A General Equilibrium in Economies with Friction and its Relation to Walrasian Equilibria

Alexis Akira Toda *†

This Version: September 20, 2010

Abstract

Statistical equilibrium is a general equilibrium concept which maximizes entropy subject to the available information, and is maximally noncommittal with regard to missing information. Although conceptually distinct from Walrasian equilibrium, I show that 1) statistical equilibrium is "efficient" in a certain sense, 2) it exists under weak conditions, and that 3) Walrasian equilibrium is mathematically a special case of statistical equilibrium. Statistical equilibrium can be applied to analyze economies with not necessarily well-organized markets, such as the labor and the real estate markets.

1 Introduction

The theory of general equilibrium when markets are well-organized—for instance every agent knows the commodities traded in full detail and has a preference over them, acts as a price taker, and optimizes—is well-known. But in reality there are markets that are not necessarily well-organized, such as the labor and the real estate markets. Since each worker or house differs from the other, it is extremely difficult or even impossible to completely describe the characteristics of a "commodity" and quantify them; hiring a worker or purchasing a house usually incurs non-negligible search effort and transaction costs; labor and real estates lack liquidity, which makes the price adjustments slow; and so on. Although not as extreme as the labor and the real estate markets, most markets seem to depart more or less from the well-organized markets assumed in the traditional Walrasian equilibrium theory. How, then, can we mathematically formulate and analyze them?

One approach is to build special models that represent some aspects of certain real world market transaction, such as random matching or trading posts, and to specify the details and analyze mathematically. However, this approach

*Department of Economics, Yale University. Email: alexisakira.toda@yale.edu

†This paper benefited from conversations with (in alphabetical order) Simone Alfarano, Sylvain Barde, Dirk Bergemann, Truman F. Bewley, Donald Brown, Duncan K. Foley, John Geanakoplos, Sander Heinsalu, Johannes Horner, Mishael Milaković, Kohta Mori, as well as the comments from the coeditor (Edward J. Green), an anonymous referee, and seminar participants at Yale and the ESHIA 2010 conference at Alessandria, Italy. The financial supports from the Cowles Foundation, the Nakajima Foundation, and Yale University are greatly acknowledged.

has several drawbacks. First, any model assumption can be criticized as being *ad hoc*, and therefore we do not have an *a priori* reason to prefer one model to another. Second, since the model assumption is usually unverifiable from real world data, when rejecting a null hypothesis about model parameters, one does not know whether certain parameter ranges are rejected or the model structure itself is rejected. Third, the specification of the details complicates the model, and once the number of agents or commodities gets large, the model becomes analytically intractable.

Another approach, which I take in this article, is to be maximally parsimonious and to assume as little as possible. What we know for sure is that agents occasionally accept some transactions, while always rejecting others, and that the final allocation of commodities must be feasible. However, assuming only acceptability and feasibility seems to get us nowhere. The assumption that I make in order to lead us somewhere—preferably to a general equilibrium concept—is the *Maximum Entropy Principle* (MaxEnt) of Jaynes (1957), which has roots in Bernoulli, Laplace (1812), Boltzmann (1877), Gibbs (2008),[1] and Shannon (1948). Given a prior density $p(x)$, the *entropy*[2] of a density $q(x)$ is the quantity

$$H(q;p) := -\int q(x)\log\frac{q(x)}{p(x)}\mathrm{d}x, \tag{1.1}$$

and is a measure of uncertainty or lack of knowledge (Shannon, 1948). MaxEnt prescribes the posterior $q(x)$ that maximizes (1.1) subject to known constraints imposed by the available information, such as the support or the moments of $q(x)$. The abstract of Jaynes (1957) very clearly summarizes MaxEnt:

> Information theory provides a constructive criterion for setting up probability distributions on the basis of partial knowledge, and leads to a type of statistical inference which is called the maximum-entropy estimate. It is the least biased estimate possible on the given information; *i.e.*, it is maximally noncommittal with regard to missing information. [...] [W]hether or not the results agree with experiment, they still represent the best estimates that could have been made on the basis of the information available.

Although entropy is currently in heavy use in statistics and econometrics, it has only been sporadically introduced into economic theory. Theil (1967) used entropy as an inequality measure of income; Krebs (1997) employs MaxEnt to resolve the indeterminacy problem of expectational equilibria; Yoshikawa (2003) determines the size distribution of sectors assuming productivity dispersion; Castaldi and Milaković (2007) analyze the return and turnover activity in wealth portfolios among wealthiest individuals in U.S. and U.K.; Alfarano and Milaković (2008) deal with firm growth rates.

The present article is closest to the general equilibrium model of Foley (1994, 1996, 2003) and its extension (Toda, 2010), where equilibrium is defined by the maximum entropy distribution over transactions subject to the acceptability

[1]Interestingly, Gibbs was one of the advisers of Irving Fisher, the first Yale Ph.D. in economics.

[2]The quantity $H(q;p)$ (ignoring the minus sign) is better known as the relative entropy, the cross entropy, and the (Jeffreys-)Kullback-Leibler divergence, etc. Since the traditional Boltzmann-Shannon entropy is a special case of the relative entropy corresponding to the uniform prior, in this article I will simply call $H(q;p)$ entropy.

and the feasibility constraints. Following Foley (1994), I call this equilibrium concept a *statistical equilibrium*.[3] The trading process is a "black box" (Foley, 2003) except for that agents have subjective perceptions—priors—about their acceptable trade opportunities. As in Toda (2010), the priors may depend on the "scarcity" of each good. By maximizing the entropy of the distribution of transactions subject to the feasibility constraint, the resulting Lagrange multiplier (termed *entropy price* by Foley (1994)) has a natural interpretation of scarcity of commodities. In order to close the model, I impose in equilibrium that we are consistent—that the scarcity parameter and the entropy price vector are parallel. This last requirement is identical to the interpretation of a typical existence proof of Walrasian equilibrium, where the price correspondence of the "auctioneer" must have a fixed point. Summarizing, we obtain the axioms as in Table 1.

Table 1. Comparison of axioms of Walrasian and statistical equilibria.

Axioms	Walrasian Equilibrium	Statistical Equilibrium
Friction	none	arbitrary
Behavior	agent optimization	agent satisfaction
Feasibility	market clearing	market clearing
Consistency	price	scarcity
Other	-	entropy maximization

While it is crucial that markets are well-organized in Walrasian equilibrium, it is inessential in statistical equilibrium. In fact, I assume nothing on the market structure. Agent optimization is replaced by the weaker notion of agent satisfaction: each agent has a prior on trade opportunities conditional on acceptance, and if the agent happen to be optimizing, then the prior must concentrate on the optimal trade. Market clearing is the same and consistency is similar in the two equilibrium concepts, but I further impose entropy maximization. Therefore, conceptually statistical equilibrium is neither stronger nor weaker than the Walrasian equilibrium.

All of the above (except perhaps Table 1) have already been pointed out in Foley (1994, 1996, 2003) and Toda (2010). The value-added of this article is twofold. First, I review the MaxEnt literature in physics, information theory, statistics, and econometrics, and employing this knowledge I interpret and justify the statistical equilibrium concept. Second, I refine the definitions and the mathematical argument in Toda (2010) and prove the following theorems: 1) statistical equilibrium is efficient in a certain sense; 2) statistical equilibrium exists under very weak conditions; 3) Walrasian equilibrium is mathematically a special case of statistical equilibrium. 1) is repeatedly mentioned in Foley (1994, 1996, 2003), and 3) is expressed in Foley (2003) as "[T]here may be a sense in which Walrasian equilibrium can be viewed as an asymptotic approximation to statistical equilibrium", but both only through a heuristic argument.

[3]The term "statistical equilibrium" has been employed by different authors in various contexts, for instance Simon (1959), Hirshleifer (1973), Grossman and Stiglitz (1980), Krebs (1997), Silver et al. (2002), Castaldi and Milaković (2007), and Alfarano and Milaković (2008). Thus if we want to be specific about the present equilibrium concept, I am equally satisfied with calling it *maximum entropy equilibrium* (because I maximize entropy), *Bayesian general equilibrium* (because of the connection between MaxEnt and Bayesian inference mentioned in Section 2, but "general" to distinguish it from Bayesian Nash equilibrium in game theory), or the judgement-free *Foley equilibrium*, etc.

My results are the first rigorous mathematical formulations. With regard to 2), existence of equilibria has been established in my earlier paper (Toda, 2010), but I relied on the assumption that the support of agents' priors are large enough. Here I make minimal assumptions, which enables me to establish the connection between Walrasian and statistical equilibria.

2 Rationale of Maximum Entropy Principle

"Why use entropy; why maximize entropy; why not some other function?" is a common reaction to MaxEnt. Before proceeding to the model, I need to provide satisfactory answers to these questions in order to convince the reader, for statistical equilibrium theory cannot be appreciated without accepting MaxEnt. There are at least four justifications.

2.1 Maximum Flexibility with Unknowns

The first rationale is that the maximum entropy distribution is achieved in the greatest number of ways, or put in terms more familiar to economists, the maximum entropy distribution maximizes likelihood subject to the available information. Historically this is the rationale recognized earliest, which dates back at least to Boltzmann (1877) and Jaynes (1968).

Let me explain by a simple example. Suppose Adrian believes that certain data come from a multinomial distribution with M possible outcomes, with prior probability $(p_1, \dots, p_M)$. Suppose that outcome m occurred N_m times and $N = N_1 + \cdots + N_M$, but that Adrian does not know these numbers. She has some other information, but it is not specified here.[4] How should she construct her posterior $(q_1, \dots, q_M)$ based on her available information?

According to Adrian's prior, the likelihood of observing $(N_1, \dots, N_M)$ is

$$\mathcal{L} = \frac{N!}{N_1! \cdots N_M!} p_1^{N_1} \cdots p_M^{N_M}. \tag{2.1}$$

If she were to update her prior p to a posterior q, it is natural to assume that N_m is approximately equal to Nq_m, at least for large enough N. Invoking the approximation $x! \approx x^x$, it follows from (2.1) that

$$\begin{aligned}
\log \mathcal{L} &\approx \log N! - \sum_{m=1}^{M} \log(Nq_m)! + \sum_{m=1}^{M} Nq_m \log p_m \\
&\approx N \log N - \sum_{m=1}^{M} Nq_m \log(Nq_m) + \sum_{m=1}^{M} Nq_m \log p_m \\
&= -N \sum_{m=1}^{M} q_m \log \frac{q_m}{p_m} = NH(q; p),
\end{aligned}$$

the (relative) entropy multiplied by a constant. Hence, by maximum likelihood, Adrian should maximize $H(q;p)$ subject to the available information.[5]

[4] Jaynes (1982) gives the example: a loaded dice was tossed 1,000 times and the mean value was 4.5, not 3.5 as we would expect from a fair dice. How do you estimate probabilities?

[5] Of course, the maximum likelihood principle (ML) and MaxEnt differ in that ML maximizes $H(q;p)$ over the *prior* p, whereas MaxEnt does over the *posterior* q. See also Akaike (1992).

This reasoning is expressed as "maximally noncommittal with regard to missing information" (Jaynes, 1957, p. 623), and "it agrees with what is known, but expresses a 'maximum uncertainty' with respect to all other matters, and thus leaves a maximum possible freedom for our final decisions to be influenced by the subsequent data" (Jaynes, 1968, p. 231). Also Jaynes (1982) shows that in the above setting, if H_{null} and $H_{\max}$ denote the entropy of the null hypothesis distribution and the maximum entropy distribution, and $\Delta H := H_{\max} - H_{\text{null}}$, then as $N \to \infty$, $2N\Delta H$ is asymptotically chi-square distributed with $M-C-1$ degrees of freedom, where M is the number of possible outcomes and C is the number of constraints. Thus, as the sample size gets large, one is overwhelmingly likely to reject distributions that do not maximize entropy.

2.2 Connection with Bayesian Inference

The second rationale is the close connection between MaxEnt and Bayesian inference.[6] Van Campenhout and Cover (1981) showed that Bayes's theorem implies MaxEnt in the following sense: the conditional distribution of a random variable X_i given the empirical observation $\frac{1}{n}\sum_{i=1}^{n} h(X_i) = \alpha$, where X_i's are i.i.d. with prior density g, converges to $f_\lambda(x) = \mathrm{e}^{\lambda' h(x)} g(x)$ (suitably normalized), where λ is chosen to satisfy the population moment constraint $\int h(x) f_\lambda(x)\mathrm{d}x = \alpha$. Note that this $f_\lambda(x)$ is the solution to

$$\max_f [-H(f;g)] \text{ subject to } \int h(x)f(x)\mathrm{d}x = \alpha,$$

i.e., the maximum entropy problem.[7] Conversely, Zellner (1988) showed that MaxEnt leads to Bayes's theorem. Thus, roughly speaking, MaxEnt and Bayesian inference are equivalent, and hence MaxEnt should be as well-received as Bayesian inference.

2.3 Axiomatic Approach

While the first two rationales are indirect (pointing out the connection between MaxEnt and maximum likelihood or Bayesian inference), the third one, which I believe the most important, is axiomatic and direct. Shore and Johnson (1980) proposed four axioms of inference: *Uniqueness* (the result should be unique); *Invariance* (the choice of coordinate system should not matter); *System Independence* (it should not matter whether one accounts for independent information about independent systems separately in terms of different densities or together in terms of a joint density); and *Subset Independence* (it should not matter whether one treats an independent subset of system states in terms of a separate conditional density or in terms of the full system density). Under these axioms, Shore and Johnson (1980) proved that maximizing (a monotone transformation of) entropy is the *unique correct way* of inference. Thus if one asks "why maximize entropy; why not some other function?", one has to provide reasons why he rejects the axioms of Shore and Johnson.

[6] I thank Sylvain Barde for bringing this point to my attention.

[7] The proof of a more general problem can be found in Borwein and Lewis (1991, 1992).

2.4 It Works

The fourth rationale is pragmatic. MaxEnt has been successfully applied for more than one century. Although MaxEnt was first proposed by Boltzmann, Gibbs, and Jaynes, who are physicists, Jaynes (1957)[8] recognized that it can be applied to any inference problem. In fact, MaxEnt has been applied to such diverse fields as statistical mechanics, thermodynamics, statistics, reliability estimation, traffic networks, queuing theory and computer system modeling, system simulation, production line decision making, group behavior, and stock market analysis. (See Shore and Johnson (1980) and the references therein.) Given the success of MaxEnt in a remarkable variety of fields, I see little reason not to introduce it to economic theory.

3 Definitions

In this section, I define the basic concepts in statistical equilibrium theory. The definition of a *statistical economy* is formally the same as in Toda (2010) but I give a new interpretation. I slightly modify the equilibrium concept proposed in Toda (2010) to derive a general theory in later sections.

3.1 Economy

Given agents, Foley (1994) and Toda (2010) start by defining *offer sets* of agents, which are interpreted as the sets consisting of all technologically feasible and *ex ante* acceptable transactions. In Foley (1994), offer sets are finite sets and he puts a weight on each point; Toda (2010) considers more general sets that can be turned into regular Borel measure spaces,[9] referred to as *offer spaces*.[10] This direction of definition has a shortcoming, for we do not know how to choose the measures from preferences.

To circumvent this ambiguity, I reverse the direction of the definition: for each agent type I prescribe a regular Borel measure (offer space), which I interpret as a prior over transaction opportunities for that agent type, taking *ex ante* acceptability into account. Then I define its support[11] to be the offer set of that type. Formally, I define as follows.

Definition 3.1 (statistical economy)**.** The object

$$\mathcal{E} = \left\{ \mathcal{I}, \{ w_i \}_{i \in \mathcal{I}}, \{ \mu_{i,p} \}_{i \in \mathcal{I}, p \in \Delta^{C-1}} \right\}$$

is called a *statistical economy* if

[8]The first two sections of Jaynes's paper are particularly illuminating and accessible to nonspecialists, which I highly recommend.

[9]See Folland (1999) for the concepts in measure theory that appear in this paper.

[10]Note that the formulation of Foley (1994) is a special case of Toda (2010) since weights on a finite set can be regarded as a counting measure, which is regular Borel.

[11]The support of a regular Borel measure μ on a second-countable topological space X is defined as follows. Let $\mathcal{U}$ be a countable base of the topology (for the case of $X = \mathbb{R}^C$, it suffices to take the family of all open balls with rational radii and centers with rational coordinates) and $S = X \setminus \bigcup_{U \in \mathcal{U} : \mu(U) = 0} U$. Since $\mathcal{U}$ is the countable base of the topology, it follows that S is closed, $\mu(S) = \mu(X)$, and $\mu(U \cap S) > 0$ whenever U is open and $U \cap S \neq \emptyset$. S is called the *support* of μ and is denoted by $S = \operatorname{supp} \mu$.

- $\mathcal{I} = \{ 1, 2, \dots, I \}$ is the set of agent types,
- w_i is the proportion of type i agents, so $w_i > 0$ and $\sum_{i=1}^{I} w_i = 1$,
- $\Delta^{C-1} = \left\{ p \in \mathbb{R}_+^C \mid \|p\|_1 = 1 \right\}$[12] is the set of the scarcity parameter p,[13] where C denotes the number of commodities,
- Given $p \in \Delta^{C-1}$, $\mu_{i,p}$ is a regular Borel measure of type i agents over their trade opportunities $x \in \mathbb{R}^C$. *Offer set* $X_{i,p}$ is the support of $\mu_{i,p}$, and the measure space $(X_{i,p}, \mu_{i,p})$ is called the *offer space.*

We can interpret $\mu_{i,p}$ as type i agents' subjective probability measure (prior) over the *ex ante* acceptable trade opportunities $x \in \mathbb{R}^C$, when the agents take the scarcity parameter p as given. The c-th coordinate of the scarcity parameter, p_c, is large if commodity c is relatively scarce. If $\mu_{i,p}$ is an infinite measure, we interpret it as an improper prior. Simply put, the primitives of a statistical economy are agents and their beliefs over the acceptable trade opportunities.

Notice the flexibility of our definition: since we have put no structure whatsoever on $\mu_{i,p}$, Definition 3.1 can be used to describe many different situations. For instance, the economy might be a pure exchange economy or a production economy; agents might be consumers, producers, arbitrageurs, or the government, etc; agents might be price takers or might have market power; the market can be well-functioning or can have a lot of friction; agents might or might not be rational, and so on.

We interpret that the measures $\{ \mu_{i,p} \}$ are the complete description of agents' prior information over the transactions, for instance the degree of perceived uncertainty or liquidity, technological feasibility, preference, likelihood of accepting a transaction that has been offered, and so on. Following Toda (2010), I do not answer to the question of how offer spaces are determined: anything goes as far as they are regular Borel measures. This attitude of not modeling how priors are formed is similar to the textbook treatment of general equilibrium in microeconomics (Mas-Collel et al., 1995). In textbooks, it is typical to treat agent (consumer and firm) behavior and general equilibrium in different chapters. One reason is because existence of equilibrium can be proved without specifying individual behavior, as far as the aggregate excess demand has certain properties (homogeneity of degree zero, Walras's law, continuity, and some boundary regularity condition). The same holds in statistical equilibrium: its existence is independent from modeling $\{ \mu_{i,p} \}$. Thus, the modeling of priors belongs to a different branch of economics, for instance decision theory.

3.2 Equilibrium

As the interpretation of the economy is modified, so is the definition of the equilibrium. Before mentioning the new definition, I remark that the *entropy* of a collection of probability density functions $f = (f_i)_{i \in \mathcal{I}}$ (relative to the priors $\mu_p = \{ \mu_{i,p} \}_{i \in \mathcal{I}}$, so the posterior is determined by $\mathrm{d}q_i = f_i \mathrm{d}\mu_{i,p}$) and the

[12] Hereafter, $\|\cdot\|_p$ will refer to the L^p norm.

[13] This set is often referred to as the "price simplex" in Walrasian equilibrium theory, but p may or may not be the price vector. For details, see Toda (2010).

associated *average transaction* are defined by

$$H[f;\mu_p] := -\sum_{i=1}^{I} w_i \int f_i \log f_i \mathrm{d}\mu_{i,p}, \tag{3.1}$$

$$\bar{x}[f;\mu_p] := \sum_{i=1}^{I} w_i \int x f_i \mathrm{d}\mu_{i,p}, \tag{3.2}$$

respectively.[14] (3.1) can be easily understood as follows. In (1.1), replace $q(x)$ by $f(x)p(x)$. Then we get $-\int [f(x)\log f(x)]p(x)\mathrm{d}x$. Note that the prior need not be absolutely continuous with respect to the Lebesgue measure, so replace $p(x)\mathrm{d}x$ by $\mathrm{d}\mu$, resulting in $-\int f \log f \mathrm{d}\mu$. Now suppose that there are n_i agents belonging to type i, and let the total number of agents be $n = \sum_i n_i$ and the proportion be $w_i = n_i/n$. In general, if two random variables X, Y are independent, we have $H(X,Y) = H(X) + H(Y)$.[15] Therefore, the economy-wide entropy is $H = -\sum_{i=1}^{I} n_i \int f_i \log f_i \mathrm{d}\mu_{i,p}$. Dividing this expression by n, we obtain the per capita entropy (3.1).[16] The rationale of (3.2) is the law of large numbers: if each type has either a large number of agents acting independently or a prior that concentrates on a single point (*i.e.*, Dirac measure), then the economy-wide per capita average transaction converges to (3.2) almost surely. Thus having a continuum of independent agents (within type) and/or prior concentration is an assumption.

Next, I give the new definition of statistical equilibrium. Here, I classify statistical equilibria into two categories: *genuine* and *degenerate* equilibria. Genuine equilibria are precisely what Toda (2010) defined to be statistical equilibria. Degenerate equilibria are intuitively the asymptotic limit of genuine ones and correspond to Walrasian equilibria when the theory is applied to standard Walrasian economies. In what follows I fix the statistical economy

$$\mathcal{E} = \left\{ \mathcal{I}, \{ w_i \}_{i\in\mathcal{I}}, \{ \mu_{i,p} \}_{i\in\mathcal{I}, p\in\Delta^{C-1}} \right\}$$

with offer sets $X_{i,p} = \operatorname{supp} \mu_{i,p}$.

Definition 3.2 (genuine statistical equilibrium)**.** A collection of densities $f = (f_i)_{i\in\mathcal{I}}$ over offer sets and a pair of vectors $(p,\pi) \in \Delta^{C-1} \times \mathbb{R}^C_+$ are called a *genuine statistical equilibrium* if

1. $(f_i)_{i\in\mathcal{I}}$ solves the maximum entropy program (MEP) associated with the scarcity parameter p, *i.e.*, $f = (f_i)_{i\in\mathcal{I}}$ solves
$$\max_f H[f;\mu_p] \text{ subject to } \bar{x}[f;\mu_p] \leq 0, \tag{3.3}$$
2. π is the *entropy price* for the MEP, *i.e.*, π is the Lagrange multiplier to (3.3),[17]

---

[14] Unless otherwise specified, the integral sign always means the integration over the whole space.

[15] See, for example, the introductory textbook of Cover and Thomas (2006).

[16] An implicit assumption here is that agents act independently conditional on p, but even if we consider the joint distribution of agents (with each agent type being not necessarily independent), after maximizing entropy we arrive at the same conclusion because MaxEnt prescribes independence without prior information violating it.

[17] Since (3.3) is an optimization problem in some functional space, it is not obvious that the standard Karush-Kuhn-Tucker theorem applies, but it does: see Luenberger (1969) and Borwein and Lewis (1991, 1992).

3. π and p are collinear.[18]

Genuine equilibria are precisely what Toda (2010) defined to be statistical equilibria. To justify it, I need to answer three questions: 1) why maximize entropy?; 2) why maximize entropy subject to the market clearing condition?; 3) why $\pi \parallel p$? The answer to question 1) is Section 2. The answer to question 2 is because market clearing is the only information we have without assuming others, and MaxEnt tells us to use the available information: nothing more, nothing less. If we happen to know something else, then of course we should incorporate that information into the maximization (3.3). This point was obviously recognized by Jaynes and also by Foley (1996) (p. 128, parenthesis by author):

> As is true of all applications of statistical equilibrium (*i.e.*, maximum entropy) reasoning, [...] the predictive relevance of statistical equilibrium is contingent on the statistical model incorporating all the constraints that produce observable regularities in the system under analysis.

Finally, the answer to question 3 lies in our custom in economics to regard equilibrium as a fixed point. Walrasian and Nash equilibria are states in which nobody has an incentive to deviate: the best response is a fixed point. In statistical equilibrium, the entropy price π is a Lagrange multiplier and hence a shadow price of entropy. Thus, $1/\pi_c$ denotes the amount of good c to be taken away from the system in order to reduce the entropy by one unit, which has an interpretation of abundance. Consequently, π is scarcity, and $\pi \parallel p$ is a state in which there is no need to update the relative scarcity—an equilibrium.

The factor $t \in [0, \infty)$ such that $\pi = tp$, which is equal to $t = \|\pi\|_1$, has the natural interpretation of *market tightness* because it is a measure of how strong the market clearing constraint is binding.

Definition 3.3 (degenerate statistical equilibrium)**.** A scarcity parameter p and a collection of points $\{ x_i \}_{i \in \mathcal{I}}$, where $x_i \in \operatorname{cl} \operatorname{co} X_{i,p}$[19] for all i, are called a *degenerate statistical equilibrium* if

4. $\sum_{i=1}^{I} w_i x_i \leq 0$, and $p_c = 0$ if $\sum_{i=1}^{I} w_i x_{ic} < 0$,
5. for all $i \in \mathcal{I}$ and $x \in X_{i,p}$, we have $p'x \geq 0$.

One rationale of the degenerate equilibrium is that market clearing (condition 4) is a minimal requirement, and that we can increase entropy if condition 5 fails by "spreading" densities, contradicting MaxEnt. Another rationale is the analogy from Walrasian equilibrium theory. If p denotes the price system and $X_{i,p}$ is the set of transactions that are at least as desirable as the Walrasian excess demand, then conditions 4 and 5 precisely define the Walrasian equilibrium.[20]

[18] Here we allow the possibility of $\pi = 0$ (thus $T = \infty$). This can happen when all goods are "bads." For instance, if the offer sets of all agents consist of single points in $-\mathbb{R}^C_{++}$, the solution of (3.3) is trivially autarkic and the entropy price (Lagrange multiplier) π is 0.

[19] $\operatorname{cl} A$ and $\operatorname{co} A$ denote the closure and the convex hull of A respectively.

[20] To see this, condition 4 implies that the market clears and the price of a commodity in excess supply is zero; condition 5 implies that agents are optimizing.

In degenerate equilibria, I define the market tightness by $t = \infty$ because the market clearing constraint does not allow to spread the density and hence it is maximally binding.

A genuine or a degenerate equilibrium is simply called a *statistical equilibrium.* A natural question arising from Definitions 3.2 and 3.3 is whether the two equilibrium concepts—genuine and degenerate equilibria—are mutually exclusive. In general the answer is negative. However, if at least one $\mu_{i,p}$ is absolutely continuous with respect to the Lebesgue measure (which is almost always the case in application), then the two concepts are mutually exclusive.[21]

4 Main Results

The main results are the answers to the following questions: 1) is statistical equilibrium "efficient" in some sense?; 2) does it exist?; 3) what is its relation to Walrasian equilibrium?

4.1 Welfare Theorem

Since $\mu_{i,p}$ is type i agents' prior over transaction opportunities, given a posterior q_i defined by $\mathrm{d}q_i = f_i \mathrm{d}\mu_{i,p}$, the quantity $\int p'xf_i \mathrm{d}\mu_{i,p}$ is the expected *ex post* value of transactions. The larger this value is, the more agents are likely to gain from arbitrage. Therefore, its economy-wide average,

$$A[f;\mu_p] := \sum_{i=1}^{I} w_i \int p'xf_i \mathrm{d}\mu_{i,p} = p'\bar{x}[f;\mu_p],$$

can be interpreted as the degree of arbitrage or that of market inefficiency. The following "welfare theorem" shows that there is a trade-off between efficiency and informational requirement: to achieve higher efficiency (lower $A[f;\mu_p]$), one has to reduce entropy.

Theorem 4.1 (Welfare Theorem). *Let* $\mathcal{E} = \left\{ \mathcal{I}, \{ w_i \}_{i\in\mathcal{I}}, \{ \mu_{i,p} \}_{i\in\mathcal{I}, p\in\Delta^{C-1}} \right\}$ *be a statistical economy with offer sets* $X_{i,p} = \operatorname{supp}\mu_{i,p}$. *Suppose that*

$$\left(\sum_{i=1}^{I} w_i \operatorname{co} X_{i,p} \right) \cap (-\mathbb{R}^C_{++}) \neq \emptyset. \tag{4.1}$$

Then a feasible and ex ante *acceptable allocation* $f = (f_i)_{i\in\mathcal{I}}$ *is a genuine statistical equilibrium distribution if and only if it minimizes the functional* $-H[g;\mu_p] + tA[g;\mu_p]$ *over unconstrained* g *for some* $t \geq 0$.

Proof. (4.1) is a sufficient condition for Slater-type constraint qualification to be satisfied.[22] By the Kuhn-Tucker theorem (Theorem C.1) and its converse (Toda, 2010, Proposition A.2), letting $\pi = tp$, by Definition 3.2 f minimizes

$$-H[g;\mu_p] + tA[g;\mu_p] = -H[g;\mu_p] + \pi'\bar{x}[g;\mu_p]$$

[21] To see this, if a statistical equilibrium is genuine as well as degenerate, then by conditions 1 and 5, we must have $p'x = 0$ almost surely with respect to the probability measure induced by the equilibrium distributions. However, this is impossible since the hyperplane $p'x = 0$ has Lebesgue measure 0.

[22] See Lemma 3.2 of Toda (2010)

over unconstrained g if and only if it is a genuine statistical equilibrium distribution. □

Thus, despite statistical equilibrium is Pareto inefficient (Foley, 1994, 1996, 2003), it is approximately efficient.

4.2 Existence of Statistical Equilibrium

The main theorems of this paper are Theorems 4.2 and 4.3 below. I first prove Theorem 4.2, which shows the existence of statistical equilibria when all measures $\{ \mu_{i,p} \}$ are finite: assuming finiteness greatly simplifies the argument. In practice, we typically wish to deal with infinite measures such as the Lebesgue measure. By assuming that the measures grow at most exponentially, Theorem 4.3 guarantees the existence of statistical equilibria in such cases.

I make the following assumptions. Each assumption is followed by a justification.

Assumption 1. *The measure $\mu_{i,p}$ is finite for all $i \in \mathcal{I}$ and $p \in \Delta^{C-1}$.*

Since $\mu_{i,p}$ is interpreted as a prior (subjective probability measure), Assumption 1 is not restrictive. This assumption will be relaxed when dealing with improper prior in Theorem 4.3.

Assumption 2 (Boundedness from below)**.** *The offer set $X_{i,p}$ is uniformly bounded below,* i.e., *there exists $a \in \mathbb{R}^C$ such that $x \geq a$ for all $i \in \mathcal{I}$, $p \in \Delta^{C-1}$, and $x \in X_{i,p}$.*

Assumption 2 means that agents' net transactions are bounded from below. This is a constrained form of free disposal: agents are able to throw away undesired commodities, but only up to a certain finite amount. It may also be interpreted as limited arbitrage, since agents cannot go arbitrarily short in one position. This assumption is not restrictive in practice at all, for in the world there is only a finite amount of everything and hence there is no reason to think that agents wish to dispose of commodities beyond that limit.

Assumption 3 (Realistic agents)**.** *For all $i \in \mathcal{I}$ and $p \in \Delta^{C-1}$, we have* $\inf \{ p'x \mid x \in X_{i,p} \} \leq 0$.

Assumption 3 implies that agents are realistic in the sense that they essentially put a positive subjective probability on the "budget-feasible transactions" $\left\{ x \in \mathbb{R}^C \,\middle|\, p'x \leq 0 \right\}$ in Walrasian environment. This assumption is crucial, for the market can never clear if all agents perceive trade opportunities only for tremendously large amounts.

Next come two continuity assumptions, which seems plausible.

Assumption 4 (Continuity of measures)**.** *The mapping $p \mapsto \mu_{i,p}$ is weakly continuous,*[23] i.e., *for every sequence $\{ p_n \}$ such that $p_n \to p$ and bounded continuous function f, we have*

$$\lim_{n\to\infty} \int f \mathrm{d}\mu_{i,p_n} = \int f \mathrm{d}\mu_{i,p}.$$

[23] Assumption 4 is satisfied, for instance, if $\{ \mu_{i,p} \}$ is absolutely continuous with respect to a common measure μ_i, the Radon-Nikodym derivative $f_i(x,p) := \frac{\mathrm{d}\mu_{i,p}}{\mathrm{d}\mu_i}$ is continuous in p for μ_i-a.e. x, and there exists a μ_i-integrable function g_i such that $|f_i(x,p)| \leq g_i(x)$ for μ_i-a.e. x. To see this, apply Lebesgue's convergence theorem.

Assumption 5 (Continuity of offer sets)**.** *The correspondence* $p \mapsto \prod_{i \in \mathcal{I}} \operatorname{cl} \operatorname{co} X_{i,p}$ *is closed at those points such that* $\sum_{i=1}^{I} w_i \inf \{ p'x \mid x \in X_{i,p} \} = 0$, *i.e.,* $p_n \to p$, $x_i^n \in \operatorname{cl} \operatorname{co} X_{i,p_n}$, *and* $x_i^n \to x_i^\infty$ *implies* $x_i^\infty \in \operatorname{cl} \operatorname{co} X_{i,p}$ *for all* $i \in \mathcal{I}$ *whenever* $\sum_{i=1}^{I} w_i \inf \{ p'x \mid x \in X_{i,p} \} = 0$.

Note that Assumption 5 is automatically satisfied if at least one inequality in Assumption 3 is strict, in which case we have $\sum_{i=1}^{I} w_i \inf \{ p'x \mid x \in X_{i,p} \} < 0$ for all p. Also if $X_{i,p}$ is convex, then $\operatorname{cl} \operatorname{co} X_{i,p} = X_{i,p}$ since $X_{i,p}$ is closed by construction.

Theorem 4.2. *Let* $\mathcal{E} = \left\{ \mathcal{I}, \{ w_i \}_{i \in \mathcal{I}}, \{ \mu_{i,p} \}_{i \in \mathcal{I}, p \in \Delta^{C-1}} \right\}$ *be a statistical economy with offer sets* $X_{i,p} = \operatorname{supp} \mu_{i,p}$ *that satisfies Assumptions 1–5. Then* $\mathcal{E}$ *has a statistical equilibrium. If Assumption 3 is replaced by*

$$\sum_{i=1}^{I} w_i \inf \{ p'x \mid x \in X_{i,p} \} < 0,$$

in which case Assumption 5 is vacuous, all statistical equilibria are genuine.

Proof. See Appendix B.1. □

Theorem 4.3 below guarantees the existence of statistical equilibria when at least one of the measures $\{ \mu_{i,p} \}$ is infinite.

Theorem 4.3. *Let everything be as in Theorem 4.2. For any measure* μ *and* $\epsilon \in \mathbb{R}_+^C$, *define the measure* μ^ϵ *by* $\mu^\epsilon(\mathrm{d}x) = \mathrm{e}^{-\epsilon' x} \mu(\mathrm{d}x)$. *Suppose that*

1. *for all* $\epsilon \gg 0$, *the economy* $\mathcal{E}^\epsilon := \left\{ \mathcal{I}, \{ w_i \}, \{ \mu_{i,p}^\epsilon \} \right\}$ *satisfies Assumptions 1–5,*
2. *for all* $i \in \mathcal{I}$, *there exists a partition* $L_i \cup N_i = \{ 1, \dots, C \}$, $L_i \cap N_i = \emptyset$ *and measures* $\lambda_{i,p}, \nu_{i,p}$ *such that*
   - *(a)* $\lambda_{i,p}$ *is supported on a closed subset of* $\mathbb{R}^{L_i}$; $\nu_{i,p}$ *is supported on a closed subset of* $\mathbb{R}^{N_i}$,
   - *(b)* $\lambda_{i,p}, \nu_{i,p}$ *are weakly continuous with respect to* p,
   - *(c)* $\lambda_{i,p}$ *is a finite measure,*
   - *(d) for all* $\epsilon \geq 0$ *such that* $\epsilon_c = 0$ *for some* $c \in \bigcup_{i \in \mathcal{I}} N_i$, *there exists an* $i \in \mathcal{I}$ *such that* $\nu_{i,p}^\epsilon$ *is an infinite measure.*

Then the conclusion of Theorem 4.2 holds.

Proof. See Appendix B.2. □

Since I have proved two existence theorems in this paper and Toda (2010) proved two others, it is worth mentioning the relationship between the four existence theorems. Theorem 4.2 is a special case of Theorem 4.3 by setting $N_i = \emptyset$ for all i. Theorem 4.3 is not implied by the two in Toda (2010) because the latter two assume $\left(\sum_{i=1}^{I} w_i \operatorname{co} X_{i,p} \right) \cap (-\mathbb{R}_{++}^C) \neq \emptyset$, which the former does not. Theorem 4.3 does not imply the two in Toda (2010) because the offer sets are uniformly bounded below in Theorem 4.3 but not necessarily so in Toda (2010). Finally, the two theorems in Toda (2010) are clearly independent.

Therefore, we have essentially three different existence theorems of statistical equilibrium and thus the reader is free to choose which one to apply to specific models. However, in my view Theorem 4.3 is the most useful because the assumptions are weak as well as economically intuitive.

4.3 Relationship between Walrasian and Statistical Equilibria

Now we are ready to answer the deep question: *how are Walrasian and statistical equilibria related?* Although Corollary 4.4 below is almost trivial by Theorem 4.3, it has a strong philosophical implication: *Walrasian equilibrium theory is contained in statistical equilibrium theory.*

Corollary 4.4. *A Walrasian equilibrium is a statistical equilibrium. More precisely, let $\mathcal{E} = \{ \mathcal{I}, \{ u_i \}, \{ e_i \} \}$ be an endowment economy, where $u_i : \mathbb{R}^C_+ \to \mathbb{R}$ is a continuous, weakly monotonic utility function of type i agents (with proportion $w_i > 0$), and the endowments satisfy $e_i \gg 0$ for all i. Then,*

1. *there exists a statistical economy $\mathcal{E}'$ such that all Walrasian equilibria of $\mathcal{E}$ are statistical equilibria of $\mathcal{E}'$,*

2. *the existence of Walrasian equilibria can be shown by using statistical equilibrium theory.*

Proof. Take $b > 0$ such that $\sum_{i=1}^I e_i \leq b\mathbf{1}$. Let $X_b = [0, b]^C$. Define the constrained indirect utility function $v_i^b(p)$ by

$$v_i^b(p) = \max \{ u_i(x) \mid x \in X_b, p'x \leq p'e_i \} .$$

Let $X_{i,p} = \left\{ y \mid u_i(y) \geq v_i^b(p) \right\} - e_i$ be the set of transactions that make type i agents at least as well off as the best consumption bundle under the b-constrained budget set. Obviously, $X_{i,p}$ is closed, bounded below by $-b\mathbf{1}$, and the correspondence $p \mapsto X_{i,p}$ is closed. Let μ be the Lebesgue measure on $\mathbb{R}^C$ and define $\mu_{i,p}$ by $\mathrm{d}\mu_{i,p} = \chi_{X_{i,p}} \mathrm{d}\mu$,[24] so $\mu_{i,p}$ is the restriction of the Lebesgue measure on $X_{i,p}$. Then, $\mathcal{E}' = \{ \mathcal{I}, \{ w_i \}, \{ \mu_{i,p} \} \}$ is a statistical economy.

Now, any Walrasian equilibrium of $\mathcal{E}$ is a degenerate statistical equilibrium of $\mathcal{E}'$ by Definition 3.2, for in Walrasian equilibrium the price of the commodity in excess supply must be zero, and by local non-satiation agents must spend all their income. Thus, the first part of Corollary 4.4 is shown.

To show the second part, it suffices to verify the assumptions of Theorem 4.3. For any $\epsilon \gg 0$, consider the measure $\mu_{i,p}^\epsilon$ as in Theorem 4.3. Assumptions 1, 2 hold because $X_{i,p} \subset -b\mathbf{1} + \mathbb{R}^C_+$ and $\epsilon \gg 0$. Assumption 3 is satisfied because the value of the b-constrained Walrasian demand is at most the budget. Since u_i is weakly monotonic, the indifference curve has measure zero, so $\chi_{X_{i,p}}$ is continuous in p for μ-a.e. x, thus assumption 4 is satisfied. (See footnote 23.) Assumption 5 is satisfied since the correspondence $p \mapsto X_{i,p}$ is closed and $X_{i,p}$ is compact.[25]

Since all assumptions of Theorem 4.2 are satisfied, there exists a statistical equilibrium. Independent of whether the equilibrium is degenerate or not, the

[24] χ_A denotes the characteristic function of A.

[25] If the correspondence $x \mapsto \Phi(x)$ is closed and $\Phi(x)$ is compact, then $x \mapsto \mathrm{co}\,\Phi(x)$ is closed. See Berge (1959) for a proof.

aggregate average transaction is nonpositive and the average transaction of type i agents belongs to $\operatorname{co} X_{i,p}$ because $X_{i,p}$ is compact. These points clearly define a Walrasian equilibrium. □

An alternative proof of Corollary 4.4 is to introduce atomic agents whose offer sets are the Walrasian excess demand functions (assuming they are unique) and apply Corollary 5.1. This proof is essentially identical to the standard ones as in Debreu (1987).

As is obvious by Corollary 4.4, we made no convexity assumptions in establishing the existence of statistical equilibria. Therefore, in proving the existence of Walrasian equilibria, we may completely dispose of convexity provided that the number of agents within each type is sufficiently large: in fact, by Carathéodory's theorem, it suffices to have at least $C+1$ agents in each type.

This result neither implies nor is implied by the result of Aumann (1966). My result rests on the assumption of a finite number of types, which Aumann (1966) does not assume; on the other hand, we do not need the assumption of preference saturation that Aumann (1966) requires.

5 Computation of Statistical Equilibrium

Since the exact evaluation of the integral $\int \mathrm{e}^{-\xi' x} \mathrm{d}\mu_i$ is impossible unless the offer set $X_i = \operatorname{supp} \mu_i$ has a simple structure such as a translation of the positive orthant $\mathbb{R}^C_+$ (or more generally, a polyhedron), Toda (2010) discusses a numerical algorithm to compute statistical equilibria.

Here I provide a result that is general enough to analyze many economic situations, yet special enough to obtain the equilibrium in an almost closed-form solution. In this case the offer spaces are simplest, namely the product measure of the counting measure and the Lebesgue measure restricted to translations of the positive orthant.

Corollary 5.1. *Let $\mathcal{E} = \{ \mathcal{I}, \{ w_i \}, \{ \mu_{i,p} \} \}$ be a statistical economy and $L_i \cup N_i = \{ 1, \dots, C \}$, $L_i \cap N_i = \emptyset$ be a partition of commodities. Suppose that for all $i \in \mathcal{I}$ and $p \in \Delta^{C-1}$, there exists a point $x_{i,p} = (x^l_{i,p}, x^n_{i,p}) \in \mathbb{R}^{L_i} \times \mathbb{R}^{N_i} = \mathbb{R}^C$ with the following properties:*

1. *the offer space $(X_{i,p}, \mu_{i,p})$ has the form $X_{i,p} = \{ x^l_{i,p} \} \times (x^n_{i,p} + \mathbb{R}^{N_i}_+)$ and $\mu_{i,p} = \lambda_{i,p} \times \nu_{i,p}$, where $\lambda_{i,p}$ is the counting measure on the point $x^l_{i,p}$ and $\nu_{i,p}$ is the restriction of the Lebesgue measure on $x^n_{i,p} + \mathbb{R}^{N_i}_+$,*
2. *for all $i \in \mathcal{I}$ and $p \in \Delta^{C-1}$, we have $p' x_{i,p} \leq 0$.*
3. *for all $i \in \mathcal{I}$, the mapping $p \mapsto x_{i,p}$ is continuous.*

Then $\mathcal{E}$ has a statistical equilibrium. If $p' \sum_{i=1}^{I} w_i x_{i,p} < 0$ holds instead of assumption 2, then all statistical equilibria are genuine.

Furthermore, if $N_i \neq \emptyset$ for some i, then the economic temperature T is finite and the equilibrium can be obtained by solving the $C+1$ equations $\sum_{c=1}^{C} p_c = 1$

and

$$\forall c \in \bigcup_{i \in \mathcal{I}} N_i, \quad T \sum_{i: c \in N_i} w_i = -p_c \sum_{i=1}^{I} w_i x_{ic,p}, \tag{5.1a}$$

$$\forall c \in \bigcap_{i \in \mathcal{I}} L_i, \quad \sum_{i=1}^{I} w_i x_{ic,p} \leq 0, \; p_c \geq 0, \; p_c \sum_{i=1}^{I} w_i x_{ic,p} = 0, \tag{5.1b}$$

with $C+1$ *unknowns* $p \in \Delta^{C-1}$ *and* $T \geq 0$. *In particular, if* $L_i = \emptyset$ *for all* $i \in \mathcal{I}$, *then* (5.1) *becomes*

$$\forall c, \quad T = -p_c \sum_{i=1}^{I} w_i x_{ic,p}. \tag{5.2}$$

Proof. We can easily see that $\mathcal{E}$ satisfies all assumptions of Theorem 4.3. (The least obvious assumption is the uniform boundedness condition (Assumption 2), but this follows from assumption 3 and the compactness of Δ^{C-1}.) Hence a statistical equilibrium exists. In the case of a degenerate equilibrium, (5.1) holds by setting $T = 0$. In the case of a genuine equilibrium, let p and π be the equilibrium scarcity parameter and entropy price, respectively. The log-partition function is

$$\begin{aligned} Q_p(\xi) &= \sum_{i=1}^{I} w_i \log \left(\prod_{c \in L_i} \mathrm{e}^{-\xi_c x_{ic,p}} \times \prod_{c \in N_i} \int_{x_{ic,p}}^{\infty} \mathrm{e}^{-\xi_c x_c} \mathrm{d}x_c \right) \\ &= \sum_{i=1}^{I} w_i \log \left(\prod_{c \in L_i} \mathrm{e}^{-\xi_c x_{ic,p}} \times \prod_{c \in N_i} \frac{1}{\xi_c} \mathrm{e}^{-\xi_c x_{ic,p}} \right) \\ &= -\sum_{i=1}^{I} w_i \left(\xi' x_{i,p} + \sum_{c \in N_i} \log \xi_c \right). \end{aligned} \tag{5.3}$$

If $N_i \neq \emptyset$ for some i, then by (5.3) we have $Q_p(0) = \infty$. Thus, the economic temperature defined by $\pi = \frac{1}{T} p$ must be finite. Then, (5.1) can be obtained by applying the Karush-Kuhn-Tucker theorem to the minimization of $Q_p(\xi)$ subject to $\xi \geq 0$ and using the relation $\pi = \frac{1}{T} p$ or $1/\pi_c = T/p_c$ imposed by Definition 3.2. □

Since $\sum_{i=1}^{I} w_i x_{ic,p}$ is the infimum average transaction of commodity c, (5.2) implies that the value of the infimum average transaction evaluated at the equilibrium scarcity parameter p is common across all commodities and that their absolute values are equal to the economic temperature. If $T = 0$, then (5.1) is the condition of degenerate statistical equilibrium. Hence degenerate equilibria are indeed "degenerate" in the sense that the economic temperature is lowest, namely absolute zero.

6 Applications

In this section, I provide a few applications of statistical equilibria.

6.1 Information Centralization as Government's Role

The commodity space is $\mathbb{R}^C$. There is only one agent type, consumers with initial endowment $e \in \mathbb{R}^C_{++}$ and Cobb-Douglas utility function $u(x_1, \dots, x_C) = \sum_{c=1}^C \alpha_c \log x_c$, where $\alpha_c > 0$ and $\sum_{c=1}^C \alpha_c = 1$. The Walrasian demand given price p is obviously $x_c(p) = \frac{\alpha_c w}{p_c}$, where $w = \sum_{c=1}^C p_c e_c$ denotes the wealth. Let $0 \leq r \leq 1$ be the "safety margin" and define the offer set by

$$X_p = \left\{ x \in \mathbb{R}^C \;\middle|\; \forall c,\ x_c \geq (1-r)x_c(p) - e_c \right\}.$$

That is, the offer set is a translation of the positive orthant $\mathbb{R}^C_+$ and agents perceive the possibility of trading quantities smaller than the Walrasian demand by the factor $1-r$ due to uncertainty, limited liquidity, or whatever reasons. Assume that the measure is the Lebesgue measure restricted on X_p.

This example satisfies all assumptions of Corollary 5.1.[26] By (5.2), we obtain

$$\forall c, \quad T = -(1-r)\alpha_c w + p_c e_c. \tag{6.1}$$

Summing up (6.1) with respect to c and using the definition of w, we obtain $T = \frac{rw}{C}$. Substituting this back into (6.1) and using $\sum_{c=1}^C p_c = 1$, we finally get

$$w = \left[\sum_{c=1}^C \frac{1}{e_c} \left(\frac{r}{C} + (1-r)\alpha_c \right) \right]^{-1}, \tag{6.2a}$$

$$T = \frac{rw}{C}, \tag{6.2b}$$

$$p_c = \frac{1}{e_c} \left(\frac{r}{C} + (1-r)\alpha_c \right) w. \tag{6.2c}$$

As r tends to 0, by (6.2), both w and p are of the order $O(1)$ and $T = O(r)$. The entropy price $\pi = \frac{1}{T}p$ is $O(r^{-1})$. Since the equilibrium distribution is exponential with the exponent $\mathrm{e}^{-\pi' x}$, its variance is of the order $O(r^2)$. Hence the smaller the "safety margin" r is, *i.e.*, the less uncertain and the more liquid the economy, the smaller horizontal inequality. In the limit of $r \to 0$, the equilibrium distribution converges in probability to the Walrasian transaction and horizontal *equality* is attained.

This simple example has an interesting policy implication. It is a common practice in many countries to impose progressive income tax to mitigate the unequal distribution of wealth, but according to statistical equilibrium theory, horizontal inequality is a necessity in an economy with some uncertainty or limited liquidity, in which the offer sets diverge from the "ideal" Walrasian ones. If we were to believe statistical equilibrium theory, redistributive tax policies fail to achieve their objectives.

If a government seeks to reduce horizontal inequality, there are two solutions. The first one—which I do not recommend—is to become a totalitarian state and enforce people specific transactions. The second one is to gather and publish information so that people can bring their offer sets as close as possible to the "ideal" situation, that is the Walrasian offer sets. Thus, from a normative point of view, an important role of the government is *information centralization.*

[26] Strictly speaking, the offer set is not well-defined when $p_c = 0$ for some c but it is easy to justify by a "box argument" as in the proof of Corollary 4.4.

6.2 Endogenous Wage Distribution in a Search Model

This example is a standard search model as in McCall (1970) with a twist of statistical equilibrium theory. Time is discrete. There are two agent types, households and Nature. At each period Nature dumps y units of consumption good per active household in the market. At period t, each active household gets a wage offer w that the household believes to come from a stationary prior P. Household's objective is to maximize $\mathrm{E}\sum_{t=0}^{\infty}\beta^t u(c_t)$, where c_t is consumption. If the household accepts the offer w, it inactivates, i.e., it receives and consumes w forever. Otherwise, it receives an unemployment compensation c and wait for another offer next period. The actual wage distribution is determined by the Maximum Entropy Principle.

An active household's Bellman equation is

$$v(w) = \max\left\{\frac{u(w)}{1-\beta}, u(c) + \beta\,\mathrm{E}[v(w')]\right\}.$$

The reservation wage $\bar{w}$ is determined by[27]

$$u(\bar{w}) = u(c) + \frac{\beta}{1-\beta}\int_{\bar{w}}^{\infty}(u(w) - u(\bar{w}))P(\mathrm{d}w). \tag{6.3}$$

Let π be the entropy price and $a(w) = \begin{cases} w, & (w \geq \bar{w}) \\ c & (w < \bar{w}) \end{cases}$ be the assignment rule. The partition functions of an active household and nature are

$$\begin{aligned} Z_h(\pi) &= \int \mathrm{e}^{-\pi a(w)}P(\mathrm{d}w) = \mathrm{e}^{-\pi c}P(c) + \int_{\bar{w}}^{\infty}\mathrm{e}^{-\pi w}P(\mathrm{d}w), \\ Z_n(\pi) &= \mathrm{e}^{\pi y}, \end{aligned}$$

respectively. Thus the log-partition function is

$$Q(\pi) = \frac{1}{2}\log\left(\mathrm{e}^{-\pi c}P(c) + \int_{\bar{w}}^{\infty}\mathrm{e}^{-\pi w}P(\mathrm{d}w)\right) + \frac{1}{2}\pi y.$$

By maximizing entropy (or minimizing the log-partition function), we obtain

$$Q'(\pi) = 0 \iff \frac{c\mathrm{e}^{-\pi c}P(\bar{w}) + \int_{\bar{w}}^{\infty} w\mathrm{e}^{-\pi w}P(\mathrm{d}w)}{\mathrm{e}^{-\pi c}P(\bar{w}) + \int_{\bar{w}}^{\infty}\mathrm{e}^{-\pi w}P(\mathrm{d}w)} = y. \tag{6.4}$$

Given the prior P, (6.3) determines the reservation wage $\bar{w}$ and (6.4) determines the entropy price π. The density of the actual wage distribution is proportional to $\mathrm{e}^{-\pi w}P(\mathrm{d}w)$.

Now suppose that P_0 is exponential and households update the prior according to the rule $P_{t+1}(\mathrm{d}w) \propto \mathrm{e}^{-\pi_t w}P_t(\mathrm{d}w)$, that is, by extrapolating the actual wage distribution to those wages under the reservation value. Then P_t is exponential with a parameter λ_t that evolves according to $\lambda_{t+1} = \pi_t + \lambda_t$, where the reservation wage w_t and the entropy price π_t are determined by (6.3), (6.4), in

[27] See, for example, Equation (6.3.3) in (Ljungqvist and Sargent, 2004, p. 144).

this case

$$u(w_t) = u(c) + \frac{\beta}{1-\beta}\int_{w_t}^{\infty}(u(w)-u(w_t))\lambda_t \mathrm{e}^{-\lambda_t w}\mathrm{d}w, \tag{6.5a}$$

$$\frac{c\mathrm{e}^{-\pi_t c}(1-\mathrm{e}^{-\lambda_t w_t}) + \frac{\lambda_t(1+(\pi_t+\lambda_t)w_t)}{(\pi_t+\lambda_t)^2}\mathrm{e}^{-(\pi_t+\lambda_t)w_t}}{\mathrm{e}^{-\pi_t c}(1-\mathrm{e}^{-\lambda_t w_t}) + \frac{\lambda_t}{\pi_t+\lambda_t}\mathrm{e}^{-(\pi_t+\lambda_t)w_t}} = y. \tag{6.5b}$$

The unemployment rate at t is $1-\mathrm{e}^{-(\pi_t+\lambda_t)w_t}$.

Since $\pi_t \geq 0$, $\{\,\lambda_t\,\}$ either converges or diverges to ∞. The interesting case is the former. Substituting $w_t = \bar{w}$, $\pi_t = 0$, and $\lambda_t = \lambda$ in (6.5), we obtain the steady state by solving

$$u(\bar{w}) = u(c) + \frac{\beta}{1-\beta}\int_{\bar{w}}^{\infty}(u(w)-u(\bar{w}))\lambda \mathrm{e}^{-\lambda w}\mathrm{d}w,$$

$$c(1-\mathrm{e}^{-\lambda\bar{w}}) + \frac{1+\lambda\bar{w}}{\lambda}\mathrm{e}^{-\lambda\bar{w}} = y.$$

As a numerical example, I set the utility function to $u(w) = \frac{1}{a}\mathrm{e}^{-aw}$ with $a = 3$, and other parameter values are: discount factor $\beta = 0.9$, unemployment compensation $c = 1$, and initial exponential parameter $\lambda_0 = 0.001$. The initial per capita consumption good is $y = 10$, but it (unexpectedly) plunges to $y = 8$ at $t = 21$.

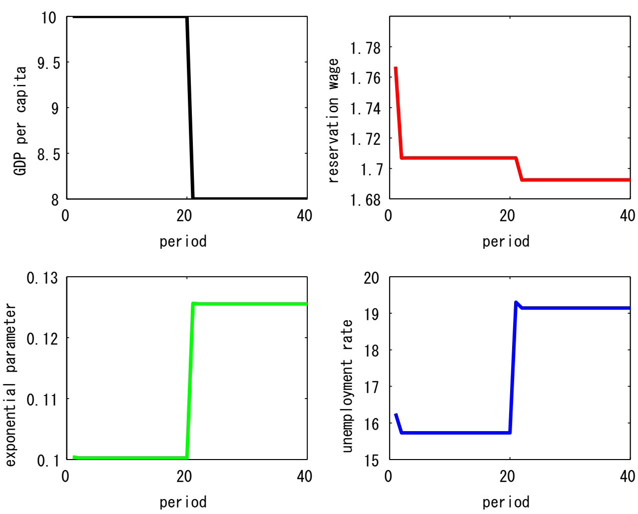

Figure 1. Time series of per capita GDP y_t (top left), reservation wage w_t (top right), exponential parameter $\lambda_{t+1} = \pi_t + \lambda_t$ (bottom left), and unemployment rate (bottom right, in percent).

Figure 1 shows the time series of per capita output y_t, reservation wage w_t, exponential parameter $\pi_t + \lambda_t$, and unemployment rate. As seen in the figure, the reservation wage hardly respond to the negative output shock. The reduction in output is adjusted through an increase in unemployment rate and decrease in the variation of wage (note that the standard deviation of the exponential distribution with parameter λ is $1/\lambda$: see the bottom left panel of Figure 1).

7 Conclusion

In this paper I reformulated the concept of statistical equilibrium advanced by Foley (1994); Toda (2010), and discussed its efficiency, existence, and the relation to Walrasian equilibrium. My theory potentially has a broad range of applicability because: my formulation of statistical equilibrium theory is abstract; the assumptions to ensure the existence of equilibria are weak as well as economically intuitive; there exists a simple algorithm to numerically obtain the equilibria as proposed in Toda (2010); my theory contains Walrasian equilibrium theory as a mathematical special case.

It is my belief that macroeconomic theory should take the heterogeneity of agents seriously, as does Aiyagari (1994). As the simple search model in Section 6 shows, statistical equilibrium theory offers an alternative to standard macroeconomic models when treating heterogeneous agents.

A Solving the Maximum Entropy Problem

In this subsection I study the maximum entropy program (MEP) using the duality theory. Its dual problem, the *minimum log-partition program* (MLPP), is finite dimensional. I give a sufficient condition for the existence of a solution.

Let $\{ \mu_i \}_{i\in\mathcal{I}}$ be a collection of regular Borel measures on $\mathbb{R}^C$. The functions

$$Z_i(\xi) := \int \mathrm{e}^{-\xi' x}\mu_i(\mathrm{d}x), \qquad Q(\xi) := \sum_{i=1}^{I} w_i \log\left(\int \mathrm{e}^{-\xi' x}\mu_i(\mathrm{d}x)\right)$$

are called the *partition function* and the *log-partition function* respectively. The optimization problem

$$\min_{\xi\geq 0} Q(\xi) \tag{A.1}$$

is called the *minimum log-partition program* (MLPP).

Theorem A.1 (Duality Theorem)**.** *Let $\{ \mu_i \}_{i\in\mathcal{I}}$ be a collection of regular Borel measures on $\mathbb{R}^C$. Suppose that there exists a collection of densities $f^1 = (f_i^1)_{i\in\mathcal{I}}$ such that $\bar{x}[f^1;\mu] \ll 0$. Then,*

$$\sup\left\{ H[f;\mu] \,\middle|\, \bar{x}[f;\mu] \leq 0,\ \forall i, \int f_i \mathrm{d}\mu_i = 1 \right\} = \inf_{\xi\geq 0} Q(\xi).$$

Furthermore, if the infimum on the right is achieved by $\xi = \pi$, then the supremum on the left is achieved by $f_i(x) = \mathrm{e}^{-\pi' x}/Z_i(\pi)$.

Proof. Similar to the proof of Theorem C.4. □

By Theorem A.1, the entropy maximizing problem reduces to the minimization of the log-partition function. Proposition A.2 below characterizes when the MLPP (A.1) has a solution.

Proposition A.2. *Let $\{ \mu_i \}_{i\in\mathcal{I}}$ be a collection of regular Borel measures on $\mathbb{R}^C$ such that μ_i is supported on $X_i \subset \mathbb{R}^C$. Suppose that $Q(\xi) < \infty$ for some $\xi \in \mathbb{R}^C_+$. Then, the MLPP* (A.1) *has a solution if*

$$\left(\sum_{i=1}^{I} w_i \operatorname{co} X_i\right) \cap (-\mathbb{R}^C_{++}) \neq \emptyset. \tag{A.2}$$

Conversely, if

$$\left(\sum_{i=1}^{I} w_i \operatorname{co} X_i \right) \cap (a - \mathbb{R}^C_{++}) = \emptyset \tag{A.3}$$

for some $a \gg 0$*, then the MLPP* (A.1) *has no solutions.*

Proof. Since Q is lower semi-continuous by Proposition C.5, we only need to know the behavior of Q as $\xi \to \infty$.

If (A.2) holds, take $\epsilon > 0$ and $x_i \in \operatorname{co} X_i$ such that $\sum_{i=1}^{I} w_i x_i \le -\epsilon \mathbf{1} \ll 0$. By Carathéodory's theorem (Rockafellar, 1970, p. 155), we can express x_i as a convex combination of at most $C+1$ points: $x_i = \sum_{k=1}^{C+1} \alpha_i^k x_i^k$, where $x_i^k \in X_i$, $\alpha_i^k \ge 0$, and $\sum_{k=1}^{C+1} \alpha_i^k = 1$. Then we obtain

$$v_i(\xi) := \inf \{ \xi' x \mid x \in X_i \} \le \min_k \xi' x_i^k \le \sum_{k=1}^{C+1} \alpha_i^k \xi' x_i^k = \xi' x_i. \tag{A.4}$$

By (A.4) and Proposition C.6 applied to $\phi(x) = -\xi' x$, it follows that if $\xi \ge 0$ and $\|\xi\|_1 = 1$, then

$$\lim_{t\to\infty} \frac{1}{t} Q(t\xi) = -\sum_{i=1}^{I} w_i v_i(\xi) \ge -\xi' \sum_{i=1}^{I} w_i x_i \ge \xi' \epsilon \mathbf{1} = \epsilon \|\xi\|_1 = \epsilon. \tag{A.5}$$

By (A.5), we get $Q(t\xi) \to \infty$ as $t \to \infty$. Therefore, we may restrict our search of the infimum of $Q(\xi)$ to a compact set, but since $Q(\xi)$ is lower semi-continuous, the minimum is attained.

Conversely, suppose that (A.3) holds. Since $\sum_{i=1}^{I} w_i \operatorname{co} X_i$ and $a - \mathbb{R}^C_{++}$ are both convex and the latter contains an interior point, by the separating hyperplane theorem (Rockafellar, 1970, p. 97), there exists $\xi \in \mathbb{R}^C_+ \setminus \{ 0 \}$ such that $\xi' x \ge \xi' a$ for all $x \in \sum_{i=1}^{I} w_i \operatorname{co} X_i$. Since $\xi > 0$ and $a \gg 0$, we have $\xi' a > 0$. By a similar argument as above, we obtain

$$\lim_{t\to\infty} \frac{1}{t} Q(t\xi) = -\sum_{i=1}^{I} w_i v_i(\xi) \le -\xi' a < 0.$$

Consequently, $Q(t\xi) \to -\infty$ as $t \to \infty$. Since by definition $Q(\xi) > -\infty$, we get $\inf_{\xi \ge 0} Q(\xi) = -\infty$; however, the infimum is never attained. □

Corollary A.3. *Let everything be as in Proposition A.2. If* (A.2) *holds, then the MEP* (3.3) *has a solution given by* $f_i(x) = \mathrm{e}^{-\pi' x}/Z_i(\pi)$*, where* π *is the solution of the MLPP* (A.1).

Proof. By Proposition A.2, the MLPP (A.1) has a solution $\xi = \pi$. By (A.2) and Lemma 3.2 of Toda (2010), there exists $f^1 = (f_i^1)_{i \in \mathcal{I}}$ such that $\bar{x}[f^1; \mu] \ll 0$. Hence by Theorem A.1 the claim is true. □

B Proofs

B.1 Proof of Theorem 4.2

For $\xi \in \mathbb{R}^C_+$, define the log-partition function by

$$Q_p(\xi) = \sum_{i=1}^{I} w_i \log \left(\int \mathrm{e}^{-\xi' x} \mathrm{d}\mu_{i,p} \right).$$

By Assumption 2, we have $e^{-\xi' x} \le e^{-\xi' a}$ for all $\xi \ge 0$ and $x \in X_{i,p}$, so $\int e^{-\xi' x} d\mu_{i,p} < \infty$ by Assumption 1. Hence $Q_p(\xi) < \infty$ for all $\xi \ge 0$.

The outline of the proof is as follows. Since we know from Corollary A.3 that solving MLPP is enough for maximizing entropy, we wish to construct a correspondence from p to the Lagrange multiplier for MLPP and show the existence of a fixed point (which is precisely the idea in Toda (2010)). However, Proposition A.2 shows that MLPP may not always have a solution. To overcome this difficulty, we bound the domain of MLPP by a constant and define a quasi equilibrium concept. Specifically, define a b-quasi equilibrium as follows.

Definition B.1 (b-quasi equilibrium)**.** Let $b > 0$. The pair of vectors $(p, \pi) \in \Delta^{C-1} \times \mathbb{R}_+^C$ is said to be a *b-quasi equilibrium* if

1. $\xi = \pi$ solves
$$\min Q_p(\xi) \text{ subject to } \xi \ge 0, \|\xi\|_1 \le b,$$
2. π and p are collinear.

Thus, the definition of a b-quasi equilibrium is the same as that of a genuine statistical equilibrium except for that the former solves the MLPP within the domain bounded by b. We show by a standard fixed point argument that a b-quasi equilibrium always exists. Thus, we can take a sequence of b-quasi equilibria such that $b \to \infty$. We then show that either some b-quasi equilibrium becomes a genuine statistical equilibrium, or a subsequence of b-quasi equilibria converges to a degenerate statistical equilibrium.

Let $(p_n, \xi_n) \subset \Delta^{C-1} \times \mathbb{R}_+^C$ be a sequence such that $(p_n, \xi_n) \to (p, \xi)$.

Step 1. *$\mu_{i,p_n} \to \mu_{i,p}$ weakly.*

Proof. Trivial by Assumption 4. □

Let $f_n(x) = e^{-\xi_n' x}$, $f(x) = e^{-\xi' x}$, $g_n(x) = \xi_n' x e^{-\xi_n' x}$, and $g(x) = \xi' x e^{-\xi' x}$.

Step 2. *$f_n \to f$ and $g_n \to g$ uniformly on compact sets, and f, g are continuous.*

Proof. That f, g are continuous is trivial. Take any compact set $K \subset \mathbb{R}^C$ and a number $C_1 > 0$ such that $\|x\| < C_1$ for all $x \in K$. Since $\xi_n \to \xi$, the sequence $\{ \xi_n \}$ is bounded; take $C_2 > 0$ such that $\|\xi_n\| < C_2$ for all n. Then $\|\xi\| \le C_2$. Applying the mean value theorem to $\phi_1(t) = e^{-t}$ and $\phi_2(t) = te^{-t}$, we obtain

$$\begin{aligned} |f_n(x) - f(x)| &= |\phi_1'(c_1)| \, |\xi_n' x - \xi' x| \le C_1 \, |\phi_1'(c_1)| \, \|\xi_n - \xi\| \,, \\ |g_n(x) - g(x)| &= |\phi_2'(c_2)| \, |\xi_n' x - \xi' x| \le C_1 \, |\phi_2'(c_2)| \, \|\xi_n - \xi\| \,, \end{aligned}$$

for some c_1, c_2 between $\xi_n' x$ and $\xi' x$. By the definition of C_1, C_2, we get $|c_j| \le \max \{ |\xi_n' x| , |\xi' x| \} \le C_1 C_2$. Thus $\left|\phi_j'(c_j)\right|$ can be bounded by a constant independent of x and n. Therefore $f_n \to f$ and $g_n \to g$ uniformly on K. □

Step 3. *$\{ f_n \}$, $\{ g_n \}$ are uniformly integrable.*

Proof. Remember the definition of uniform integrability (C.15). Clearly $\{ F_n \}$ is uniformly integrable if F_n is uniformly bounded (take α larger than $\sup_n |F_n|$ in (C.15)). Since $\xi_n' x \ge \xi_n' a \ge -C_2 \|a\|$ and $e^{-t}, te^{-t} \to 0$ as $t \to \infty$, it follows that f_n, g_n are uniformly bounded. Therefore $\{ f_n \}$, $\{ g_n \}$ are uniformly integrable. □

Step 4. *The integrals* $\int e^{-\xi' x} d\mu_{i,p}$, $\int \xi' x e^{-\xi' x} d\mu_{i,p}$ *are continuous in* (p, ξ) *on* $\Delta^{C-1} \times \mathbb{R}_+^C$.

Proof. This follows from Steps 1–3 and Theorem C.7. □

Step 5. $Q_p(\xi)$ *is continuous in* $(p, \xi) \in \Delta^{C-1} \times \mathbb{R}_+^C$ *and continuously differentiable in* ξ.

Proof. $Q_p(\xi)$ is continuous in (p, ξ) by Step 4. It is continuously differentiable in ξ by applying Lebesgue's convergence theorem. □

Step 6. *For all* $b > 0$, *a* b*-quasi equilibrium exists.*

Proof. As we saw above, $Q_p(\xi) < \infty$ for all $\xi \geq 0$. By Proposition C.5, $Q_p(\xi)$ is convex and lower semi-continuous in ξ. Since the set $X_b := \{ \xi \in \mathbb{R}_+^C \mid \|\xi\|_1 \leq b \}$ is compact and convex, the set $\Pi(p) := \arg\min_{\xi \in X_b} Q_p(\xi)$ is nonempty, compact and convex. If $0 \in \Pi(p)$ for some $p \in \Delta^{C-1}$, since p and 0 are collinear, obviously $(p, 0)$ is a b-quasi equilibrium (actually a genuine statistical equilibrium).

Assume $0 \notin \Pi(p)$ for all p. By Step 5, $Q_p(\xi)$ is continuous in (p, ξ) on $\Delta^{C-1} \times \mathbb{R}_+^C$. Thus, by Berge's maximum theorem (Berge, 1959, p. 116), $\Pi : \Delta^{C-1} \rightrightarrows X_b$ is upper semi-continuous. Define $\Phi(p) := \{ \xi / \|\xi\|_1 \mid \xi \in \Pi(p) \}$. Since $0 \notin \Pi(p)$, $\Phi(p)$ is well-defined.

Let us show that $\Phi : \Delta^{C-1} \rightrightarrows \Delta^{C-1}$ is nonempty, compact, convex and upper semi-continuous. $\Phi(p) \neq \emptyset$ is trivial. Since $\Phi(p)$ is the intersection of Δ^{C-1} (a convex set) and the convex cone generated by $\Phi(p)$, it is convex. If $p_n \to p$, $q_n \in \Phi(p_n)$, and $q_n \to q$, take a sequence $\{ \xi_n \} \subset \Pi(p_n)$ such that $q_n = \xi_n / \|\xi_n\|_1$. Since $\{ \xi_n \} \subset X_b$ and X_b is compact, $\{ \xi_n \}$ has a convergent subsequence $\xi_{n_k} \to \xi$. Since $p \mapsto \Pi(p)$ is upper semi-continuous, we have $\xi \in \Pi(p)$, so $q = \xi / \|\xi\|_1 \in \Phi(p)$. Thus, $p \mapsto \Phi(p)$ is upper semi-continuous. In particular, by letting $p_n = p$ for all n, it follows that $\Phi(p)$ is closed, but since $\Phi(p) \subset \Delta^{C-1}$, it is compact.

By Kakutani's fixed point theorem, there exists $p^* \in \Delta^{C-1}$ such that $p^* \in \Phi(p^*)$. Thus, there exists $t > 0$ such that $tp^* \in \Pi(p)$, so (p^*, tp^*) is a b-quasi equilibrium. □

Let $\{ t_n \}_{n=1}^{\infty} \subset (0, \infty)$ be a monotone increasing sequence tending to ∞. By passing to a subsequence if necessary, we may assume that for each n, there exists a t_n-quasi equilibrium (p_n, π_n) such that p_n converges to some $p \in \Delta^{C-1}$. If $\pi_n \in \arg\min_{\xi \geq 0} Q_{p_n}(\xi)$ for some n, then by Corollary A.3 (p_n, π_n) is a genuine statistical equilibrium. Therefore, without loss of generality we may assume that for all n, $\min_{\xi \geq 0} Q_{p_n}(\xi)$ has no solutions. Let

$$\mathcal{L}(\xi, \lambda_n, \theta_n) = \sum_{i=1}^{I} w_i \log \left(\int e^{-\xi' x} d\mu_{i,p_n} \right) - \lambda_n' \xi + \theta_n (\|\xi\|_1 - t_n)$$

be the Lagrangian of $\min_{\xi \in X_{t_n}} Q_{p_n}(\xi)$, where $\lambda_n \in \mathbb{R}_+^C$ and $\theta_n \geq 0$ are Lagrange multipliers. By the Karush-Kuhn-Tucker theorem, we obtain

$$\sum_{i=1}^{I} w_i \frac{\int -x e^{-\pi_n' x} d\mu_{i,p_n}}{\int e^{-\pi_n' x} d\mu_{i,p_n}} - \lambda_n + \theta_n \mathbf{1} = 0 \tag{B.1}$$

and $\lambda_n' \pi_n = 0$. Since (p_n, π_n) is not a statistical equilibrium but a t_n-quasi equilibrium, $\|\pi_n\|_1 \leq t_n$ is binding for all n. Hence we have $\pi_n = t_n p_n$.

Step 7. $\lim_{n\to\infty}\theta_n = 0$.

Proof. Multiplying $\pi_n = t_n p_n$ as an inner product to (B.1) and dividing both sides by $t_n > 0$, it follows from $\lambda_n'\pi_n = 0$ that

$$\sum_{i=1}^{I} w_i \frac{\int -p_n' x \mathrm{e}^{-t_n p_n' x} \mathrm{d}\mu_{i,p_n}}{\int \mathrm{e}^{-t_n p_n' x} \mathrm{d}\mu_{i,p_n}} + \theta_n = 0. \tag{B.2}$$

Regard $Q_p(t\xi) = \sum_{i=1}^{I} w_i \log\left(\int \mathrm{e}^{-t\xi' x}\mathrm{d}\mu_{i,p}\right)$ as a function of t. Since Q_p is convex, $Q_p'(t\xi)$ is increasing in t. Take any $t > 0$ and choose n sufficiently large such that $t_n > t$. Then, by (B.2) we obtain

$$\theta_n = -Q_{p_n}'(t_n p_n) \le -Q_{p_n}'(t p_n) = \sum_{i=1}^{I} w_i \frac{\int p_n' x \mathrm{e}^{-t p_n' x} \mathrm{d}\mu_{i,p_n}}{\int \mathrm{e}^{-t p_n' x} \mathrm{d}\mu_{i,p_n}}.$$

Letting $n \to \infty$, it follows from Step 4 that

$$\limsup_{n\to\infty} \theta_n \le \sum_{i=1}^{I} w_i \frac{\int p' x \mathrm{e}^{-t p' x} \mathrm{d}\mu_{i,p}}{\int \mathrm{e}^{-t p' x} \mathrm{d}\mu_{i,p}} = -Q_p'(tp). \tag{B.3}$$

Since t is arbitrary in (B.3), for any $s > 0$, take $t_s > 0$ such that

$$\frac{Q_p(sp) - Q_p(0)}{s} = Q_p'(t_s p),$$

which is of course possible by the mean value theorem. Then, (B.3) becomes

$$\limsup_{n\to\infty} \theta_n \le -\frac{Q_p(sp) - Q_p(0)}{s}.$$

Letting $s \to \infty$, by Proposition C.6 and Assumption 3 we obtain

$$\limsup_{n\to\infty} \theta_n \le \sum_{i=1}^{I} w_i \inf \{\, p' x \mid x \in X_{i,p} \,\} \le 0. \tag{B.4}$$

Since $\theta_n \ge 0$, we have $\theta_n \to 0$. ☐

Step 8. *The correspondence* $p \mapsto \operatorname{cl}\operatorname{co} X_{i,p}$ *is closed at* $p = \lim p_n$. *Furthermore,* $p' x \ge 0$ *for all* $x \in X_{i,p}$.

Proof. By Step 7 and (B.4), we have $\sum_{i=1}^{I} w_i \inf \{\, p' x \mid x \in X_{i,p} \,\} = 0$. By Assumption 5, the correspondence $p \mapsto \operatorname{cl}\operatorname{co} X_{i,p}$ is closed at p. By Assumption 3 it must be $\inf \{\, p' x \mid x \in X_{i,p} \,\} = 0$ for all $i \in \mathcal{I}$ or $p' x \ge 0$ for all $x \in X_{i,p}$. ☐

Step 9. *If* $\mathcal{E}$ *has no genuine statistical equilibria, then* $\mathcal{E}$ *has a degenerate statistical equilibrium.*

Proof. Define the sequence $\{\, x_i^n \,\}_{n=1}^{\infty} \subset \mathbb{R}^C$ by $x_i^n = \frac{\int x \mathrm{e}^{-t_n p_n' x} \mathrm{d}\mu_{i,p_n}}{\int \mathrm{e}^{-t_n p_n' x} \mathrm{d}\mu_{i,p_n}}$. Since $\mathrm{e}^{-t_n p_n' x} / \int \mathrm{e}^{-t_n p_n' x} \mathrm{d}\mu_{i,p_n}$ is a probability density function, we have $x_i^n \in \operatorname{cl}\operatorname{co} X_{i,p_n}$. Since by Assumption 2 we have $x_i^n \ge a$, it follows from $\lambda_n \ge 0$ and (B.1) that

$$a \le (1 - w_i)a + w_i x_i^n \le \sum_{i=1}^{I} w_i x_i^n \le \theta_n \mathbf{1} \implies a \le x_i^n \le \frac{\theta_n \mathbf{1} - (1 - w_i)a}{w_i}.$$

Thus $\{ x_i^n \}$ is bounded.

Since $\theta_n \to 0$, by taking a subsequence we may assume $x_i^n \to x_i$ for some $x_i \in \mathbb{R}^C$ for all $i \in \mathcal{I}$; hence $\sum_{i=1}^I w_i x_i^n \to \sum_{i=1}^I w_i x_i \leq 0$. Since by Step 8 the correspondence $p \mapsto \operatorname{cl}\operatorname{co} X_{i,p}$ is closed, it follows that $x_i \in \operatorname{cl}\operatorname{co} X_{i,p}$. By (B.2), we have $\sum_{i=1}^I w_i p_n' x_i^n = \theta_n$, so letting $n \to \infty$ we get $p' \sum_{i=1}^I w_i x_i = 0$. Hence, if $\sum_{i=1}^I w_i x_{ic} < 0$, it must be $p_c = 0$ and condition 4 of Definition 3.2 holds. Condition 5 of Definition 3.2 holds by Step 8. Therefore $(p, \{ x_i \}_{i \in \mathcal{I}})$ is a degenerate statistical equilibrium. □

Step 10. *Suppose that Assumptions 1–5 hold. If*

$$\sum_{i=1}^I w_i \inf \{ p'x \mid x \in X_{i,p} \} < 0$$

for all p, then all statistical equilibria of $\mathcal{E}$ are genuine.

Proof. If $\mathcal{E}$ has a degenerate statistical equilibrium $(p, \{ x_i \})$, by Definition 3.2 we have $p'x \geq 0$ for all $x \in X_{i,p}$, or $\inf \{ p'x \mid x \in X_{i,p} \} \geq 0$. Thus $\sum_{i=1}^I w_i \inf \{ p'x \mid x \in X_{i,p} \} \geq 0$, contradiction. □

B.2 Proof of Theorem 4.3

Proof. For all $\epsilon \gg 0$, by Theorem 4.2 the economy $\mathcal{E}^\epsilon$ has a statistical equilibrium. If the equilibrium is degenerate, since $\mu_{i,p}$ and $\mu_{i,p}^\epsilon$ have a common support and the support (offer set) is all that matters in the definition of degenerate equilibria (see Definition 3.2), $\mathcal{E}$ also has a degenerate equilibrium. Hence, for all $\epsilon \gg 0$, we may assume that $\mathcal{E}^\epsilon$ has a genuine statistical equilibrium.

Let $\{ \epsilon_n \} \subset \mathbb{R}_{++}^C$ be a sequence such that $\epsilon_n \to 0$ as $n \to \infty$. Let p_n and $\pi_n = t_n p_n$ be the equilibrium scarcity parameter and the entropy price of $\mathcal{E}^{\epsilon_n}$, respectively. By passing to a subsequence if necessary, we may assume $p_n \to p \in \Delta^{C-1}$ and $t_n \to t \in [0, \infty]$ as $n \to \infty$.

If $t_n \to \infty$, by the same argument as in Step 9 of the proof of Theorem 4.2, we can show that a degenerate statistical equilibrium exists.

If $t_n \to 0$ or $t_n \to t < \infty$ and $p_c = 0$ for some $c \in \bigcup_{i \in \mathcal{I}} N_i$, take any $\xi \in \mathbb{R}_+^C$ such that $\xi_c > 0$ for all $c \in \bigcup_{i \in \mathcal{I}} N_i$. Since $\pi_n = t_n p_n$ solves the MLPP, we obtain

$$\sum_{i=1}^I w_i \log \left(\int \mathrm{e}^{-t_n p_n' x - \epsilon_n' x} \mathrm{d}\mu_{i,p_n} \right) \leq \sum_{i=1}^I w_i \log \left(\int \mathrm{e}^{-\xi' x - \epsilon_n' x} \mathrm{d}\mu_{i,p_n} \right). \quad \text{(B.5)}$$

Taking lim inf of (B.5), by Fatou's lemma and assumption 2d the left-hand side of (B.5) is bounded below by ∞ as $n \to \infty$. By Step 5 of the proof of Theorem 4.2, we can take the limit of the right-hand side of (B.5) to obtain

$$\infty \leq \sum_{i=1}^I w_i \log \left(\int \mathrm{e}^{-\xi' x} \mathrm{d}\mu_{i,p} \right).$$

However, this is a contradiction because by assumptions 1 and 2c, the measure $\mu_{i,p}^\xi$ is finite. Therefore, it must be $t_n \to t \in (0, \infty)$ and $p_c > 0$ for all $c \in$

$\bigcup_{i \in \mathcal{I}} N_i$. Again, letting $n \to \infty$ in (B.5), we obtain

$$\sum_{i=1}^{I} w_i \log \left(\int \mathrm{e}^{-tp'x} \mathrm{d}\mu_{i,p} \right) \leq \sum_{i=1}^{I} w_i \log \left(\int \mathrm{e}^{-\xi' x} \mathrm{d}\mu_{i,p} \right), \tag{B.6}$$

or $Q_p(tp) \leq Q_p(\xi)$. (B.6) holds for all $\xi \in \mathbb{R}_+^C$ such that $\xi_c > 0$ for all $c \in \bigcup_{i \in \mathcal{I}} N_i$, but by assumption 2d, it trivially holds even if $\xi_c = 0$ for some $c \in \bigcup_{i \in \mathcal{I}} N_i$. Thus, (B.6) holds for all $\xi \in \mathbb{R}_+^C$, and as such tp solves the MLPP associated with p. Therefore, (p, tp) is a genuine statistical equilibrium.

If $\sum_{i=1}^{I} w_i \inf \{ p'x \mid x \in X_{i,p} \} < 0$ holds instead of Assumption 3, by the same argument as in the proof of Theorem 4.2, $\mathcal{E}$ has no degenerate statistical equilibria. □

C Mathematical Results

First I mention two general theorems in optimization that shall be used later.

Theorem C.1 (Generalized Kuhn-Tucker)**.** *Let X be a linear vector space, Z_1, Z_2 normed spaces, Ω a convex subset of X, and P the positive cone in Z_1. Assume that P contains an interior point.*

Let f be a real-valued convex functional on Ω, $G_1 : \Omega \to Z_1$ a convex mapping, and $G_2 : X \to Z_2$ an affine mapping. Assume the existence of a point $x_1 \in \Omega$ for which $G_1(x_1) < 0$ (i.e., $G_1(x_1)$ is an interior point of $N = -P$) and $G_2(x_1) = 0$, and that 0 is an interior point of $G_2(\Omega)$. Let

$$\mu_0 = \inf f(x) \text{ subject to } x \in \Omega,\ G_1(x) \leq 0,\ G_2(x) = 0 \tag{C.1}$$

and assume μ_0 is finite. Then there exist $z_1^* \geq 0$ in Z_1^* and $z_2^* \in Z_2^*$ such that

$$\mu_0 = \inf_{x \in \Omega} \left[f(x) + \langle G_1(x), z_1^* \rangle + \langle G_2(x), z_2^* \rangle \right]. \tag{C.2}$$

Furthermore, if the infimum is achieved in (C.1) *by $x_0 \in \Omega$, it is achieved by x_0 in* (C.2) *and $\langle G_2(x_0), z_2^* \rangle = 0$.*

Proof. Similar to (Luenberger, 1969, Theorem 1, p. 217). □

Theorem C.2 (Lagrange Duality)**.** *Let f be a real-valued convex functional defined on a convex subset Ω of a vector space X, and let G be a convex mapping of X into a normed space Z with a positive cone P. Suppose there exists an x_1 such that $G(x_1) < 0$ and that $\mu_0 = \inf \{ f(x) : G(x) \leq 0, x \in \Omega \}$ is finite. Let $\phi(z^*) := \inf_{x \in \Omega} [f(x) + \langle G(x), z^* \rangle]$ be the dual function of f. Then*

$$\inf_{G(x) \leq 0, x \in \Omega} f(x) = \max_{z^* \geq 0} \phi(z^*)$$

and the maximum on the right is achieved by some $z_0^* \geq 0$.

Proof. See (Luenberger, 1969, Theorem 1, p. 224). □

Next I study the entropy maximization problem subject to equality and inequality moment constraints.

Lemma C.3. *Let (X,μ) be a measure space, $g: X \to \mathbb{R}_+$ measurable and $\int g\mathrm{d}\mu = 1$. Then g is the solution of*

$$\min_f \int f \log \frac{f}{g} \mathrm{d}\mu \text{ subject to } f \geq 0, \quad \int f\mathrm{d}\mu = 1,^{28} \tag{C.3}$$

and the minimum is zero.

Proof. In order for $\int f \log(f/g)\mathrm{d}\mu$ to be finite, it is necessary that $f = g$ μ-a.e. on the set $\{ x \in X \mid g(x) = 0 \}$. Let $\Omega = \{ x \in X \mid f(x) > 0 \}$. Then $\int_\Omega f = 1$ and $\int_\Omega g \leq 1$. Since $\log t \leq t - 1$ for all $t \geq 0$, we obtain

$$\begin{aligned} \int f \log(f/g) &= \int_\Omega f \log(f/g) = -\int_\Omega f \log(g/f) \\ &\geq -\int_\Omega f(g/f - 1) = 1 - \int_\Omega g \geq 0, \end{aligned}$$

with equality if and only if $f = g$ μ-a.e. Hence the minimum of (C.3) is attained by $f = g$ and the minimum is zero. □

Theorem C.4 is an adaptation of Borwein and Lewis (1991, 1992).

Theorem C.4 (Duality Theorem). *Let (X,μ) be a σ-finite measure space, Z_1, Z_2 normed spaces, and P the positive cone in Z_1. Assume that P contains an interior point. For $i = 1, 2$ let $c_i \in Z_i$, $g_i : X \to Z_i$ be measurable, and define $G_i : L^1(X,\mu) \to Z_i$ by*

$$G_i(f) = \int f(x) g_i(x) \mu(\mathrm{d}x) - c_i$$

whenever the Bochner integral is defined. Let

$$\Omega = \left\{ f \in L^1(X,\mu) \;\middle|\; f \geq 0, \int f\mathrm{d}\mu = 1 \right\}.$$

Suppose that there exists an $f_1 \in \Omega$ such that $G_1(f_1) < 0$ and $G_2(f_1) = 0$. Assume also that 0 is an interior point of $G_2(\Omega)$. Define the entropy of $f \in \Omega$ by $H(f) = -\int f \log f \mathrm{d}\mu$ whenever the integral exists. Then

$$\begin{aligned} \alpha &:= \sup \{ H(f) \mid f \in \Omega, G_1(f) \leq 0, G_2(f) = 0 \} \\ &= \inf_{z_1^* \geq 0, z_2^* \in Z_2^*} \log \left(\int \mathrm{e}^{-z_1^* g_1(x) - z_2^* g_2(x)} \mu(\mathrm{d}x) \right). \end{aligned} \tag{C.4}$$

Furthermore, if the infimum on the right is achieved by some (z_1^*, z_2^*), then the supremum on the left is achieved by

$$f_0(x) = \frac{\mathrm{e}^{-z_1^* g_1(x) - z_2^* g_2(x)}}{\int \mathrm{e}^{-z_1^* g_1(x) - z_2^* g_2(x)} \mu(\mathrm{d}x)}. \tag{C.5}$$

[28] By convention we define $0 \log 0 = 0$, $0 \log(0/0) = 0$, and $x \log(x/0) = \infty$ if $x > 0$.

Proof. Let $\Omega' = \{ f \in \Omega \mid G_2(f) = 0 \}$. Since Ω is convex and G_2 is affine, Ω' is convex. Let α be the supremum of the entropy maximization problem (C.4). If α is finite, by the Lagrange duality theorem C.2 we obtain

$$\begin{aligned}
&\sup \{ H(f) \mid f \in \Omega, G_1(f) \le 0, G_2(f) = 0 \} \\
&= - \inf \{ -H(f) \mid f \in \Omega, G_1(f) \le 0, G_2(f) = 0 \} \\
&= - \inf \{ -H(f) \mid f \in \Omega', G_1(f) \le 0 \} = - \max_{z_1^* \ge 0} \phi(z_1^*),
\end{aligned} \tag{C.6}$$

where $\phi(z_1^*) = \inf_{f\in\Omega'} [-H(f) + \langle G_1(f), z_1^* \rangle]$ is the dual functional. Fix z_1^* and let $A(f) = -H(f) + \langle G_1(f), z_1^*\rangle$. Let us show that

$$\inf_{f\in\Omega'} A(f) = \max_{z_2^* \in Z_2^*} \inf_{f\in\Omega} [A(f) + \langle G_2(f), z_2^*\rangle] . \tag{C.7}$$

To show this, first note that for all z_2^* we have

$$\begin{aligned}
\inf_{f\in\Omega'} A(f) &= \inf_{f\in\Omega'} [A(f) + \langle G_2(f), z_2^*\rangle] \\
&\ge \inf_{f\in\Omega} [A(f) + \langle G_2(f), z_2^*\rangle]
\end{aligned}$$

because $\Omega' \subset \Omega$. Take the maximum on the right with respect to z_2^* and we obtain

$$\inf_{f\in\Omega'} A(f) \ge \max_{z_2^* \in Z_2^*} \inf_{f\in\Omega} [A(f) + \langle G_2(f), z_2^*\rangle] .$$

To show the reverse inequality, by the generalized Kuhn-Tucker theorem C.1, there exists a $z_2^* \in Z_2^*$ such that

$$\inf_{f\in\Omega'} A(f) = \inf_{f\in\Omega} [A(f) + \langle G_2(f), z_2^*\rangle] ,$$

so replacing the specific z_2^* by any such one we obtain

$$\inf_{f\in\Omega'} A(f) \le \max_{z_2^* \in Z_2^*} \inf_{f\in\Omega} [A(f) + \langle G_2(f), z_2^*\rangle] .$$

By (C.6), (C.7), we obtain

$$\begin{aligned}
&\sup \{ H(f) \mid f \in \Omega, G_1(f) \le 0, G_2(f) = 0 \} \\
&= - \max_{z_1^*\ge 0, z_2^* \in Z_2^*} \inf_{f\in\Omega} [-H(f) + \langle G_1(f), z_1^*\rangle + \langle G_2(f), z_2^*\rangle] \\
&= - \max_{z_1^*\ge 0, z_2^* \in Z_2^*} \inf_{f\in\Omega} \int [f \log f + z_1^* g_1 f + z_2^* g_2 f] \, \mathrm{d}\mu.
\end{aligned} \tag{C.8}$$

Suppose that for fixed z_1^*, z_2^* we have

$$Z(z_1^*, z_2^*) := \int \mathrm{e}^{-z_1^* g_1(x) - z_2^* g_2(x)} \mu(\mathrm{d}x) < \infty.$$

Define f_0 as in (C.5). Then by Lemma C.3 for $f \in \Omega$ we get

$$\begin{aligned}
\int [f \log f + z_1^* g_1 f + z_2^* g_2 f] \, \mathrm{d}\mu &= \int [f \log(f/f_0) + (\log Z) f] \, \mathrm{d}\mu \\
&= \int f \log(f/f_0) \mathrm{d}\mu - \log Z \ge - \log Z,
\end{aligned}$$

with equality if and only if $f = f_0$ μ-a.e. Thus

$$(\text{C.8}) = -\max_{z_1^* \geq 0, z_2^* \in Z_2^*} [-\log Z] = \min_{z_1^* \geq 0, z_2^* \in Z_2^*} \log \left(\int \mathrm{e}^{-z_1^* g_1(x) - z_2^* g_2(x)} \mu(\mathrm{d}x) \right).$$

If $Z(z_1^*, z_2^*) = \infty$ for all $z_1^* \geq 0$ and $z_2^* \in Z_2^*$, then define

$$Z_n = \int_{X_n} \mathrm{e}^{-z_1^* g_1(x) - z_2^* g_2(x)} \mu(\mathrm{d}x)$$

and $f_n(x) = \mathrm{e}^{-z_1^* g_1(x) - z_2^* g_2(x)} / Z_n$, where $\{ X_n \}$ is an increasing sequence of measurable sets such that $\mu(X_n) < \infty$ and $X_n \uparrow X$. Then by a similar argument and invoking the monotone convergence theorem, we obtain

$$\alpha = (\text{C.8}) \geq -\max_{z_1^* \geq 0, z_2^* \in Z_2^*} [-\log Z_n] = \min_{z_1^* \geq 0, z_2^* \in Z_2^*} \log Z_n \to \infty,$$

which contradicts the finiteness of α. Hence if α is finite, then (C.4) holds, the infimum on the right is achieved by some (z_1^*, z_2^*), and the supremum on the left is achieved by the corresponding f_0 defined in (C.5). □

Proposition C.5. *Let $(X, \mathcal{B}, \mu)$ be a measure space, where X is a topological space, $\mathcal{B}$ is the Borel σ-algebra, and $\mu(X) > 0$. Let $T : X \to \mathbb{R}^C$ be measurable. Then,*

$$f(\xi) := \log \left(\int \mathrm{e}^{-\xi' T(x)} \mu(\mathrm{d}x) \right)$$

is convex and lower semi-continuous on $\operatorname{dom} f$.[29] *Furthermore, f is strictly convex if* $\dim T(\operatorname{supp} \mu) = C$.[30]

Proof. See Proposition B.4 in Toda (2010). □

Proposition C.6. *Let $(X, \mathcal{B}, \mu)$ be as in Proposition C.5 and $\phi : X \to \mathbb{R}$ be measurable. If $\int \mathrm{e}^{t\phi(x)} \mu(\mathrm{d}x) < \infty$ for some $t > 0$, then*

$$\lim_{t \to \infty} \frac{1}{t} \log \left(\int \mathrm{e}^{t\phi} \mathrm{d}\mu \right) = \operatorname{ess\,sup} \phi. \tag{C.9}$$

If ϕ is upper semi-continuous, then (C.9) *is equal to* $\sup \{ \phi(x) \mid x \in \operatorname{supp} \mu \}$.

Proof. Let $v = \operatorname{ess\,sup} \phi = \sup \{ c \mid \mu(\{ x \in X \mid \phi(x) \geq c \}) > 0 \}$ and $E_n = \{ x \in X \mid \phi(x) \geq -n \}$. If $\mu(E_n) = 0$ for all n, then

$$\mu(X) = \mu \left(\bigcup_{n=1}^{\infty} E_n \right) = \lim_{n \to \infty} \mu(E_n) = 0,$$

which contradicts $\operatorname{supp} \mu \neq \emptyset$. Therefore $\mu(E_n) > 0$ for some n, hence $v > -\infty$.

Let us first prove (C.9) when $v < \infty$. Define

$$X_+ = \{ x \in X \mid \phi(x) > v \}, \ X_- = \{ x \in X \mid \phi(x) \leq v \}, \text{ and}$$
$$X_n = \left\{ x \in X \,\middle|\, \phi(x) \geq v + \frac{1}{n} \right\}.$$

[29] $\operatorname{dom} f = \{ x \in X \mid f(x) < \infty \}$ is the domain of f.

[30] For a subset A of a vector space, $\dim A$ denotes the dimension of the smallest affine space that contains A.

Since $X_+ = \bigcup_{n=1}^{\infty} X_n$ and $\mu(X_n) = 0$ by the definition of v, we have $\mu(X_+) = 0$. Obviously, $X_\pm$ are disjoint and $X_+ \cup X_- = X$, so $\mu(X_-) = \mu(X) > 0$. Fix $t_0 > 0$ such that $\int \mathrm{e}^{t_0\phi(x)}\mathrm{d}\mu < \infty$. Then, for all $t > 0$ we obtain

$$\int \mathrm{e}^{t\phi(x)}\mathrm{d}\mu = \mathrm{e}^{tv} \int_{X_-} \mathrm{e}^{t(\phi(x)-v)}\mathrm{d}\mu. \tag{C.10}$$

Denote the integral over X_- in (C.10) by $I(t)$. Since $\phi(x) \le v$ for $x \in X_-$, for each $x \in X_-$ the integrand $\mathrm{e}^{t(\phi(x)-v)}$ is decreasing in t, so $I(t)$ is decreasing in t. (In particular, $0 < I(t) < \infty$ for $t \ge t_0$.) Hence for $t \ge t_0$ we obtain

$$\frac{1}{t}\log\left(\int \mathrm{e}^{t\phi(x)}\mathrm{d}\mu\right) = v + \frac{1}{t}\log I(t) \le v + \frac{1}{t}\log I(t_0). \tag{C.11}$$

Letting $t \to \infty$ in (C.11), we obtain

$$\limsup_{t\to\infty} \frac{1}{t}\log\left(\int \mathrm{e}^{t\phi(x)}\mathrm{d}\mu\right) \le v. \tag{C.12}$$

To show the reverse inequality, take any $\epsilon > 0$ and let $A = \{\, x \in X_- \mid \phi(x) \ge v - \epsilon \,\}$. By assumption and the definition of $X_\pm$, we have

$$\mu(A) = \mu(\{\, x \in X \mid \phi(x) \ge v - \epsilon \,\}) > 0.$$

By taking a compact subset of A if necessary, by the regularity of μ we may assume $0 < \mu(A) < \infty$. Therefore we obtain

$$\begin{aligned}\frac{1}{t}\log\left(\int \mathrm{e}^{t\phi(x)}\mathrm{d}\mu\right) &\ge \frac{1}{t}\log\left(\mathrm{e}^{t(v-\epsilon)}\int_A \mathrm{e}^{t(\phi(x)-v+\epsilon)}\mathrm{d}\mu\right)\\ &\ge v - \epsilon + \frac{1}{t}\log\mu(A).\end{aligned} \tag{C.13}$$

Letting $t \to \infty$ in (C.13) and then $\epsilon \to 0$, we obtain

$$\liminf_{t\to\infty} \frac{1}{t}\log\left(\int \mathrm{e}^{t\phi(x)}\mathrm{d}\mu\right) \ge v. \tag{C.14}$$

(C.9) follows by (C.12) and (C.14).

If $v = \infty$, let $F_n = \{\, x \in X \mid \phi(x) \ge n \,\}$. By the definition of v, we have $\mu(F_n) > 0$. Then we obtain the same result as (C.13) with A replaced by F_n and $v - \epsilon$ replaced by n. Letting $n \to \infty$ we get (C.9).

Finally let us show $\operatorname{ess\,sup}\phi = \sup\{\, \phi(x) \mid x \in \operatorname{supp}\mu \,\}$ if ϕ is upper semi-continuous. Let $u = \sup\{\, \phi(x) \mid x \in \operatorname{supp}\mu \,\}$. If $u < \infty$, for all $\epsilon > 0$ there exists an $x_0 \in X$ such that $u - \epsilon < \phi(x_0)$. Since ϕ is upper semi-continuous, there exists an open neighborhood U of x_0 such that $x \in U$ implies $\phi(x) > u - \epsilon$. Since $\mu(U \cap X) > 0$ by assumption, it follows that $v \ge u - \epsilon$. Since $\epsilon > 0$ is arbitrary, we obtain $v \ge u$. A similar reasoning holds for the case $u = \infty$.

To show the reverse inequality, take any $\epsilon > 0$. By the definition of v, we have $\mu(\{\, x \in X \mid \phi(x) \ge v - \epsilon \,\}) > 0$. In particular, there exists an $x_0 \in \operatorname{supp}\mu$ such that $\phi(x_0) \ge v - \epsilon$. Therefore,

$$u = \sup\{\, \phi(x) \mid x \in \operatorname{supp}\mu \,\} \ge \phi(x_0) \ge v - \epsilon.$$

Since $\epsilon > 0$ is arbitrary, we obtain $u \ge v$. Therefore $u = v$. □

Theorem C.7. *Let X be a metric space with Borel σ-algebra $\mathcal{B}$. Let μ, $\{ \mu_n \}$ be finite measures defined on $(X, \mathcal{B})$. Let f, $\{ f_n \}$ be measurable functions. Suppose that*

1. *$\mu_n \to \mu$ weakly,* i.e., *for every bounded continuous function g, we have*

$$\lim_{n\to\infty} \int g \mathrm{d}\mu_n = \int g \mathrm{d}\mu,$$

2. *$f_n \to f$ uniformly on compact sets, and f is continuous,*

3. *$\{ f_n \}$ is uniformly integrable,* i.e.,

$$\lim_{\alpha\to\infty} \sup_n \int_{|f_n|>\alpha} |f_n| \, \mathrm{d}\mu_n = 0. \tag{C.15}$$

Then f is μ-integrable and we have

$$\lim_{n\to\infty} \int f_n \mathrm{d}\mu_n = \int f \mathrm{d}\mu.$$

Proof. The claim is mentioned in (Hildenbrand, 1974, pp. 50–52) without proof. The essential steps of the proof (Skorohod's representation theorem and the generalized Lebesgue's convergence theorem) can be found in Billingsley (1999). □

References

Aiyagari, S. R. (1994): "Uninsured Idiosyncratic Risk and Aggregate Saving," *Quarterly Journal of Economics*, 109, 659–684.

Akaike, H. (1992): "Information Theory and an Extension of the Maximum Likelihood Principle," in *Breakthroughs in Statistics*, ed. by S. Kotz and N. L. Johnson, New York: Springer-Verlag, vol. 1, 610–624.

Alfarano, S. and M. Milaković (2008): "Does Classical Competition Explain the Statistical Features of Firm Growth?" *Economic Letters*, 101, 272–274.

Aumann, R. J. (1966): "Existence of Competitive Equilibria in Markets with a Continuum of Traders," *Econometrica*, 34, 1–17.

Berge, C. (1959): *Espaces Topologiques*, Paris: Dunod, English edition: *Topological Spaces*. MacMillan, 1963.

Billingsley, P. (1999): *Convergence of Probability Measures*, Wiley Series in Probability and Statistics, New York: John Wiley & Sons, second ed.

Boltzmann, L. (1877): "On the Relation Between the Second Law of the Mechanical Theory of Heat and the Probability Calculus with Respect to the Theorems on Thermal Equilibrium," *Kais. Akad. Wiss. Wien Math. Naturmiss. Classe*, 76, 373–435.

Borwein, J. M. and A. S. Lewis (1991): "Duality Relationships for Entropy-like Minimization Problems," *SIAM Journal of Control and Optimization*, 29, 325–338.

——— (1992): "Partially Finite Convex Programming, Part I: Quasi Relative Interiors and Duality Theory," *Mathematical Programming*, 57, 15–48.

Castaldi, C. and M. Milaković (2007): "Turnover Activity in Wealth Portfolios," *Journal of Economic Behavior and Organization*, 63, 537–552.

Cover, T. M. and J. A. Thomas (2006): *Elements of Information Theory*, Hoboken, NJ: John Wiley & Sons, second ed.

Debreu, G. (1987): *Theory of Value*, Cowles Foundation Monograph 17, Yale University Press.

Foley, D. K. (1994): "A Statistical Equilibrium Theory of Markets," *Journal of Economic Theory*, 62, 321–345.

——— (1996): "Statistical Equilibrium in a Simple Labor Market," *Metroeconomica*, 47, 125–147.

——— (2003): "Statistical Equilibrium in Economics: Method, Interpretation, and an Example," in *General Equilibrium: Problems and Prospects*, ed. by F. Petri and F. Hahn, London and New York: Routledge, chap. 4.

Folland, G. B. (1999): *Real analysis: modern techniques and their applications*, Hoboken, NJ: John Wiley & Sons, second ed.

Gibbs, J. W. (2008): *Elementary Principles in Statistical Mechanics*, BiblioBazaar, LLC, (reprint of 1901 ed.).

Grossman, S. J. and J. E. Stiglitz (1980): "On the Impossibility of Informationally Efficient Markets," *American Economic Review*, 70, 393–408.

Hildenbrand, W. (1974): *Core and Equilibria of a Large Economy*, Princeton, NJ: Princeton University Press.

Hirshleifer, J. (1973): "Where Are We in the Theory of Information?" *American Economic Review*, 63, 31–39.

Jaynes, E. T. (1957): "Information Theory and Statistical Mechanics, I," *Physical Review*, 106, 620–630.

——— (1968): "Prior Probabilities," *IEEE Transactions on Systems Science and Cybernetics*, SSC-4, 227–241.

——— (1982): "On the Rationale of Maximum-Entropy Methods," *Proceedings of the IEEE*, 70, 939–952.

Krebs, T. (1997): "Statistical Equilibrium in One-Step Forward Looking Economic Models," *Journal of Economic Theory*, 73, 365–394.

Laplace, P. S. (1812): *Théorie Analytique des Probabilités*, Paris: Courcier.

Ljungqvist, L. and T. J. Sargent (2004): *Recursive Macroeconomic Theory*, Cambridge, Massachusetts: MIT Press, second ed.

Luenberger, D. G. (1969): *Optimization by Vector Space Methods*, John Wiley & Sons.

Mas-Collel, A., M. D. Whinston, and J. R. Green (1995): *Microeconomic Theory*, Cambridge, Massachusetts: Oxford University Press.

McCall, J. J. (1970): "Economics of Information and Job Search," *Quarterly Journal of Economics*, 84, 113–126.

Rockafellar, R. T. (1970): *Convex Analysis*, Princeton, NJ: Princeton University Press.

Shannon, C. E. (1948): "A Mathematical Theory of Communication," *Bell System Technical Journal*, 27, 379–423, 623–656.

Shore, J. E. and R. W. Johnson (1980): "Axiomatic Derivation of the Principle of Maximum Entropy and the Principle of Minimum Cross-Entropy," *IEEE Transactions on Information Theory*, IT-26, 26–37.

Silver, J., E. Slud, and K. Takamoto (2002): "Statistical Equilibrium Wealth Distributions in an Exchange Economy with Stochastic Preferences," *Journal of Economic Theory*, 106, 417–435.

Simon, H. A. (1959): "Theories of Decision-Making in Economics and Behavioral Science," *American Economic Review*, 49, 253–283.

Theil, H. (1967): *Economics and Information Theory*, Amsterdam: North-Holland.

Toda, A. A. (2010): "Existence of a Statistical Equilibrium for an Economy with Endogenous Offer Sets," *Economic Theory*, `http://www.springerlink.com/content/n2r33513r5g1225m/`.

Van Campenhout, J. M. and T. M. Cover (1981): "Maximum Entropy and Conditional Probability," *IEEE Transactions on Information Theory*, IT-27, 483–489.

Yoshikawa, H. (2003): "The Role of Demand in Macroeconomics," *Japanese Economic Review*, 54, 1–27.

Zellner, A. (1988): "Optimal Information Processing and Bayes's Theorem," *American Statistician*, 42, 278–280.