the dominant efficiency of the decentralized scheme comes from its adaptation to local consumer-voters (Tiebout [**63**]); the decentralization theorem dominates citizen-oriented decision-making processes (Oates [**48, 49**]); the decentralized scheme promotes the emergence and maintenance of an efficient market (Weingast [**67**]); the decentralized scheme preserves market incentives (Qian and Weingast [**53**]). Even if we do not develop a normative principle to compare decentralized and centralized governance structures, we do obtain, under a set of restricted assumptions, a same efficiency judgement on the decentralized and centralized public good provision schemes to the popularly accepted one.

What's more, this chapter might also contribute partly to studies on a number of other topics. To wit, positive applications of the concept authority in organizations and partnerships, *e.g.*, the authority relationship in employment contracts (Simon [**58**]), and the optimal authority allocation in complex partnerships (Francesconi and Muthoo [**23**]); social network game and communication, *e.g.*, information (as a public good) acquisition and communication structure on social networks (Galeotti and Goyal [**24**]); private provision of public good and market institution design, *e.g.*, truthful revelation mechanism that can admit Pareto optima (Groves and Loeb [**27**]), and noncooperative solution with a neutral property of income redistribution (Bergstrom, Blume, and Varian [**7**]).

The writing of the present chapter shall be organized as follows. This section is a general introduction, while the last section provides some additional remarks.

In Section 2, we give formal definitions to related economic and political objects in the political system. Concretely, a government network is defined to be a graph that admits a hierarchical structure in whole. By introducing an authority allocation rule to the government network, we can then accordingly define centralized, autonomous, and decentralized governance structures. Besides, a local public good is characterized by its partial provision capacity (*viz.*, the ability of benefit exclusion) with regards its provision district, and thus it will be defined as a two-dimensional economic object and measured by its provision level and provision capacity.

In Section 3, each government in the political system will be simplified to be a political agency with a utilitarian welfare function. We could then show two basic propositions on optimal local public good provision states under the centralized and decentralized provision schemes. These arguments will also become our fundamental knowledge prepared for the studies in the next three sections, which constitute the most significant parts of this chapter.

Section 4 and 5 will be devoted to extensive studies of the H-form game and M-form game, respectively. Roughly speaking, an H-form game is another word of a negotiation game with "hierarchical interactions", and an M-form game can be represented by a collection of parallel negotiation games which are connected by a noncooperative game with "coordinated interactions". In Section 4, we show that any Pareto efficient solution to an H-form game is determined by two hierarchically consecutive governments, which therefore have the greatest enforceable authorities in the political system. Moreover, the utilitarian solution of an H-form game can be characterized by some polynomial in the political discounting. In Section 5, we mainly study "degenerate" and "analytic" M-form games under the assumption that the provision cost function is separable. The solution to a degenerate M-form game is actually such a strategic equilibrium that is determined by a system of marginal conditions. Although the (utilitarian) solution to an analytic M-form game does not directly have computational properties, we still propose an interval representation, and develop a number of propositions on its bounds.

In Section 6, the more general mixed-form game (usually with great complexity) will be discussed. We define a mixed-form game in such a way that we could have a complete classification of games on a compact government network, *viz.*, any game on a compact government network must be in one of the following three classes — H-form game, analytic M-form game, and mixed-form game. Since any incompact government network could be reduced to a compact one by identifying and removing its authority gaps, we actually have developed a full knowledge of games on any government network. After defining the indicator matrix for a compact government network, we clarify the equivalence relations among the three mathematical objects, that is, graph, game, and matrix. Lastly, we propose a pragmatic solution concept to the mixed-form game, which hopefully can help capture complex interactions in a complex political system.

2. Basic Definitions

2.1. Government Network. Consider a group of n governments

$$N = \{r_0, r_1, r_2, \ldots, r_{n-1}\},$$

in which r_0 is the *central government*, while $r_1, r_2, \ldots, r_{n-1}$ are all *local governments*. Let $c = r_0$, then we can also write

$$N = \{c, r_1, r_2, \ldots, r_{n-1}\}.$$

Let $I = \{0, 1, \ldots, n-1\}$ be the index set of N. Define a mapping $f : I \times I \to \{0, 1\}$, and let $f_{ij} = f(i, j)$ for all $i, j \in I$. Then we have a graph (N, f), whose structure is determined by the $n \times n$ matrix (f_{ij}). We shall say that

(i) (N, f) has no *loop*, if $f_{ii} = 0$ for all $i \in I$.
(ii) (N, f) is *undirected*, if $f_{ij} = f_{ji}$ for all $i, j \in I$.
(iii) $r_{i_1} r_{i_2} \cdots r_{i_l}$ is an *l-cycle* in (N, f), if $f_{i_1 i_2} = f_{i_2 i_3} = \cdots = f_{i_l i_1} = 1$, where $l \geq 3$ and $i_1, i_2, \ldots, i_l$ are distinct.
(iv) (N, f) is *acyclic*, if it does not have any l-cycle for $3 \leq l \leq n$.

Definition. (N, f) is called a *government network*, if it is undirected, acyclic, and without loops.

For all government $r_i, r_j \in N$, the distance (that is, the number of edges on the shortest path) between them is denoted by d_{ij}. Define the *hierarchical level* of a government $r_i \in N$ to be the value of d_{i0}. Let $d_{00} = 0$, and $d_{i0} = \infty$ if there does not exist any path between r_i and c. Thus $d_{i0} \in \{0, 1, \ldots, n-1, \infty\}$ for all $r_i \in N$. If (N, f) is connected, $d_{i0} \neq \infty$ for all $r_i \in N$.

There are two fundamental connected government networks which can be defined as follows:

(i) If there is a unique r_i such that $d_{i0} = k$ for all $k \in \{0, 1, \ldots, n-1\}$, we say (N, f) is an *H-form government network* or *government line*.
(ii) If $d_{i0} = 1$ for all $r_i \neq c$, we say (N, f) is an *M-form government network* or *government star*.

Evidently, any connected government network (N, f) has a collection of blocks, each of which is an H-form or M-form government network.

On a government network (N, f), the group of level-k governments is

$$N_k = \{r_i \in N : d_{i0} = k\},$$

where $k \in \{0, 1, \ldots, n-1, \infty\}$. Note that $N_0 = \{c\}$, and N_∞ contains all the *isolated governments* which do not connect with the central government c. Clearly, $N \setminus N_\infty$ is the connected part of N. If (N, f) is connected, then $N_\infty = \emptyset$.

Let the group of the *directly subordinate governments* to $r_i \in N$ be D_i, then

$$D_i = \{r_j \in N : d_{ij} = 1 \text{ and } d_{j0} \geq d_{i0}\}.$$

Note that $D_0 = N_1$. If $r_i \in N_k$, then $D_i \subseteq N_{k+1}$, where $k \leq n-2$. As for any $r_i, r_j \in N_\infty$ with $d_{ij} = 1$, by definition we have both $r_j \in D_i$ and $r_i \in D_j$, and thus we shall say all the governments in N_∞ have identical roles on the hierarchical structure of (N, f).

Let the group of all the *subordinate governments* to $r_i \in N$ be $\mathcal{D}_i$, then

$$\mathcal{D}_i = \{r_j \in N : 0 < d_{ij} < \infty \text{ and } d_{j0} \geq d_{i0}\}.$$

Note that $\{c\} \cup \mathcal{D}_0 = N \setminus N_\infty$, and $\mathcal{D}_i \supseteq D_i$ for all $r_i \in N$. If $r_i \in N_\infty$, then $\mathcal{D}_i$ is either empty or the connected block (in N_∞) to r_i.

Let the group of *bottom governments* be N_b, then

$$N_b = \{r_i \notin N_\infty : D_i = \emptyset\}.$$

If (N, f) is connected, we can write $N_b = \{r_i \in N : D_i = \emptyset\}$. Note that $\mathcal{D}_i = \emptyset$ for all $r_i \in N_b$.

Example 4.1. Let $N = \{c, r_1, r_2, \ldots, r_{12}\}$, and define a graph (N, f) as shown in Figure 4.1. Observe that (N, f) is a connected government network. We have

$$N_0 = \{c\}, \quad N_1 = \{r_1, r_2, r_3, r_4\}, \quad N_2 = \{r_5, r_6, r_7, r_8, r_9, r_{10}\}, \quad N_3 = \{r_{11}, r_{12}\}.$$

Notice that $D_0 = N_1$, $D_2 = \{r_5\}$, $D_3 = \{r_6, r_7, r_8\}$, $D_4 = \{r_9, r_{10}\}$, and $D_{10} = \{r_{11}, r_{12}\}$. Since $D_i = \emptyset$ for all $r_i \notin \{c, r_2, r_3, r_4, r_{10}\}$, we have

$$N_b = \{r_1, r_5, r_6, r_7, r_8, r_9, r_{11}, r_{12}\}.$$

Finally, we have $\mathcal{D}_i = D_i$ for all $r_i \notin \{c, r_4\}$, while

$$\mathcal{D}_0 = \{r_1, r_2, \ldots, r_{12}\}, \quad \mathcal{D}_4 = \{r_9, r_{10}, r_{11}, r_{12}\}.$$

Example 4.2. Let $N = \{c, r_1, r_2, r_3, r_4, r_5\}$, and define an H-form government network (N, f) as shown in Figure 4.2. Observe that $N_0 = \{c\}$, and $N_i = \{r_i\}$ for all $i \in \{1, 2, 3, 4, 5\}$. Note that for all $i \in \{0, 1, 2, 3, 4\}$,

$$D_i = \{r_{i+1}\}, \quad \mathcal{D}_i = \{r_{i+1}, r_{i+2}, \ldots, r_5\},$$

while $N_b = N_5 = \{r_5\}$.

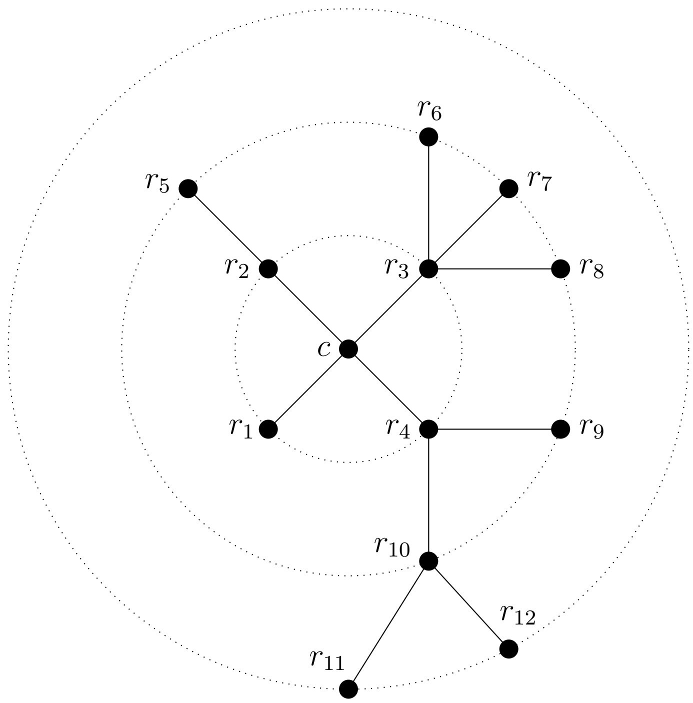


FIGURE 4.1.

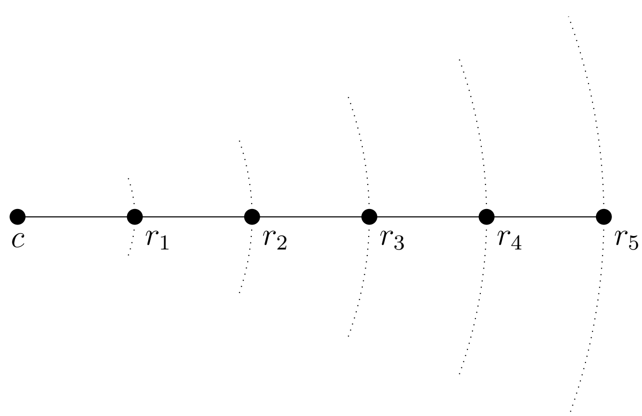


FIGURE 4.2.

EXAMPLE 4.3. Let $N = \{c, r_1, r_2, \ldots, r_6\}$, and define an M-form government network (N, f) as shown in Figure 4.3. Observe that

$$N_0 = \{c\}, \quad N_1 = \{r_1, r_2, r_3, r_4, r_5, r_6\}.$$

Note that $D_0 = N_1$, $\mathcal{D}_0 = D_0$, and $N_b = N_1$; in effect, they are all generic for any M-form government network.

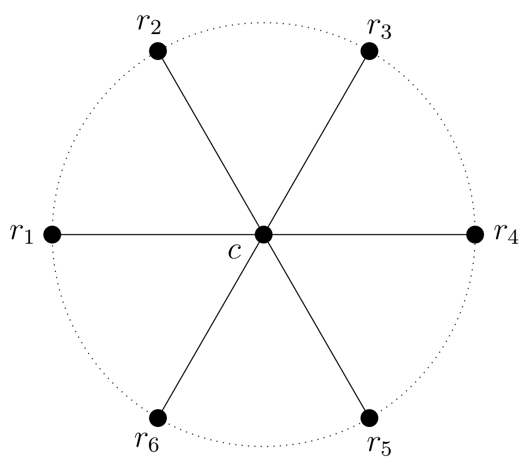


FIGURE 4.3.

2.2. Governance Structure. In order to make formal definitions and develop identification principles for different governance structures, we need a prerequisite notion, authority. That's because the governance arrangements would not set up stable institutional regulations for collective actions in the political system, unless the distributed authorities have been clarified. Here, we shall interpret the authority as the power to make favorable decisions in a decision-making process, and any decision-making process ought to be designed to decide actions for a certain government on the government network.

If a decision-making process aims to decide actions for $r_i \notin N_\infty$, the authority in that process will be shared by a group of governments that participate in the process, such that it is a connected block, and it contains r_i and governments with hierarchical levels lower than that of r_i. In particular, we are interested in the finest group of participating governments which minimizes the *ex ante* transaction costs, though the *ex post* negotiation is assumed to be Coasian. Let L_i denote such a unique group, then

$$L_i = \{r_j \in N : r_i \in \mathcal{D}_j\} \cup \{r_i\}, \tag{4.1}$$

which actually represents a path connecting r_i to c.

If a decision-making process aims to decide actions for $r_i \in N_\infty$, we then directly set $L_i = \{r_i\}$, by virtue of all the governments in N_∞ having identical roles on the hierarchical structure of (N, f) (*cf.*, the notion of directly subordinate government).

Properly speaking, any decision-making process in the political system could then be represented by its participating government path in (N, f). For example, with regards a decision-making process deciding actions for a level-k government, $r_i \in N_k$ for $k \neq \infty$, we can write

$$L_i = \{c, r_{i_1}, r_{i_2}, \ldots, r_{i_{k-1}}, r_i\},$$

where $r_{i_x} \in N_x$ for all $x \in \{1, 2, \ldots, k-1\}$. Let the authority shared by r_{i_x} be γ_x for all x, and the authorities shared by c and r_i be γ_c and γ_k, respectively. Suppose the absolute authority for any decision-making process in the political system is normalized to 1, and its distributed parts to distinct governments are homogeneous. It then appears that

$$\gamma_c + \sum_{x=1}^{k} \gamma_x = 1, \tag{4.2}$$

where $\gamma_x \in [0, 1]$ for all $x \in \{c, 1, 2, \ldots, k\}$.

In general, distributed authorities might not be normally homogeneous in the decision-making process, so there should be a nonlinear authority allocation rule, say Γ, such that

$$\Gamma(\gamma_c, \gamma_1, \gamma_2, \ldots, \gamma_k) = 1. \tag{4.3}$$

However, in the present studies, we will simply assume Γ is linearly additive, and use γ to denote such an authority allocation rule.

Definition. The governance structure of (N, f) is called *centralized*, if $N_\infty = \emptyset$ and $\gamma_c = 1$ for all decision-making process in (N, f).

Once (N, f) has a centralized governance structure, it must be connected, and the authority will be exclusively allocated to the central government c, so that all the local governments are confined to the monolithic authority.

Definition. The governance structure of (N, f) is called *autonomous*, if $N_\infty = N \setminus \{c\}$.

Once (N, f) has an autonomous governance structure, all the local governments should be on the ∞-level, and hence isolated from c. Moreover, any decision-making process aiming to decide actions for $r_i \neq c$ has $L_i = \{r_i\}$ and thus $\gamma_i = 1$, which therefore means that each government in N_∞ has full sovereignty of its own.

Definition. The governance structure of (N, f) is *decentralized*, if it is neither centralized nor autonomous.

If (N, f) has a decentralized governance structure, two cases could be classified as follows:

(i) $N_\infty \neq \emptyset$, and $N \setminus (\{c\} \cup N_\infty) \neq \emptyset$.
(ii) $N_\infty = \emptyset$, but there exists at least one decision-making process with the group L_i, such that $\gamma_c \neq 1$.

In case (i), (N, f) is disconnected, and there exists at least one local government $r_i \in N_\infty$ with $\gamma_i = 1$ for the decision-making process deciding its own actions. In case (ii), (N, f) is connected, and there exists at least one local government in L_i with $\gamma_i \neq 0$ for the decision-making process represented by L_i. Thus in any case the authority will be always allocated to at least one local government, when (N, f) adopts a decentralized governance structure. In addition, it might be noted that the governance structure of a connected government network is either centralized or decentralized.

Let's consider a decision-making process in (N, f) with a decentralized governance structure, and suppose it can be represented by the path

$$L_i = \{c, r_{i_1}, r_{i_2}, \ldots, r_{i_k}\},$$

where $r_{i_x} \in N_x$ for all $x \in \{1, 2, \ldots, k\}$. If the authority is allocated solely to the uppermost-level government r_{i_k} in L_i, then the decision-making process is called *purely decentralized*.

Definition. A decentralized governance structure of (N, f) is called *pure*, if all decision-making process in (N, f) is purely decentralized.

Notice that any decision-making process aiming to decide actions for $r_i \in N_\infty$ is naïvely a purely decentralized one, as its participating group is $L_i = \{r_i\}$. Thus it's sufficient to just use the verification that all decision-making process in $N \setminus N_\infty$ is purely decentralized.

As a further remark on all these above definitions, we would like to comment on the stability of a governance structure which would be implicitly adopted in our investigations. Although the formation process of a governance structure could have a few complex features (see *e.g.*, Wibbels [**68**]), the structure, once it had been shaped, will be assumed to be stable for a sufficiently long time. Thus such a statement that a political system adopts a certain governance structure means that the structure itself is exogenous and will be stable at least in our reasoning framework.

2.3. Local Public Good. Let R denote the region collectively administered by the government network (N, f). Each element $s \in R$ denotes a place in the region R, and each subset $S \subseteq R$ denotes a district of R. Suppose there is a topology, *i.e.*, a collection of subsets of R, such that for every two distinct places s and s' in R, there exist two disjoint districts S and S' of R satisfying $s \in S$ and $s' \in S'$. The topology on R is actually Hausdorff, and the region R described as above is thus a Hausdorff space.

Let $|N_k|$ denote the number of level-k governments for all $k \in \{0, 1, \ldots, n-1, \infty\}$, then

$$\sum_{k=0}^{n-1} |N_k| + |N_\infty| = n,$$

where $|N_0| = 1$ always. If (N, f) is connected, then we have $|N_\infty| = 0$. If (N, f) has an autonomous governance structure, then $|N_\infty| = n - 1$, and $|N_k| = 0$ for all $k \notin \{0, \infty\}$.

Let the district directly administered by $r_i \in N$ be R_i. On each hierarchical level k with $N_k \neq \emptyset$, there are $|N_k|$ governments sharing the control rights over the district $\bigcup_{r_i \in N_k} R_i$, where $R_i \cap R_j = \emptyset$ for all distinct r_i, r_j. It is direct to see that

$$\bigcup_{r_i \in N_b \cup N_\infty} R_i = R.$$

If (N, f) is connected, we have $\bigcup_{r_i \in N_b} R_i = R$, which thus means all the bottom governments make a partition of R.

With regards a level-k government $r_i \in N_k$ having a nonempty $D_i \subseteq N_{k+1}$, we have a partition of R_i,

$$R_i = \bigcup_{r_j \in D_i} R_j,$$

where $R_j \cap R_{j'} = \emptyset$ for all distinct $r_j, r_{j'}$. The district R_i administered by r_i thus contains $|D_i|$ (no greater than $|N_{k+1}|$) disjoint districts, which are administered by the governments in D_i, separately. In particular, in a connected (N, f), the central government c administers the whole region R which is also fully administered by all the level-1 governments in N_1, by virtue of $R = \bigcup_{r_i \in N_1} R_i$ for $D_0 = N_1$.

Let the population of the region R be P. Define a mapping

$$P : S \mapsto P(S),$$

so that it generates a population $P(S) \subseteq P$ for all district $S \subseteq R$. Since the population of a district is an invariant representation for the district, the mapping should satisfy

(i) $P(\emptyset) = \emptyset$ and $P(R) = P$,
(ii) $P(S) \subseteq P(T)$ if $S \subseteq T$,
(iii) $P(S \cup T) = P(S) \cup P(T)$ for all $S, T \subseteq R$.

Let $P(R_i) = P_i$ for all $r_i \in N$, then each district R_i has a population P_i. It can be claimed that $P_i \neq \emptyset$ for all $r_i \in N$, because there is no reason to place a government for a district without citizen, and hence without decision-making process. When (N, f) is connected, we have $\bigcup_{r_i \in N_b} P_i = P$, where $P_i \cap P_j = \emptyset$ for all distinct r_i, r_j.

Definition. A public good G is called *local*, if there exist nonempty districts $S, T \subset R$, such that $P(S)$ can consume G but $P(T)$ is excluded from the benefits of G.

A pure (or global) public good can be defined by such properties as "jointness of supply" and "impossibility of exclusion" (see Hardin [**28**], p. 17). Our definition of local public good generalizes the assumption about the exclusion ability, in such a way that the partial exclusion impossibility replaces the absolute one adopted in the definition of public good.

By means of the definition of local public good, we are able to bridge the gap between private good and public good, and think of private good and pure public good as two special cases of regular local public good. In fact, if G is provided to cover the region R, and no citizen in P is excluded from its benefits, then G is clearly a pure public good. On the other hand, if G is provided to cover a (virtual) district S with a single citizen, and all the population P except that citizen living in S can be excluded from its benefits, then G is actually a typical private good. In general, G is a regular local public good if it is neither private nor purely public; here, "local" means it partially excludes a proper and nonempty subset of P from its benefits.

To treat the definition of local public good analytically, we gauge the provision ability of a local public good using a measure of analytic sets of the Hausdorff space R, which is in essence a capacity monotone of order 2 in Choquet's [**12**] term. Let $\wp(R)$ denote the power set of R, and define a mapping $\lambda : \wp(R) \to [0, 1]$ such that

(i) $\lambda(\emptyset) = 0$ and $\lambda(R) = 1$,
(ii) $\lambda(S) \leq \lambda(T)$ if $S \subseteq T$,
(iii) $\lambda(S \cup T) + \lambda(S \cap T) \geq \lambda(S) + \lambda(T)$ for all $S, T \subseteq R$.

If λ satisfies property (iii), it is called *superadditive*. It can be noted that for any superadditive λ, $\lambda(S \cup T) \geq \lambda(S) + \lambda(T)$ for all disjoint $S, T \subseteq R$.

The capacity λ as a measure of provision ability is assumed to be superadditive rather than *subadditive*, because a local public good provided to $S \cup T$ seems to need a less exclusion ability than that needed by two local public goods, which are provided to S and T, separately.

The provision state of a local public good G can thus be characterized by the pair (g, S), where $g \in \mathbb{R}^+$ denotes the provision level, and $S \subseteq R$ denotes the provision district. Since a district S can now be measured by its provision capacity $\lambda_S = \lambda(S)$, the provision state of G can be equivalently characterized by the real pair (g, λ_S), where $\lambda_S \in [0, 1]$. We shall see that

(i) If $g = 0$, G is in fact not provided at all.
(ii) If $g > 0$ and $\lambda_S = 1$, then G is actually a pure public good.

(iii) If $g > 0$ and $\lambda_S = 0$[1], then G turns out to be a private good.
(iv) If $g > 0$ and $\lambda_S \in (0, 1)$, then G is a regular local public good which partially excludes $P \setminus P(S)$ from its benefits.

Suppose the local public good provision is implemented by the governments in $N_b \cup N_\infty$ (*i.e.*, either bottom or isolated), and all the implementing governments have a same real provision cost function[2] $M(g, \lambda)$, where $(g, \lambda) \in \mathbb{R}^+ \times [0, 1]$ measures the provision state. Once the real pair (g, λ) is replaced with its initial pair (g, S), the provision cost function should then be written as $M(g, \lambda(S))$, or $M(g, \lambda_S)$ for $\lambda_S = \lambda(S)$.

Moreover, we shall assume

$$M_g \geq 0, \quad M_\lambda \geq 0, \quad M_{gg} \geq 0, \quad M_{\lambda\lambda} \geq 0,$$

where all these inequalities will be strict if $g > 0$ and $\lambda > 0$. It thus directly suggests that $M(g, \lambda)$ is convex in both g and λ. Additionally, to get rid of such an optimal provision state as (∞, λ) or $(g, 1)$, that is, an infinite provision level or a perfect provision capacity, we ought to set the following boundary conditions,

$$M_g(0, \lambda) = 0, \quad M_g(\infty, \lambda) = \infty, \quad M_\lambda(g, 0) = 0, \quad M_\lambda(g, 1) \leq \infty.$$

EXAMPLE 4.4. Consider $M(g, \lambda) = \varphi(g) + \psi(\lambda)$, where $\varphi(g)$ and $\psi(\lambda)$ are both convex functions. Assume $\varphi'(0) = 0$, $\varphi'(\infty) = \infty$, $\psi'(0) = 0$, and $\psi'(1) \leq \infty$, then $M(g, \lambda)$ is clearly a provision cost function. Note that $M_{g\lambda} = 0$. For example, let $\varphi(g) = \beta_1 g^2$ and $\psi(\lambda) = \beta_2 \exp(\lambda^2)$ for $\beta_1, \beta_2 > 0$, then the provision cost function is written in the form

$$M(g, \lambda) = \beta_1 g^2 + \beta_2 \exp(\lambda^2). \tag{4.4}$$

EXAMPLE 4.5. Consider $M(g, \lambda) = \varphi(g)\psi(\lambda)$, where $\varphi(g)$ and $\psi(\lambda)$ are both real and convex. Assume $\varphi'(0) = \psi'(0) = 0$ and $\psi'(1) \leq \varphi'(\infty) = \infty$. In this case, we have $M_{g\lambda} = \varphi'(g)\psi'(\lambda)$ which is normally nonzero. For example, let $\varphi(g) = \beta_1 g^2/2$ and $\psi(\lambda) = \beta_2 \lambda^2/2$ for $\beta_1, \beta_2 > 0$, then

$$M(g, \lambda) = \beta(\lambda g)^2/4, \tag{4.5}$$

where $\beta = \beta_1 \beta_2$. Note that $M_{g\lambda} = \beta \lambda g \geq 0$.

[1] More precisely, we should write $\lambda_S = \mathrm{o}(1)$, as λ_S is an infinitesimal quantity but should not be equal to zero.

[2] It implicitly implies the technology for providing the local public good is identical for all the implementing governments. On a connected government network with a hierarchical (or vertical) structure, it is evidently true by the fact that the diffusion of a technology could be thought of to be immediate, for example, its spread from the level k to the level $k+1$, or from the level k to the level 0 and then directly to the level $k-1$, where $1 \leq k \leq n-2$.

3. Local Provision

3.1. Political Agency. A government $r_i \in N$ in the political system will be considered as a governmental agent of the population P_i living in its administered district R_i. Since we have already set up the economic context in the political system, *viz.*, local public good provision, a decision-making process that aims to decide the provision state for r_i should exclusively have economic natures. After clarifying the political aggregation law on P_i adopted by r_i, we will perceive, at the end of this subsection, that r_i is actually thought of as an economic agent which therefore replaces the initial governmental one.

Suppose the population P of the region R can be partitioned into a collection of disjoint groups according to a combination of social attributes of the citizens, such as age, profession, income, and so on. Let J denote a finite index set such that

$$P = \bigcup_{j \in J} P_j,$$

where $|J| \ll |P|$. Each population P_i of the district R_i will then also be partitioned into a collection of groups $\{P_i \cap P_j : j \in J\}$. Let $P_i^j = P_i \cap P_j$ for all $j \in J$, then $P_i = \bigcup_{j \in J} P_i^j$ for all $r_i \in N$. Notice that $P_c = P$, and hence $P_c^j = P_j$ for all $j \in J$.

Assume each government $r_i \in N$ adopts a political aggregation law on P_i in order to generate a unique citizen $p_i^j \in P_i^j$ as a representative of P_i^j for all j with $P_i^j \neq \emptyset$. Let P_i^r denote the committee of all the representatives of P_i, then

$$P_i^r = \{p_i^j \in P_i^j : j \in J \text{ and } P_i^j \neq \emptyset\}.$$

In case r_i happens to be the central government c, we will take the notation P^r instead of P_c^r, and p_j instead of p_c^j.

Recall that we claim $P_i \neq \emptyset$ for all $r_i \in N$, so any r_i has at least one nonempty P_i^j, and thus $P_i^r \neq \emptyset$ for all $r_i \in N$. Evidently, $1 \leq |P_i^r| \leq |J|$ for all $r_i \in N$, and $|P^r| = |J|$. Once $|P_i^r| = |J|$, we simply set $P_i^r = P^r$, so that the political aggregation laws of r_i and c should in principle generate a same representative committee.

Suppose every citizen in P_i^j has a rational preference for the local public good, in the sense that the representative $p_i^j \in P_i^j$ has a utility function of local public good provision state in R_i,

$$v_i^j(g, S) = \begin{cases} u_i^j(g) & \text{if } p_i^j \in P(S) \\ 0 & \text{if } p_i^j \notin P(S) \end{cases} \tag{4.6}$$

where $P(S)$ is the population of the district S, and $u_i^j(g)$ is a concave utility function. If $P_i^r = P^r$, we use the notations $v_j(g, S)$ and $u_j(g)$ to replace $v_i^j(g, S)$ and $u_i^j(g)$, respectively. As usual, we assume $u_j'(0) = \infty$ and $u_j'(\infty) = 0$, so that any p_j should require a nonzero and finite provision level to have a reasonable marginal utility balancing her satiable wants.

Each government r_i is supposed to take a normative utilitarianism perspective towards its committee P_i^r. In particular, the welfare of r_i is assumed to be equal to the weighted average of its representatives' utilities, in other words, it is determined by the societal utility function of the committee P_i^r which admits a functional form as was proposed by Harsanyi [**30**]. Consequently, the welfare function of r_i can be written as

$$V_i(g,S) = \sum_{p_i^j \in P_i^r} \alpha_i^j v_i^j(g,S), \tag{4.7}$$

where $\alpha_i^j \in (0,1)$ for all p_i^j, and $\sum_{p_i^j \in P_i^r} \alpha_i^j = 1$. Let

$$U_i(g) = \sum_{p_i^j \in P_i^r} \alpha_i^j u_i^j(g). \tag{4.8}$$

If $P(S) \supseteq P_i$ or $S \supseteq R_i$, we will definitely have $V_i(g,S) = U_i(g)$. Note that $V_i(g,S) \leq U_i(g)$ for all $g \geq 0$ and all $S \subseteq R$.

In case $P_i^r = P^r$, we will use the notations α_j, $V(g,S)$, and $U(g)$ instead of α_i^j, $V_i(g,S)$, and $U_i(g)$, simply as the committee of r_i now coincides with that of c. So we will have

$$V(g,S) = \sum_{p_j \in P^r} \alpha_j v_j(g,S), \quad U(g) = \sum_{p_j \in P^r} \alpha_j u_j(g).$$

As a result, a government $r_i \in N_b \cup N_\infty$, which implements provision actions to provide the local public good, will have an objective function of the provision state (g,S),

$$\Pi_i(g,S) = V_i(g,S) - M(g,\lambda(S)). \tag{4.9}$$

If $S \supseteq R_i$ and $P_i^r = P^r$, the above objective function (4.9) will become a uniform one,

$$\Pi_i(g,S) = U(g) - M(g,\lambda(S)). \tag{4.9$'$}$$

However, the objective function of a government $r_i \notin N_b \cup N_\infty$ can not be specified by a general functional form at this moment, because its (shared, but not direct) provision cost function should depend on the structure of (N,f).

3.2. Optimal Provision State. The main goal of this subsection is to show two propositions on optimal states of the local public good provision in a government network when its governance structure is centralized or purely decentralized. With either governance structure, the authority for any decision-making process will be absolutely allocated to a single government, that is, either the central government c if centralized, or a government in $N_b \cup N_\infty$ if purely decentralized.

PROPOSITION 4.1. *If $r_i \in N_b$ has $\gamma_i = 1$ in the local public good provision decision-making process for r_i, the provision district decision will be exactly R_i.*

PROOF. Let the decision on the local public good provision state for r_i be (g, S), where $g \geq 0$ and $S \subseteq R$. We first show $S \supseteq R_i$. Suppose not, we would have $S \subset R_i$ and thus $R_i \setminus S \neq \emptyset$. But then $P(R_i \setminus S) \neq \emptyset$, otherwise the administered district of r_i would be S rather than R_i. In the meantime, we should see $P(S) = P_i$, otherwise the population in $R_i \setminus S$ would be attracted to the provision district S. It therefore appears that $P(R_i \setminus S) = P_i \setminus P(S) = \emptyset$, a contradiction.

Now that $S \supseteq R_i$, we have $V_i(g, S) = U_i(g)$. Let $\lambda = \lambda(S)$, then the objective function of r_i can be written as

$$\Pi_i(g, S) = U_i(g) - M(g, \lambda).$$

If $\gamma_i = 1$ in the provision decision-making process for r_i, its decision should directly maximize $\Pi_i(g, S)$. Recall that $U_i'(0) = \infty$ and $M_g(0, \lambda) = 0$, thus $(\partial \Pi_i/\partial g)(0, S) = \infty$ which implies the provision level decision $g > 0$. Since $M_\lambda \geq 0$, $\partial \Pi_i/\partial \lambda = -M_\lambda \leq 0$, where the inequality will be strict if $\lambda \neq 0$. By $S \supseteq R_i$ we have $\lambda = \lambda(S) \geq \lambda(R_i)$, and hence $\Pi_i(g, S) \leq \Pi_i(g, R_i)$. Moreover, $\Pi_i(g, S) < \Pi_i(g, R_i)$ if $S \supset R_i$ for $\lambda(R_i) \neq 0$. Consequently, the provision district decision $S = R_i$, which then completes the proof. □

A direct corollary of Proposition 4.1 is that the provision district decision for any $r_i \in N_\infty$ must be R_i. The reason is that the group of participating governments in the provision decision-making process for r_i is the singleton $L_i = \{r_i\}$, in which r_i should naturally have the absolute authority and hence $\gamma_i = 1$.

It comes to be realized that any government $r_i \in N_b \cup N_\infty$ with $\gamma_i = 1$ in the decision-making process for r_i will make a provision decision (g, R_i), or (g, λ_i) for $\lambda_i = \lambda(R_i)$. Clearly, the objective function of $r_i \in N_b \cup N_\infty$ is now written as a real function in g,

$$\Lambda_i(g) = \Pi_i(g, R_i) = U_i(g) - M(g, \lambda_i). \tag{4.10}$$

Since $U_i'(0) = \infty$, $M_g(0, \lambda_i) = 0$, $U_i'(\infty) = 0$, and $M_g(\infty, \lambda_i) = \infty$, we have

$$\Lambda_i'(0) = \infty, \quad \Lambda_i'(\infty) = -\infty.$$

Recall also that $U_i'' \leq 0$ and $M_{gg} \geq 0$, so $\Lambda_i'(g)$ should be monotonic in g, and thus there exists a unique $g_i \in (0, \infty)$ such that $\Lambda_i'(g_i) = 0$, or equivalently,

$$U_i'(g_i) = M_g(g_i, \lambda_i). \tag{4.11}$$

So far, we could arrive at a significant conclusion that the unique equilibrium provision states in (N, f) with a pure decentralized governance structure can be described by the collection

$$\big\{(g_i, R_i) : r_i \in N_b \cup N_\infty\big\},$$

where $\bigcup_{r_i \in N_b \cup N_\infty} R_i = R$ and $g_i \in (0, \infty)$ for all r_i. In this respect, such a systematic provision arrangement is called the *decentralized provision scheme*.

If a government network takes a centralized governance structure, then the central government c has the absolute authority (*i.e.*, $\gamma_c = 1$) in any provision decision-making process for $r_i \in N_b$. In practice, c makes a provision decision for each $r_i \in N_b$, and each $r_i \in N_b$ will then implement provision actions according to the decision approved by c. Suppose c decides a uniform provision level for all $r_i \in N_b$, and perfectly identifies provision costs as if c itself implemented provision actions in the whole region R, then c would have an objective function of provision state (g, S),

$$\Pi_c(g, S) = V(g, S) - M(g, \lambda(S)). \tag{4.12}$$

The provision decision procedure behind (4.12) is called the *centralized provision scheme*.

Proposition 4.2. *In a connected government network (N, f) the decentralized provision scheme is more efficient than the centralized provision scheme, if $P_i^r = P^r$ for all $r_i \in N_b$ and $M_{g\lambda} = 0$.*

Proof. By similar arguments developed in the proof of Proposition 4.1, the provision district decision under the centralized provision scheme is R which has a capacity $\lambda(R) = 1$. c thus has the objective function $\Pi_c(g, R) = U(g) - M(g, 1)$ as per (4.12), so the provision decision is (g_c, R) such that $U'(g_c) = M_g(g_c, 1)$.

Since $P_i^r = P^r$ for all $r_i \in N_b$, we have $U_i(g) = U(g)$ for all $r_i \in N_b$. The provision decision for $r_i \in N_b$ under the decentralized provision scheme is (g_i, R_i) such that $U'(g_i) = M_g(g_i, \lambda(R_i))$ as per (4.11), where $\bigcup_{r_i \in N_b} R_i = R$. By $M_{g\lambda} = 0$ we should see

$$U'(g_i) = M_g(g_i, \lambda(R_i)) = M_g(g_i, 1),$$

and thus $g_i = g_c$ for all $r_i \in N_b$. Recall that the capacity λ is superadditive, so

$$\sum_{r_i \in N_b} \lambda(R_i) \le \lambda\Big(\bigcup_{r_i \in N_b} R_i\Big) = \lambda(R) = 1.$$

By virtue of $M_\lambda \ge 0$ and $M_{\lambda\lambda} \ge 0$,

$$\sum_{r_i \in N_b} M(g_i, \lambda(R_i)) = \sum_{r_i \in N_b} M(g_c, \lambda(R_i)) \le M\Big(g_c, \sum_{r_i \in N_b} \lambda(R_i)\Big) \le M(g_c, 1).$$

It is now evident that these two provision schemes yield a same social welfare $U(g_c)$, but different provision costs. The judgement on their efficiencies should then be direct to derive. $\square$

In addition, at the equilibrium provision state (g_i, R_i) the following comparative statics can be acquired:

$$\frac{\mathrm{d}g_i}{\mathrm{d}\lambda_i} = \frac{M_{g\lambda}(g_i, \lambda_i)}{U''(g_i) - M_{gg}(g_i, \lambda_i)}, \tag{4.13}$$

where $\lambda_i = \lambda(R_i)$. In Proposition 4.2, we assume $M_{g\lambda} = 0$ to have $\mathrm{d}g_i/\mathrm{d}\lambda_i = 0$, so that $g_i = g_c$ for all $r_i \in N_b$. If $M_{g\lambda} < 0$, then $\mathrm{d}g_i/\mathrm{d}\lambda_i > 0$ and hence $g_c > g_i$ for all $r_i \in N_b$, which means the local public good will be over-provided and require much more provision costs under the centralized provision scheme. On the other hand, if $M_{g\lambda} > 0$, then $\mathrm{d}g_i/\mathrm{d}\lambda_i < 0$ and hence $g_i > g_c$ for all $r_i \in N_b$, which means the local public good will be over-provided under the decentralized provision scheme. However, the efficiency judgement on these two provision schemes when $M_{g\lambda} \neq 0$ seems unlikely in general.

4. Hierarchical Interaction

We now advance to study collective decision-making processes on a government network in which at least two governments share the authority. A collective decision-making process is clearly much more complicated than what we studied in the last "preparation" section, in which only one government has the authority and provision decisions are always kept to the local administration. As we have mentioned in the very beginning, two quite fundamental connected government network will be studied in the present chapter, *viz.*, the H-form government line and the M-form government star. This section will be devoted to a formal study of the H-form government line and games emerging on it, while the next section will investigate in detail the M-form government star and its associated games. Properly speaking, we shall call a game emerging on an H-form government line an *H-form game*, and similarly, a game on an M-form government star can be named as an *M-form game*.

4.1. H-form Game. Recall that $N = \{c, r_1, r_2, \ldots, r_{n-1}\}$, where $c = r_0$ and $I = \{0, 1, \ldots, n-1\}$ denotes the index set of N. Suppose (N, f) is an H-form government line, then $|N_i| = 1$ for all $i \in I$. Without loss of generality, we assume $N_i = \{r_i\}$ for all $i \in I$, so that $D_i = \{r_{i+1}\}$ for all $i \in I \setminus \{n-1\}$, and $D_{n-1} = \emptyset$, and hence $N_b = \{r_{n-1}\}$. Evidently, an H-form government line is actually a chain with a simple and complete order over N, for this reason, we shall simply use N to denote it. Besides, we should notice that $R_i = R$ and $P_i = P$ for all $r_i \in N$ in the H-form government line N.

If any decision-making process in the political system is intentionally designed to decide the local public good provision for the bottom government r_{n-1} which in practice implements provision actions thereafter the decision, there will be a unique decision-making process with a group of participating governments $L_{n-1} = N$. The authority for that decision-making process is distributed over N, such that r_i has an authority γ_i, and $\sum_{r_i \in N} \gamma_i = 1$. Clearly, there is at least one government with a nonzero authority, and possibly $\gamma_i \neq 0$ for all $r_i \in N$. If $\gamma_c = 1$, the governance structure of N will be centralized; if $\gamma_{n-1} = 1$, its governance structure will be purely decentralized; if there are at least two governments with nonzero authorities, its governance structure will be decentralized but impure.

Let r_t and r_b be the lowermost-level government and uppermost-level government among the governments with nonzero authorities in N, then the indices t, b can be defined by

$$t = \min\{i \in I : \gamma_i \neq 0\}, \quad b = \max\{i \in I : \gamma_i \neq 0\},$$

where $0 \leq t \leq b \leq n-1$. Once $t = b$, the authority will be solely allocated to the unique government $r_t = r_b$. If $t \neq b$, there should be at least two governments with nonzero authorities.

Definition. If there is a government r_m in the H-form government line N such that $\gamma_m = 0$ for $t < m < b$, N is then said to have an *authority gap*.

Notice that if $t \neq 0$ and $b \neq n-1$, all government r_m for $0 \leq m < t$ and $b < m \leq n-1$ will have no authority at all, however, they do not shape any authority gap in N by the above definition. In a much stricter sense, we shall say N has *no authority gaps*, if $\gamma_i \neq 0$ for all $0 \leq i \leq n-1$.

More generally, we say a government network (N, f) has an authority gap on a path, if there is a government on the path (as a local H-form government line) satisfying the above definition. And we say (N, f) has no authority gaps, if the authority for any decision-making process which aims to decide actions for some r_i, will be allocated to all the governments in L_i.

To simplify our analysis, we shall claim that an authority gap does not affect negotiations or interactions in the H-form government line N at all. Now suppose N has some authority gaps (between r_t and r_b), then the participating group in the collective decision-making process must have at least two governments with nonzero authorities, and some "active" governments in the group are not connected directly but through certain intermediary "inactive" governments which in this case have zero authorities. Since the decision-making process does not depend on any private information of the intermediary governments situated in authority gaps, information transmission over N should be naïvely equivalent to transmissions over just the group of these governments with nonzero authorities. Consequently, a game on the H-form N should be completely determined by a certain authority allocation rule γ, and two games on N would have a similar structure if the numbers of nonzero-authority governments of these two games are same. It then appears to us that there are in total n classes of game structure on N, and each can be characterized by the fact that the authority for the collective decision-making process is shared by k consecutive governments for $k \in \{1, 2, \ldots, n\}$.

If there is a single government $r_i \in N$ with $\gamma_i = 1$, then the provision district decision will be R with a capacity $\lambda(R) = 1$. So the objective function of r_i is written as $U_i(g) - M(g, 1)$, whether the provision scheme is decentralized or centralized, and the provision level decision is g_i such that for all $g \geq 0$,

$$U_i(g_i) - M(g_i, 1) \geq U_i(g) - M(g, 1),$$

where $U_i = U$, as $P_i = P$ and hence $P_i^r = P^r$. By the marginal condition (4.11), g_i should be the unique solution of the following equation,

$$U'(g) - M_g(g, 1) = 0.$$

The provision decision of r_i will be the pair (g_i, R), where $g_i = g_{n-1}$ for all $r_i \in N$. However, if the provision decision of $r_i \neq r_{n-1}$ is not simply determined through the decentralized or centralized provision scheme, it is very likely that $g_i \neq g_{n-1}$.

If there are at least two governments with nonzero authorities, the collective decision-making process will turn out to be a negotiation game, in which governments bargain to reach agreement on the local public good provision decision, and the bargaining power of each government is assumed to be equal to its allocated authority.

Suppose k consecutive governments in N share the authority for the negotiation game, where $2 \leq k \leq n$. There are $n - k + 1$ possible combinations,

$$\{r_i, r_{i+1}, \ldots, r_{i+k-1}\} \quad \text{for} \quad i \in \{0, 1, \ldots, n-k\}.$$

Since they are all in a same class of game structure, we can focus on the specific one, to which the bottom government r_{n-1} always belongs, that is,

$$\{r_{n-k}, r_{n-k+1}, \ldots, r_{n-1}\}.$$

It comes out to be an H-form government line as well, and clearly equivalent to $\{c, r_1, r_2, \ldots, r_{k-1}\}$, in which $\gamma_i \neq 0$ for all r_i. By taking a different notation, we can therefore directly concentrate on the H-form government line $N = \{c, r_1, r_2, \ldots, r_{n-1}\}$ for $n \geq 2$, in which $\gamma_i \neq 0$ for all r_i.

4.2. Negotiation Game. Note that $P_{n-1} = P$ and hence $P_{n-1}^r = P^r$, then the welfare function of the bottom government r_{n-1} can be written in the form

$$V(g, S) = \sum_{p_j \in P^r} \alpha_j v_j(g, S),$$

where $\alpha_j \in (0,1)$ for all p_j, and $\sum_{p_j \in P^r} \alpha_j = 1$. Since r_{n-1} is the only government that really implements local public good provision actions in the political system, its (direct) provision cost function ought to be $M(g, \lambda(S))$. Finally, we obtain its objective function,

$$\Pi_{n-1}(g,S) = V(g,S) - M(g,\lambda(S)). \tag{4.14}$$

Recall that $D_i = \{r_{i+1}\}$ for all $r_i \neq r_{n-1}$. Assume each government $r_i \neq r_{n-1}$ has such an objective function that is a convex combination of its own welfare function and the objective function of its directly subordinate government r_{i+1}. Since $P_i^r = P^r$ for all r_i, the welfare function of r_i is also $V(g,S)$. The objective function of r_i can therefore be expressed as

$$\Pi_i(g,S) = (1-\delta)V(g,S) + \delta\Pi_{i+1}(g,S), \tag{4.15}$$

where $0 \leq i \leq n-2$ and $\delta \in [0,1]$.

Here, the constant δ is called the *political discounting* in the political system, in the sense that any government can only partially identify the welfare states of its directly subordinate governments. If $\delta = 1$, then $\Pi_i(g,S) = \Pi_{i+1}(g,S)$, and thus all the governments in N would have a same objective function $\Pi_{n-1}(g,S)$. If $\delta = 0$, then $\Pi_i(g,S) = V(g,S)$ for all $r_i \neq r_{n-1}$, which thus means all the nonbottom governments would directly take the welfare function $V(g,S)$ as their own objective functions. In most discussions, a *regular* political discounting will be assumed in advance however, in order to keep $\delta \in (0,1)$.

According to the recursive formula (4.15) and the initial condition (4.14), we can obtain

$$\Pi_i(g,S) = V(g,S) - \delta^{n-1-i}M(g,\lambda(S)), \tag{4.16}$$

where $0 \leq i \leq n-1$. In particular, the objective function of the central government c now takes the form

$$\Pi_c(g,S) = V(g,S) - \delta^{n-1}M(g,\lambda(S)).$$

Despite none of the governments has the absolute authority, the provision district decision of the negotiation game should be R, as all the governments unanimously agree on that (*cf.*, arguments in the proof of Proposition 4.1). The welfare function $V(g,R)$ then becomes

$$U(g) = \sum_{p_j \in P^r} \alpha_j u_j(g),$$

and the provision cost function of r_{n-1} becomes $M(g,1)$. The objective function of each $r_i \in N$ can thus be rewritten as

$$\Pi_i(g,R) = U(g) - \delta^{n-1-i}M(g,1). \tag{4.16$'$}$$

The status quo provision level of each r_i denoted by g_i can be directly thought of as the provision state maximizing its objective function $\Pi_i(g, R)$. It is clear that g_i is the unique provision level that satisfies the following marginal condition,

$$U'(g_i) = \delta^{n-1-i} M_g(g_i, 1). \tag{4.17}$$

In case the political discounting δ is irregular, *i.e.*, $\delta \in \{0, 1\}$, the status quo provision level would either be trivial or not make sense in the political system. In more detail, if $\delta = 1$, we would have $g_i = g_{n-1}$ for all $r_i \neq r_{n-1}$, yet the negotiation game would then have no reason to exist; if $\delta = 0$, we would have $g_i = \infty$ for all $r_i \neq r_{n-1}$, but it were clearly impossible for them to stand as the status quo. This gives us another reason to set $\delta \in (0, 1)$ so as to keep the political discounting regular.

If $\delta \in (0, 1)$, it then finely appears that $g_i \neq g_{i+1}$ for all $0 \leq i \leq n-2$. So for each $r_i \neq r_{n-1}$, there exist two provision levels between g_i and g_{i+1}, say $g_\imath$ and $g_{\hat\imath}$ such that

$$U'(g_i) - U'(g_{i+1}) = U''(g_\imath)(g_i - g_{i+1}),$$
$$M_g(g_i, 1) - M_g(g_{i+1}, 1) = M_{gg}(g_{\hat\imath}, 1)(g_i - g_{i+1}).$$

By the formula (4.17), $U'(g_i)$ and $U'(g_{i+1})$ can be substituted in terms of $M_g(g_i, 1)$ and $M_g(g_{i+1}, 1)$, thus

$$g_i - g_{i+1} = \frac{(\delta - 1)\delta^{n-2-i} M_g(g_i, 1)}{U''(g_\imath) - \delta^{n-2-i} M_{gg}(g_{\hat\imath}, 1)}.$$

Since $U'' < 0$, $M_{gg} > 0$, and $M_g > 0$ for all positive provision level, $g_i - g_{i+1} > 0$, and therefore $g_i > g_{i+1}$ for all $0 \leq i \leq n-2$, which then yields

$$g_c > g_1 > g_2 > \cdots > g_{n-2} > g_{n-1} > 0.$$

LEMMA 4.3. *Any Pareto efficient bargaining solution to the negotiation game on N must be a convex combination of g_{i+1} and g_i for some $0 \leq i \leq n-2$.*

PROOF. Note that all $\Pi_i(g, R)$ is concave in g, and achieves its maximum at g_i, so any provision level $g > g_c$ or $g < g_{n-1}$ is clearly not Pareto efficient. Thus any Pareto efficient bargaining solution to the negotiation game should be a convex combination of g_{n-1} and g_c. Since all the n status quo provision levels $g_c, g_1, g_2, \ldots, g_{n-1}$ are pairwise distinct, the interval (g_{i+1}, g_i) for $0 \leq i \leq n-2$ is dense.

A Pareto efficient bargaining solution is therefore either equal to such a status quo provision level as g_i, or located in such an interval as (g_{i+1}, g_i). In either case, there is a constant $\varepsilon_i \in [0, 1]$ such that this bargaining solution is equal to $(1 - \varepsilon_i)g_{i+1} + \varepsilon_i g_i$, where $0 \leq i \leq n-2$ fixed, which then completes the proof. $\square$

No matter which Pareto efficient bargaining solution concept we take, by Lemma 4.3, there is always an interval $[g_{i+1}, g_i]$ for some $0 \leq i \leq n-2$, which can serve as the negotiation set for the game. It helps us notice that a negotiation game with more than two government players can be in practice reduced to an equivalent one with just two government players, in case we could identify its ad hoc negotiation set. The basic but rather critical feature of a negotiation game thus lies in political interactions between each pair of consecutive governments in N. In consideration of this fact, we shall first study in detail the negotiation game that appears on $\{c, r_1\}$.

4.3. Utilitarian Solution. Let's now take the utilitarian bargaining solution concept for the two-government negotiation game on $\{c, r_1\}$. The solution to that negotiation game can be captured by a procedure that c and r_1 jointly minimize the weighted average of their welfare losses with respect to their status quo welfare states, where a status quo welfare state is determined by its corresponding status quo provision level, and the weighting factor of a government is equal to its corresponding allocated authority.

The status quo provision levels of c and r_1 are g_c and g_1 with $g_c > g_1 > 0$, and according to (4.17) for $n = 2$ here, they should satisfy

$$U'(g_c) = \delta M_g(g_c, 1), \quad U'(g_1) = M_g(g_1, 1).$$

These two status quo states can be denoted by $\Pi_c^* = \Pi_c(g_c, R)$ and $\Pi_1^* = \Pi_1(g_1, R)$. In addition, the authority allocation rule γ over $\{c, r_1\}$ delivers γ_c and γ_1 to c and r_1, where $\gamma_c, \gamma_1 > 0$ and $\gamma_c + \gamma_1 = 1$.

The collective strategy space of the negotiation game between c and r_1 is in effect the Pareto efficient negotiation set $[g_1, g_c]$. Let g^u denote the utilitarian bargaining solution to that negotiation game, then it should be the argument

$$\min_{g \in [g_1, g_c]} \gamma_c\big(\Pi_c^* - \Pi_c(g, R)\big) + \gamma_1\big(\Pi_1^* - \Pi_1(g, R)\big), \tag{4.18}$$

or equivalently,

$$\max_{g \in [g_1, g_c]} \gamma_c \Pi_c(g, R) + \gamma_1 \Pi_1(g, R). \tag{4.18$'$}$$

Proposition 4.4. *$g^u = (1 - \varepsilon)g_1 + \varepsilon g_c$, in which $\varepsilon \in (0, 1)$ and moreover, ε increases when $\gamma_c \delta + \gamma_1$ decreases.*

Proof. Recall that

$$\Pi_c(g, R) = U(g) - \delta M(g, 1), \quad \Pi_1(g, R) = U(g) - M(g, 1).$$

g^u is then the argument

$$\max_{g \in [g_1, g_c]} U(g) - (\gamma_c \delta + \gamma_1) M(g, 1), \tag{4.18$''$}$$

and thus g^u satisfies the following marginal condition,

$$U'(g^u) = (\gamma_c \delta + \gamma_1) M_g(g^u, 1). \tag{4.19}$$

Since $\delta \in (0,1)$, $\gamma_c, \gamma_1 > 0$ and $\gamma_c + \gamma_1 = 1$, we have

$$\delta M_g(g^u, 1) < U'(g^u) < M_g(g^u, 1).$$

Notice that

$$U'(g_c) = \delta M_g(g_c, 1), \quad U'(g_1) = M_g(g_1, 1),$$

so $g^u \neq g_c, g_1$, simply as $U'' < 0$ and $M_{gg} > 0$ at all $g > 0$.

Since a utilitarian bargaining solution is Pareto efficient, by Lemma 4.3, g^u must be a convex combination of g_1 and g_c. Thus there exists a unique constant $\varepsilon \in (0,1)$ such that

$$g^u = (1-\varepsilon)g_1 + \varepsilon g_c,$$

which yields $\mathrm{d}g^u/\mathrm{d}\varepsilon = g_c - g_1 > 0$. By (4.19), it is not hard to see that

$$\frac{\mathrm{d}g^u}{\mathrm{d}(\gamma_c\delta + \gamma_1)} = \frac{M_g(g^u, 1)}{U''(g^u) - (\gamma_c\delta + \gamma_1)M_{gg}(g^u, 1)} < 0,$$

which then completes the proof. □

Let us return back to the two-government negotiation game, as if Proposition 4.4 had not been developed to characterize its utilitarian solution. Alternatively, we shall propose an intuitive method to find such a solution.

Note that at any $g \in [g_1, g_c]$, the marginal welfare states of c and r_1 are

$$U'(g) - \delta M_g(g, 1) \geq 0, \quad U'(g) - M_g(g, 1) \leq 0,$$

whose difference is equal to $(1-\delta)M_g(g, 1)$. Any collective strategy g will divide that difference into two parts to be shared by c and r_1 separately. Define a "parameter" function

$$\beta : [g_1, g_c] \to [0, 1],$$

such that at any g the ratio of the part shared by c to the remaining part shared by r_1 is exactly $\beta(g)$ to $1 - \beta(g)$. By (4.18), g^u is the equilibrium if and only if $\beta(g^u) = \gamma_1$ and $1 - \beta(g^u) = \gamma_c$. In other words,

$$U'(g^u) - \delta M_g(g^u, 1) = \gamma_1(1-\delta)M_g(g^u, 1),$$
$$U'(g^u) - M_g(g^u, 1) = -\gamma_c(1-\delta)M_g(g^u, 1),$$

in fact, both of which are equivalent to the marginal condition (4.19). The following Figure 4.4 gives a graphic illustration of this method.

With regards some negotiation game emerging on the H-form government line N with $n \geq 3$ governments, thought Lemma 4.3 is a quite useful statement, finding the critical r_i should be very likely hard. However, if we consider a general n-government negotiation game simply as a "linear" extension of a 2-government negotiation game, we could still have a few interesting results which might be unexpected at this stage.

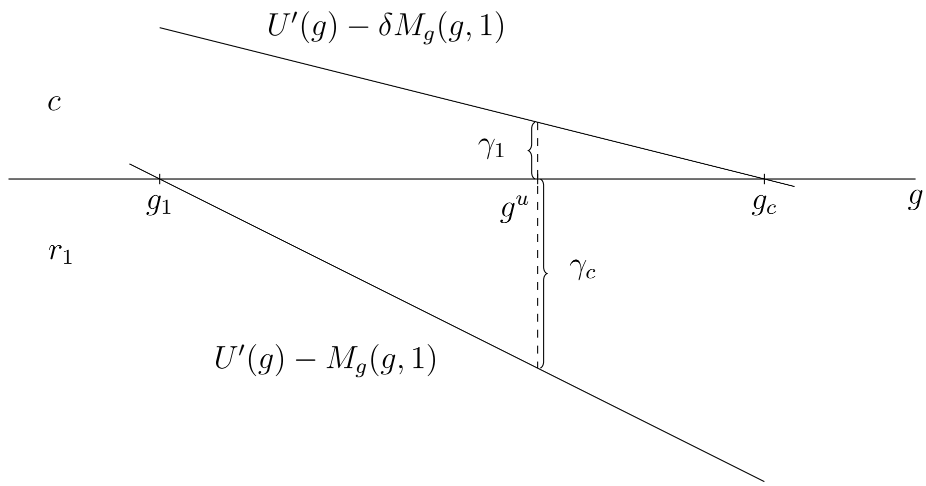


FIGURE 4.4.

Directly, we shall say that the utilitarian bargaining solution to the n-government negotiation game on $N = \{c, r_1, r_2, \ldots, r_{n-1}\}$ is $g^u \in [g_{n-1}, g_c]$, which maximizes

$$\gamma_c \Pi_c(g, R) + \sum_{i=1}^{n-1} \gamma_i \Pi_i(g, R), \tag{4.20}$$

where $\gamma_i > 0$ for all i, and $\gamma_c + \sum_{i=1}^{n-1} \gamma_i = 1$. Equivalently, g^u should then maximize

$$U(g) - h(\delta, \gamma) M(g, 1), \tag{4.21}$$

where $h(\delta, \gamma)$ is a polynomial in δ of degree $n-1$, that is,

$$h(\delta, \gamma) = \gamma_c \delta^{n-1} + \gamma_1 \delta^{n-2} + \cdots + \gamma_{n-2}\delta + \gamma_{n-1}. \tag{4.22}$$

Let's define the characteristic of $h(\delta, \gamma)$ as

$$\chi = \log_\delta h(\delta, \gamma). \tag{4.23}$$

Note that $\delta^{n-1} < h(\delta, \gamma) < 1$ for all authority allocation rule γ, thus $0 < \chi < n-1$, and hence $0 \le \lfloor \chi \rfloor \le n-2$ as $\lfloor \chi \rfloor = \max\{z \in \mathbb{Z} : z \le \chi\}$. Next, let's define the conjugate value of $\lfloor \chi \rfloor$ (with respect to $n-2$) as

$$\chi^* = n - 2 - \lfloor \chi \rfloor. \tag{4.24}$$

It is clear that $0 \le \chi^* \le n-2$.

PROPOSITION 4.5. *g^u is a convex combination of g_{χ^*+1} and g_{χ^*}.*

PROOF. Since $\lfloor \chi \rfloor \leq \chi < \lfloor \chi \rfloor + 1$, we have $n-2-\chi^* \leq \chi < n-1-\chi^*$. By the definition of χ, we have $h(\delta,\gamma) = \delta^{\chi}$, thus

$$\delta^{n-1-\chi^*} < h(\delta,\gamma) \leq \delta^{n-2-\chi^*}.$$

By (4.17), there are two consecutive governments r_{χ^*} and r_{χ^*+1}, whose status quo provision levels are g_{χ^*} and g_{χ^*+1} such that

$$U'(g_{\chi^*}) = \delta^{n-1-\chi^*} M_g(g_{\chi^*}, 1), \quad U'(g_{\chi^*+1}) = \delta^{n-2-\chi^*} M_g(g_{\chi^*+1}, 1).$$

In the meantime, by (4.21), g^u should satisfy

$$U'(g^u) = h(\delta,\gamma) M_g(g^u, 1). \tag{4.25}$$

Let $Z(g) = U'(g)/M_g(g,1)$, then $Z(g)$ is a decreasing function of g, as $U''(g) < 0$ and $M_{gg}(g,1) > 0$ at all $g > 0$. Note that

$$Z(g^u) = h(\delta,\gamma); \quad Z(g_{\chi^*}) = \delta^{n-1-\chi^*}, \quad Z(g_{\chi^*+1}) = \delta^{n-2-\chi^*},$$

which yield

$$Z(g_{\chi^*}) < Z(g^u) \leq Z(g_{\chi^*+1}),$$

thus $g_{\chi^*+1} \leq g^u < g_{\chi^*}$, which completes the proof. □

By Proposition 4.5, it could be concluded that the 2-government negotiation game on $\{r_{\chi^*}, r_{\chi^*+1}\}$ stands out to be a reasonable approximation for the original n-government negotiation game on N. Thus χ^* appears to be a meaningful indicator for the focal hierarchical level of a long H-form government line, and these two governments on the hierarchical levels χ^* and χ^*+1 evidently have the greatest (enforceable) authorities in the political system.

EXAMPLE 4.6. Assume $u_j(g) = \log(g+1)$ and $\alpha_j = 1/|J|$ for all $p_j \in P^r$, where $|J| = |P^r|$, thus

$$U(g) = \log(g+1), \quad U'(g) = 1/(g+1).$$

Assume $M(g,\lambda) = (\lambda g)^2/2$, then

$$M(g,1) = g^2/2, \quad M_g(g,1) = g.$$

In the H-form government line $\{c, r_1\}$, set $\gamma_c = 0.4$, $\gamma_1 = 0.6$, and $\delta = 0.5$. By solving the following two equations,

$$1/(g+1) = 0.5g, \quad 1/(g+1) = g,$$

we obtain $g_c = 1$ and $g_1 = (\sqrt{5}-1)/2 \approx 0.618$. The utilitarian solution to the negotiation game on $\{c, r_1\}$ is $g^u > 0$ such that

$$1/(g^u+1) = 0.8g^u,$$

which yields $g^u = (\sqrt{6}-1)/2 \approx 0.725$.

Let $g^u = (1-\varepsilon)g_1 + \varepsilon g_c$, then

$$\varepsilon = \frac{g^u - g_1}{g_c - g_1} = \frac{\sqrt{6}-\sqrt{5}}{3-\sqrt{5}} \approx 0.28.$$

When $\gamma_c\delta + \gamma_1$ decreases, or equivalently $\gamma_c(1-\delta)$ increases, ε and g^u will both increase, which means g^u gets closer to g_c.

Example 4.7. In the H-form government line $\{c, r_1, r_2, r_3, r_4\}$, set $\gamma_i = 0.2$ for all i, and $\delta = 0.8$. We have $h(\delta,\gamma) = 1 - 0.8^5$, thus

$$\chi = \log_{0.8}(1-0.8^5) \approx 1.78,$$

and hence $\chi^* = 3 - \lfloor 1.78 \rfloor = 2$. By Proposition 4.5, we see that the utilitarian solution of the 5-government negotiation game which is denoted by g^u, must be a convex combination of g_3 and g_2. In effect, it does happen that $g_3 < g^u < g_2$, which thus suggests r_2 and r_3 have the most powerful authorities in the political system.

Suppose again $U'(g) = 1/(g+1)$ and $M_g(g,1) = g$, then we should have

$$1/(g_3+1) = 0.8g_3, \quad 1/(g_2+1) = 0.8^2 g_2,$$

which yield $g_3 = (\sqrt{6}-1)/2 \approx 0.725$ and $g_2 = (\sqrt{29}-2)/4 \approx 0.846$. Meanwhile, g^u satisfies

$$1/(g^u+1) = (1-0.8^5)g^u,$$

which shows $g^u = (\sqrt{14601/2101} - 1)/2 \approx 0.818$. It seems that r_2 plays a more powerful role than r_3 in the negotiation game, as we might observe that g^u is much closer to g_2 than to g_3.

5. Coordinated Interaction

5.1. M-form Game. Suppose now (N, f) is an M-form government star, then there are only two hierarchical levels, the level 0 and level 1, and $N = N_0 \cup N_1$, where $N_0 = \{c\}$ and $N_1 = N \setminus \{c\}$. Notice that $D_0 = N_1$ and $D_i = \emptyset$ for all $r_i \in N_1$, thus $N_b = N_1 = N \setminus \{c\}$. Similar to the case of H-form government line, we shall also use N to directly denote an M-form (N, f) by virtue of its simple structure. Note that, related to an M-form government star, the region R is partitioned into $n-1$ pairwise disjoint districts, *i.e.*, R_i for all $r_i \in N_1$, each of which holds a population $P(R_i) = P_i$, and is administered by a bottom government $r_i \in N_1$ which is also local here.

In an H-form government line, all the governments have a common committee of representatives, *viz.*, $P_i^r = P^r$ for all $r_i \in N$, because $P_i = P$ for all $r_i \in N$. But for an M-form government star, P_i^r is not necessarily identical to P^r, simply as now P_i is only a proper subgroup of P. Recall that $P^r = \{p_j \in P_j : j \in J\}$ and $|P^r| = |J|$. Since $P_i^r \neq \emptyset$ and highly likely $P_i^r \subseteq P^r$ for all $r_i \in N_1$, any P_i^r could be one of the nonempty subgroups of P^r, and thus it has $2^{|J|}-1$ possible combinations. The number of composite arrangements of those $n-1$ committees for N_1 is thus equal to $(2^{|J|}-1)^{n-1}$. In case $|J|$ is designed to be equal to $|N_1|$ in the political system, that is, $|J| = n-1$, then that value would be

$$(2^{n-1}-1)^{n-1} \approx 2^{(n-1)^2}.$$

Evidently, it grows exponentially fast with n, and it will become rather great for even $n > 5$.

In spite of the existence of enormous possible committee arrangements due to various population distributions over R, two crucial and yet quite specific cases could be imaged:

(i) any P_j for $j \in J$ is naïvely distributed over R, so that $P_i^r = P^r$ for all $r_i \in N$;

(ii) any P_j for $j \in J$ is distributed over one district R_i, and any R_i for $r_i \in N_1$ holds only one population P_j, so that $P^r = \{p_1, p_2, \ldots, p_{n-1}\}$ and $P_i^r = \{p_i\}$ for all $r_i \in N_1$.

In case (i), $|P_i^r| = |P^r| = |J|$ for all $r_i \in N_1$; in case (ii), $|J|$ is designed to be equal to $n-1$, so that $|P^r| = n-1$, and $|P_i^r| = 1$ for all $r_i \in N_1$.

Suppose a single decision-making process in N is again designed to decide the local public good provision for one certain bottom government, and moreover, distinct decision-making processes do not have any overlapping procedure, then there should be $n-1$ distinct collective decision-making processes in the political system for $|N_b| = n-1$. With regards a specific decision-making process that aims to decide actions for some $r_i \in N_b$, there is a group of participating governments, *i.e.*, $L_i = \{c, r_i\}$, and an authority distribution over L_i. It is noteworthy that a local government in N_1 is merely involved in one decision-making process, and the central government c is involved in all the $n-1$ decision-making processes. It then turns to be clear that there are in total $n-1$ distinct authority distributions over $L_i = \{c, r_i\}$ for all $r_i \in N_1$, which on the whole constitute the authority rule γ in the political system.

Let the authority allocated to $r_i \in N_1$ be γ_i, then the central government c should be allocated all the remaining authorities which can be ordered as an $(n-1)$-tuple,

$$\Gamma_c = (1-\gamma_1, 1-\gamma_2, \ldots, 1-\gamma_{n-1}),$$

where $1-\gamma_i$ is the authority allocated to c concerning the decision-making process for $r_i \in N_1$. Note that $\gamma_i \in [0,1]$ for all $r_i \in N_1$, and $\Gamma_c \in [0,1]^{n-1}$. If $\Gamma_c = (1, 1, \ldots, 1)$, or $\gamma_i = 0$ for all $r_i \in N_1$, N will have a centralized governance structure; if $\Gamma_c = (0, 0, \ldots, 0)$, or $\gamma_i = 1$ for all $r_i \in N_1$, N will have a pure decentralized governance structure.

An isolated collective decision-making process for some $r_i \in N_1$ stands out to be a negotiation game on the H-form government line $L_i = \{c, r_i\}$, and interactions in

N_1, in view of that it can be separated from $N_0 = \{c\}$, appear to be a noncooperative game, in which local public good provision is dominated by free-rider inertia. In a quite practical sense, the M-form game N can then be thought of as a collection of n normal games, *viz.*, $n-1$ two-government negotiation games on $\{c, r_i\}$ for all $r_i \in N_1$, and one noncooperative game on N_1. Note that the appearances of these $n-1$ negotiation games and the noncooperative game on N_1 is fairly determined by Γ_c. For example, if $\Gamma_c = (0, 0, \ldots, 0)$, there will be no negotiation games, but only the noncooperative game on N_1; if $\Gamma_c = (1, 1, \ldots, 1)$, all the n normal games will disappear, and hence the M-form game will degenerate into a decision situation for the central government c.

What's more, two main (and nearly complete) classes of M-form game can be discerned by means of Γ_c, so that M-form games in either class would have a similar analytical structure. They are

(i) $\Gamma_c \in \{0, 1\}^{n-1}$, so that each decision-making process allocates its absolute authority to one single government, either central or local, and we call it the *degenerate* class;
(ii) $\Gamma_c \in (0, 1)^{n-1}$, so that each decision-making process allocates its nonzero authorities to the central government and a local government, and we call it the *analytic*[3] class.

Let (g_i, S_i) denote the local provision state decided by the negotiation game on $\{c, r_i\}$ for all $r_i \in N_1$. Let (g_c, S_c) denote the global provision state identifiable to the central government c. We shall assume g_c is equal to the weighted average of these $n-1$ local provision levels, where the weighting factor of g_i is set as its relative provision capacity $\mu_i = \lambda(S_i)/\sum_{r_j \in N_1} \lambda(S_j)$, thus

$$g_c = \sum_{r_i \in N_1} \mu_i g_i. \tag{4.26}$$

It is direct that $S_c = \bigcup_{r_i \in N_1} S_i$, as there ought to be no difficulty in provision district identifications.

Depending on all (variable) local provision state (g_i, S_i) for $r_i \in N_1$, the welfare function of the local government $r_i \in N_1$ takes the form $V_i(g_i, S_c \cap R_i)$, which will be $U_i(g_i)$ if $S_c \cap R_i = R_i$. And the welfare function of the central government c is $V(g_c, S_c)$, which will be $U(g_c)$ if $S_c = R$. In particular, we assume $V(g_c, S_c)$ is linear in g_c, and the welfare state $V(g_i, S_c)$ of c is equivalent to the welfare state

[3]Here, we use the term "analytic class" rather than directly "nondegenerate class" as is the negation of degenerate class, because this class only contains some nondegenerate M-form games, though most of them.

$V_i(g_i, S_c \cap R_i)$ of r_i, it therefore follows that

$$V(g_c, S_c) = \sum_{r_i \in N_1} \mu_i V(g_i, S_c) = \sum_{r_i \in N_1} \mu_i V_i(g_i, S_c \cap R_i). \tag{4.27}$$

If $S_c = R$, we have $V(g_c, S_c) = U(g_c)$ and $V_i(g_i, S_c \cap R_i) = U_i(g_i)$, and thus

$$U(g_c) = \sum_{r_i \in N_1} \mu_i U_i(g_i). \tag{4.28}$$

It might be emphasized that the relevant provision district to the welfare function of r_i is in effect $S_c \cap R_i$ rather than S_i. Clearly, $S_i \subseteq S_c \cap R_i$. In case $S_i \subset S_c \cap R_i$, there must be some other local governments with provisions to cover parts of the district $R_i \setminus S_i$, and the population in R_i who can benefit from the local public good is $P(S_c \cap R_i)$ which satisfies

$$P(S_i) \subset P(S_c \cap R_i) \subseteq P_i.$$

Although the local public good provisions in $R_i \setminus S_i$ could then be very likely varied and different from g_i, the identifiable provision level entering the welfare function of r_i is simply its own provision level g_i, because the utility functions of its representatives should basically reflect the provision level of r_i.

The objective function of the local government $r_i \in N_1$ can now be expressed as

$$\Pi_i(g_i, S_c) = V_i(g_i, S_c \cap R_i) - M(g_i, \lambda(S_i)). \tag{4.29}$$

The objective function of the central government c is again assumed to be a convex combination of its own welfare function and the objective functions of its directly subordinate governments in $D_0 = N_1$. It can thus be written as

$$\Pi_c(g_c, S_c) = (1 - \delta)V(g_c, S_c) + \delta \sum_{r_i \in N_1} \mu_i \Pi_i(g_i, S_c), \tag{4.30}$$

where $\delta \in (0, 1)$ is a regular political discounting. By the condition (4.27), we have

$$\sum_{r_i \in N_1} \mu_i \Pi_i(g_i, S_c) = V(g_c, S_c) - \sum_{r_i \in N_1} \mu_i M(g_i, \lambda(S_i)),$$

and therefore we reach the function

$$\Pi_c(g_c, S_c) = V(g_c, S_c) - \delta \sum_{r_i \in N_1} \mu_i M(g_i, \lambda(S_i)). \tag{4.31}$$

Note that the condition (4.27) is sufficient for c having an objective function like (4.31), and yet it is surely not necessary. In the remaining discussion of this section, we shall directly use (4.31), but not require that V be a convex combination of all V_i for $r_i \in N_1$. To put it clearer, (4.31) will be supposed to be independent of (4.27), so that V and V_i's could have the potential to take forms freely.

5.2. Degenerate Game. Suppose $\Gamma_c \in \{0,1\}^{n-1}$, then the M-form game on N will be degenerate. In this subsection, we shall develop three formal assertions on its solutions.

PROPOSITION 4.6. *If* $\Gamma_c = (1,1,\ldots,1)$ *and* $M_{g\lambda} = 0$*, the solution of the M-form game will be the provision decision* (g_c^c, R) *such that* $U'(g_c^c) = \delta M_g(g_c^c, 1)$.

PROOF. If $\Gamma_c = (1,1,\ldots,1)$, the M-form game will degenerate into a decision situation for the central government c, and its solution should then maximize the objective function of c. Let (g_i^c, S_i^c) denote the provision state for $r_i \in N_1$ decided by c, then $g_c^c = \sum_{r_i\in N_1} \mu_i g_i^c$. Since $S_i^c = R_i$ for all $r_i \in N_1$ (*cf.*, Proposition 4.1), $S_c^c = \bigcup_{r_i\in N_1} S_i^c = R$. The objective function of c defined by (4.31) will then be rewritten as a real function in g_c,

$$\Lambda_c(g_c) = \Pi_c(g_c, R) = U(g_c) - \delta \sum_{r_i\in N_1} \mu_i M(g_i, \lambda(R_i)), \tag{4.31$'$}$$

where $g_c = \sum_{r_i\in N_1} \mu_i g_i$.

It is clear that all g_i^c for $r_i \in N_1$ should satisfy

$$\mathrm{d}\Lambda_c(g_c^c)/\mathrm{d}g_i = \mu_i U'(g_c^c) - \delta\mu_i M_g(g_i^c, \lambda(R_i)) = 0,$$

where $\mu_i > 0$ as $\lambda(R_i) > 0$, so it follows that

$$U'(g_c^c) = \delta M_g(g_i^c, \lambda(R_i)).$$

Since $M_{g\lambda} = 0$, it clearly happens that $M_g(g_i^c, \lambda(R_i)) = M_g(g_i^c, 1)$, and thus we have the following marginal condition for all r_i,

$$U'(g_c^c) = \delta M_g(g_i^c, 1). \tag{4.32}$$

Recall that $M_{gg} > 0$ and $U'' < 0$ at all positive provision level, then g_i^c is unique and also constant for all $r_i \in N_1$. So $g_i^c = g_c^c$ for all $r_i \in N_1$, and hence the provision state for all $r_i \in N_1$ is (g_c^c, R_i), which means the provision decision of c is (g_c^c, R) such that $U'(g_c^c) = \delta M_g(g_c^c, 1)$. $\square$

If $M_{g\lambda} \neq 0$, it still be true that $S_c^c = \bigcup_{r_i\in N_1} S_i^c = R$, however, it now could happen that $S_i^c \neq R_i$ for some $r_i \in N_1$. The provision state (g_i^c, S_i^c) for $r_i \in N_1$ satisfies

$$U'(g_c^c) = \delta M_g(g_i^c, \lambda(S_i^c)). \tag{4.33}$$

Two artificial decision procedures without great computing complexity could be imagined. First, suppose c decides $g_i^c = g_c^c$ for all $r_i \in N_1$, then $\lambda(S_i^c)$ must be constant, which implies these $n-1$ provision districts should be fairly divided[4].

[4]Here, the fairness means that each S_i^c has a same provision capacity, say λ^c. Since the provision capacity is superadditive, we must have $\lambda^c \leq 1/(n-1)$.

Next, suppose c decides $S_i^c = R_i$ for all $r_i \in N_1$, then for all distinct r_i, r_j, the following relation should hold,

$$(g_i^c - g_j^c)\big(\lambda(R_i) - \lambda(R_j)\big)M_{g\lambda} \leq 0. \tag{4.34}$$

In other words, if $M_{g\lambda} < 0$, then $g_i^c \geq g_j^c$ for $\lambda(R_i) \geq \lambda(R_j)$, and if $M_{g\lambda} > 0$, then $g_i^c \geq g_j^c$ for $\lambda(R_i) \leq \lambda(R_j)$.

PROPOSITION 4.7. *If $\Gamma_c = (0, 0, \ldots, 0)$, the solution of the M-form game will be the strategic equilibrium (g_i^r, R_i) for all $r_i \in N_1$, such that $U_i'(g_i^r) = M_g(g_i^r, \lambda(R_i))$.*

PROOF. If $\Gamma_c = (0, 0, \ldots, 0)$, the M-form game will plainly degenerate into a noncooperative game on N_1, and its solution should then be the strategic equilibrium of that noncooperative game, which can be denoted by the collection

$$\big\{(g_i^r, S_i^r) : r_i \in N_1\big\}.$$

Note that $S_i^r \cap S_j^r = \emptyset$ for all distinct r_i, r_j, otherwise some local government would have incentives to decrease its provision capacity. Meanwhile, $S_i^r \supseteq R_i$ for all $r_i \in N_1$ (*cf.*, Proposition 4.1), so $S_i^r = R_i$ for all $r_i \in N_1$, and hence $\bigcup_{r_i \in N_1} S_i^r = R$. g_i^r for all $r_i \in N_1$ will then be the optimal provision level in an isolated decision situation for r_i, to wit g_i^r maximizes $\Pi_i(g_i, R)$. Recall that

$$\Pi_i(g_i, R) = U_i(g_i) - M(g_i, \lambda(R_i)),$$

so g_i^r is the unique provision level satisfying

$$U_i'(g_i^r) = M_g(g_i^r, \lambda(R_i)), \tag{4.35}$$

which completes the proof. □

Define

$$N_1^c = \{r_i \in N_1 : \gamma_i = 0\}, \quad N_1^r = \{r_i \in N_1 : \gamma_i = 1\}.$$

It should be evident that N_1^c and N_1^r form a bipartition of N_1. If $\Gamma_c = (1, 1, \ldots, 1)$, then $\gamma_i = 0$ for all $r_i \in N_1$, and hence $N_1^r = \emptyset$. If $\Gamma_c = (0, 0, \ldots, 0)$, we will have $N_1^c = \emptyset$. We shall next consider such a general tuple Γ_c that N_1^c and N_1^r could be both nonempty.

COROLLARY 4.8. *If $\Gamma_c \in \{0, 1\}^{n-1}$ and $M_{g\lambda} = 0$, the solution of the M-form game will be (g_i^d, R_i) for all $r_i \in N_1$, such that $U_i'(g_i^d) = M_g(g_i^d, 1)$ for all $r_i \in N_1^r$, and $U'(g_c^d) = \delta M_g(g_i^d, 1)$ for all $r_i \in N_1^c$.*

PROOF. If $\Gamma_c = (1, 1, \ldots, 1)$, then $N_1^c = N_1$, and Proposition 4.6 shows that $g_i^d = g_c^d = g_c^c$ such that $U'(g_c^c) = \delta M_g(g_c^c, 1)$ for all $r_i \in N_1$. If $\Gamma_c = (0, 0, \ldots, 0)$, then $N_1^r = N_1$, and Proposition 4.7 with the condition $M_{g\lambda} = 0$ shows that $g_i^d = g_i^r$ such that $U_i'(g_i^r) = M_g(g_i^r, 1)$ for all $r_i \in N_1$.

With regards all the other $\Gamma_c \in \{0,1\}^{n-1}$, we have $N_1^c \neq \emptyset$ and $N_1^r \neq \emptyset$. The M-form game will degenerate into a noncooperative game on N_1^r, and a decision situation concerning N_1^c for c. By Proposition 4.7 with the condition $M_{g\lambda} = 0$, the provision decision concerning N_1^r will be (g_i^d, R_i) such that $U_i'(g_i^d) = M_g(g_i^d, 1)$ for all $r_i \in N_1^r$. By Proposition 4.6, the provision decision concerning N_1^c will be (g_i^d, R_i) such that $U'(g_c^d) = \delta M_g(g_i^d, 1)$ for all $r_i \in N_1^c$, where $g_c^d = \sum_{r_i \in N_1} \mu_i g_i^d$. □

Let g^c and g^r denote the average provision levels in N_1^c and N_1^r, respectively. Note that $g_i^d = g^c$ for all $r_i \in N_1^c$, and $g^r = \sum_{r_i \in N_1^r} \mu_i g_i^d / \sum_{r_i \in N_1^r} \mu_i$, then it follows that

$$g_c^d = \sum_{r_i \in N_1} \mu_i g_i^d = \sum_{r_i \in N_1^c} \mu_i \cdot g^c + \sum_{r_i \in N_1^r} \mu_i \cdot g^r, \tag{4.36}$$

which means that g_c^d is a convex combination of g^c and g^r.

It might be noticed that g_c^c is also a convex combination of g^c and g^r. To show this fact, suppose $g_c^d \leq g_c^c$, then $U'(g_c^c) \leq U'(g_c^d)$, and thus the following relation holds,

$$\delta M_g(g_c^c, 1) = U'(g_c^c) \leq U'(g_c^d) = \delta M_g(g_i^d, 1) = \delta M_g(g^c, 1),$$

which clearly implies $g_c^c \leq g^c$. It then follows that $g_c^d \leq g^c$, and hence $g_c^d \geq g^r$ according to (4.36). In consequence, it appears to us that

$$g^r \leq g_c^d \leq g_c^c \leq g^c.$$

On the other hand, if $g_c^d \geq g_c^c$, we will have $g_c^c \geq g^c$, and hence

$$g^r \geq g_c^d \geq g_c^c \geq g^c.$$

Observe that the order on the 3-tuple (g^r, g_c^d, g^c) has a bifurcation at g_c^c, *viz.*, it would change from "$\leq$" to its inverse "$\geq$" when g_c^d moves from the left of g_c^c to its right in a real line.

5.3. Analytic Game. If $\Gamma_c \in (0,1)^{n-1}$, the M-form games on N will be clearly nondegenerate and of the analytic class. We shall assume $M_{g\lambda} = 0$ throughout this subsection, except the examples thereafter our theoretical investigations.

As we have so far developed, an analytic M-form game can be decomposed of $n-1$ negotiation games on $L_i = \{c, r_i\}$ for all $r_i \in N_1$ and one noncooperative game on N_1. When $M_{g\lambda} = 0$ as is assumed above, the decisions on provision capacity and provision level are in effect separable, as M_g is independent of λ and also M_λ is independent of g. Thus the equilibrium arrangement of provision capacities in the noncooperative game on N_1 should comply with the administration regions, that's to say, a local government $r_i \in N_1$ will set its optimal provision district exactly as its administration region R_i. In this respect, the arrangement of provision levels in the noncooperative game on N_1 will be reduced to $n-1$ decision situations without any strategic interaction, therefore it should be sufficient to only consider the remaining $n-1$ "parallel" two-government negotiation games. As usual, the utilitarian solution concept will be adopted for any negotiation game, thus the solution to the analytic M-form game will have utilitarian meanings per se, and for this reason, we shall use the term utilitarian solution when referring to it.

Let the collection of these $n-1$ provision states $\{(g_i^u, S_i^u) : r_i \in N_1\}$ denote the utilitarian solution to the analytic M-form game, then $g_c^u = \sum_{r_i \in N_1} \mu_i g_i^u$ and $S_c^u = \bigcup_{r_i \in N_1} S_i^u$. It is direct by the above treatment of the M-form game, that $S_i^u = R_i$ for all $r_i \in N_1$, which then yields $S_c^u = R$. Recall that the following conditions for g_c^c and g_i^r where $r_i \in N_1$ always holds:

$$U'(g_c^c) = \delta M_g(g_c^c, 1), \quad U_i'(g_i^r) = M_g(g_i^r, 1).$$

Proposition 4.9. *Suppose $U(g)$ and $\delta U_i(g)$ have no crossing point on the plane, then*

$$\max\{g_i^u, g_c^u\} \geq \min\{g_i^r, g_c^c\}, \quad \min\{g_i^u, g_c^u\} \leq \max\{g_i^r, g_c^c\}.$$

Proof. The coordinated decision problem delivered by the negotiation game on $\{c, r_i\}$ is

$$\max_{g_i > 0} \gamma_i \Pi_i(g_i, R) + (1-\gamma_i)\Pi_c(g_c, R),$$

or equivalently,

$$\max_{g_i > 0} \gamma_i \big(U_i(g_i) - M(g_i, \lambda_i)\big) + (1-\gamma_i)\big(U(g_c) - \delta \sum_{r_j \in N_1} \mu_j M(g_j, \lambda_j)\big),$$

where $\lambda_i = \lambda(R_i)$ for all r_i. g_i^u should then satisfy

$$\gamma_i\big(U_i'(g_i^u) - M_g(g_i^u, \lambda_i)\big) + \mu_i(1-\gamma_i)\big(U'(g_c^u) - \delta M_g(g_i^u, \lambda_i)\big) = 0,$$

which directly yields the following marginal condition,

$$\gamma_i U_i'(g_i^u) + \mu_i(1-\gamma_i)U'(g_c^u) = \big(\gamma_i + \delta\mu_i(1-\gamma_i)\big)M_g(g_i^u, 1), \tag{4.37}$$

where $M_g(g_i^u, 1) = M_g(g_i^u, \lambda_i)$ as $M_{g\lambda} = 0$, and $\delta, \gamma_i, \mu_i(1-\gamma_i) > 0$.

First, suppose $g_i^u \leq g_c^u$, then

$$U'(g_i^u) \geq U'(g_c^u); \quad U_i'(g_c^u) \leq U_i'(g_i^u), \quad M_g(g_i^u, 1) \leq M_g(g_c^u, 1).$$

Thus there is a system of inequalities,

$$\gamma_i U_i'(g_i^u) + \mu_i(1-\gamma_i)U'(g_i^u) \geq \big(\gamma_i + \delta\mu_i(1-\gamma_i)\big)M_g(g_i^u, 1) \tag{4.38}$$

$$\gamma_i U_i'(g_c^u) + \mu_i(1-\gamma_i)U'(g_c^u) \leq \big(\gamma_i + \delta\mu_i(1-\gamma_i)\big)M_g(g_c^u, 1) \tag{4.39}$$

Since $U(g)$ and $\delta U_i(g)$ have no crossing point, either $U(g) \geq \delta U_i(g)$ or $U(g) \leq \delta U_i(g)$ for all $g \geq 0$. If $U(g) \geq \delta U_i(g)$, then $U'(g) \geq \delta U_i'(g)$ and thus $U_i'(g_c^c) \leq M_g(g_c^c, 1)$, but $U_i'(g_i^r) = M_g(g_i^r, 1)$. Notice that $Z_i(g) = U_i'(g)/M_g(g, 1)$ is decreasing with g, so $g_c^c \geq g_i^r$. By (4.38), $U'(g_i^u) \geq \delta M_g(g_i^u, 1)$, and hence $g_i^u \leq g_c^c$; by (4.39), $U_i'(g_c^u) \leq M_g(g_c^u, 1)$, and hence $g_c^u \geq g_i^r$. It then comes to us that

$$g_c^u \geq g_i^r, \quad g_i^u \leq g_c^c.$$

If $U(g) \leq \delta U_i(g)$, then $g_c^c \leq g_i^r$. By (4.38), $U_i'(g_i^u) \geq M_g(g_i^u, 1)$, and hence $g_i^u \leq g_i^r$; by (4.39), $U'(g_c^u) \leq \delta M_g(g_c^u, 1)$, and hence $g_c^u \geq g_c^c$. We thus have

$$g_c^u \geq g_c^c, \quad g_i^u \leq g_i^r.$$

Next, suppose $g_i^u \geq g_c^u$, then

$$U'(g_i^u) \leq U'(g_c^u); \quad U_i'(g_c^u) \geq U_i'(g_i^u), \quad M_g(g_i^u, 1) \geq M_g(g_c^u, 1).$$

Thus there is another system of inequalities,

$$\gamma_i U_i'(g_i^u) + \mu_i(1-\gamma_i)U'(g_i^u) \leq \big(\gamma_i + \delta\mu_i(1-\gamma_i)\big) M_g(g_i^u, 1) \tag{4.40}$$

$$\gamma_i U_i'(g_c^u) + \mu_i(1-\gamma_i)U'(g_c^u) \geq \big(\gamma_i + \delta\mu_i(1-\gamma_i)\big) M_g(g_c^u, 1) \tag{4.41}$$

If $U(g) \geq \delta U_i(g)$, we have $g_c^c \geq g_i^r$. By (4.40), $U_i'(g_i^u) \leq M_g(g_i^u, 1)$, and hence $g_i^u \geq g_i^r$; by (4.41), $U'(g_c^u) \geq \delta M_g(g_c^u, 1)$, and hence $g_c^u \leq g_c^c$. So

$$g_i^u \geq g_i^r, \quad g_c^u \leq g_c^c.$$

If $U(g) \leq \delta U_i(g)$, we have $g_c^c \leq g_i^r$. By (4.40), $U'(g_i^u) \leq \delta M_g(g_i^u, 1)$, and hence $g_i^u \geq g_c^c$; by (4.41), $U_i'(g_c^u) \geq M_g(g_c^u, 1)$, and hence $g_c^u \leq g_i^r$. Thus

$$g_i^u \geq g_c^c, \quad g_c^u \leq g_i^r.$$

It is clear that the above classification about the pairs (g_i^u, g_c^u) and (g_i^r, g_c^c) is complete, thus the assertion is proven. $\square$

If $\max\{g_i^u, g_c^u\} \geq \min\{g_i^r, g_c^c\}$, it must be true that $\max\{g_i^r, g_c^c\} \geq \min\{g_i^u, g_c^u\}$, and vice versa. It hence shows that the real interval bounded by g_i^u and g_c^u and the real interval bounded by g_i^r and g_c^c always have a nonempty intersection. To have a graphical impression, we can consider one of the four possibilities, $g_i^u \geq g_c^u$ and $g_i^r \geq g_c^c$. In the following diagram (see Figure 4.5), we set the interval $[g_c^c, g_i^r]$ fixed on the real line, and show three generic cases of $[g_c^u, g_i^u]$ which have different relative positions to $[g_c^c, g_i^r]$.

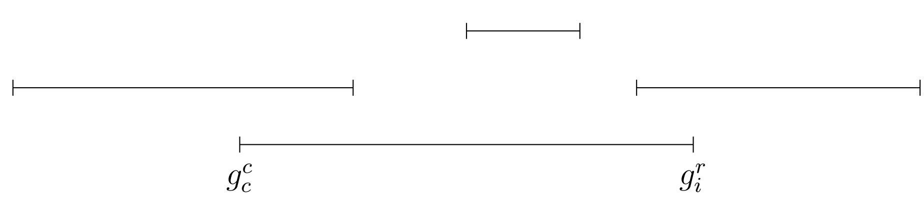


Figure 4.5.

Moreover, if there exists some $r_i \in N_1$ such that $g_i^u = g_c^u$, then $\max\{g_i^u, g_c^u\} = \min\{g_i^u, g_c^u\}$, and thus g_c^u (and also that g_i^u) must be located between g_i^r and g_c^c, that is,

$$\min\{g_i^r, g_c^c\} \leq g_c^u \leq \max\{g_i^r, g_c^c\}. \tag{4.42}$$

However, in many cases we may find that $g_i^u \neq g_c^u$ for all $r_i \in N_1$. Since $g_c^u = \sum_{r_i \in N_1} \mu_i g_i^u$, there must exist at least two distinct local governments r_i and r_j, such that $g_i^u < g_c^u < g_j^u$. It thus appears that

$$g_c^u = \min\{g_j^u, g_c^u\} \leq \max\{g_j^r, g_c^c\},$$

and meanwhile,

$$g_c^u = \max\{g_i^u, g_c^u\} \geq \min\{g_i^r, g_c^c\}.$$

Without loss of generality, we assume the sequence $(g_i^r,\ r_i \in N_1)$ could be ordered as

$$g_1^r \geq g_2^r \geq \cdots \geq g_{n-1}^r > 0.$$

It then follows that

$$\max\{g_j^r, g_c^c\} \leq \max\{g_1^r, g_c^c\}, \quad \min\{g_i^r, g_c^c\} \geq \min\{g_{n-1}^r, g_c^c\},$$

and therefore the following assertion has been proven:

Corollary 4.10.

$$\min\{g_{n-1}^r, g_c^c\} \leq g_c^u \leq \max\{g_1^r, g_c^c\} \tag{4.43}$$

In a quite ethical sense, $U(g)$ should be neither greater nor smaller than all $\delta U_i(g)$ for $r_i \in N_1$, *viz.*, the central government should neither overestimate nor underestimate all local government's politically discounted welfare state. Due to that presumption, there must be two local governments in N_1, say again r_i and r_j, such that $\delta U_i(g) \leq U(g) \leq \delta U_j(g)$ for all $g \geq 0$. In other words, it should always come out that $g_i^r \leq g_c^c \leq g_j^r$ (*cf.*, the proof of Proposition 4.9), and furthermore, $g_{n-1}^r \leq g_c^c \leq g_1^r$, which thus gives (4.43) a finer expression,

$$g_{n-1}^r \leq g_c^u \leq g_1^r. \tag{4.43$'$}$$

Surprising as it might seem, the global provision level g_c^u in the analytic M-form game is always confined by the lower and upper bounds of local provision levels decided through the decentralized provision scheme. As for a specific Γ_c, there must be some $1 \leq i \leq n-2$ such that $g_{i+1}^r \leq g_c^u \leq g_i^r$, which thus means, as we have long been familiar, that g_c^u is a convex combination of g_{i+1}^r and g_i^r.

Corollary 4.11. *If* $\min\{g_c^u, g_c^c\} \leq g_i^r \leq \max\{g_c^u, g_c^c\}$, *then*

$$(g_i^u - g_i^r)(g_c^u - g_c^c) \leq 0. \tag{4.44}$$

PROOF. Suppose $g_c^u \leq g_c^c$, then $g_c^u \leq g_i^r \leq g_c^c$. By Proposition 4.9, we have

$$\max\{g_i^u, g_c^u\} \geq \min\{g_i^r, g_c^c\} = g_i^r \geq g_c^u,$$

thus $g_i^u \geq g_c^u$, and hence $\max\{g_i^u, g_c^u\} = g_i^u$ which yields $g_i^u \geq g_i^r$.

On the other hand, suppose $g_c^u \geq g_c^c$, then $g_c^c \leq g_i^r \leq g_c^u$. By Proposition 4.9, we have

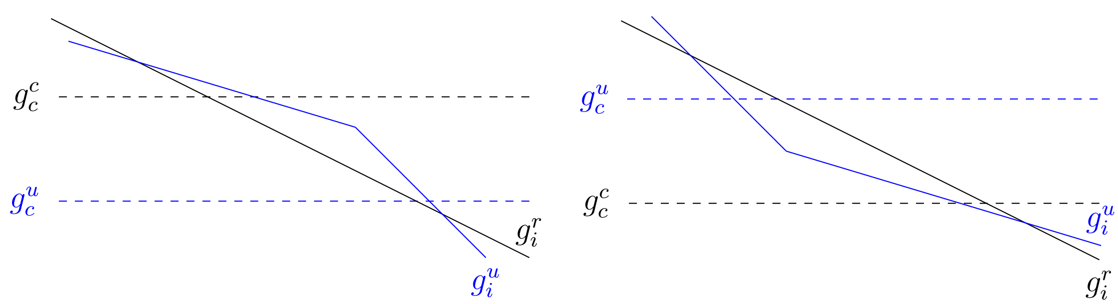
$$\min\{g_i^u, g_c^u\} \leq \max\{g_i^r, g_c^c\} = g_i^r \leq g_c^u,$$

thus $g_i^u \leq g_c^u$, and hence $\min\{g_i^u, g_c^u\} = g_i^u$ which yields $g_i^u \leq g_i^r$. $\square$

Suppose there are nearly infinite local governments in N_1, then the index set I should be dense almost as $\mathbb{Z}^+$. Let's use a bounded interval $\mathcal{I}$ on the real line to represent I, then g_i^u and g_i^r will be some almost continuous functions of i defined on $\mathcal{I}$. Note that g_i^r is determinate, while g_i^u is indeterminate but depends on the value Γ_c. Just as we have assumed above, g_i^r can be similarly thought of to be decreasing with i. Then by Corollary 4.11, there is a subinterval of $\mathcal{I}$,

$$\mathcal{I}_m = \big\{i \in \mathcal{I} : \min\{g_c^u, g_c^c\} \leq g_i^r \leq \max\{g_c^u, g_c^c\}\big\},$$

such that g_i^u is always bounded by g_i^r which could serve as its floor or ceiling. Figure 4.6 shows such two opposite cases, in which we simply set g_i^r linear, and g_i^u partly linear with only one cusp to g_i^r.



(i) $g_c^u \leq g_c^c$ (ii) $g_c^u \geq g_c^c$

FIGURE 4.6.

Once γ_i is designed to be sufficiently close to 1 for all $i \notin \mathcal{I}_m$, g_i^u would be nearly equal to g_i^r for all $i \notin \mathcal{I}_m$. In consequence, if $g_c^u \leq g_c^c$, then $g_c^u \geq g_c^r$ as $g_i^u \geq g_i^r$ for all $i \in \mathcal{I}_m$ (see case (i) in Figure 4.6); if $g_c^u \geq g_c^c$, then $g_c^u \leq g_c^r$ as $g_i^u \leq g_i^r$ for all $i \in \mathcal{I}_m$ (see case (ii) in Figure 4.6). Hence

$$\min\{g_c^r, g_c^c\} \leq g_c^u \leq \max\{g_c^r, g_c^c\}, \tag{4.45}$$

and immediately, we can see there might exist some $\Gamma_c \in (0,1)^{n-1}$ with $\gamma_i \uparrow 1$ for all $i \notin \mathcal{I}_m$, such that g_c^u is a certain convex combination of g_c^r and g_c^c.

To close this section, we shall propose two computational examples, in which the committee arrangement is described as either one of the two specific cases as was imagined at the beginning of this section.

EXAMPLE 4.8. Consider the case that $P_i^r = P^r$ for all $r_i \in N_1$, then $U_i(g) = U(g)$ for all $r_i \in N_1$. If $\Gamma_c = (1, 1, \ldots, 1)$, the equilibrium provision states of the degenerate M-form game will be (g_i^c, S_i^c) for all $r_i \in N_1$, such that

$$U'(g_c^c) = \delta M_g(g_i^c, \lambda_i^c), \quad M_\lambda(g_i^c, \lambda_i^c) = M_\lambda(g_j^c, \lambda_j^c),$$

where $\lambda_i^c = \lambda(S_i^c)$ for all r_i, and r_i, r_j are distinct. If $\Gamma_c = (0, 0, \ldots, 0)$, the equilibrium provision states will be (g_i^r, R_i) for all $r_i \in N_1$, such that $U'(g_i^r) = M_g(g_i^r, \lambda(R_i))$.

Let $N = \{c, r_1, r_2\}$, and assume

$$U(g) = \log(g + 1), \quad M(g, \lambda) = (\lambda g)^2,$$

then the marginal utility and marginal provision cost functions take the forms

$$U'(g) = 1/(g + 1), \quad M_g(g, \lambda) = 2\lambda^2 g.$$

It should be noted that $M_{g\lambda} = 4\lambda g \neq 0$ for all $g, \lambda > 0$. Suppose $\lambda(R_1) = 1/2$ and $\lambda(R_2) = 1/4$, so $\mu_1 = 2/3$ and $\mu_2 = 1/3$, and hence $g_c = (2g_1 + g_2)/3$. Besides, set $\delta = 0.5$ arbitrarily.

If $\Gamma_c = (1, 1)$, then we should have $\lambda_1^c = \lambda_2^c$ and $g_1^c = g_2^c$. Suppose the two fairly divided regions for r_1 and r_2 both have a provision capacity 1/3, that is, $\lambda_1^c = \lambda_2^c = 1/3$, then g_1^c, g_2^c should be the positive solution of the equation

$$1/(g + 1) = g/9,$$

which yields $g_1^c = g_2^c = (\sqrt{37} - 1)/2 \approx 2.541$, and hence $g_c^c \approx 2.541$. Observe that $\lambda_1^c < \lambda(R_1)$ and $\lambda_2^c > \lambda(R_2)$, thus there is a nonempty $\Delta \subset R_1$ such that $S_1^c = R_1 \setminus \Delta$ and $S_2^c = R_2 \cup \Delta$.

If $\Gamma_c = (0, 0)$, then $S_1^r = R_1$ and $S_2^r = R_2$. g_1^r and g_2^r then satisfy the following two marginal conditions,

$$1/(g_1^r + 1) = g_1^r/2, \quad 1/(g_2^r + 1) = g_2^r/8,$$

which yield $g_1^r = 1$ and $g_2^r = (\sqrt{33} - 1)/2 \approx 2.372$, and hence $g_c^r = (\sqrt{33} + 3)/6 \approx 1.457$.

Now keep all the above assumptions unaltered except $M(g, \lambda)$. Let's assume $M(g, \lambda)$ is separable such that $M_{g\lambda} = 0$ for all g, λ, for example, $M(g, \lambda) = g^2/2 + \psi(\lambda)$, and thus $M_g(g, \lambda) = g$. It is clear that $g_1^c = g_2^c = g_c^c$ and $g_1^r = g_2^r = g_c^r$, which should satisfy

$$1/(g_c^c + 1) = g_c^c/2, \quad 1/(g_c^r + 1) = g_c^r,$$

so $g_c^c = 1$ and $g_c^r = (\sqrt{5} - 1)/2 \approx 0.618$.

If $\Gamma_c \in (0,1)^2$, for example, $\Gamma_c = (1/3, 2/3)$ which says $\gamma_1 = 2/3$ and $\gamma_2 = 1/3$. By (4.37), there is a system of equations

$$\begin{cases} 6/(1+g_1) + 2/(1+g_c) = 7g_1 \\ 3/(1+g_2) + 2/(1+g_c) = 4g_2 \end{cases} \tag{4.46}$$

which yields $g_1^u \approx 0.679$ and $g_2^u \approx 0.729$, and hence $g_c^u \approx 0.696$. Note that

$$g_c^r < g_c^u < g_c^c; \quad g_1^u > g_1^r, \quad g_2^u > g_2^r.$$

EXAMPLE 4.9. Consider the case that $P_i^r = \{p_i\}$ for all $r_i \in N_1$, and $P^r = \{p_1, p_2, \ldots, p_{n-1}\}$, then we should have $U_i(g) = u_i(g)$ for all $r_i \in N_1$, and $U(g) = \sum_{i=1}^{n-1} \alpha_i u_i(g)$, where $\alpha_i \in (0,1)$ and $\sum_{i=1}^{n-1} \alpha_i = 1$.

Let $N = \{c, r_1, r_2\}$, so $N_1 = \{r_1, r_2\}$. Assume

$$u_1(g) = 2\log(g+1), \quad u_2(g) = \log(g+1)/2,$$

and set $\alpha_1 = 1/3$ and $\alpha_2 = 2/3$, then

$$U(g) = \log(g+1).$$

Assume $M(g,\lambda) = g^2/2 + \psi(\lambda)$ so that $M_{g\lambda}(g,\lambda) = 0$ for all g, λ, then $M_g(g,\lambda) = g$. Let $\lambda(R_1) = 1/2$ and $\lambda(R_2) = 1/4$, then $\mu_1 = 2/3$ and $\mu_2 = 1/3$, and hence $g_c = (2g_1 + g_2)/3$. Set once again $\delta = 0.5$.

Clearly, $g_1^c = g_2^c = g_c^c$, and they should satisfy

$$1/(g_c^c + 1) = g_c^c/2,$$

which yields $g_c^c = 1$. With regards g_1^r and g_2^r, they are the solutions of the following two equations,

$$2/(g_1+1) = g_1, \quad 1/(g_2+1) = 2g_2,$$

which yield $g_1^r = 1$ and $g_2^r = (\sqrt{3}-1)/2 \approx 0.366$, and hence $g_c^r = (\sqrt{3}+3)/6 \approx 0.789$.

As for $\Gamma_c \in (0,1)^2$, we can take for example $\Gamma_c = (1/3, 2/3)$ which says $\gamma_1 = 2/3$ and $\gamma_2 = 1/3$. By (4.37), there is a system of equations

$$\begin{cases} 12/(1+g_1) + 2/(1+g_c) = 7g_1 \\ \ 3/(1+g_2) + 4/(1+g_c) = 8g_2 \end{cases} \tag{4.47}$$

which yields $g_1^u \approx 1.008$ and $g_2^u \approx 0.518$, and hence $g_c^u \approx 0.845$.

Observe that $g_c^r < g_c^u < g_c^c$ and $g_2^r < g_c^u < g_1^r$. Notice also that $g_1^u > g_1^r$ and $g_2^u > g_2^r$, it thus follows that

$$g_2^r < g_2^u < g_c^r < g_c^u < g_1^r = g_c^c < g_1^u.$$

It is not hard to see that

$$[g_2^u, g_c^u] \cap [g_2^r, g_c^c] = [g_2^u, g_c^u] \neq \emptyset, \quad [g_c^u, g_1^u] \cap [g_1^r, g_c^c] = \{g_c^c\} \neq \emptyset,$$

which both comply with Proposition 4.9, and that $g_1^u > g_1^r$ for $g_c^u < g_1^r \leq g_c^c$ should comply with Corollary 4.11.

6. Complex Interaction

When one studies dynamics and processes on a network carrying randomness, the disorder of the network very often leads to the emergence of complexity (see for example, Barrat, Barthélemy, and Vespignani [**6**]). In this respect, the notion complexity which is usually associated with nonlinear dynamics seems to be closely related with time and randomness, and thus a complex network means that its behavior or any performance that could be investigated and gauged should essentially show complicated and sometimes even chaotic dynamics. In this section, we will however take a different perspective towards complexity, that's to say, it could also exist in a static and deterministic government network, particularly when the number of governments is extremely great.

6.1. Mixed-form Game. Recall that any isolated government in (N, f) only participates in the decision-making process deciding actions for itself, so it's sufficient to just consider the connected part of (N, f). Also, any government in an authority gap does not participate in any decision-making process, and even more, holds no information relevant to any decision-making process, thus a government network can be reduced to an equivalent one that has no authority gaps, in which the authority allocation rule should cover all the governments concerning at least one decision-making process[5].

We shall thus focus on government networks that are not only *physically connected* (*i.e.*, with no isolated governments), but also *politically connected* (*i.e.*, with no authority gaps).

Definition. A government network (N, f) with an authority allocation rule γ is called *compact*, if $N_\infty = \emptyset$ and it has no authority gap under γ.

The compactness of a government network implies its physical connectedness and political connectedness. Note that the governance structure of a compact government network must be decentralized, and yet not purely decentralized.

Consider a (physically) connected government network (N, f), then $N_\infty = \emptyset$ and $N_b \neq \emptyset$. Recall that $|N| = n$ and $c \notin N_b$, so $1 \leq |N_b| \leq n - 1$. If $|N_b| = 1$, then (N, f) is an H-form government network, and if $|N_b| = n - 1$, then it is an M-form

[5]Here, a government r_i is in an authority gap means that there exist some governments with nonzero authorities on both higher and lower hierarchical levels than that of r_i. On the other hand, there being no authority gaps, however conveys a much stronger condition, that is, any participating group L_i in (N, f) for a certain decision-making process has no zero-authority government.

government network. If $2 \leq |N_b| \leq n-2$, we shall say it is a *mixed-form government network*.

Definition. A game emerging from a compact mixed-form government network is called a *mixed-form game*.

The definition of mixed-form game appears to need much stronger conditions than the former terms, H-form game and M-form game, as we do not require an H-form or M-form game should happen on a compact H-form government line or compact M-form government star. However, we can still focus on the H-form games on a compact H-form government line, as authority gaps in a government line could be omitted without loss of generality. Also, we can point up the analytic M-form games which must be on a compact M-form government star, though the degenerate M-form games should happen on a politically disconnected and hence incompact M-form government star. Consequently, the game emerging on a compact government network must be one of the three possibilities: H-form game, analytic M-form game, and more generally, mixed-form game. Without generating any confusion, we will plainly use the group of government players N to denote a certain game on (N, f).

Recall that the group of directly subordinate governments to any $r_i \in N$ is denoted by D_i. In a mixed-form government network, we have $D_i \neq \emptyset$ for all $r_i \notin N_b$.

Definition. $\{r_i\} \cup D_i$ in a mixed-form game N is called a *simple game* for all $r_i \notin N_b$.

In particular, if $|D_i| = 1$, we call $\{r_i\} \cup D_i$ a *simple H-form game*, and if $|D_i| \geq 2$, we call it a *simple M-form game*. Since there are $n-|N_b|$ governments not in N_b, the number of simple games in N is exactly equal to $n-|N_b|$, where $2 \leq n-|N_b| \leq n-2$, as $2 \leq |N_b| \leq n-2$ for a mixed-form government network.

Notice that, in a simple game $\{r_i\} \cup D_i$, the government r_i is a lower-level one to all government in D_i, on the other hand, each government in D_i is an upper-level one to r_i. Here, an upper-level government to r_i means it is further to c than r_i, while a lower-level government to r_i means it is closer to c than r_i.

Lemma 4.12. *Any government player r_i in a mixed-form game N is engaged in either one simple game or two distinct simple games.*

Proof. We first show that any government player in N should be engaged in at least one simple game. Suppose not, and there is a government r_i not engaged in any simple game, then we must have $D_i = \emptyset$, and hence $r_i \in N_b$. So there is a lower-level government to r_i, say r_j such that $r_i \in D_j$, otherwise $r_i = c$ would contradict $r_i \in N_b$. It hence appears that $D_j \neq \emptyset$ and $r_j \notin N_b$, thus $\{r_j\} \cup D_j$ must be a simple game. But r_i is then engaged in the simple game $\{r_j\} \cup D_j$, a contradiction.

We next show that any government player can not be engaged in more than two distinct simple games, which would complete the proof. Suppose not, and there is a government r_i engaged in more than two distinct simple games. Notice that r_i has a unique D_i, so there is at most one simple game $\{r_i\} \cup D_i$, in which r_i is on the lower level. Then there should be at least two distinct simple games, in which r_i is on the upper level, for instance, $\{r_j\} \cup D_j$ and $\{r_{j'}\} \cup D_{j'}$, where $j \neq j'$, and $r_i \in D_j$ as well as $r_i \in D_{j'}$.

Since (N, f) as a mixed-form government network must be (physically) connected, there exist two paths,

$$L_j = \{c, r_\jmath, \ldots, r_{\hat\jmath}, r_j\}, \quad L_{j'} = \{c, r_{\jmath'}, \ldots, r_{\hat\jmath'}, r_{j'}\},$$

where $|L_j|, |L_{j'}| \geq 2$, and thus

$$f_{j\hat\jmath} = \cdots = f_{\jmath 0} = 1, \quad f_{0\jmath'} = \cdots = f_{\hat\jmath' j'} = 1.$$

Since $r_i \in D_j$ and $r_i \in D_{j'}$, we have $f_{ij} = f_{j'i} = 1$. Immediately, there would be an l-cycle in (N, f) for $l \geq 4$, *i.e.*, $r_i r_j \cdots c \cdots r_{j'}$, as

$$f_{ij} = f_{j\hat\jmath} = \cdots = f_{\jmath 0} = f_{0\jmath'} = \cdots = f_{\hat\jmath' j'} = f_{j'i} = 1.$$

But (N, f) is defined to be acyclic, a contradiction. □

In a mixed-form government network (N, f), there should be $|N_b|$ distinct paths connecting the central government to the bottom governments, *i.e.*, $L_i = \{c, \ldots, r_i\}$ for all $r_i \in N_b$, where $|L_i| \geq 2$. Note that the central government c is always on the lowermost level 0 in any path L_i, while each bottom government $r_i \in N_b$ is on the uppermost level $|L_i| - 1$ in the path L_i. It plainly implies that merely the governments in $\{c\} \cup N_b$ are engaged in one simple game, and hence by Lemma 4.12, all the other governments which are neither central nor bottom, should be engaged in exactly two distinct simple games.

Recall that the group of all the subordinate governments to any $r_i \in N$ is denoted by $\mathcal{D}_i$. Properly stating, any $\mathcal{D}_i$ should be completely determined by all path L_j for $r_j \in N_b$, such that $r_i \in L_j$, and more precisely, for all $r_i \in N$,

$$\mathcal{D}_i = \bigcup_{L_j \ni r_i} L_j \setminus \bigcap_{L_j \ni r_i} L_j. \tag{4.48}$$

Directly, we can hence verify that $\mathcal{D}_0 = N \setminus \{c\}$, and $\mathcal{D}_i = \emptyset$ for all $r_i \in N_b$.

Definition. $\{r_i\} \cup \mathcal{D}_i$ in a mixed-form game N is called an *induced subgame* for all $r_i \notin N_b$.

If $r_i = c$, the induced subgame will then be $\{c\} \cup \mathcal{D}_0 = N$, which is actually the mixed-form game N itself. If $\mathcal{D}_i \subseteq N_b$ for some r_i, then we must have $\mathcal{D}_i = D_i$, and hence the induced subgame $\{r_i\} \cup \mathcal{D}_i$ is also a simple game. If $\mathcal{D}_i \neq D_i$, or hence $\mathcal{D}_i \supset D_i$ for some r_i, we shall call the induced subgame $\{r_i\} \cup \mathcal{D}_i$ *nonsimple*.

Lemma 4.13. *Any nonsimple induced subgame $\{r_i\} \cup \mathcal{D}_i$ can be hierarchically decomposed as a simple game $\{r_i\} \cup D_i$ and a collection of induced subgames $\{r_j\} \cup \mathcal{D}_j$ for all $r_j \in D_i$.*

Proof. Suppose $\{r_i\} \cup \mathcal{D}_i$ is a nonsimple induced subgame of N, then we must have $r_i \notin N_b$ and $\mathcal{D}_i \supset D_i$, and of course, $\mathcal{D}_i \neq \emptyset$. Notice that $\{r_i\}$ and D_i form a simple game $\{r_i\} \cup D_i$, and there is still a nonempty group of government players not engaged in the simple game, *i.e.*, $\mathcal{D}_i \setminus D_i$, which is the group of all the governments subordinate to some government in D_i. We thus have

$$\mathcal{D}_i \setminus D_i = \bigcup_{r_j \in D_i} \mathcal{D}_j,$$

and as a result of that, each $r_j \in D_i$ serves as a government hub connecting the induced subgame $\{r_j\} \cup \mathcal{D}_j$ to the simple game $\{r_i\} \cup D_i$, which then completes our proof. □

Proposition 4.14. *A mixed-form game N can be hierarchically decomposed as $n - |N_b|$ simple games.*

Proof. Note that $N = \{c\} \cup \mathcal{D}_0$ is an induced subgame itself, and it is clearly nonsimple, otherwise N would be an analytic M-form game. By Lemma 4.13, N can be hierarchically decomposed as a simple game $\{c\} \cup D_0$, or equivalently $\{c\} \cup N_1$ as $D_0 = N_1$, and a collection of induced subgames $\{r_i\} \cup \mathcal{D}_i$ for all $r_i \in N_1$. Note that $\mathcal{D}_i \supseteq D_i$ for all $r_i \in N_1$, so either $\mathcal{D}_i = D_i$, or $\mathcal{D}_i \supset D_i$ for each $r_i \in N_1$. If $\mathcal{D}_i = D_i$ for some $r_i \in N_1$, the induced subgame $\{r_i\} \cup \mathcal{D}_i$ is actually simple, and hence the hierarchical decomposition stops there.

If $\mathcal{D}_i \supset D_i$ for some $r_i \in N_1$, the induced subgame $\{r_i\} \cup \mathcal{D}_i$ is nonsimple, and again by Lemma 4.13, it can be hierarchically decomposed as a simple game $\{r_i\} \cup D_i$ and a collection of induced subgames $\{r_j\} \cup \mathcal{D}_j$ for all $r_j \in D_i$. Notice that r_i is now engaged in two distinct simple games, *i.e.*, $\{c\} \cup N_1$ and $\{r_i\} \cup D_i$. By Lemma 4.12, any government can be engaged in at most two distinct simple games, so the hierarchical decomposition process at any $r_i \in N_1$ should be complete.

In general, the hierarchical decomposition process is complete and would continue until it reaches all government r_j with $\mathcal{D}_j = D_j$, or $\{r_j\} \cup \mathcal{D}_j$ is simple. Evidently, $\mathcal{D}_j = D_j$ happens only if $\mathcal{D}_j \subseteq N_b$. Therefore, the hierarchical decomposition process passes through all the governments except these bottom ones, and hence we will have $1 + 2(n - |N_b| - 1)$ simple games, among which $n - |N_b| - 1$ ones are duplicated. It finally shows that the number of distinct simple games emerging in the hierarchical decomposition of the mixed-form game N should be equal to

$$1 + 2(n - |N_b| - 1) - (n - |N_b| - 1) = n - |N_b|,$$

which then completes the proof. □

We should admit that the hierarchical decomposition of a mixed-form game is a rather artificial and yet quite useful notion. A complex mixed-form game could then have a separable structure with fair simplicity, as any simple game in a mixed-form game is either H-form or analytic M-form.

Example 4.10. Consider a small mixed-form game on the government network as defined in Figure 4.1, in which $n = 13$ and $|N_b| = 8$. Such a mixed-form game can be hierarchically decomposed as the following 5 simple games,

$$\{c, r_1, r_2, r_3, r_4\}, \quad \{r_2, r_5\}, \quad \{r_3, r_6, r_7, r_8\}, \quad \{r_4, r_9, r_{10}\}, \quad \{r_{10}, r_{11}, r_{12}\}.$$

Note that $\{r_2, r_5\}$, $\{r_3, r_6, r_7, r_8\}$, and $\{r_{10}, r_{11}, r_{12}\}$ are also induced subgames, as we have

$$\mathcal{D}_2 = D_2 = \{r_5\}, \quad \mathcal{D}_3 = D_3 = \{r_6, r_7, r_8\}, \quad \mathcal{D}_{10} = D_{10} = \{r_{11}, r_{12}\}.$$

Observe that $\{r_4, r_9, r_{10}\}$ and $\{r_{10}, r_{11}, r_{12}\}$ are the outcomes of the hierarchical decomposition of the (nonsimple) induced subgame $\{r_4, r_9, r_{10}, r_{11}, r_{12}\}$.

6.2. Political Structure. As we have frequently written, any decision-making process in (N, f) is in practice designed towards local public good provision decisions for some bottom government in N_b, and distinct decision-making processes have no overlapping procedures. In that sense, there should be $|N_b|$ distinct decision-making processes in total, and the authority allocation rule γ is then entirely determined by $|N_b|$ authority distributions over such path as L_j for all $r_j \in N_b$.

Suppose each $r_i \in N$ would be allocated with an authority γ_i^j concerning the decision-making process aiming to decide provision actions for $r_j \in N_b$, then $\gamma_i^j \in [0, 1)$, or more precisely, γ_i^j could be either 0 for $r_i \notin L_j$, or within $(0, 1)$ for $r_i \in L_j$. Let

$$\Gamma_i = (\gamma_i^j,\ r_j \in N_b),$$

thus Γ_i is an $|N_b|$-tuple. Let z_i denote the number of nonzero entries in Γ_i. Since (N, f) is compact and hence politically connected, we have $1 \leq z_i \leq |N_b|$ for all r_i.

For example, for any $r_j \in N_b$, we have $r_j \in L_j$ with $\gamma_j^j \neq 0$, and $r_j \notin L_\jmath$ for all $r_\jmath \in N_b \setminus \{r_j\}$ and hence $\gamma_j^\jmath = 0$, thus $z_j = 1$ for all $r_j \in N_b$. As for c, we have $c \in L_j$ for all $r_j \in N_b$ and hence $\gamma_0^j \neq 0$, thus $z_0 = |N_b|$. Moreover, the authority distribution over a path $L_j = \{c, \ldots, r_j\}$ for some $r_j \in N_b$ should satisfy

$$\gamma_0^j + \gamma_j^j + \sum_{r_i \notin N_b \cup \{c\}} \gamma_i^j = 1, \tag{4.49}$$

where $\gamma_0^j, \gamma_j^j \neq 0$.

We now have an $|N_b| \times n$ matrix

$$Q = \left(\Gamma_0^{\mathrm{T}}, \Gamma_1^{\mathrm{T}}, \ldots, \Gamma_{n-1}^{\mathrm{T}}\right),$$

where Γ_i^{T} denotes the transpose of Γ_i for all r_i. Recall that $I = \{0, 1, \ldots, n-1\}$ is the index set of N, we then define a permutation rule $\sigma : I \to I$, such that

(i) $\sigma(i) \in \{i \in I : r_i \in N \setminus N_b\}$ for all $0 \le i \le n - |N_b| - 1$,

(ii) $\sigma(i) \in \{i \in I : r_i \in N_b\}$ for all $n - |N_b| \le i \le n - 1$,

(iii) $\sigma(i) < \sigma(j)$, if $0 \le i < j \le n - |N_b| - 1$ or $n - |N_b| \le i < j \le n - 1$.

Intuitively saying, σ allows us to rearrange Q in such a way that any government in $N \setminus N_b$ will be placed before all the governments in N_b, and governments in $N \setminus N_b$ and N_b will both keep their original orders. Let

$$Q^\sigma = \left(\Gamma_{\sigma(0)}^{\mathrm{T}}, \Gamma_{\sigma(1)}^{\mathrm{T}}, \ldots, \Gamma_{\sigma(n-1)}^{\mathrm{T}}\right). \tag{4.50}$$

And we shall partition Q^σ into a pair of submatrices, $Q^\sigma = (Q_a^\sigma, Q_b^\sigma)$, such that

$$Q_a^\sigma = \left(\Gamma_{\sigma(0)}^{\mathrm{T}}, \Gamma_{\sigma(1)}^{\mathrm{T}}, \ldots, \Gamma_{\sigma(n-|N_b|-1)}^{\mathrm{T}}\right),$$
$$Q_b^\sigma = \left(\Gamma_{\sigma(n-|N_b|)}^{\mathrm{T}}, \Gamma_{\sigma(n-|N_b|+1)}^{\mathrm{T}}, \ldots, \Gamma_{\sigma(n-1)}^{\mathrm{T}}\right).$$

Clearly, Q_a^σ is an $|N_b| \times (n - |N_b|)$ matrix, and Q_b^σ is an invertible $|N_b| \times |N_b|$ diagonal matrix.

Let $\lceil\Gamma_i\rceil = (\lceil\gamma_i^j\rceil,\ r_j \in N_b)$, where $\lceil\gamma_i^j\rceil = \min\{x \in \mathbb{Z} : x \ge \gamma_i^j\}$. Define the ceiling of Q to be

$$\lceil Q\rceil = \left(\lceil\Gamma_0\rceil^{\mathrm{T}}, \lceil\Gamma_1\rceil^{\mathrm{T}}, \ldots, \lceil\Gamma_{n-1}\rceil^{\mathrm{T}}\right). \tag{4.51}$$

Since $\gamma_i^j \in [0, 1)$, we have $\lceil\gamma_i^j\rceil \in \{0, 1\}$, and in particular, $\lceil\gamma_i^j\rceil = 1$ if $\gamma_i^j \in (0, 1)$, and $\lceil\gamma_i^j\rceil = 0$ if $\gamma_i^j = 0$. Note that z_i now also counts the number of entries equal to 1 in $\lceil\Gamma_i\rceil$.

Notice that $\lceil Q\rceil^\sigma = \lceil Q^\sigma\rceil$. Let $Q^* = \lceil Q^\sigma\rceil$ and $Q_a^* = \lceil Q_a^\sigma\rceil$. We thus have

$$Q^* = \left(\lceil Q_a^\sigma\rceil, \lceil Q_b^\sigma\rceil\right) = \left(Q_a^*, I_{|N_b|}\right), \tag{4.52}$$

where $I_{|N_b|}$ is the identity matrix of order $|N_b|$. Evidently, Q_a^* has all the information about γ carried by Q^*, and thus we could use it to replace Q^*.

Definition. Q_a^* is called the *indicator matrix* of the compact (N, f).

Q_a^* contains all the column vectors $\lceil\Gamma_i\rceil^{\mathrm{T}}$ for $r_i \notin N_b$, and the $(i+1)$-th column of Q_a^* for $0 \le i \le n - |N_b| - 1$ is fixed by $\lceil\Gamma_{\sigma(i)}\rceil^{\mathrm{T}}$, or inversely stating, each $\lceil\Gamma_i\rceil^{\mathrm{T}}$ for $r_i \notin N_b$ is placed as the $(\sigma^{-1}(i) + 1)$-th column of Q_a^*.

If (N, f) is H-form, then $|N_b| = 1$, and hence Q_a^* is an $(n-1)$-dimensional row vector. Since $z_i = 1$ for all $r_i \in N$, we have $Q_a^* = (1, 1, \ldots, 1)$. If (N, f) is M-form, then $|N_b| = n - 1$, and hence Q_a^* is an $(n-1)$-dimensional column vector. Notice that $z_0 = n - 1$, so $Q_a^* = (1, 1, \ldots, 1)^{\mathrm{T}}$.

If (N, f) is mixed-form, then $2 \leq |N_b| \leq n-2$, and Q_a^* is an $|N_b| \times (n - |N_b|)$ matrix. Note that $z_0 = |N_b|$, and $1 \leq z_i \leq |N_b|$ for all $r_i \notin N_b \cup \{c\}$, Q_a^* should thus satisfy

$$n - 1 \leq \sum_{r_i \notin N_b} z_i \leq |N_b|(n - |N_b|). \tag{4.53}$$

Example 4.11. We can consider the mixed-form game on (N, f) as defined in Figure 4.1 once more. The permutation rule σ is defined by $\sigma(0) = 0$, $\sigma(4) = 10$, $\sigma(5) = 1$, and

$$\sigma(i) = \begin{cases} i+1 & \text{if } i = 1, 2, 3 \\ i-1 & \text{if } i = 6, 7, 8, 9, 10 \\ i & \text{if } i = 11, 12 \end{cases}$$

Since $n = 13$ and $|N_b| = 8$, Q_a^* is an 8×5 matrix, and moreover,

$$Q_a^* = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 & 1 \end{pmatrix}$$

The $(i+1)$-th column of Q_a^* corresponds to the government $r_{\sigma(i)}$ for all $0 \leq i \leq 4$, that's to say, these 5 ordered columns of Q_a^* are determined by c, r_2, r_3, r_4, and r_{10}, respectively. Observe that $z_0 = 8$, $z_2 = 1$, $z_3 = z_4 = 3$, and $z_{10} = 2$, thus there are 17 entries in Q_a^* with the value 1.

Let N_i^a denote the group of governments in N_b with entries equal to 1 in the $\sigma^{-1}(i)$-th column of Q_a^* corresponding to $r_i \in N \setminus N_b$. Let L_j^a denote the group of governments in $N \setminus N_b$ with entries equal to 1 in the $(\sigma^{-1}(j) - N + |N_b| + 1)$-th row of Q_a^* corresponding to $r_j \in N_b$. Thus for each $r_i \in N \setminus N_b$, we have

$$N_i^a = \{r_j \in N_b : \lceil \gamma_i^j \rceil = 1\}, \tag{4.54}$$

and in particular, $N_0^a = N_b$. Note that $N_i^a \neq \emptyset$ for all $r_i \in N \setminus N_b$, as $|N_i^a| = z_i \geq 1$. And for each $r_j \in N_b$, we have

$$L_j^a = \{r_i \notin N_b : \lceil \gamma_i^j \rceil = 1\}. \tag{4.55}$$

Note that $c \in L_j^a$ and $L_j^a \cup \{r_j\} = L_j$ for all $r_j \in N_b$.

Lemma 4.15. *The following two dual statements hold for all Q_a^*:*

(i) *If* $N_i^a \cap N_{i'}^a \neq \emptyset$*, then either* $N_i^a \subseteq N_{i'}^a$ *or* $N_{i'}^a \subseteq N_i^a$*;*
(ii) *If* $L_j^a \cap L_{j'}^a \supset \{c\}$*, then either* $L_j^a \subseteq L_{j'}^a$ *or* $L_{j'}^a \subseteq L_j^a$*.*

Proof. Suppose $N_i^a \cap N_{i'}^a \neq \emptyset$, but neither $N_i^a \subseteq N_{i'}^a$ nor $N_{i'}^a \subseteq N_i^a$. Then $N_i^a \setminus N_{i'}^a \neq \emptyset$ and also $N_{i'}^a \setminus N_i^a \neq \emptyset$, thus there should be at least two distinct bottom governments, r_j and $r_{j'}$, such that $r_j \in N_i^a$ but $r_j \notin N_{i'}^a$, and $r_{j'} \in N_{i'}^a$ but $r_{j'} \notin N_i^a$. Let $L_j \cap L_{j'} = L_\imath$, so $L_\imath \supseteq \{c\}$. If $r_i \in L_\imath$, then $\mathcal{D}_i \supseteq \mathcal{D}_\imath$. Since $r_{j'} \in \mathcal{D}_\imath$, we then have $r_{j'} \in \mathcal{D}_i$, and hence $r_{j'} \in N_i^a$, a contradiction. If $r_{i'} \in L_\imath$, we would have $r_j \in N_{i'}^a$, again a contradiction. Thus $r_i, r_{i'} \notin L_\imath$. Since $N_i^a \cap N_{i'}^a \neq \emptyset$, there is at least one bottom government $r_\jmath \in N_b$, such that $r_\jmath \in N_i^a$ and $r_\jmath \in N_{i'}^a$. But then the path $L_\jmath$ would have two distinct representations,

$$L_\imath \cup \{\dots, r_i, \dots, r_\jmath\}, \quad L_\imath \cup \{\dots, r_{i'}, \dots, r_\jmath\},$$

and evidently, that fact contradicts $L_\jmath$ should be unique. Therefore, N_i^a must be either a subset or a superset of $N_{i'}^a$, which completes the proof of statement (i).

Since $r_j \in N_i^a$ if and only if $r_i \in L_j^a$ for all $r_i \notin N_b$ and $r_j \in N_b$, the proof of statement (ii) is direct. □

Proposition 4.16. *The rank of* Q_a^* *is equal to the number of governments that directly connect with some bottom government.*

Proof. Let's partition N_b into two disjoint groups, N_b^c and N_b^r, such that $L_j^a = \{c\}$ for all $r_j \in N_b^c$, and $L_j^a \supset \{c\}$ for all $r_j \in N_b^r$. The rows of Q_a^* corresponding to N_b^c are thus all $(1, 0, 0, \dots, 0)$, while the row of Q_a^* corresponding to some $r_j \in N_b^r$ can be written as

$$(1, 0, 0, \dots, 0) + q_j,$$

where $q_j \neq (0, 0, \dots, 0)$. It should be direct that $(1, 0, 0, \dots, 0)$ is linearly independent to all the rows corresponding to N_b^r.

Consider $r_j, r_{j'} \in N_b^r$ with $L_j^a \cap L_{j'}^a = \{c\}$, then the rows that correspond to r_j and $r_{j'}$ must be linearly independent, as $q_j \neq q_{j'}$. Consider $r_j, r_{j'} \in N_b^r$ with $L_j^a \cap L_{j'}^a \supset \{c\}$, then by statement (ii) of Lemma 4.15, either $L_j^a \subseteq L_{j'}^a$ or $L_{j'}^a \subseteq L_j^a$, and hence the rows corresponding to r_j and $r_{j'}$ are again linearly independent. In any case, if the rows corresponding to r_j and $r_{j'}$ are not same, they then must be linearly independent.

In sum, the rows corresponding to all distinct $r_j, r_{j'} \in N_b$ are linearly dependent, if and only if they are exactly same. Consequently, the rank of Q_a^* is equal to the number of distinct, and hence linearly independent, rows in Q_a^*. Note that the rows corresponding to r_j and $r_{j'}$ are same, if and only if $L_j^a = L_{j'}^a$, in other words,

$$L_j \setminus \{r_j\} = L_{j'} \setminus \{r_{j'}\},$$

which means r_j and $r_{j'}$ connect to a same government. Thus the rank of Q_a^* is equal to the number of distinct governments next to the bottom governments in all L_j for $r_j \in N_b$, which completes the proof. $\square$

The objects of mixed-form game N and indicator matrix Q_a^* are both derived from a compact government network (N, f). In fact, when the compact government network (N, f) is mixed-form, *i.e.*, $2 \leq |N_b| \leq n-2$, the game is a mixed-form one. Since the authority allocation rule γ in a compact government network fully covers all L_j for $r_j \in N_b$, Q_a^* as a Boolean matrix is effectively determined by the network structure of (N, f), which can be characterized by all value $\lceil\gamma_i^j\rceil$ for $r_i \notin N_b$ and $r_j \in N_b$. In addition, we should see that the mixed-form game N can be derived from the indicator matrix Q_a^* by means of N_i^a and L_j^a for $r_i \notin N_b$ and $r_j \in N_b$, and vice versa. It thus suggests that Q_a^* should have some invariants with respect to its corresponding mixed-form game N, and Proposition 4.16 does show that the rank of Q_a^* is such an invariant.

The following diagram (see Figure 4.7) briefly presents these relations among the three mathematical objects, *i.e.*, the graph (N, f), the game N, and the matrix Q_a^*.

$$\begin{tikzcd}
(N,f) \arrow[r, "2\leq|N_b|\leq n-2"] \arrow[d, "\lceil\gamma_i^j\rceil"'] & N \arrow[dl, leftrightarrow, "N_i^a, L_j^a"] \\
Q_a^* &
\end{tikzcd}$$

FIGURE 4.7.

Without any more formal arguments, we can make some observations about the mixed-form game N based on the indicator matrix Q_a^*. First of all, for all $r_i \neq c$ (and not in N_b), if there is no $r_{i'} \neq r_i, c$ with $N_{i'}^a \subseteq N_i^a$, then

$$N_i^a = D_i = \mathcal{D}_i,$$

and hence $\{r_i\} \cup N_i^a$ is a simple game.

Next, for all $r_i \notin N_b$, if there are governments $r_{i_1}, r_{i_2}, \ldots, r_{i_l}$, such that $N_{\hat{\imath}}^a \subseteq N_i^a$ for all $\hat{\imath} \in \{i_1, i_2, \ldots, i_l\}$, and there is no other government $r_{i'}$ with $N_{i'}^a \subseteq N_i^a$, then

$$\mathcal{D}_i = \{r_{i_1}, r_{i_2}, \ldots, r_{i_l}\} \cup N_i^a, \tag{4.56}$$

and thus

$$\{r_{i_1}, r_{i_2}, \ldots, r_{i_l}, r_i\} \cup N_i^a = \{r_i\} \cup \mathcal{D}_i$$

is an induced subgame. In case $\{r_{i_1}, r_{i_2}, \ldots, r_{i_l}\} \neq \emptyset$, such an induced subgame would be nonsimple.

Last but not least, by statement (i) of Lemma 4.15, N_b can be partitioned into pairwise disjoint nonempty groups of governments, say,

$$N_\imath^a, \quad N_{i_1}^a, \quad N_{i_2}^a, \quad \ldots, \quad N_{i_l}^a,$$

such that for all $r_i \neq c$, $N_i^a \subseteq N_{\hat{\imath}}^a$ for some $\hat{\imath} \in \{i_1, i_2, \ldots, i_l\}$, and

$$N_\imath^a = \big\{r_j \in N_b : L_j^a = \{c\}\big\}. \tag{4.57}$$

In a mixed-form government network, $r_\imath$ can not be c, as $N_0^a = N_b$ but $N_\imath^a \subset N_b$, and can not be any other $r_i \neq c$ either, as $N_i^a \cap N_\imath^a = \emptyset$ for all $r_i \neq c$. It therefore implies that $r_\imath$ does not really exist, and for that reason, we shall call $r_\imath$ an *imaginary government*.

It might be noted that $N_\imath^a = N_b^c$ (*cf.*, the proof of Proposition 4.16), and

$$N_\imath^a \cup \{r_{i_1}, r_{i_2}, \ldots, r_{i_l}\} = D_0 = N_1. \tag{4.58}$$

In particular, if $N_\imath^a = \emptyset$, then N_1 will contain no bottom government, and hence $N_1 \cap N_b = \emptyset$; if $N_\imath^a = N_b$, the partition of N_b will then become trivial, and hence $N_1 = N_b$, which means the mixed-form game N would degenerate into an analytic M-form one.

6.3. Pragmatic Solution. In this subsection, we are about to propose one solution to the mixed-form game N on a large compact government network (N, f). To deal with the complexity of the game, we shall impose a simplified reasoning structure on the game, so that its solution could emerge in a rather natural way. Even though the imposed reasoning structure might have some satisfying features, we are still unable to predict that complex interactions in a real government network should generate the same outcome just as we propose.

Let's introduce a small time interval into our modelling framework, and assume the mixed-form game would then exist on that time interval. Hence the mixed-form game would be essentially dynamic, and by Proposition 4.14, the mixed-form game N could be hierarchically decomposed as $n - |N_b|$ simple games, which can now be viewed as existing at distinct time points of the time interval. In case a simple game could be represented by its solution, or more generally, its (integrated) objective function, then the simple game itself would turn out to be an integrated government with a certain common objective characterized by its solution. Suppose a simple game on an upper hierarchical level should also appear relatively earlier in the time interval, as it is closer to the bottom governments which implement provision actions in the political system. The mixed-form game will thus become an updating process, which goes through all the $n - |N_b|$ simple games from N_b to $\{c\}$.

Let k denote the uppermost hierarchical level of (N, f). Since (N, f) is compact, and hence physically connected, we have $2 \leq k \leq n - |N_b|$, where $2 \leq |N_b| \leq n - 2$. To see this fact, suppose $k = 1$, then there were $n - |N_b| \geq 2$ central governments, a contradiction; suppose $k = n - |N_b| + 1$, then there were at most $n - k = |N_b| - 1$ bottom governments, simply as there were at least k governments not in N_b, again a contradiction.

Each government in N should be on a certain hierarchical level l for $0 \leq l \leq k$, and directly, N can thus be partitioned into the following $k + 1$ pairwise disjoint groups,

$$N_0, \quad N_1, \quad N_2, \quad \ldots, \quad N_k,$$

where $N_0 = \{c\}$ and $N_k \subseteq N_b$.

Let N_b^l denote the group of bottom governments on the hierarchical level l for $1 \leq l \leq k$, then we have

$$N_b^l = \{r_j \in N_b : r_j \in N_l\}, \tag{4.59}$$

where $N_b^k = N_k$, and $N_b^l \subset N_l$ for all $1 \leq l \leq k - 1$.

It thus appears that N_b can be partitioned into k pairwise disjoint groups,

$$N_b^1, \quad N_b^2, \quad \ldots, \quad N_b^k.$$

Notice that N_b^l can be equally expressed as

$$N_b^l = \{r_j \in N_b : |L_j^a| = l\}, \tag{4.59$'$}$$

thus we have $N_b^1 = N_b^c$ and $N_b \setminus N_b^1 = N_b^r$ (*cf.*, the proof of Proposition 4.16), which yield

$$N_b^r = N_b^2 \cup N_b^3 \cup \cdots \cup N_b^k.$$

The group of the directly subordinate governments to $N_l \setminus N_b^l$ is now another perspective to interpret the group of governments on the next hierarchical level $l + 1$ for all $1 \leq l \leq k - 1$. Accordingly, the following recursive relation between N_l and N_{l+1} can be derived,

$$N_{l+1} = \bigcup_{r_i \in N_l \setminus N_b^l} D_i, \tag{4.60}$$

where $1 \leq l \leq k - 1$ and for $l = 0$ we have $N_1 = D_0$.

By (4.60), we should see the updating process has k time steps, and the τ-th time step for $1 \leq \tau \leq k$, will be occupied with such a collection of simple games that $N_{k+1-\tau}$ is exactly the group of all their upper-level government players. Let $\varepsilon[\tau, \tau+1)$ denote the τ-th time step, then the introduced short time interval is $\varepsilon[1, k+1)$, where $\varepsilon > 0$ is close enough to 0. The timing representation of the updating process is shown in the following diagram (see Figure 4.8), in which $\varepsilon[1, k + 1)$ is uniformly scaled by $1/\varepsilon$ for transparency.

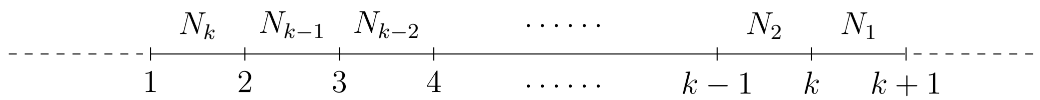


FIGURE 4.8.

At the first time step, $N_k = N_b^k$ is directly subordinate to $N_{k-1} \setminus N_b^{k-1}$, so the involved simple games are $\{r_i\} \cup D_i$ for all $r_i \in N_{k-1} \setminus N_b^{k-1}$. Each $r_i \in N_{k-1} \setminus N_b^{k-1}$ will then become integrated at the second time step, in the sense that each r_i would represent the simple game $\{r_i\} \cup D_i$.

In general, at the τ-th time step, for all $1 \leq \tau \leq k-1$, the involved simple games are $\{r_i\} \cup D_i$ for all $r_i \in N_{k-\tau} \setminus N_b^{k-\tau}$, while each $r_i \in N_{k-\tau} \setminus N_b^{k-\tau}$ will then represent the simple game $\{r_i\} \cup D_i$ at the $(\tau+1)$-th time step.

At the last time step, there is only one simple game remains, that is, $\{c\} \cup D_0 = \{c\} \cup N_1$, in which all the governments in $N_1 \setminus N_b^1$ have already been integrated.

Since any simple game of N is either an H-form game (with only two governments) or an analytic M-form game, its solutions and properties should be fairly accessible, especially when they are treated as utilitarian ones in Harsanyi's [**30**] sense. Assume that the collection of simple games at the τ-th time step admits a solution, (g_j^τ, S_j^τ) for all $r_j \in N_b$ and $1 \leq \tau \leq k$. The updating process will thus yield the solution to the original mixed-form game, (g_j^k, S_j^k) for all $r_j \in N_b$, such that $S_j^k \cap S_{j'}^k = \emptyset$ for all distinct $r_j, r_{j'}$, and $\bigcup_{r_j \in N_b} S_j^k \subseteq R$, for the reason that all the provision costs are aggregated into the objective function of c by the updating process. However, we can not say anything more about that solution, especially if a general provision cost function with nonzero $M_{g\lambda}$ has been adopted.

7. Final Remarks

We shall close this chapter with three remarks. First of all, we would propose another pragmatic solution for a complex mixed-form game. Once we take the mixed-form game N itself as an integrated government r_c, the game would then be reduced to a decision situation for r_c. Similar to the polynomial $h(\delta, \gamma)$ in an H-form game, we can introduce a mapping

$$H : (\delta, Q) \mapsto H(\delta, Q),$$

where $Q = (\gamma_i^j)$ denotes the $|N_b| \times n$ authority matrix, and $H(\delta, Q) \in (0, 1)$. Hence the objective function of r_c takes the form

$$\Pi(g, S) = V(g, S) - H(\delta, Q) \sum_{r_j \in N_b} \mu_j M(g_j, \lambda(S_j)), \tag{4.61}$$

where $g = \sum_{r_j \in N_b} \mu_j g_j$ and $S = \bigcup_{r_j \in N_b} S_j$. Clearly, the solution to the game N is the collection of (g_j^c, S_j^c) for all $r_j \in N_b$ which jointly maximize $\Pi(g, S)$.

Secondly, it should be stressed that the roles of communication and information sharing in the games were intentionally neglected. In fact, we have assumed that any authority gap in a government network does not affect any decision-making process and hence the entire game. However, governments in some authority gap, thought with no authority at all, might still play nontrivial roles in some collective decision-making process, especially when they serve as hubs connecting governments with enforceable political powers.

Last but not least, the authority allocation rule in a political system has been set as fixed in our studies. Nevertheless, it seems fairly natural that different political systems should adopt different authority allocation rules according to their different normative initiations per se. For example, if a political system is supposed to rely on economic efficiency and direct democracy more seriously, then more authorities should be distributed to the bottom governments. If a political system has developed a rather dominant tendency towards justice, then more authorities might be better allocated to such governments that have greatest enforceable political powers. As was demonstrated in an H-form government line, they should be the governments on the hierarchical levels χ^* and $\chi^* + 1$, where χ^* is defined to be the conjugate value of $\lfloor \log_\delta h(\delta, \gamma) \rfloor$. Overall, it seems quite hard to determine and even to design an appropriate authority allocation rule for a specific political system.

Bibliography

1. M. A. Akivis and V. V. Goldberg, *Algebraic aspects of web geometry*, Comment. Math. Univ. Carolin. **41** (2000), no. 2, 205–236.
2. M. Al-Suhaibani and L. Kryzanowski, *An exploratory analysis of the order book, and order flow and execution on the Saudi Stock Market*, J. Bank. Finance **24** (2000), no. 8, 1323–1357.
3. L. Bachelier, *Théorie de la spéculation*, Ann. Sci. Éc. Norm. Supér. **3** (1900), no. 17, 21–86.
4. J. Banks, J. Brooks, G. Cairns, G. Davis, and P. Stacey, *On Devaney's definition of chaos*, Amer. Math. Mon. **99** (1992), no. 4, 332–334.
5. M. F. Barnsley and S. Demko, *Iterated function systems and the global construction of fractals*, Proc. R. Soc. Lond. A **399** (1985), no. 1817, 243–275.
6. A. Barrat, M. Barthelémy, and A. Vespignani, *Dynamical processes on complex networks*, Cambridge University Press, New York, 2008.
7. T. Bergstrom, L. Blume, and H. Varian, *On the private provision of public goods*, J. Public Econ. **29** (1986), no. 1, 25–49.
8. N. P. Bhatia and G. P. Szegö, *Dynamical systems: Stability theory and applications*, Lect. Notes Math., vol. 35, Springer-Verlag, Berlin, 1967.
9. B. Biais, P. Hillion, and C. Spatt, *An empirical analysis of the limit order book and the order flow in the Paris Bourse*, J. Finance **50** (1995), no. 5, 1655–1689.
10. D. Bouyssou and T. Marchant, *Additive conjoint measurement with ordered categories*, Eur. J. Oper. Res. **203** (2010), no. 1, 195–204.
11. L. Calvet and A. Fisher, *Multifractality in asset returns: theory and evidence*, Rev. Econ. Stat. **84** (2002), no. 3, 381–406.
12. G. Choquet, *Theory of capacities*, Ann. Inst. Fourier **5** (1954), 131–295.
13. G. de Rham, *On some curves defined by functional equations*, Classics on Fractals (G. A. Edgar, ed.), Westview Press, Boulder, CO, 2004, pp. 285–297.
14. G. Debreu, *Topological methods in cardinal utility theory*, Mathematical Methods in the Social Sciences (K. J. Arrow, S. Karlin, and P. Suppes, eds.), Stanford University Press, Stanford, CA, 1960, pp. 16–26.
15. H. Demsetz, *The cost of transacting*, Q. J. Econ. **82** (1968), no. 1, 33–53.
16. R. L. Devaney, *An introduction to chaotic dynamical systems*, Benjamin Cummings, Menlo Park, CA, 1986.
17. M. Ezekiel, *The cobweb theorem*, Q. J. Econ. **52** (1938), no. 2, 255–280.
18. J. D. Farmer, P. Patelli, and I. I. Zovko, *The predictive power of zero intelligence in financial markets*, Proc. Natl. Acad. Sci. USA **102** (2005), no. 6, 2254–2259.
19. P. C. Fishburn, *Utility theory for decision making*, John Wiley & Sons, New York, 1970.
20. ______, *Nontransitive additive conjoint measurement*, J. Math. Psychol. **35** (1991), no. 1, 1–40.
21. P. C. Fishburn and A. Rubinstein, *Time preference*, Int. Econ. Rev. **23** (1982), no. 3, 677–694.

22. T. Foucault, O. Kadan, and E. Kandel, *Limit order book as a market for liquidity*, Rev. Financ. Stud. **18** (2005), no. 4, 1171–1217.
23. M. Francesconi and A. Muthoo, *Control rights in complex partnership*, J. Eur. Econ. Assoc. **9** (2011), no. 3, 551–589.
24. A. Galeotti and S. Goyal, *The law of the few*, Am. Econ. Rev. **100** (2010), no. 4, 1468–1492.
25. L. R. Glosten and P. R. Milgrom, *Bid, ask and transaction prices in a specialist market with heterogeneously informed traders*, J. Financ. Econ. **14** (1985), no. 1, 71–100.
26. W. M. Gorman, *Separable utility and aggregation*, Econometrica **27** (1959), no. 3, 469–481.
27. T. Groves and M. Loeb, *Incentives and public inputs*, J. Public Econ. **4** (1975), no. 3, 211–226.
28. R. Hardin, *Collective action*, John Hopkins University Press, Baltimore, MA, 1982.
29. L. Harris and J. Hasbrouck, *Market vs. limit orders: the SuperDOT evidence on order submission strategy*, J. Financ. Quant. Anal. **31** (1996), no. 2, 213–231.
30. J. C. Harsanyi, *Cardinal welfare, individual ethics, and interpersonal comparisons of utility*, J. Polit. Econ. **63** (1955), no. 4, 309–321.
31. N. Kaldor, *A classificatory note on the determinateness of equilibrium*, Rev. Econ. Stud. **1** (1934), no. 2, 122–136.
32. E. Karni and Z. Safra, *The hexagon condition and additive representation for two dimensions: An algebraic approach*, J. Math. Psychol. **42** (1998), no. 4, 393–399.
33. W. Krabs and S. Pickl, *Dynamical systems: Stability, controllability and chaotic behavior*, Springer-Verlag, Berlin, 2010.
34. D. H. Krantz, R. D. Luce, P. Suppes, and A. Tversky, *Foundations of measurement*, vol. I: Additive and polynomial representations, Academic Press, New York, 1971.
35. A. S. Kyle, *Continuous auctions and insider trading*, Econometrica **53** (1985), no. 6, 1315–1335.
36. B. N. Lehmann and D. M. Modest, *Trading and liquidity on the Tokyo Stock Exchange: a bird's eye view*, J. Finance **49** (1994), no. 3, 951–984.
37. W. Leontief, *Delayed adjustment of supply and partial equilibrium*, J. Econ. **5** (1934), no. 5, 670–676.
38. T.-Y. Li and J. A. Yorke, *Period three implies chaos*, Amer. Math. Mon. **82** (1975), no. 10, 985–992.
39. R. D. Luce and J. W. Tukey, *Simultaneous conjoint measurement: a new type of fundamental measurement*, J. Math. Psychol. **1** (1964), no. 1, 1–27.
40. B. B. Mandelbrot, *The variation of certain speculative prices*, J. Bus. **36** (1963), no. 4, 394–419.
41. ———, *Les objets fractals: forme, hasard et dimension*, Flammarion, Paris, 1975.
42. ———, *The fractal geometry of nature*, W. H. Freeman and Company, New York, 1982.
43. B. B. Mandelbrot and J. W. van Ness, *Fractional Brownian motions, fractional noises and applications*, SIAM Rev. **10** (1968), no. 4, 422–437.
44. C. F. Manski, *Maximum score estimation of the stochastic utility model of choice*, J. Econom. **3** (1975), no. 3, 205–228.
45. J. Milnor, *On the concept of attractor*, Comm. Math. Phys. **99** (1985), no. 2, 177–195.
46. T. Mitra, L. Montrucchio, and F. Privileggi, *The nature of the steady state in models of optimal growth under uncertainty*, Econ. Theory **23** (2004), no. 1, 39–71.
47. T. Mitra and F. Privileggi, *On Lipschitz continuity of the iterated function system in a stochastic optimal growth model*, J. Math. Econ. **45** (2009), no. 1–2, 185–198.
48. W. E. Oates, *Fiscal federalism*, Harcourt Brace Jovanovich, New York, 1972.
49. ———, *An essay on fiscal federalism*, J. Econ. Lit. **37** (1999), no. 3, 1120–1149.
50. M. F. M. Osborne, *Brownian motion in the stock market*, Oper. Res. **7** (1959), no. 2, 145–173.

51. C. A. Parlour, *Price dynamics in limit order markets*, Rev. Financ. Stud. **11** (1998), no. 4, 789–816.
52. H. Poincaré, *Science and method*, (F. Maitland trans.), Thomas Nelson & Sons, London, 1914.
53. Y. Qian and B. R. Weingast, *Federalism as a commitment to preserving market incentives*, J. Econ. Perspect. **11** (1997), no. 4, 83–92.
54. F. S. Roberts, *Measurement theory: With applications to decisionmaking, utility, and the social sciences*, Addison-Wesley, Reading, MA, 1979.
55. D. Ruelle, *Strange attractors*, Math. Intell. **2** (1980), no. 3, 126–137.
56. D. Ruelle and F. Takens, *On the nature of turbulence*, Comm. Math. Phys. **20** (1971), no. 3, 167–192.
57. P. A. Samuelson, *The pure theory of public expenditure*, Rev. Econ. Stat. **36** (1954), no. 4, 387–389.
58. H. A. Simon, *A formal theory of the employment relationship*, Econometrica **19** (1951), no. 3, 293–305.
59. E. Smith, J. D. Farmer, L. Gillemot, and S. Krishnamurthy, *Statistical theory of the continuous double auction*, Quant. Finance **3** (2003), no. 6, 481–514.
60. D. Sornette, *Why stock markets crash: Critical events in complex financial systems*, Princeton University Press, Princeton, NJ, 2003.
61. G. Teschl, *Ordinary differential equations and dynamical systems*, Grad. Stud. Math., vol. 140, Amer. Math. Soc., Providence, RI, 2012.
62. G. Thomsen, *Un teoreme topologico sulle schiere di curve e una caratterizzazione geometrica delle superficie isotermo-asintotiche*, Boll. Un. Mat. Ital. Bologna **6** (1927), 80–85.
63. C. M. Tiebout, *A pure theory of local expenditure*, J. Polit. Econ. **64** (1956), no. 5, 416–424.
64. H. Uzawa, *On a two-sector model of economic growth*, Rev. Econ. Stud. **29** (1961), no. 1, 40–47.
65. P. P. Wakker, *Additive representations of preferences: A new foundation of decision analysis*, Kluwer Academic Publishers, Dordrecht, 1989.
66. S. Wang, *Dynamical trading mechanisms in limit order markets*, Algorithm. Finance **2** (2013), no. 3–4, 213–231.
67. B. R. Weingast, *The economic role of political institutions: market-preserving federalism and economic development*, J. Law Econ. Organ. **11** (1995), no. 1, 1–31.
68. E. Wibbels, *Decentralized governance, constitution formation, and redistribution*, Constitut. Polit. Econ. **16** (2005), no. 2, 161–188.

Index

Estratto per riassunto della tesi di dottorato

Studente	Shilei Wang
Matricola	955949
Dottorato	Economia
Ciclo	XXVII
Titolo della tesi	*Group Structure and Collective Behavior*

Abstract:

The present dissertation consists of four chapters, which are loosely connected by the abstract relation between group structure and collective behavior. Three main abstractions of group structure are proposed, *viz.*, time-dependent group which can be characterized by dynamical system, preference-dependent group that depends on aggregation of individual preferences, and relation-dependent group which is known as network.

In this respect, some interesting problems are investigated so as to demonstrate the effects of group structure on collective behavior. In detail, Chapter 1 studies such properties as dynamics, chaos, and stability in a typical time-dependent group, which also lays some solid foundations for the studies in the other three chapters. Chapter 2 investigates separable and additive aggregation rules in a preference-dependent group, with the aim of developing algebraic-geometric principles to efficiently identify additive aggregation rules. In particular, the Thomsen condition has been revisited and shown to be a both sufficient and necessary requirement for an aggregation rule being additive on a quite general two-dimensional domain.

Chapter 3 can be thought of as an application of time-dependent group in equity markets, in which quotation formation and dynamics are discussed in great detail. Some significant assertions on stability and stochastic stability are developed by means of market conditions. Chapter 4 relies on collective decision-making process on a government network to show how decisions are contingent on different network structures, which can be classified as H-form government line, M-form government star, and more general mixed-form network.

JEL Classification. C69, C79, D69, D79, G19

2000 Mathematics Subject Classification. 91B15, 91B16, 91B18, 91B55, 91G99

Firma dello studente

Shilei Wang