Title
Bayesian general equilibrium

Journal
Economic Theory, 58(2)

ISSN
0938-2259

Author
Toda, AA

Publication Date
2015-02-01

DOI
10.1007/s00199-014-0849-4

Peer reviewed

Bayesian General Equilibrium

Alexis Akira Toda*†

This Version: August 21, 2014

Abstract

In this paper I build a general equilibrium model of non-optimizing agents that respond to aggregate variables (prices and average demand profile of agent types) by putting a 'prior' on their demand. An interim equilibrium is defined by the posterior demand distribution of agent types conditional on market clearing. A Bayesian general equilibrium (BGE) is an interim equilibrium such that aggregate variables are correctly anticipated. Under weak conditions I prove the existence and the informational efficiency of BGE. I discuss the conditions under which the set of Bayesian and Walrasian equilibria coincide and show that the Walrasian equilibrium arises from a large class of non-optimizing behavior.

Keywords: Bayes rule; distribution; Kullback-Leibler information; maximum entropy.

JEL codes: C11, D03, D3, D51, D83.

1 Introduction

Most of economic theory centers around optimizing behavior, but there is now a growing evidence in the behavioral economics literature that real human beings do not necessarily fully optimize (Kahneman and Tversky, 1979). What if agents do not optimize but only behave according to a satisficing behavioral rule (Simon, 1959)? In a classic paper Becker (1962) has shown that the downward-sloping aggregate demand can be obtained even under a large class of non-optimizing behavior at the individual level. In this paper I develop a general equilibrium model of satisficing behavior and study the conditions under which the equilibrium coincides to the Walrasian equilibrium.

I model a behavioral rule of an agent type by a correspondence that maps aggregate variables such as the price vector and the average demand profile of all agent types (which can be regarded as a sort of 'reference point' as in Kahneman and Tversky (1979)) to a prior probability distribution of demand

*Department of Economics, University of California San Diego. Email: atoda@ucsd.edu

†This paper has been first drafted in 2009 and circulated under various titles. I thank Sylvain Barde, Andrew Barron, Truman Bewley, Gaetano Bloise, Donald Brown, Alessandro Citanna, Duncan Foley, Xavier Gabaix, John Geanakoplos, Sander Heinsalu, Johannes Hörner, Kazuya Kamiya, Mishael Milaković, Kohta Mori, Herakles Polemarchakis, Larry Samuelson, Paolo Siconolfi, and seminar participants at Yale and ESHIA 2010 for comments and feedback. I am grateful to two anonymous referees for comments and suggestions that significantly improved the paper. The financial supports from the Cowles Foundation, the Nakajima Foundation, and Yale University are greatly acknowledged.

over the consumption set. This correspondence (which depends on the price vector) can be interpreted as a generalization of the usual demand function, the only difference being that agents need not optimize and the demand may be random. Because agents may also violate their budget constraints (nothing prevent them from dreaming of whatever trade opportunities they like), these demands may not be feasible. I assume that agents are sophisticated enough so that if the demands are infeasible, the agents revise their demand distribution by applying the Bayes rule. An interim equilibrium is defined by the posterior demand distributions of all types conditional on market clearing, which is achieved by minimizing the Kullback-Leibler information of the distributions subject to the feasibility constraint. Such interim equilibria naturally give rise to the shadow price of each commodity and an updated average demand profile. I define a Bayesian general equilibrium (correct expectations equilibrium) by an interim equilibrium that is correctly anticipated—it is a pair of the price vector, the average demand profile, and posterior demand distributions such that the price and the average demand profile are self-fulfilling and markets clear.

This paper has three main contributions. First, I prove the existence of Bayesian general equilibrium under much weaker assumptions than in the previous literature, which will be discussed shortly. Second, I prove that the Bayesian general equilibrium is informationally efficient in the sense that it achieves the best possible trade-off between the lack of information gain (surprise) and the lack of arbitrage (getting a good deal by chance). Third, I prove that if for each agent type the support of the prior probability distribution (called *offer set*) is contained in the upper contour set of the average demand and coincide to the upper contour set within the budget set whenever the average demand is budget feasible, then the set of Bayesian and Walrasian equilibria exactly coincide. Therefore the Walrasian equilibrium is robust in the sense that it arises from a large class of non-optimizing behavior.

This paper is broadly related to two strands of literature. The first is the statistical equilibrium model of markets developed by Foley (1994, 1996, 2003) and extended by Toda (2010). According to Foley (1994), a statistical market equilibrium is defined by the maximum entropy distribution over transactions subject to the acceptability and the feasibility constraints. In this paper I clarify the relation between maximum entropy and Bayesian inference, which seems to be still relatively unknown in the economics literature. Since information-theoretic techniques are recently more and more applied to economics (Krebs, 1997; Sims, 2003; Veldkamp, 2011; Cabrales et al., 2013), the results on maximum entropy and Bayesian inference presented in this paper might be of interest to economists. With this clarification, Foley's statistical equilibrium is precisely an interim equilibrium that I define in this paper. In Toda (2010), agents' prior demand distributions are not fixed but depend on price, but in order to prove the existence of equilibrium I assumed in my earlier paper that a strictly feasible allocation always exists, which is a strong assumption that is incompatible with the Arrow-Debreu model. In this paper I refine the equilibrium concept and weaken the assumptions of equilibrium existence. Most importantly, I dispose of the assumption that a strictly feasible allocation always exists, which enables me to include standard Walrasian economies as special cases of my model.

Second, this paper is also related to the literature of getting the Nash and Walras equilibria from satisficing behavior (McKelvey and Palfrey, 1995; Geanakoplos, 2003; Becker and Chakrabarti, 2005; Chakrabarti, 2013). Roughly speaking, in this paper I prove that when (within the budget set) the offer set of each agent type coincides to the upper contour set of the average demand, then the set of Walrasian and Bayesian equilibria exactly coincide. Therefore Walrasian equilibrium is robust in the sense that it arises from a large class of non-optimizing behavior. My result is also the first rigorous formulation and proof of Foley (2003)'s conjecture that "there may be a sense in which Walrasian equilibrium can be viewed as an asymptotic approximation to statistical equilibrium." In a recent paper, Chakrabarti (2013) shows that the set of his equilibria exactly coincides to that of Walrasian equilibria under different assumptions and equilibrium concept than mine. The similarities and differences between my result and his are discussed in Section 4.

2 Model

2.1 Economy

There are I agent types indexed by $i = 1, 2, \dots, I$. Type i consists of a continuum of agents with mass $n_i > 0$, where $\sum_i n_i = 1$. There are C commodities labeled by $c = 1, 2, \dots, C$, and the commodity space is X, a nonempty, closed, convex subset of $\mathbb{R}^C_+$. Agents of the same type are ex ante identical and are endowed with a commodity bundle $e_i \in X$.

Let $\Delta^{C-1} = \left\{ p \in \mathbb{R}^C_+ \,\middle|\, \sum_{c=1}^C p_c = 1 \right\}$ be the usual price simplex. The only difference of my model from the standard Walrasian (Arrow-Debreu) model is that agents do not necessarily respond to prices optimally. Instead, agents act probabilistically (and independently) according to a satisficing behavioral rule, taking aggregate variables as given. I assume that the aggregate variables upon which agents base their behavior are the price vector $p \in \Delta^{C-1}$ and the profile of the average demand of each type $x = (x_i)_{i \in I} \in X^I$. x_i can be regarded as a sort of 'reference point' of type i agents.

In the context of the standard Arrow-Debreu model (utility maximization), agents respond only to prices, not on average demand. The decision rule or the "new" demand is merely the utility maximizing bundle. In my non-optimizing ("satisficing") context, type i agents' decision rule is characterized by a probability measure $\mu_i(p, x)$ that is supported on a subset of the commodity space X, where p is the price vector and $x = (x_i) \in X^I$ is the profile of the average demand of each type. A type i agent will draw a random demand from the distribution μ_i. I refer to the probability measure $\mu_i(p, x)$ as the *prior*, because it is the distribution of actions (demands) that type i agents take given the aggregate variables (p, x) before updating any information.

Let $\mathcal{M}(X)$ be the set of all probability measures on the commodity space X. Then $\mu_i(p, x) \in \mathcal{M}(X)$. I formally define an economy as follows.

Definition 2.1. An economy is a quadruple $\mathcal{E} = \{I, \{n_i\}, \{e_i\}, \{\mu_i\}\}$, where $I = \{1, \dots, I\}$ is the set of agent types, $n_i > 0$ is the mass of type i agents with $\sum_{i=1}^I n_i = 1$, $e_i \in X$ is the endowment of type i, and $\mu_i : \Delta^{C-1} \times X^I \to \mathcal{M}(X)$ is a mapping that maps the price vector p and average demand profile $x = (x_i)_{i \in I}$ to a (prior) probability measure $\mu_i(p, x)$ on the commodity space X. The support[1] of the prior, $X_i(p, x) = \operatorname{supp} \mu_i(p, x)$, is called the *offer set* of

[1] The *support* of a Borel measure μ on a second-countable topological space X, denoted

type i given price p and average demand x.[2]

Thus the offer set $X_i(p,x)$ consists of demands $y \in X$ that type i agents take with positive probability given aggregate variables (p,x). If $y \notin X_i(p,x)$, then the agent never demands y.

The reason why I allow the prior measure $\mu_i(p,x)$ to depend on price p as well as the average demand profile x is intuitive. First, the price affects the value of agents' initial endowment and thus their ability to spend. Second, agents may change their behavior by learning what other agents are doing on average, such as "catching up with Joneses."

Clearly the classic utility maximization (Walrasian) framework is a special case of my model.

Example 1. Consider the standard Walrasian economy with I agent types, where type i has mass $n_i > 0$ and a continuous, strictly quasi-concave utility function u_i. The optimal behavior is

$$x_{i,p} = \arg\max_{y \in \mathbb{R}^C_+} \{u_i(y) \mid p \cdot y \le p \cdot e_i\}.$$

Setting $\mu_i(p,x) =$ "counting measure on $x_{i,p}$" (thus the average demand profile x does not influence agent behavior), $\mathcal{E} = \{I, \{n_i\}, \{e_i\}, \{\mu_i\}\}$ is an economy in the sense of Definition 2.1.

One way to think of the prior μ_i is the subjective probability measure over bundles that type i agents are willing to hold. In the Walrasian case of Example 1, agents completely trust the market mechanism and demand only (puts the Dirac measure on) the utility maximizing bundle. In my model, agents are not so certain and their beliefs are more spread. Another view is that agents observe their utility function subject to some noise, which translates into their behavior. The quantal response equilibrium model of McKelvey and Palfrey (1995) is such an example.

In this paper I do not address the question of how the priors $\{\mu_i\}$ are formed from more primitive notions such as preferences. There are two reasons for this approach. First, there can be many ways to model satisficing behavior, each of them leading to different priors. But at the abstract level of modeling in this paper, no particular modeling choice is more plausible than another. Second, in this paper I am concerned with existence and informational efficiency of equilibrium and its relation to the Walrasian equilibrium. It turns out that I can answer these questions without specifying a particular behavioral rule, so there is no need to model the agent behavior from a more primitive level. This point contrasts with the classical general equilibrium theory in which the preferences of agents are primitives of the model. In my model, the primitive is the prior (behavioral rule) $\mu_i(p,x)$, which corresponds to the demand function in the classical setting.

by $S = \operatorname{supp}\mu$, consists of the points that do not have a neighborhood with measure zero. Let $\mathcal{U}$ be the 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 $U \cap S \neq \emptyset$. Hence $S = \operatorname{supp}\mu$.

[2] I follow Foley (1994) in calling $X_i(p,x)$ the offer set, but a more appropriate term might be *offer correspondence*, since sets are parametrized by the price vector p and the average demand profile x.

2.2 Equilibrium

Given an economy $\mathcal{E} = \{I, \{n_i\}, \{e_i\}, \{\mu_i\}\}$, how should we define the equilibrium concept? As in any equilibrium concept, markets have to clear. However, since so far there is nothing that keeps agents from violating their budget constraints (they are free to believe whatever trade opportunities they want), these demands may not be met. This implies that agents' belief are typically incompatible with market clearing, so in order to clear the market agents need to update their beliefs (prior over demand). Thus the question is *how* agents update their beliefs. The obvious and natural answer is that agents should revise their demand distribution by applying the Bayes rule conditional on market clearing.

In our context, since there are a continuum of agents, the calculation of posterior distributions using the Bayes rule is not obvious. However, this can be accomplished by using the "equivalence" between Bayesian inference and the minimization of the Kullback-Leibler information (maximum entropy), as explained in Appendix A.

To define the equilibrium, I proceed in two steps. First, I define the updating rules for price and average demand. This part is similar to Foley (1994)'s "statistical equilibrium" model of markets, which Foley (2003) interprets as a temporary equilibrium concept. Second, I define the equilibrium (Bayesian general equilibrium) by self-fulfilling price and average demand.

2.2.1 Interim equilibrium

Let $\mathcal{E} = \{I, \{n_i\}, \{e_i\}, \{\mu_i\}\}$ be an economy and $f_i(y)$ be the probability density function of the posterior (to be computed) with respect to the prior $\mu_i(p, x)$. Then

$$\int y f_i(y) \mu_i(\mathrm{d}y; p, x)$$

is the expected *ex post* demand of type i. Since I assumed that agents act independently conditional on aggregate variables (p, x) and that there are a continuum of agents in each type, by the law of large numbers for a continuum of random variables (Uhlig, 1996) the total demand of type i agents is

$$n_i \int y f_i(y) \mu_i(\mathrm{d}y; p, x)$$

almost surely. Hence by letting $f = (f_i)_{i \in I}$ and $\mu(p, x) = (\mu_i(p, x))_{i \in I}$, we can define the aggregate demand in the economy by

$$\bar{x}[f; \mu(p, x)] := \sum_{i=1}^{I} n_i \int y f_i(y) \mu_i(\mathrm{d}y; p, x). \tag{2.1}$$

Next, in order to use the equivalence between the Bayes rule and maximum entropy, we need to define the Kullback-Leibler information. In our context, the Kullback-Leibler information of the densities $f = (f_i)_{i \in I}$ relative to the priors $\mu(p, x) = (\mu_i(p, x))_{i \in I}$ is defined by

$$H[f; \mu(p, x)] := \sum_{i=1}^{I} n_i \int f_i(y) \log f_i(y) \mu_i(\mathrm{d}y; p, x).^3 \tag{2.2}$$

[3] This definition can be understood as follows. If we take the prior measure $\mu_i(p, x)$ as the

With notations and terminologies adapted to my situation, Foley (1994) defined his "statistical equilibrium" concept as follows.

Definition 2.2 (Interim equilibrium)**.** Let $\mathcal{E} = \{I, \{n_i\}, \{e_i\}, \{\mu_i\}\}$ be an economy with offer sets $X_i(p,x) = \operatorname{supp} \mu_i(p,x)$. Given price $p \in \Delta^{C-1}$ and the average demand profile $x = (x_i)_{i \in I} \in X^I$, the collection of densities $f = (f_i)$ and the vector $\pi \in \mathbb{R}^C_+$ constitute a *interim equilibrium* if f_i's are posterior densities conditional on market clearing, *i.e.*, $f = (f_i)$ solves

$$\min_f H[f; \mu(p,x)] \text{ subject to } \bar{x}[f; \mu(p,x)] \leq \bar{e}, \tag{2.3}$$

where $H[\cdot; \mu(p,x)]$ is the Kullback-Leibler information defined in (2.2), $\bar{x}[\cdot; \mu(p,x)]$ is the aggregate demand defined in (2.1), $\bar{e} = \sum_{i=1}^I n_i e_i$ is the aggregate endowment, and π is the corresponding Lagrange multiplier on the feasibility constraint in (2.3).

The idea of the interim equilibrium is as follows. Agents come to the market with their prior behavioral rule $\mu_i(p,x)$. But they realize that markets may not clear with these rules, so we need some kinds of "rationing scheme" in order to clear the market. Specifically, our agents revise their behavior by the Bayes rule conditional on market clearing, which is equivalent to minimizing the Kullback-Leibler information of the entire economy subject to the feasibility constraint according to the results in Appendix A. Therefore although agents are not sophisticated in that they do not fully optimize, they are sophisticated enough to apply the Bayes rule.

Implicit in the minimization problem (2.3) is that $f = (f_i)$ is constrained to be a collection of probability density functions, so $f_i \geq 0$ and $\int f_i \mathrm{d}\mu_i(p,x) = 1$ for all i. The interim equilibrium in Definition 2.2 is essentially a partial equilibrium concept in which the price and the average demand profile are exogenously given. It is exactly the statistical equilibrium defined by Foley (1994) except that he uses finite discrete sets as offer sets and motivates the equilibrium by entropy maximization, not Bayesian updating or minimum information principle. Foley (1994) calls the Lagrange multiplier π the *entropy price* because it is the shadow price of commodities in units of entropy (Kullback-Leibler information).

In the literature of general equilibrium with price rigidities, many rationing schemes have been proposed (Drèze, 1975; Herings, 1996). These schemes typically have upper bounds on demand and supply, whereas in my model agents do not change the support of the distribution but only tilt the probabilities of demanding particular consumption bundles. Any rationing scheme is ad hoc in

---

reference measure μ in (A.2) and noting that f_i corresponds to the posterior density, we get $q = 1$ and hence the Kullback-Leibler information of a single agent of type i is

$$H_i = H(f_i; 1) = \int f_i \log f_i \mathrm{d}\mu_i(p,x).$$

Now suppose that there are N_i agents of type i, and let the total number of agents be $N = \sum_i N_i$ and the proportion be $n_i = N_i/N$. In general, the entropy of the joint distribution of two independent random variables is the sum of the entropy of each variable.[4] The additivity of the entropy carries over to the Kullback-Leibler information. Therefore, the economy-wide information is

$$H = \sum_{i=1}^I N_i H_i = \sum_{i=1}^I N_i \int f_i \log f_i \mathrm{d}\mu_i(p,x).$$

Dividing this expression by N, we obtain the per capita information (2.2).

the sense that none is more convincing than any other. Foley's choice of using (only) maximum entropy principle (Bayes rule) is natural because it is the only rational way of updating beliefs in the presence of new information according to the axiomatization of Jaynes (2003) and Knuth and Skilling (2012).

Of course, in order to solve the minimum information problem (2.3) agents need to know a great deal about the economy. Not only they need to know the quoted price p and the average demand profile $x \in X^I$, but also the priors of all agent types $\{\mu_i(p,x)\}_{i\in I}$. This is of course informationally demanding and unrealistic, but agents can learn a great deal from rounds of the following experiments. At each round, each agent expresses his demand randomly drawn from his prior. Then the "auctioneer" calculates the aggregate demand and tell the agents whether the allocation was feasible or not. By repeating these rounds, which are similar to the setting in Van Campenhout and Cover (1981), agents will obtain their posterior conditional on market clearing.

2.2.2 Bayesian general equilibrium

An interim equilibrium is a rule for updating demand, taking price and average demand profile as given. In order to define the full (general) equilibrium, we need updating rules for price and average demand profile. Let $((f_i),\pi)$ be an interim equilibrium corresponding to price $p \in \Delta^{C-1}$ and average demand profile $x = (x_i) \in X^I$. Since f_i is a density,

$$x_i' = \int y f_i(y)\mu_i(\mathrm{d}y;p,x)$$

is a natural candidate for updating the average demand. Since $\pi \in \mathbb{R}^C_+$ is the Lagrange multiplier corresponding to the feasibility constraint, π_c is the shadow price of commodity c (the amount of Kullback-Leibler information reduced by injecting one unit of commodity c in the economy). Therefore normalizing π such that

$$p' = \frac{\pi}{\sum_{c=1}^C \pi_c}$$

is a natural candidate for updating the price.

Based on this argument I define a (non-degenerate) correct expectations equilibrium as follows. (The definition is essentially due to Toda (2010).)

Definition 2.3 (Correct expectations equilibrium). Let $\mathcal{E} = \{I, \{n_i\}, \{e_i\}, \{\mu_i\}\}$ be an economy with offer sets $X_i(p,x) = \operatorname{supp}\mu_i(p,x)$. The price $p \in \Delta^{C-1}$, average demand profile $x = (x_i) \in X^I$, and collection of densities (f_i) constitute a (non-degenerate) *correct expectations equilibrium* if

1. f_i's are posterior densities conditional on market clearing, *i.e.*, $f = (f_i)$ solves (2.3),

2. the Lagrange multiplier π to (2.3) is proportional to p,

3. the average demand profile is consistent, *i.e.*, for all i

$$x_i = \int y f_i(y)\mu_i(\mathrm{d}y;p,x).$$

The idea of the correct expectations equilibrium is to push the degree of sophistication of agents one step further than the interim equilibrium. Here agents not only use the Bayes rule for updating their demand. They are aware that their change in behavior will shift the price and average demand profile, so they agree on a price and average demand profile that are self-fulfilling. In many institutions there is a mechanism for revising prices that plays the role of the market auctioneer. For instance, in the (equilibrium existence proof of the Arrow-Debreu model), agents respond optimally to the quoted price, and the Walrasian auctioneer updates the price given the revealed demand. As a result, the equilibrium is defined by a rest point where prices are correctly anticipated. Thus we can interpret the definition of the correct expectations equilibrium as a rest point where there is no further need of updating.

In cases that the market clearing condition is extremely tight, it may happen that we cannot find any feasible posteriors that are absolutely continuous with respect to the priors. The following definition of degenerate equilibria handles those cases. Degenerate equilibria are intuitively (and mathematically: see Proposition 3.5) the asymptotic limit of non-degenerate ones.

Definition 2.4 (degenerate equilibrium)**.** Let $\mathcal{E} = \{I, \{n_i\}, \{e_i\}, \{\mu_i\}\}$ be an economy with offer sets $X_i(p,x) = \operatorname{supp} \mu_i(p,x)$. A price vector p and an average demand profile $x = (x_i) \in X^I$ constitute a *degenerate equilibrium* if

4. $x_i \in \operatorname{cl} \operatorname{co} X_i(p,x)$ for all i,[5]
5. $\sum_{i=1}^I n_i(x_i - e_i) \leq 0$, and $p_c = 0$ if $\sum_{i=1}^I n_i(x_{ic} - e_{ic}) < 0$,
6. for all i and $y \in X_i(p,x)$, we have $p \cdot y \geq p \cdot x_i$.

Condition 4 means that the average demand profile $x = (x_i)$ is consistent. Since the demand y of a type i agent must lie on the offer set $X_i(p,x)$, by averaging y across all type i agents the average demand x_i must belong to $\operatorname{cl} \operatorname{co} X_i(p,x)$, since averaging has convexifying effects. Condition 5 means that the average demand profile is feasible, and that the price of a commodity in excess supply is zero. Condition 6 means that the offer set $X_i(p,x)$ is supported by the hyperplane $p \cdot y = p \cdot x_i$.

A rationale of the degenerate equilibrium is that consistency (condition 4) and market clearing (condition 5) are minimum requirements, and that we can reduce the Kullback-Leibler information if condition 6 fails by "spreading" densities over the offer sets, contradicting the minimum information principle.

A degenerate or non-degenerate correct expectations equilibrium is simply referred to as a *Bayesian general equilibrium*, for the obvious reason that agents use the Bayes rule for updating their beliefs.

3 Existence and informational efficiency

3.1 Existence and uniqueness of interim equilibrium

Since a Bayesian general equilibrium is an interim equilibrium with correct expectations (self-fulfilling price and average demand profile), the natural starting point for studying existence is that of interim equilibrium. For subsets X, Y of a

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

vector space, let $\alpha X + \beta Y := \{\alpha x + \beta y \mid x \in X, y \in Y\}$. The following theorem shows the existence and uniqueness of an interim equilibrium when there is a strictly feasible allocation.

Theorem 3.1 (Existence, uniqueness, duality)**.** *Let price $p \in \Delta^{C-1}$ and average demand profile $x = (x_i) \in X^I$ be given. Let $\mathcal{E} = \{I, \{n_i\}, \{e_i\}, \{\mu_i\}\}$ be an economy with offer sets $X_i(p,x) = \operatorname{supp} \mu_i(p,x)$ and define the dual function $H^* : \mathbb{R}^C_+ \to \mathbb{R}$ by*

$$H^*(\xi) = \sum_{i=1}^I n_i \left[\xi' e_i + \log \left(\int \mathrm{e}^{-\xi' y} \mu_i(\mathrm{d}y; p, x) \right) \right].$$

If $\int \|y\| \, \mu_i(\mathrm{d}y; p, x) < \infty$ and

$$\left(\sum_{i=1}^I n_i (\operatorname{co} X_i(p,x) - e_i) \right) \cap (-\mathbb{R}^C_{++}) \neq \emptyset, \tag{3.1}$$

then there exists a unique interim equilibrium $f = (f_i)$ and

$$\min_f \{H[f; \mu(p,x)] \mid \bar{x}[f; \mu(p,x)] \le \bar{e}\} = - \min_{\xi \in \mathbb{R}^C_+} H^*(\xi), \tag{3.2}$$

where $\bar{e} = \sum_{i=1}^I n_i e_i$ is the aggregate endowment. Furthermore, letting $\pi \in \mathbb{R}^C_+$ be a Lagrange multiplier for the left-hand side minimization of (3.2), π is also a solution to the right-hand side minimization of (3.2), and the (unique) solution to the left-hand side minimization is $f_i(y) = \mathrm{e}^{-\pi' y} / \int \mathrm{e}^{-\pi' y} \mu_i(\mathrm{d}y; p, x)$.

Proof. Since the price p and the average demand profile $x = (x_i) \in X^I$ are fixed, in what follows I suppress p, x.

Step 1. Proof of duality (3.2)*.*

Consider the primal problem

$$\inf_{f=(f_i)} \sum_{i=1}^I n_i \int f_i \log f_i \mathrm{d}\mu_i \text{ subject to } \sum_{i=1}^I n_i \int y f_i(y) \mu_i(\mathrm{d}y) \le \bar{e}. \tag{P}$$

By Lemma 3.2 of Toda (2010) and (3.1), there exists a strictly feasible $f^1 = (f_i^1)$, that is,

$$\sum_{i=1}^I n_i \int f_i^1 \mathrm{d}\mu_i \ll \bar{e} = \sum_{i=1}^I n_i e_i,$$

so the regularity condition of Fenchel duality (Borwein and Lewis, 1991, Corollary 2.6) holds. Letting

$$\phi(t) = \begin{cases} t \log t, & (t > 0) \\ 0, & (t = 0) \\ \infty, & (t < 0) \end{cases}$$

we can ignore the non-negativity constraints $f_i \ge 0$ in (P). Since the convex conjugate function of ϕ is

$$\phi^*(s) = \sup_{t \in \mathbb{R}} [st - \phi(t)] = \sup_{t \ge 0} [st - t \log t] = \mathrm{e}^{s-1},$$

the dual problem of (P) is

$$\sup_{\nu\in\mathbb{R}^I,\xi\in\mathbb{R}^C_+}\sum_{i=1}^I n_i\left[\nu_i-\xi'e_i-\int \mathrm{e}^{\nu_i-\xi'y-1}\mu_i(\mathrm{d}y)\right], \tag{D}$$

where $\xi\in\mathbb{R}^C_+$ is the Lagrange multiplier to the feasibility constraint $\bar{x}[f;\mu]\le\bar{e}$ and $\nu_i\in\mathbb{R}$ is the Lagrange multiplier to $\int f_i\mathrm{d}\mu_i=1$ (accounting of probability). By the Fenchel duality theorem, the optimal value $H_{\min}$ of (P) and (D) coincide and (D) has a solution. Taking the partial derivative of the objective function in (D) with respect to ν_i, we obtain

$$1-\int \mathrm{e}^{\nu_i-\xi'y-1}\mu_i(\mathrm{d}y)=0 \iff \nu_i=1-\log\left(\int \mathrm{e}^{-\xi'y}\mu_i(\mathrm{d}y)\right).$$

Substituting ν_i into (D), after some algebra we get

$$H_{\min}=-\min_{\xi\in\mathbb{R}^C_+}\sum_{i=1}^I n_i\left[\xi'e_i+\log\left(\int \mathrm{e}^{-\xi'y}\mu_i(\mathrm{d}y)\right)\right], \tag{D'}$$

which is precisely the right-hand side of (3.2).

Step 2. Proof that $f_i(y)=\mathrm{e}^{-\pi'y}/\int \mathrm{e}^{-\pi'y}\mu_i(\mathrm{d}y)$ is a solution to (3.2).

Since μ_i is a finite measure and $\int\|y\|\,\mu_i(\mathrm{d}y)<\infty$, by Lebesgue's dominated convergence theorem $\int \mathrm{e}^{-\xi'y}\mu_i(\mathrm{d}y)$ is C^1 as a function of $\xi\in\mathbb{R}^C_+$. Letting $\pi\in\mathbb{R}^C_+$ be a solution to (3.2), by the Karush-Kuhn-Tucker theorem there exists a Lagrange multiplier $\lambda\in\mathbb{R}^C_+$ such that $\lambda_c\pi_c=0$ for all c and

$$0=\sum_{i=1}^I n_i\left[e_i-\frac{\int y\mathrm{e}^{-\pi'y}\mu_i(\mathrm{d}y)}{\int \mathrm{e}^{-\pi'y}\mu_i(\mathrm{d}y)}-\lambda\right]. \tag{3.3}$$

Letting $f_i(y)=\mathrm{e}^{-\pi'y}/\int \mathrm{e}^{-\pi'y}\mu_i(\mathrm{d}y)$, (3.3) shows that $f=(f_i)$ is feasible. Multiplying π to (3.3) as an inner product, we get

$$0=\pi'\sum_{i=1}^I n_i\left[e_i-\int yf_i\mathrm{d}\mu_i\right]. \tag{3.4}$$

Substituting $f=(f_i)$ into the objective function in (P) and invoking (3.4), the Kullback-Leibler information is

$$\begin{aligned}H[f;\mu]&=\sum_{i=1}^I n_i\int f_i\log f_i\mathrm{d}\mu_i=\sum_{i=1}^I n_i\int f_i\left[-\pi'y-\log\left(\int \mathrm{e}^{-\pi'y}\mathrm{d}\mu_i\right)\right]\mathrm{d}\mu_i\\&=-\sum_{i=1}^I n_i\left[\pi'e_i+\log\left(\int \mathrm{e}^{-\pi'y}\mathrm{d}\mu_i\right)\right],\end{aligned}$$

which is precisely the value in (D'). Therefore the minimum of the left-hand side of (3.2) is attained by $f=(f_i)$, which shows that $f=(f_i)$ is an interim equilibrium. (Almost everywhere) uniqueness follows by the strict convexity of the function $\phi(t)=t\log t$. □

Theorem 3.1 is essentially due to Toda (2010). The novelty is to show that Fenchel duality makes clear how to solve the minimum information problem. By Theorem 3.1, the minimum information problem (minimization over a functional space) reduces to the minimization of the dual function H^* (over a Euclidean space), as long as the regularity condition (3.1) is satisfied. Some kind of regularity condition is necessary in order to apply the Fenchel duality theorem in optimization problems like (2.3). However, the regularity condition (3.1) is rather strong from an economic point of view since it implies that there is a strictly feasible allocation under the priors $\mu = \{\mu_i\}_{i\in I}$, but in the Walrasian case (Example 1) there is never a strictly feasible allocation. Fortunately, the following proposition shows that we do not need to assume this rather strong regularity condition.

Proposition 3.2. *Let everything be as in Theorem 3.1 except that we do not assume the strict feasibility* (3.1)*. If the right-hand minimization of* (3.2) *has a solution* $\pi \in \mathbb{R}^C_+$*, then* $f = (f_i)$ *given by* $f_i(y) = \mathrm{e}^{-\pi' y} / \int \mathrm{e}^{-\pi' y} \mu_i(\mathrm{d}y; p, x)$ *is the unique solution of the left-hand minimization of* (3.2)*, and* π *is the corresponding Lagrange multiplier.*

Proof. Since $\phi(t) = t \log t$ is strictly convex, for $s, t > 0$ we have

$$\phi(t) - \phi(s) \geq \phi'(s)(t - s) \iff t \log t \geq (\log s + 1)t - s, \tag{3.5}$$

with equality if and only if $s = t$. The same inequality holds even if $t = 0$ or $s = 0$ provided that we define $0 \log 0 = 0$.

Suppose that $\pi \in \mathbb{R}^C_+$ minimizes H^*. Letting $f_i(y) = \mathrm{e}^{-\pi' y} / \int \mathrm{e}^{-\pi' y} \mu_i(\mathrm{d}y)$, by (3.3) and (3.4) we have $\bar{x}[f; \mu] \leq \bar{e}$ and $\pi'(\bar{x}[f; \mu] - \bar{e}) = 0$. Let $g = (g_i)$ be any (not necessarily feasible) collection of densities. Then setting $t = g_i$ and $s = f_i$ in (3.5) and using $\int f_i \mathrm{d}\mu_i = \int g_i \mathrm{d}\mu_i = 1$, we get

$$\begin{aligned} H[g; \mu] &= \sum_{i=1}^{I} n_i \int g_i \log g_i \mathrm{d}\mu_i \geq \sum_{i=1}^{I} n_i \int [(\log f_i + 1) g_i - f_i] \mathrm{d}\mu_i \\ &= \sum_{i=1}^{I} n_i \int g_i \log f_i \mathrm{d}\mu_i = -\pi'(\bar{x}[g; \mu] - \bar{e}) - H^*(\pi) \\ \iff \quad H[g; \mu] &+ \pi'(\bar{x}[g; \mu] - \bar{e}) \geq -H^*(\pi), \end{aligned} \tag{3.6}$$

with equality if $g = f$. (3.6) shows that $f = (f_i)$ minimizes the Lagrangian corresponding to the left-hand minimization of (3.2) with Lagrange multiplier π. If $g = (g_i)$ is feasible, since $\pi \geq 0$ and $\bar{x}[g; \mu] \leq \bar{e}$, by (3.6) we get

$$H[g; \mu] \geq H[g; \mu] + \pi'(\bar{x}[g; \mu] - \bar{e}) \geq -H^*(\pi).$$

Since $\bar{x}[f; \mu] \leq \bar{e}$ and $\pi'(\bar{x}[f; \mu] - \bar{e}) = 0$, it follows that $H[f; \mu] = -H^*(\pi)$, so $f = (f_i)$ minimizes the Kullback-Leibler information among all feasible densities. Uniqueness follows by the strict convexity of $\phi(t) = t \log t$. □

Before proceeding to the existence of Bayesian general equilibrium, it is important to understand what are assumed and proved in Theorem 3.1 and Proposition 3.2. Theorem 3.1 shows that under a Slater-type constraint qualification (that a strictly feasible allocation exists), the minimum information problem (the left-hand minimization of (3.2)) has a unique solution and it corresponds to a solution to the dual problem (right-hand minimization of (3.2)). Although this result motivates using the dual function, we do not want to assume that there exists a strictly feasible allocation. Then Proposition 3.2 shows that if the dual problem has a solution, then the primal problem (the minimum information problem, the left-hand minimization of (3.2)) has a (unique) solution, and it has the form $f_i(y) = \mathrm{e}^{-\pi' y} / \int \mathrm{e}^{-\pi' y} \mu_i(\mathrm{d}y; p, x)$. Thus in order to obtain a solution to the primal problem, we do not *need* to assume the existence of a strictly feasible allocation; it *suffices* to solve the dual problem. The sole purpose of presenting Theorem 3.1 is to motivate the dual problem.

3.2 Existence of Bayesian general equilibrium

In this subsection I prove the existence of Bayesian general equilibrium under the following assumptions. Each assumption is followed by a justification.

Assumption 1. *For all agent type $i \in I$, price $p \in \Delta^{C-1}$, and average demand profile $x \in X^I$, the prior $\mu_i(p, x)$ is a finite measure. Furthermore,*

$$\sup_{\substack{p \in \Delta^{C-1} \\ x \in X^I}} \int y \mu_i(\mathrm{d}y; p, x) < \infty. \tag{3.7}$$

Since $\mu_i(p, x)$ is a prior (subjective probability measure, hence $\mu_i(X; p, x) = 1$), $\mu_i(p, x)$ is necessarily a finite measure. (3.7) means that the average demand implied by p, x is uniformly bounded. This condition *does not* say that the demand is finite: the offer set $X_i(p, x)$ may well be unbounded, but agents must put less and less probability on $y \in X_i(p, x)$ as y tends to infinity. This assumption seems natural, since agents are aware that aggregate endowment is finite. Assumption 1 trivially holds if the consumption set X is compact and μ_i is a regular Borel measure.

Assumption 2 (Budget feasibility)**.** *For all agent type $i \in I$, price $p \in \Delta^{C-1}$, and average demand profile $x \in X^I$, we have*

$$\inf \{ p \cdot (y - e_i) \mid y \in X_i(p, x) \} \leq 0.$$

Assumption 2 implies that agents are "realistic" in the sense that they always put some probability on allocations that are budget feasible. If agents "aim too high" by demanding only allocations that exceed their budgets, it is obvious that markets cannot clear.

Actually the individual budget feasibility is not necessary for the existence of an equilibrium. We can relax Assumption 2 to

$$\sum_{i=1}^{I} n_i \inf \{ p \cdot (y - e_i) \mid y \in X_i(p, x) \} \leq 0,$$

which means that the total value of acceptable allocations with lowest expenditures is at most the total value of the endowment. Thus, there can be agents that "aim too high" as long as there are enough agents that are conservative enough. (For an analogy, little kids can get what they want if they insist stubbornly and their parents give in.)

Next come two continuity assumptions, which are technical conditions necessary for the proof. These conditions are similar to the properties of the demand correspondences in the classical general equilibrium theory.

Assumption 3 (Continuity of priors)**.** *For all agent type $i \in I$, the mapping*

$$\mu_i : \Delta^{C-1} \times X^I \to \mathcal{M}(X)$$

is weakly continuous, i.e.*, for every bounded continuous function f and sequences $\{p_n\} \subset \Delta^{C-1}$ and $\{x_n\} \subset X^I$ such that $p_n \to p$ and $x_n \to x$, we have*

$$\lim_{n\to\infty} \int f(y)\mu_i(\mathrm{d}y; p_n, x_n) = \int f(y)\mu_i(\mathrm{d}y; p, x).$$

Assumption 3 is satisfied, for instance, if $\{\mu_i(p, x)\}$ is absolutely continuous with respect to a common measure ν_i, the Radon-Nikodym derivative $f_i(y; p, x) := \frac{\mathrm{d}\mu_i(p,x)}{\mathrm{d}\nu_i}$ is continuous in p, x for ν_i-a.e. y, and there exists a ν_i-integrable function g_i such that $|f_i(y; p, x)| \le g_i(y)$ for ν_i-a.e. y. To see this, apply the dominated convergence theorem.

Assumption 4 (Continuity of offer sets)**.** *The correspondence*

$$(p, x) \mapsto \prod_{i\in I} \operatorname{cl}\operatorname{co} X_i(p, x)$$

is closed at points (p, x) such that

$$\sum_{i=1}^{I} n_i \inf \{p \cdot (y - e_i) \mid y \in X_i(p, x)\} = 0, \tag{3.8}$$

i.e.*, $p_n \to p$, $x_n \to x$, $y_i^n \in \operatorname{cl}\operatorname{co} X_i(p_n, x_n)$, and $y_i^n \to y_i^\infty$ implies $y_i^\infty \in \operatorname{cl}\operatorname{co} X_i(p, x)$ for all $i \in I$ whenever* (3.8) *holds.*

Note that Assumption 4 is automatically satisfied if the inequality in Assumption 2 is strict, since I require Assumption 4 only when the equality (3.8) holds. Also if $X_i(p, x)$ is convex, then $\operatorname{cl}\operatorname{co} X_i(p, x) = X_i(p, x)$ since $X_i(p, x)$ is closed by construction.

Under the aforementioned assumptions, we can prove the first main result of this paper: a Bayesian general equilibrium exists.

Theorem 3.3. *Let $\mathcal{E} = \{I, \{n_i\}, \{e_i\}, \{\mu_i\}\}$ be an economy with offer sets $X_i(p, x) = \operatorname{supp} \mu_i(p, x)$ that satisfies Assumptions 1–4. Then $\mathcal{E}$ has a Bayesian general equilibrium. If Assumption 2 is replaced by*

$$\sum_{i=1}^{I} n_i \inf \{p \cdot (y - e_i) \mid y \in X_i(p, x)\} < 0,$$

in which case Assumption 4 is vacuous, all equilibria are non-degenerate.

Proof. For $\xi \in \mathbb{R}^C_+$, $p \in \Delta^{C-1}$, and $x \in X^I$, define the dual function H^* by

$$H^*(\xi; p, x) = \sum_{i=1}^{I} n_i \left[\xi' e_i + \log \left(\int \mathrm{e}^{-\xi' y} \mu_i(\mathrm{d}y; p, x) \right) \right].$$

Write $H^*(\xi) = H^*(\xi; p, x)$ when the choice of (p, x) is obvious. Since

$$X_i(p,x) = \operatorname{supp} \mu_i(p,x) \subset X \subset \mathbb{R}^C_+,$$

where X is the consumption set, we have $\mathrm{e}^{-\xi' y} \leq 1$ for all $\xi \geq 0$ and $y \in X_i(p, x)$. Then $\int \mathrm{e}^{-\xi' y} \mu_i(\mathrm{d}y; p, x) < \infty$ by Assumption 1, so $H^*(\xi) < \infty$ for all $\xi \geq 0$.

The outline of the proof is as follows. Since we know from Proposition 3.2 that minimizing H^* is enough for minimizing the Kullback-Leibler information, we wish to construct a correspondence from $(p, x) \in \Delta^{C-1} \times X^I$ to the Lagrange multiplier and implied average demand and show the existence of a fixed point (which is precisely the idea in Toda (2010)). However, H^* may not attain a minimum if the regularity condition (3.1) is violated. To overcome this difficulty, I restrict the domain of H^* to a compact set and define a quasi equilibrium concept. Specifically, define a t-quasi equilibrium as follows. Let $\|\cdot\|_1$ denote the L^1 norm, so $\|\xi\|_1 = \sum_{c=1}^C |\xi_c|$.

Definition 3.4 (t-quasi equilibrium)**.** Let $t > 0$. The pair of vectors $(p, x, \pi) \in \Delta^{C-1} \times X^I \times \mathbb{R}^C_+$ is said to be a *t-quasi equilibrium* if

1. $\xi = \pi$ solves

$$\min H^*(\xi; p, x) \text{ subject to } \xi \geq 0, \|\xi\|_1 \leq t,$$

2. π is proportional to p, and

3. $x = (x_i)$ is self-fulfilling, *i.e.*, $x_i = \int y f_i(y) \mu_i(\mathrm{d}y; p, x)$, where $f_i(y) = \mathrm{e}^{-\pi' y} / \int \mathrm{e}^{-\pi' y} \mu_i(\mathrm{d}y; p, x)$.

Thus, in view of Proposition 3.2, the definition of a t-quasi equilibrium is the same as that of a non-degenerate equilibrium except that the former minimizes H^* within a domain bounded by t. I show by a standard fixed point argument that a t-quasi equilibrium always exists. Thus, we can take a sequence of t-quasi equilibria such that $t \to \infty$. I then show that either some t-quasi equilibrium happens to be a non-degenerate equilibrium, or a subsequence of t-quasi equilibria converges to a degenerate equilibrium.

Step 1. $H^*(\xi; p, x)$ is C^1 in ξ and is differentiable under the integral sign. $H^*(\xi; p, x)$ and its first derivatives are continuous in $(p, x, \xi) \in \Delta^{C-1} \times X^I \times \mathbb{R}^C_+$.

(Proof in Appendix B.)

Step 2. Construction of a fixed-point correspondence.

As we saw above, $H^*(\xi) < \infty$ for all $\xi \geq 0$. By Proposition C.3, $H^*(\xi)$ is convex and lower semi-continuous (indeed, continuous) in ξ. Since the set $t\Delta^C := \left\{ \xi \in \mathbb{R}^C_+ \,\middle|\, \|\xi\|_1 \leq t \right\}$ is nonempty, compact, and convex, so is the set

$$\Pi(p, x) := \operatorname*{arg\,min}_{\xi \in t\Delta^C} H^*(\xi; p, x).$$

By Step 1, $H^*(\xi; p, x)$ is continuous in (p, x, ξ) on $\Delta^{C-1} \times X^I \times \mathbb{R}^C_+$. Thus, by the Maximum Theorem, $\Pi : \Delta^{C-1} \times X^I \rightrightarrows t\Delta^C$ is upper semi-continuous.

Using $\Pi(p,x)$, define the set $\Phi(p,x) \subset \Delta^{C-1}$ by

$$\Phi(p,x) := \begin{cases} \{\xi / \|\xi\|_1 \mid \xi \in \Pi(p,x)\}, & (0 \notin \Pi(p,x)) \\ \Delta^{C-1}. & (0 \in \Pi(p,x)) \end{cases}$$

Let us show that $\Phi : \Delta^{C-1} \times X^I \rightrightarrows \Delta^{C-1}$ is nonempty, compact, convex and upper semi-continuous.

1. $\Phi(p,x) \neq \emptyset$ is trivial. Since $\Phi(p,x)$ is either Δ^{C-1} itself (a convex set) or the intersection of Δ^{C-1} and the convex cone generated by $\Pi(p,x)$, it is convex.

2. Suppose that $p_n \to p$, $x_n \to x$, $q_n \in \Phi(p_n, x_n)$, and $q_n \to q$. If $0 \in \Pi(p,x)$, then $\Phi(p,x) = \Delta^{C-1} \ni q$. Therefore without loss of generality we may assume $0 \notin \Pi(p,x)$.

3. If $0 \notin \Pi(p_n, x_n)$ infinitely often, by taking a subsequence we may assume $0 \notin \Pi(p_n, x_n)$ for all n. Take $\xi_n \in \Pi(p_n, x_n)$ such that $q_n = \xi_n / \|\xi_n\|_1$. Since $\{\xi_n\} \subset t\Delta^C$ and $t\Delta^C$ is compact, $\{\xi_n\}$ has a convergent subsequence $\xi_{n_k} \to \xi$. Since $(p,x) \mapsto \Pi(p,x)$ is upper semi-continuous, we have $\xi \in \Pi(p,x)$, so $q = \xi / \|\xi\|_1 \in \Phi(p,x)$.

4. If $0 \in \Pi(p_n, x_n)$ eventually, by the definition of $\Pi(p,x)$ for large enough n we have $H^*(0; p_n, x_n) \leq H^*(\xi; p_n, x_n)$ for all $\xi \in t\Delta^C$. Letting $n \to \infty$ and using the continuity of H^*, we get $H^*(0; p, x) \leq H^*(\xi; p, x)$, so $0 \in \Pi(p,x)$. Hence $\Phi(p,x) = \Delta^{C-1} \ni q$.

Thus, $(p,x) \mapsto \Phi(p,x)$ is upper semi-continuous. In particular, by letting $p_n = p$ for all n, it follows that $\Phi(p,x)$ is closed, but since $\Phi(p,x) \subset \Delta^{C-1}$, it is compact.

Finally, define $\Psi : \Delta^{C-1} \times X^I \to X^I$ by

$$\Psi_i(p,x) = \int y f_i(y) \mu_i(\mathrm{d}y; p, x),$$

where $f_i(y) = \mathrm{e}^{-\xi' y} / \int \mathrm{e}^{-\xi' y} \mu_i(\mathrm{d}y; p, x)$ for $\xi \in \Pi(p,x)$. (Any such ξ will give the same f_i almost surely.) By Assumption 1, we can take $b > 0$ such that $\int y \mu_i(\mathrm{d}y; p, x) \leq b$ for all p, x. Let $X_b = X \cap [0,b]^C$ be the consumption set bounded by b.

Step 3. Ψ is well-defined, continuous, and $\Psi : \Delta^{C-1} \times X_b^I \to X_b^I$.

Proof. Let

$$g_{ic}(y_c) = \int f_i(y_c, y_{-c}) \mu_i(\mathrm{d}y_{-c}; p, x)$$

be the marginal density of y_c, where $y = (y_1, \dots, y_C) \in \mathbb{R}_+^C$. Since $\xi \geq 0$, $g_{ic}(\cdot)$ is a decreasing function. On the other hand, y_c is an increasing function of y_c. Hence by Chebyshev's inequality (Lemma C.1), we obtain

$$\begin{aligned} \Psi_{ic}(p,x) &= \int y_c f_i(y) \mu_i(\mathrm{d}y; p, x) = \mathrm{E}[Y_c f_i(Y)] \\ &= \mathrm{E}[Y_c g_{ic}(Y_c)] \leq \mathrm{E}[Y_c]\, \mathrm{E}[g_{ic}(Y_c)] \\ &= \mathrm{E}[Y_c] = \int y \mu_i(\mathrm{d}y; p, x) \leq b. \end{aligned}$$

Clearly Ψ is a single-valued upper semi-continuous correspondence, so it is a continuous function. $\square$

Step 4. For all $t > 0$, a t-quasi equilibrium exists. Either there exists a non-degenerate equilibrium, or we can take a sequence of t_n-quasi equilibrium with $t_n \to \infty$ as $n \to \infty$.

Proof. Since Δ^{C-1} and X_b^I are both nonempty, compact, and convex, by Kakutani's fixed point theorem there exists $(p, x) \in \Delta^{C-1} \times X_b^I$ such that

$$(p, x) \in (\Phi \times \Psi)(p, x).$$

Since $p \in \Phi(p, x)$, there exists $k \geq 0$ such that $\pi = kp \in \Pi(p, x)$, where $k = 0$ if $0 \in \Pi(p, x)$ and $k > 0$ if $0 \notin \Pi(p, x)$. Since $x \in \Psi(p, x)$, we have $x_i = \int y f_i(y) \mu_i(\mathrm{d}y; p, x)$. Therefore (p, x, π) is a t-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, x_n, π_n) such that $p_n \to p \in \Delta^{C-1}$ and $x_n \to x \in X_b^I$. If $\pi_n \in \arg\min_{\xi \geq 0} H^*(\xi; p_n, x_n)$ for some n, then by Proposition 3.2 (p_n, x_n, π_n) is a non-degenerate equilibrium.

Suppose that for all n, $\min_{\xi \geq 0} H^*(\xi; p_n, x_n)$ has no solution. For notational simplicity let $\mu_i^n := \mu_i(p_n, x_n)$, $H_n^*(\xi) = H^*(\xi; p_n, x_n)$, and

$$L(\xi, \lambda_n, \theta_n) = \sum_{i=1}^{I} n_i \left[\xi' e_i + \log \left(\int \mathrm{e}^{-\xi' y} \mu_i^n(\mathrm{d}y) \right) \right] - \lambda_n' \xi + \theta_n (\|\xi\|_1 - t_n)$$

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

$$\sum_{i=1}^{I} n_i \left[e_i - \frac{\int y \mathrm{e}^{-\pi_n' y} \mu_i^n(\mathrm{d}y)}{\int \mathrm{e}^{-\pi_n' y} \mu_i^n(\mathrm{d}y)} \right] - \lambda_n + \theta_n \mathbf{1} = 0 \tag{3.9}$$

and $\lambda_{cn} \pi_{cn} = 0$ for all c. Since (p_n, x_n, π_n) is not a Bayesian general equilibrium but a t_n-quasi equilibrium, the constraint $\|\pi_n\|_1 \leq t_n$ binds for all n. To see this, suppose $\|\pi_n\|_1 < t_n$. Since $\pi_n \notin \arg\min_{\xi \geq 0} H_n^*(\xi)$, we can take $\xi \geq 0$ such that $H_n^*(\pi_n) > H_n^*(\xi)$. Let $\xi_\alpha := (1 - \alpha)\pi_n + \alpha \xi$ for $\alpha \in [0, 1]$. Then for sufficiently small $\alpha > 0$ we have $\|\xi_\alpha\|_1 < t_n$, and by the convexity of H^* we have

$$H_n^*(\xi_\alpha) \leq (1 - \alpha) H_n^*(\pi_n) + \alpha H_n^*(\xi) < H_n^*(\pi_n),$$

which contradicts the optimality of π_n in $t_n \Delta^C$. Hence we have $\pi_n = t_n p_n$. $\square$

Step 5. $\lim_{n \to \infty} \theta_n = 0$.

Proof. Multiplying $\pi_n = t_n p_n$ as an inner product to (3.9) and dividing both sides by $t_n > 0$, it follows from $\lambda_{cn} \pi_{cn} = 0$ for all c that

$$\sum_{i=1}^{I} n_i \left[p_n' e_i - \frac{\int p_n' y \mathrm{e}^{-t_n p_n' y} \mu_i^n(\mathrm{d}y)}{\int \mathrm{e}^{-t_n p_n' y} \mu_i^n(\mathrm{d}y)} \right] + \theta_n = 0. \tag{3.10}$$

Regard

$$H_n^*(t\xi) = \sum_{i=1}^{I} n_i \left[p_n' e_i + \log \left(\int \mathrm{e}^{-t \xi' y} \mu_i^n(\mathrm{d}y) \right) \right]$$

as a function of t. Since by Proposition C.3 H_n^* is convex, $\frac{\mathrm{d}}{\mathrm{d}t}H_n^*(t\xi)$ is increasing in t. Fix any $t > 0$ and choose n sufficiently large such that $t_n > t$. Then, by (3.10) we obtain

$$\theta_n = -\frac{\mathrm{d}}{\mathrm{d}t}H_n^*(t_n p_n) \le -\frac{\mathrm{d}}{\mathrm{d}t}H_n^*(tp_n).$$

Letting $n \to \infty$, it follows from Step 1 that

$$\limsup_{n\to\infty} \theta_n \le \limsup_{n\to\infty} \left[-\frac{\mathrm{d}}{\mathrm{d}t}H_n^*(tp_n)\right] = -\frac{\mathrm{d}}{\mathrm{d}t}H_\infty^*(tp), \tag{3.11}$$

where $H_\infty^*(\xi) = H^*(\xi; p, x)$. Since t is arbitrary in (3.11), for any $s > 0$, by the mean value theorem take $t_s > 0$ such that

$$\frac{H_\infty^*(sp) - H_\infty^*(0)}{s} = \frac{\mathrm{d}}{\mathrm{d}t}H_\infty^*(t_s p).$$

Then (3.11) becomes

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

Letting $s \to \infty$, by Proposition C.4 and Assumption 2 we obtain

$$\limsup_{n\to\infty} \theta_n \le \sum_{i=1}^{I} n_i \inf \{p \cdot (y - e_i) \mid y \in X_i(p, x)\} \le 0. \tag{3.12}$$

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

Step 6. $\mathcal{E}$ has a degenerate equilibrium.

Proof. By Step 5 and (3.12), we have

$$\sum_{i=1}^{I} n_i \inf \{p \cdot (y - e_i) \mid y \in X_i(p, x)\} = 0.$$

By Assumption 4, the correspondence $(p, x) \mapsto \operatorname{cl}\operatorname{co} X_i(p, x)$ is closed at (p, x).

Since (p_n, x_n) is a fixed point of $\Phi \times \Psi$, in particular we have

$$x_i^n = \int y f_i^n(y) \mu_i^n(\mathrm{d}y) \in \operatorname{cl}\operatorname{co} X_{i,p_n},$$

where $x_n = (x_i^n)$ and $f_i^n(y) = \mathrm{e}^{-t_n p_n' y} / \int \mathrm{e}^{-t_n p_n' y} \mu_i^n(\mathrm{d}y)$. Therefore letting $n \to \infty$ we have $x_i \in \operatorname{cl}\operatorname{co} X_i(p, x)$, so condition 4 of Definition 2.4 holds. By (3.9), $\{x_i^n\}_{i=1}^I$ satisfies

$$\sum_{i=1}^{I} n_i (x_i^n - e_i) = -\lambda_n + \theta_n \mathbf{1} = 0. \tag{3.13}$$

Since $\lambda_n \ge 0$ and $\theta_n \to 0$ by Step 5, letting $n \to \infty$ in (3.13) we get

$$\sum_{i=1}^{I} n_i (x_i^n - e_i) \to \sum_{i=1}^{I} n_i (x_i - e_i) \le 0.$$

Thus, $\{x_i\}$ is feasible. By (3.10), we have $\sum_{i=1}^I n_i p_n'(x_i^n - e_i) = \theta_n$, so letting $n \to \infty$ we get $p \cdot \sum_{i=1}^I n_i(x_i - e_i) = 0$. Hence, if $\sum_{i=1}^I n_i(x_{ic} - e_{ic}) < 0$, it must be $p_c = 0$ and condition 5 of Definition 2.4 holds. Condition 6 of Definition 2.4 holds because Assumption 2 and $p \cdot \sum_{i=1}^I n_i(x_i - e_i) = 0$ imply $p \cdot y \geq p \cdot x_i$ for all $y \in X_i(p, x)$. Therefore (p, x) is a degenerate equilibrium. ☐

Step 7. Suppose that Assumptions 1–4 hold. If

$$\sum_{i=1}^I n_i \inf \{p \cdot (y - e_i) \,|\, y \in X_i(p, x)\} < 0$$

for all $p \in \Delta^{C-1}$ *and* $x \in X^I$, *then all Bayesian general equilibria of* $\mathcal{E}$ *are non-degenerate.*

Proof. If $\mathcal{E}$ has a degenerate equilibrium (p, x), by Definition 2.4 we have $p \cdot y \geq p \cdot x_i$ for all $y \in X_i(p, x)$. Thus

$$\sum_{i=1}^I n_i \inf \{p \cdot (y - e_i) \,|\, x \in X_i(p, x)\} \geq p \cdot \sum_{i=1}^I n_i(x_i - e_i) = 0,$$

a contradiction. ☐

The above steps complete the proof. ☐

The value added of this existence theorem compared with those in my previous paper (Toda, 2010) is that I do not assume that there is a strictly feasible allocation for all p, *i.e.*, the regularity condition (3.1). We know that there must be a feasible allocation *in the Bayesian general equilibrium* but not necessarily so otherwise. Therefore the assumptions of Theorem 3.3 are considerably weaker than my earlier existence theorems. Disposing the regularity condition (3.1) is important because the Walrasian economy in Example 1 is never compatible with (3.1).

Proposition 3.5 below strengthens the conclusion of Theorem 3.3. Its proof shows that degenerate equilibria are indeed the asymptotic limit of non-degenerate ones.

Proposition 3.5. *Let* $\mathcal{E} = \{I, \{n_i\}, \{e_i\}, \{\mu_i\}\}$ *be an economy with offer sets* $X_i(p, x) = \operatorname{supp} \mu_i(p, x)$ *that satisfies Assumptions 1, 3, 4, and*

$$\sum_{i=1}^I n_i \inf \{p \cdot (y - e_i) \,|\, y \in X_i(p, x)\} = 0 \tag{3.14}$$

for all $(p, x) \in \Delta^{C-1} \times X^I$. *Then* $\mathcal{E}$ *has a Bayesian general equilibrium, and all equilibria can be interpreted as degenerate ones. Furthermore, we can take a sequence of economies* $\{\mathcal{E}^n\}$ *such that all equilibria of* $\mathcal{E}^n$ *are non-degenerate and they converge to degenerate equilibria of* $\mathcal{E}$.

Proof. See Appendix B. ☐

3.3 Informational efficiency

the Kullback-Leibler information is a measure of information gain or "surprise" from updating the prior to the posterior. Let

$$\alpha_i(p, x) := \inf \{p \cdot y \mid y \in X_i(p, x)\}$$

be the minimum value of demand of type i agents, given price $p \in \Delta^{C-1}$ and average demand profile $x \in X^I$. If a type i agent demands y, then $p \cdot y - \alpha_i(p, x)$ is the *ex post* value of trade relative to the worst case scenario. The larger this value is, the more the agent is likely to gain from arbitrage.[6] Therefore, its economy-wide *ex post* average,

$$\begin{aligned} A[f; \mu(p, x)] &:= \sum_{i=1}^{I} n_i \int (p \cdot y - \alpha_i(p, x)) f_i(y) \mu_i(\mathrm{d}y; p, x) \\ &= p \cdot \bar{x}[f; \mu(p, x)] - \sum_{i=1}^{I} n_i \alpha_i(p, x), \end{aligned}$$

can be interpreted as the degree of arbitrage or that of market imperfection.

The following "informational efficiency theorem" states that there is a trade-off between the lack of arbitrage and the lack of surprise: in a market with low arbitrage (low $A[f; \mu]$), the agents have to gain a lot of information or be surprised (high $H[f; \mu]$). The Bayesian general equilibrium attains the best possible trade-off.

Theorem 3.6 (Informational efficiency theorem)**.** *Let $\mathcal{E} = \{I, \{n_i\}, \{e_i\}, \{\mu_i\}\}$ be an economy with offer sets $X_i(p, x) = \operatorname{supp} \mu_i(p, x)$, and let $(p, (x_i), (f_i))$ be a non-degenerate Bayesian general equilibrium. If* (3.1) *holds, so there exists a strictly feasible allocation, then there exists $t \geq 0$ such that $f = (f_i)$ minimizes the functional $H[g; \mu(p, x)] + tA[g; \mu(p, x)]$ over unconstrained $g = (g_i)$.*

Proof. Let us suppress p, x. Let $\pi = tp$ and $\bar{e} = \sum_{i=1}^{I} n_i e_i$ be the aggregate endowment. Since

$$H[g; \mu] + tA[g; \mu] = H[g; \mu] + \pi'(\bar{x}[g; \mu] - \bar{e}) - t \sum_{i=1}^{I} n_i \alpha_i + \pi' \bar{e}$$

and the last two terms are additive constants, g minimizes the left-hand side if and only if g minimizes $H[g; \mu] + \pi'(\bar{x}[g; \mu] - \bar{e})$. By the generalized Kuhn-Tucker theorem (Theorem C.2), f minimizes $H[g; \mu] + \pi'(\bar{x}[g; \mu] - \bar{e})$ over unconstrained g's if it is a non-degenerate Bayesian general equilibrium distribution, provided that the regularity condition of the Kuhn-Tucker theorem is satisfied. However, by Lemma 3.2 of Toda (2010), (3.1) is a sufficient condition for the Slater-type constraint qualification to be satisfied. □

The informational efficiency theorem should not be confounded with Pareto (or allocative) efficiency (first welfare theorem) in standard Walrasian economies.

[6] Here I am using the term 'arbitrage' informally to refer to a situation that agents execute trades that yielded higher values than they had expected. Thus 'arbitrage' here is synonymous to "getting a good deal by luck".

In Walrasian economies, the arbitrage measure A is (roughly) a measure of welfare, and we want to make A large. Nevertheless, the first welfare theorem states that A is nonpositive for any feasible allocation and is zero for the Walrasian equilibrium allocation. In Theorem 3.6, on the other hand, A is viewed as the degree of arbitrage. Large A means that there is a lot of room for agents to improve their position compared to the worst case scenario, so their worst case scenario were too pessimistic. A perfect market is one that there is no room for improvement, so the smaller A is, the closer the market is to perfection.

4 Relation between Walrasian and Bayesian equilibria

Now we can answer the question: how are Walrasian and Bayesian equilibria related? Somewhat surprisingly, for any economy in which each offer set is contained in the upper contour set to the average demand and coincides to the upper contour set within the budget set, the sets of Walrasian and Bayesian equilibria exactly coincide. Thus the Walrasian equilibrium is robust in the sense that it arises from a large class of non-optimizing behavior.

In order to state and prove the theorem, I introduce some notations. Let $X \subset \mathbb{R}^C_+$ be the nonempty, closed, convex consumption set and

$$\mathcal{E}^W = \{I, \{n_i\}, \{e_i\}, \{u_i\}\}$$

be a Walrasian economy, where n_i is the fraction of type i agents, e_i is the endowment of a type i agent, and $u_i : X \to \mathbb{R}$ is a continuous, quasi-concave, locally non-satiated utility function. The budget and upper contour sets of type i given average demand $x_i \in X$ are

$$\begin{aligned} B_i(p) &= \{y \in X \mid p \cdot y \leq p \cdot e_i\}, \\ U_i(p, x_i) &= \{y \in X \mid u_i(y) \geq u_i(x_i)\}. \end{aligned}$$

With the usual topology on $\mathbb{R}^C$ relative to X, the interior of the budget set is

$$\operatorname{int} B_i(p) = \{y \in X \mid p \cdot y < p \cdot e_i\}.$$

Theorem 4.1. *Let $\mathcal{E}^W$ be a Walrasian economy such that $\operatorname{int} B_i(p) \neq \emptyset$ for all i, p and $\mathcal{E}^B = \{I, \{n_i\}, \{e_i\}, \{\mu_i\}\}$ be a Bayesian economy with offer sets $X_i(p, x) = \operatorname{supp} \mu_i(p, x)$ that satisfies Assumptions 1–4. Suppose that for all $i \in I$, $p \in \Delta^{C-1}$, and $x_i \in B_i(p)$, we have*

(i) the offer set is contained in the upper contour set, i.e., *$X_i(p, x) \subset U_i(p, x_i)$,*

(ii) within the budget set, the upper contour set is locally contained in the offer set, i.e., *there exists $\epsilon > 0$ such that if $y \in B_i(p)$, $\|y - x_i\| < \epsilon$, and $u_i(y) > u_i(x_i)$, then $y \in X_i(p, x)$.*

Then all Bayesian general equilibria of $\mathcal{E}^B$ can be interpreted as degenerate, and they are Walrasian equilibria of $\mathcal{E}^W$. If in addition

$$X_i(p, x) \cap B_i(p) = U_i(p, x_i) \cap B_i(p), \tag{4.1}$$

in which case condition (ii) *automatically holds, then all Walrasian equilibria of $\mathcal{E}^W$ are Bayesian general equilibria of $\mathcal{E}^B$.*

Proof. By Theorem 3.3, $\mathcal{E}^B$ has a Bayesian general equilibrium. Assume that $(p, (x_i), (f_i))$ is a non-degenerate equilibrium. Since $x_i = \int y f_i(y) \mu_i(\mathrm{d}y; p, x)$, clearly $x_i \in \operatorname{cl} \operatorname{co} X_i(p, x)$ since integration has a convexifying effect.

Step 1. $\sum_{i=1}^I n_i(x_i - e_i) \leq 0$, and $p_c = 0$ if $\sum_{i=1}^I n_i(x_{ic} - e_{ic}) < 0$.

By market clearing we have

$$\sum_{i=1}^I n_i x_i = \sum_{i=1}^I n_i \int y f_i(y) \mu_i(\mathrm{d}y; p, x) \leq \sum_{i=1}^I n_i e_i,$$

so $\sum_{i=1}^I n_i(x_i - e_i) \leq 0$. Multiplying p as an inner product, we get $\sum_{i=1}^I n_i p \cdot (x_i - e_i) \leq 0$. If $p \cdot x_i < p \cdot e_i$ for some i, by local non-satiation we can take $y \in B_i(p)$ such that $\|y - x_i\| < \epsilon$ and $u_i(y) > u_i(x_i)$. By condition (ii), we have $y \in X_i(p, x)$. Since $X_i(p, x)$ is the support of the measure $\mu_i(p, x)$, for any neighborhood V of y in $X_i(p, x)$, we have $\mu_i(V; p, x) > 0$. By the continuity of u_i, we may assume $u_i(z) > u_i(x_i)$ for all $z \in V$. Then by the quasi-concavity of u_i, we obtain

$$u_i(x_i) = u_i\left(\int y f_i(y) \mu_i(\mathrm{d}y)\right) = u_i\left(\int_V + \int_{X_i(p,x) \setminus V}\right) > u(x_i),$$

which is a contradiction. Therefore $p \cdot x_i \geq p \cdot e_i$ for all i. Combining with the previous inequality, we get $p \cdot \sum_{i=1}^I n_i(x_i - e_i) = 0$. Therefore the claim holds.

Step 2. If $y \in X_i(p, x)$, then $p \cdot y \geq p \cdot x_i$.

Since by condition (i) $y \in X_i(p, x) \subset U_i(p, x_i)$, we have $u_i(y) \geq u_i(x_i)$. If $p \cdot y < p \cdot x_i = p \cdot e_i$, by local non-satiation there exists y' such that $p \cdot y' < p \cdot e_i$ and $u_i(y') > u_i(x_i)$. Let $y'' = (1 - \alpha)x_i + \alpha y'$. By the continuity and quasi-concavity of u_i, for sufficiently small α we have $\|y'' - x_i\| < \epsilon$, $p \cdot y'' < p \cdot e_i$, and $u_i(y'') > u_i(x_i)$. By condition (ii), we have $y'' \in X_i(p, x)$. Then we get a contradiction by the same argument as in Step 1. Therefore $p \cdot y \geq p \cdot x_i$.

By Steps 1 and 2, $(p, (x_i))$ is a degenerate equilibrium.

Step 3. If $(p, (x_i))$ is a degenerate equilibrium, it is also a Walrasian equilibrium.

By Definition 2.4 $p \cdot y \geq p \cdot x_i$ for all $y \in X_i(p, x)$, so $p \cdot (x_i - e_i) \leq p \cdot (y - e_i)$. Taking the infimum over y, by Assumption 2 we get $p \cdot (x_i - e_i) \leq 0$. Again by Definition 2.4 we have $p \cdot \sum_{i=1}^I n_i(x_i - e_i) = 0$, so it must be $p \cdot x_i = p \cdot e_i$ for all i. In order to show that (p, x) is a Walrasian equilibrium, it remains to show that $u_i(y) \leq u_i(x_i)$ whenever $p \cdot y \leq p \cdot e_i$. Suppose on the contrary that there exists y such that $u_i(y) > u_i(x_i)$ and $p \cdot y \leq p \cdot e_i$. Then by the continuity of u_i and $\operatorname{int} B_i(p) \neq \emptyset$ there exists y' such that $u_i(y') > u_i(x_i)$ and $p \cdot y' < p \cdot e_i$. Taking y'' as in Step 2, we get $y'' \in X_i(p, x)$ and $p \cdot y'' < p \cdot e_i = p \cdot x_i$, contradicting the definition of a degenerate equilibrium. This completes the proof that all Bayesian equilibria are Walrasian.

Step 4. If (4.1) *holds, then a Walrasian equilibrium is a Bayesian equilibrium.*

Let $(p, (x_i))$ be a Walrasian equilibrium. By market clearing, we have $\sum_{i=1}^I n_i(x_i - e_i) \leq 0$. By local non-satiation, we have $p \cdot x_i = p \cdot e_i$, so

$p \cdot \sum_{i=1}^I n_i(x_i - e_i) = 0$. By (4.1), clearly $x_i \in X_i(p,x)$. If there exists $y \in X_i(p,x)$ such that $p \cdot y < p \cdot e_i$ (so $u_i(y) \geq u_i(x_i)$ by (4.1)), by local non-satiation there exists z such that $u_i(z) > u_i(x_i)$ and $p \cdot z < p \cdot e_i$, contradicting that $(p,(x_i))$ is a Walrasian equilibrium. Therefore $p \cdot y \geq p \cdot e_i = p \cdot x_i$ for all $y \in X_i(p,x)$. Hence $(p,(x_i))$ is a degenerate Bayesian equilibrium. ☐

Theorem 4.1 is vacuous unless we show that a Bayesian economy satisfying the assumptions of Theorem 4.1 exists. The following proposition constructs such an economy.

Proposition 4.2. *A Bayesian economy satisfying the assumptions of Theorem 4.1 exists.*

Proof. Let $X = \mathbb{R}_+^C$ be the consumption set. For any $x = (x_i) \in X^I$ and $p \in \Delta^{C-1}$, define the probability measures

$$\nu_i^1(A;p,x) = \int_A \mathrm{e}^{-\epsilon' y} \mathrm{d}y \bigg/ \int_{U_i(p,x_i)} \mathrm{e}^{-\epsilon' y} \mathrm{d}y,$$
$$\nu_i^2(A;p,x) = \int_A \mathrm{e}^{-\epsilon' y} \mathrm{d}y \bigg/ \int_{U_i(p,e_i)} \mathrm{e}^{-\epsilon' y} \mathrm{d}y,$$

where A is any Borel set on X, $\epsilon \in \mathbb{R}_{++}^C$ is arbitrary, and $\mathrm{d}y$ refers to the Lebesgue measure. Take $b > 0$ such that $\sum_{i=1}^I n_i e_i \ll n_i b \mathbf{1}$ and let $X_b = [0,b]^C$. For each i take a continuous function $\alpha_i : \Delta^{C-1} \times X^I \to [0,1]$ such that $\alpha_i(p,x) = 1$ whenever $x = (x_i) \in X_b^I$ and $p \cdot x_i \leq p \cdot e_i$ and $\alpha_i(p,x) < 1$ otherwise. For instance,

$$\alpha_i(p,x) = \frac{1}{1 + \mathrm{dist}(x, X_b^I) + \mathrm{dist}(x_i, B_i(p))}$$

will do, where $\mathrm{dist}(x,S)$ is the distance from x to a closed convex set S. Define

$$\mu_i(p,x) = \alpha_i(p,x)\nu_i^1(p,x) + (1 - \alpha_i(p,x))\nu_i^2(p,x).$$

Then Assumptions 1–4 clearly hold. If $x = (x_i) \in X^I$ is feasible, then $x \in X_b^I$. Hence if $p \cdot x_i \leq p \cdot e_i$, then by the definition $\alpha_i(p,x) = 1$ and hence $\mu_i(p,x) = \nu_i^1(p,x)$, so $X_i(p,x) = U_i(p,x_i)$. Therefore (4.1) holds. ☐

The conclusion of Theorem 4.1 that the set of equilibria arising from a large class of non-optimizing behavior coincides to the set of Walrasian equilibria is qualitatively the same as Chakrabarti (2013). For comparison I describe his model below.

In his model, given the average demand profile $x = (x_i)$ (which can be interpreted as the current position of each agent type), agents put a probability measure $\mu_i(p,x)$ on preferred bundles that are affordable. Thus $\mu_i(p,x)$ is supported on $U_i(p,x_i) \cap B_i(p)$. Then agents update the demand by taking the average

$$x_i' = \int y \mu_i(\mathrm{d}y; p, x).$$

The "market" (or auctioneer) also puts a probability measure $\mu_m(p,x)$ on prices that make the value of demand larger. Thus $\mu_m(p,x)$ is supported on

$$U_m(p,x) := \left\{ q \in \Delta^{C-1} \,\middle|\, q \cdot \sum_{i=1}^{I} (x_i - e_i) \geq p \cdot \sum_{i=1}^{I} (x_i - e_i) \right\}.$$

Then the market updates the price vector by taking the average

$$p' = \int_{\Delta^{C-1}} q \mu_m(\mathrm{d}q; p, x).$$

A *social equilibrium* is defined by a fixed point of the map $(p,x) \mapsto (p',x')$. With a particular choice of the measures $\mu_i(p,x)$ and $\mu_m(p,x)$, Chakrabarti (2013) shows that a social equilibrium exists and that the sets of social and Walrasian equilibria coincide.

The difference between my model and Chakrabarti (2013)'s is the following. In my model agents are free to demand bundles outside their budget as long as they put some probability within their budget, whereas in Chakrabarti's model agents demand bundles only within their budget. My assumption has some realism since we are free to dream of becoming rich. The price to be paid with this weaker assumption is that the resulting average demand may not be feasible. In my model, agents apply the Bayes rule to update their demand distribution so that the demand becomes feasible. This aspect is what replaces utility maximization in the Walrasian model and the better response demand in Chakrabarti's model.

5 Application: endogenous wage distribution in a search model

The example is a standard search model as in McCall (1970) with a twist of Bayesian general equilibrium theory. While in a typical search model we assume that the distribution of wage offer is exogenous, by using Bayesian general equilibrium theory we can make sharp predictions about it. Time is discrete. There is only one agent type, workers, who can be either employed or unemployed. At period t, each unemployed worker gets an i.i.d. wage offer w that he believes to come from a stationary prior P. If the worker accepts the offer w, he becomes employed and consumes w forever (and goes out of the model because for simplicity I do not explicitly model firing). Otherwise, he receives an unemployment compensation c and waits for another offer next period. The supply of the consumption good is exogenously given at y per unemployed worker ($y > c$). The actual wage distribution (posterior) is determined by Bayes's rule. A worker's objective is to maximize $\mathrm{E}_P \sum_{t=0}^{\infty} \beta^t u(c_t)$, where c_t is consumption and E_P denotes the expectation under the prior P.

An unemployed worker's Bellman equation is

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

The reservation wage $\bar{w}$ is determined by[7]

$$u(\bar{w}) = u(c) + \frac{\beta}{1-\beta} \int_{\bar{w}}^{\infty} (u(w) - u(\bar{w})) P(\mathrm{d}w). \tag{5.1}$$

Since the worker consumes c if $w < \bar{w}$ and w otherwise, the offer set is $X = \{c\} \cup [\bar{w}, \infty)$. The measure μ on the offer set consists of the probability mass $\Pr(w \leq \bar{w}) = P(\bar{w})$ on the point c and the probability measure P restricted on the set $[\bar{w}, \infty)$. For this economy let us show that a genuine Bayesian general equilibrium exists. Assumption 1 is automatic if P is a probability measure. Assumption 2 hold because $x \geq c$ for $x \in X$ and $c < y$. Assumption 3 is trivial because since there is only one commodity, the relative price is always $p = 1$. Assumption 4 is vacuous. Hence by Theorem 3.3, there exists a genuine equilibrium.

Let π be the entropy price and define the dual function H^* by

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

By minimizing the Kullback-Leibler information (or minimizing the dual function H^*; see Theorem 3.1), we obtain

$$\frac{\mathrm{d}}{\mathrm{d}\pi} H^*(\pi) = 0 \iff \frac{c\mathrm{e}^{-\pi c} P(\bar{w}) + \int_{\bar{w}}^{\infty} w \mathrm{e}^{-\pi w} P(\mathrm{d}w)}{\mathrm{e}^{-\pi c} P(\bar{w}) + \int_{\bar{w}}^{\infty} \mathrm{e}^{-\pi w} P(\mathrm{d}w)} = y. \tag{5.2}$$

Given the prior P, (5.1) determines the reservation wage $\bar{w}$ and (5.2) 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 the initial prior P_0 is exponential and at each period workers 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 (posterior) 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 (5.1) and (5.2), which reads in this case

$$u(w_t) = u(c) + \frac{\beta}{1-\beta} \int_{w_t}^{\infty} (u(w) - u(w_t)) \lambda_t \mathrm{e}^{-\lambda_t w} \mathrm{d}w, \tag{5.3a}$$

$$\frac{c\mathrm{e}^{-\pi_t c}(1 - \mathrm{e}^{-\lambda_t w_t}) + \frac{\lambda_t (1 + (\pi_t + \lambda_t) w_t)}{(\pi_t + \lambda_t)^2} \mathrm{e}^{-(\pi_t + \lambda_t) w_t}}{\mathrm{e}^{-\pi_t c}(1 - \mathrm{e}^{-\lambda_t w_t}) + \frac{\lambda_t}{\pi_t + \lambda_t} \mathrm{e}^{-(\pi_t + \lambda_t) w_t}} = y. \tag{5.3b}$$

The unemployment rate at period t is $1 - \mathrm{e}^{-(\pi_t + \lambda_t) w_t}$. At first glance the rule $P_{t+1}(\mathrm{d}w) \propto \mathrm{e}^{-\pi_t w} P_t(\mathrm{d}w)$ might seem too strong, for workers learn the wage *distribution* itself at the end of each period. However, in order for this rule to be feasible, workers only need to know the current GDP, unemployment rate, unemployment compensation, and reservation wage because the wage distribution (which is exponential) can be recovered by these information.

Since the entropy price π_t is nonnegative, the sequence of exponential parameter $\{\lambda_t\}$ either converges or diverges to ∞. The interesting case is the

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

former. Substituting $w_t = \bar{w}$, $\pi_t = 0$, and $\lambda_t = \lambda$ in (5.3), we obtain the steady state by solving

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

As a numerical example, I choose the constant absolute risk aversion (CARA) utility function $u(w) = \frac{1}{a}e^{-aw}$ with absolute risk aversion $a = 3$, 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 $t = 21$.

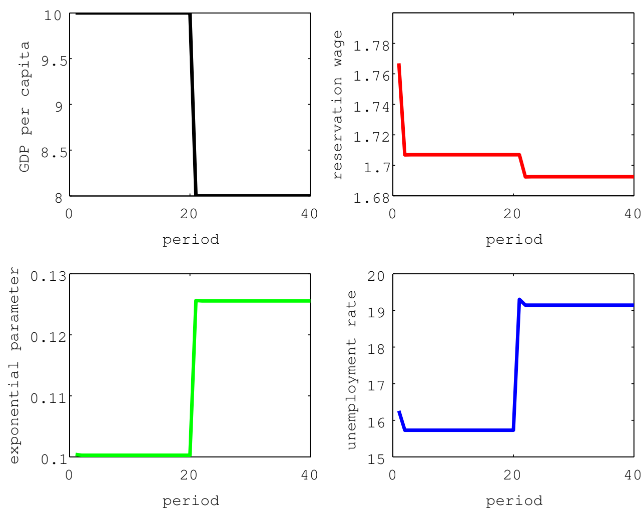

Figure 1. Time series of per capita GDP y_t (top left), reservation wage w_t (top right), exponential parameter $\lambda_{t+1} = \pi_t + \lambda_t$ (bottom left), and unemployment rate (bottom right, in percent).

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). This result—quantity adjustment rather than price adjustment—is qualitatively in accordance with empirical findings (Bewley, 1999).

6 Conclusion

This paper has reinterpreted Foley (1994)'s statistical equilibrium as an equilibrium concept for an economy in which agents demand commodities according to a not necessarily optimizing but satisficing behavioral rule. The Bayesian general equilibrium is a situation such that the price is correctly anticipated and agents no longer revise their behavior. I proved the existence of equilibrium under much weaker assumptions than previous works. I also showed that the equilibrium is informationally efficient and that if each offer set is contained in the upper contour set to the average demand and coincide to the upper contour set within the budget set, then the set of Bayesian and Walrasian equilibria coincide. Therefore the Walrasian equilibrium is robust in the sense that it arises from a large class of non-optimizing behavior. Although still not well known in the economics literature, the fact that applying the Bayes rule implies the minimization of the Kullback-Leibler information in large samples (Van Campenhout and Cover, 1981) seems an important property that deserves more attention.

An interesting (and challenging) future research topic might be to analyze the relation between BGE and recursive equilibria (Markov equilibria), the standard equilibrium concept used in modern macroeconomics. In many applications, recursive equilibria are unique but only because the assumption of point expectations is made (price function) and distributions over exogenous variables are taken as given. If, in contrast, we use the concept of Markov equilibria defined as a joint Markov process over exogenous and endogenous variables (Duffie et al., 1994), then indeterminacy of equilibria is the rule. Krebs (1997) has shown how to use maximum entropy to resolve this indeterminacy issue by using the concept of statistical expectational equilibrium (SEE), which is in a sense the most likely rational expectations equilibrium. In the example of Section 5, BGE and SEE coincide. Whether this property is robust is left for future research.

A Bayesian inference and maximum entropy

Given a multinomial distribution $\mathbf{p} = (p_1, \dots, p_K)$, where $p_k \geq 0$ and $\sum_{k=1}^{K} p_k = 1$, Shannon (1948) defined its *entropy* by

$$H(\mathbf{p}) = -\sum_{k=1}^{K} p_k \log p_k. \tag{A.1}$$

Jaynes (1957) proposed that when we want to assign probabilities $\mathbf{p} = (p_1, \dots, p_K)$ given some background information (such as moment constraints), we should maximize the Shannon entropy (A.1) subject to the constraints imposed by the background information. This is the original maximum entropy principle (MaxEnt). A prototypical example of such an inference problem is the die problem in Jaynes (1978), which dates back to a 1962 lecture:

> Suppose a die has been tossed N times and we are told only that the average number of spots up was not 3.5 as one might expect for an "honest" die but 4.5. Given this information, *and nothing else*, what probability should we assign to i spots in the next toss?

One drawback of the Shannon entropy is that it is not clear how to define it for distributions on a continuous space (say, Euclidean space). To circumvent this difficulty, Kullback and Leibler (1951) introduced the information measure

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

where $q(x)$ is the "prior" and $p(x)$ is the "posterior" density. Here we are implicitly using the Lebesgue measure $\mathrm{d}x$ and the density functions with respect to the Lebesgue measure, but it does not need to be so. A big advantage of the Kullback-Leibler information is that it is invariant to the choice of the reference measure: if measures P, Q, μ_1, μ_2 are mutually absolutely continuous and $p_i = \mathrm{d}P/\mathrm{d}\mu_i$, $q_i = \mathrm{d}Q/\mathrm{d}\mu_i$ are Radon-Nikodym derivatives (which correspond to density functions of probability distributions), then

$$H(p_1; q_1) = \int p_1 \log \frac{p_1}{q_1} \mathrm{d}\mu_1 = \int p_2 \log \frac{p_2}{q_2} \mathrm{d}\mu_2 = H(p_2, ; q_2),$$

so the choice of μ_1, μ_2 is irrelevant. Thus given any "prior" measure Q and "posterior" measure P, we can define the Kullback-Leibler information[8] of P with respect to Q by

$$H(P; Q) = \int \frac{\mathrm{d}P}{\mathrm{d}Q} \log \frac{\mathrm{d}P}{\mathrm{d}Q} \mathrm{d}Q = \int p \log \frac{p}{q} \mathrm{d}\mu, \tag{A.2}$$

where μ is any reference measure and $p = \mathrm{d}P/\mathrm{d}\mu$, $q = \mathrm{d}Q/\mathrm{d}\mu$ are Radon-Nikodym derivatives (density functions).

The Shannon entropy (A.1) corresponds to the Kullback-Leibler information (A.2) with respect to the uniform prior on a discrete set modulo the sign and an additive constant. Thus the maximum entropy principle of Jaynes (1957) can be generalized to what I refer to as the *minimum information principle*, which prescribes to minimize the Kullback-Leibler information (A.2) subject to the given constraints. Axiomatizations of the minimum information principle have been obtained by Shore and Johnson (1980), Caticha and Giffin (2006), and Knuth and Skilling (2012).

Returning to Bayesian inference, Van Campenhout and Cover (1981) showed that Bayes's theorem implies the minimum information principle in the following sense: the conditional distribution of a random variable X_n given the empirical observation

$$\frac{1}{N} \sum_{n=1}^{N} T(X_n) = \bar{T},$$

where X_n's are i.i.d. with prior density g, converges to $f_\lambda(x) = \mathrm{e}^{\lambda' T(x)} g(x)$ (suitably normalized) as $N \to \infty$, where λ is chosen to satisfy the population moment constraint

$$\int T(x) f_\lambda(x) \mathrm{d}x = \bar{T}.$$

This $f_\lambda(x)$ turns out to be the solution to

$$\min_f H(f; g) \text{ subject to } \int T(x) f(x) \mathrm{d}x = \bar{T},$$

i.e., the minimum information problem, and that λ is the corresponding Lagrange multiplier.[9] Although Van Campenhout and Cover (1981) proved the

[8] In the literature this quantity is also known as the relative entropy, cross entropy, information gain, I-divergence, etc.

[9] The solution of the more general entropy-like minimization problem can be found in Borwein and Lewis (1991, 1992).

statement only for the case T is a real function, Csiszár (1984) showed that the same conclusion holds even if T is vector-valued and the sample moment constraints $\frac{1}{N}\sum_{n=1}^{N} T(X_n) = \bar{T}$ are replaced by the condition that the sample moments belong to a specified convex set, in particular with inequality constraints. Thus computing the posterior distribution (in the Bayesian sense) reduces to solving the minimum Kullback-Leibler information problem, at least in the large sample limit.

B Proof of equilibrium existence

Proof of Step 1. That $H^*(\xi; p, x)$ is C^1 in ξ and is differentiable under the integral sign follow by Assumption 1 and Lebesgue's dominated convergence theorem. I only show that $H^*(\xi; p, x)$ is continuous in (p, x, ξ) since the case for its first derivatives is similar. Let $(p_n, x_n, \xi_n) \to (p, x, \xi)$ as $n \to \infty$, $f_n(y) = \mathrm{e}^{-\xi_n' y}$, and $f(y) = \mathrm{e}^{-\xi' y}$ for $y \in \mathbb{R}_+^C$. Then for any sequence such that $y_n \to y$, we have $f_n(y_n) \to f(y)$. (This property is referred to as "f_n continuously converges to f".) Since $f_n \leq 1$, $\{f_n\}$ are uniformly $\mu_i(p_n, x_n)$-integrable, *i.e.*,

$$\lim_{\alpha\to\infty} \sup_n \int_{f_n>\alpha} f_n(y)\mu_i(\mathrm{d}y; p_n, x_n) = 0.$$

(Just take $\alpha \geq 1$.) Therefore by Theorem C.5, we have

$$\lim_{n\to\infty} \int f_n(y)\mu_i(\mathrm{d}y; p_n, x_n) = \int f(y)\mu_i(\mathrm{d}y; p, x).$$

Hence $\int \mathrm{e}^{-\xi' y}\mu_i(\mathrm{d}y; p, x)$ is continuous in (p, x, ξ), and so is $H^*(\xi; p, x)$. □

Proof of Proposition 3.5. $\mathcal{E}$ has a Bayesian general equilibrium because (3.14) is stronger than Assumption 2. Suppose that $\mathcal{E}$ has a non-degenerate equilibrium $(p, x, (f_i))$. Then

$$x_i = \int y f_i(y)\mu_i(\mathrm{d}y; p, x) \in \operatorname{cl}\operatorname{co} X_i(p, x).$$

By market clearing and the nature of Lagrange multipliers, we have

$$\sum_{i=1}^{I} n_i(x_i - e_i) = \bar{x}[f; \mu(p, x)] - \bar{e} \leq 0$$

and $p \cdot \sum_{i=1}^{I} n_i(x_i - e_i) = 0$. Then for any $y = (y_i)$ with $y_i \in X_i(p, x)$, by (3.14) we get

$$p \cdot \sum_{i=1}^{I} n_i(y_i - e_i) \geq 0 = p \cdot \sum_{i=1}^{I} n_i(x_i - e_i),$$

so $p \cdot y \geq p \cdot x_i$ for all i and $y \in X_i(p, x)$. Hence, by Definition 2.4, (p, x) is a degenerate equilibrium.

Let $\mathcal{E}^n = \{I, \{n_i\}, \{e_i\}, \{\mu_i^n\}\}$ be an economy such that

$$\mu_i^n(B; p, x) = \mu_i\left((1 - 1/n)^{-1}B; p, x\right)$$

for any Borel set B, that is, the prior $\mu_i^n(p,x)$ is obtained by shrinking the domain of $\mu_i(p,x)$ by $1-1/n$ about the origin. Then the offer set is $X_i^n(p,x) = \operatorname{supp}\mu_i^n(p,x) = (1-1/n)X_i(p,x)$, so

$$\begin{aligned}\sum_{i=1}^I n_i \inf\left\{p\cdot y \,|\, y\in X_i^n(p,x)\right\} &= \left(1-\frac{1}{n}\right)\sum_{i=1}^I n_i \inf\left\{p\cdot y \,|\, y\in X_i(p,x)\right\} \\ &< \sum_{i=1}^I n_i \inf\left\{p\cdot y \,|\, y\in X_i(p,x)\right\}.\end{aligned}$$

Hence by (3.14) we get

$$\sum_{i=1}^I n_i \inf\left\{p\cdot (y-e_i) \,|\, y\in X_i^n(p,x)\right\} < \sum_{i=1}^I n_i \inf\left\{p\cdot (y-e_i) \,|\, y\in X_i(p,x)\right\} = 0.$$

By Theorem 3.3, the economy $\mathcal{E}^n$ has a non-degenerate equilibrium $(p^n,(x_i^n),(f_i^n))$, where

$$x_i^n = \int y f_i^n(y)\mathrm{d}\mu_i(p^n,x^n) \in \operatorname{cl}\operatorname{co} X_i(p^n,x^n).$$

Then by market clearing and the nature of Lagrange multipliers, after some algebra we obtain

$$\sum_{i=1}^I n_i(x_i^n - e_i) \le 0 \text{ and } p^n\cdot\sum_{i=1}^I n_i(x_i^n-e_i) = 0, \tag{B.1}$$

where $\bar{e} = \sum_{i=1}^I n_i e_i$ is the aggregate endowment. Since $x_i^n \in \operatorname{cl}\operatorname{co} X_i(p^n,x^n) \subset X \subset \mathbb{R}_+^C$ and $\{x_i^n\}$ is bounded ($\because$ market clearing), by taking a subsequence if necessary we may assume x_i^n converges to some x_i as $n\to\infty$. Since Δ^{C-1} is compact, we may also assume $p^n\to p$. Letting $n\to\infty$ in (B.1), we get $\sum_{i=1}^I n_i(x_i-e_i)\le 0$ and $p\cdot\sum_{i=1}^I n_i(x_i-e_i)=0$. By Assumption 4, we have $x_i\in\operatorname{cl}\operatorname{co}X_i(p,x)$, where $x=(x_i)$. By the same argument as above, (p,x) is a degenerate equilibrium. □

C Mathematical Results

Lemma C.1 (Chebyshev's inequality)**.** *If $f,g:\mathbb{R}\to\mathbb{R}$ are increasing (decreasing) functions and X is a random variable, then*

$$\mathrm{E}[f(X)g(X)] \ge \mathrm{E}[f(X)]\,\mathrm{E}[g(X)].$$

Proof. Let X' be an i.i.d. copy of X. Since f,g are monotone, we have

$$(f(X)-f(X'))(g(X)-g(X'))\ge 0.$$

Taking expectations of both sides, noting that X,X' are i.i.d., and rearranging terms, we obtain $\mathrm{E}[f(X)g(X)]\ge\mathrm{E}[f(X)]\,\mathrm{E}[g(X)]$. □

Clearly if one of f,g is increasing and the other is decreasing, the reverse inequality holds.

Theorem C.2 (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

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

and assume μ_0 is finite. Then there exist $z_1^* \ge 0$ 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). □

Proposition C.3. *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 $\operatorname{dom} f$.[10] *Furthermore, f is strictly convex if* $\dim T(\operatorname{supp} \mu) = C$.[11]

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

Proposition C.4. *Let $(X, \mathcal{B}, \mu)$ be as in Proposition C.3 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*

$$\lim_{t \to \infty} \frac{1}{t} \log \left(\int \mathrm{e}^{t\phi} \mathrm{d}\mu \right) = \operatorname{ess\,sup} \phi. \tag{C.3}$$

If ϕ is upper semi-continuous, then (C.3) *is equal to* $\sup \{\phi(x) \mid x \in \operatorname{supp} \mu\}$.

Proof. Let $E_n = \{x \in X \mid \phi(x) \ge -n\}$. 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 $\operatorname{supp} \mu \ne \emptyset$. Therefore $\mu(E_n) > 0$ for some n. Letting

$$v = \operatorname{ess\,sup} \phi = \sup \{c \mid \mu(\{x \in X \mid \phi(x) \ge c\}) > 0\},$$

we have $v > -\infty$.

Let us first prove (C.3) when $v < \infty$. Define

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

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

[11] 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 $\int e^{t_0\phi(x)} d\mu < \infty$. Then, for all $t > 0$ we obtain

$$\int e^{t\phi(x)} d\mu = e^{tv} \int_{X_-} e^{t(\phi(x)-v)} d\mu. \tag{C.4}$$

Denote the integral over X_- in (C.4) by $I(t)$. Since $\phi(x) \le v$ for $x \in X_-$, for each $x \in X_-$ the integrand $e^{t(\phi(x)-v)}$ 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

$$\frac{1}{t} \log \left(\int e^{t\phi(x)} d\mu \right) = v + \frac{1}{t} \log I(t) \le v + \frac{1}{t} \log I(t_0). \tag{C.5}$$

Letting $t \to \infty$ in (C.5), we obtain

$$\limsup_{t\to\infty} \frac{1}{t} \log \left(\int e^{t\phi(x)} d\mu \right) \le v. \tag{C.6}$$

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, we may assume $0 < \mu(A) < \infty$ since μ is regular. Therefore we obtain

$$\begin{aligned}\frac{1}{t} \log \left(\int e^{t\phi(x)} d\mu \right) &\ge \frac{1}{t} \log \left(e^{t(v-\epsilon)} \int_A e^{t(\phi(x)-v+\epsilon)} d\mu \right) \\ &\ge v - \epsilon + \frac{1}{t} \log \mu(A). \end{aligned} \tag{C.7}$$

Letting $t \to \infty$ in (C.7) and then $\epsilon \to 0$, we obtain

$$\liminf_{t\to\infty} \frac{1}{t} \log \left(\int e^{t\phi(x)} d\mu \right) \ge v. \tag{C.8}$$

(C.3) follows by (C.6) and (C.8).

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.7) with A replaced by F_n and $v - \epsilon$ replaced by n. Letting $n \to \infty$ we get (C.3).

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 $u = v$. □

Finally, I need a convergence theorem of Lebesgue integrals with varying measures. Let X be a locally compact second countable Hausdorff space (*e.g.*, Euclidean space) with Borel σ-algebra $\mathcal{B}$. We say that f_n continuously converges to f, denoted by $f_n \to_c f$, if $\lim f_n(x_n) \to f(x)$ for any $x_n \to x$.[12]

Theorem C.5. *Let $\mu, \{\mu_n\}$ be finite Borel measures on X. Suppose that $f_n \geq 0$, $f_n \to_c f$, $\mu_n \to \mu$ weakly, and $\int f_n \mathrm{d}\mu_n < \infty$ for all n. Then*

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

if and only if $\{f_n\}$ is uniformly $\{\mu_n\}$-integrable, i.e.,

$$\lim_{\alpha\to\infty} \sup_n \int_{f_n > \alpha} f_n \mathrm{d}\mu_n = 0.$$

Proof. See (Serfozo, 1982, Theorem 3.5). □

References

Gary S. Becker. Irrational behavior and economic theory. *Journal of Political Economy*, 70(1):1–13, February 1962.

Robert A. Becker and Subir K. Chakrabarti. Satisficing behavior, Brouwer's fixed point theorem and Nash equilibrium. *Economic Theory*, 26(1):63–83, 2005. doi:10.1007/s00199-004-0519-z.

Truman F. Bewley. *Why Wages Don't Fall During a Recession?* Harvard University Press, Cambridge, Massachusetts, 1999.

Jonathan M. Borwein and Adrian S. Lewis. Duality relationships for entropy-like minimization problems. *SIAM Journal on Control and Optimization*, 29 (2):325–338, March 1991. doi:10.1137/0329017.

Jonathan M. Borwein and Adrian S. Lewis. Partially finite convex programming, part I: Quasi relative interiors and duality theory. *Mathematical Programming*, 57(1-3):15–48, May 1992. doi:10.1007/BF01581072.

Antonio Cabrales, Olivier Gossner, and Roberto Serrano. Entropy and the value of information for investors. *American Economic Review*, 103(1):360–377, February 2013. doi:10.1257/aer.103.1.360.

Ariel Caticha and Adom Giffin. Updating probabilities. In Ali Mohammad-Djafari, editor, *Bayesian Inference and Maximum Entropy Methods in Science and Engineering*, volume 872 of *AIP Conference Proceedings*, pages 31–42, 2006. doi:10.1063/1.2423258.

Subir K. Chakrabarti. On the robustness of the competitive equilibrium: Utility-improvements and equilibrium points. 2013.

[12] For example, $f_n \to_c f$ if $f_n \to f$ uniformly on compact sets and f is continuous. To see this, let K a compact neighborhood of x and take N such that $n > N$ implies $x_n \in K$. Then

$$|f_n(x_n) - f(x)| \leq |f_n(x_n) - f(x_n)| + |f(x_n) - f(x)| \to 0.$$

Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. John Wiley & Sons, Hoboken, NJ, second edition, 2006.

Imre Csiszár. Sanov property, generalized I-projection and a conditional limit theorem. *Annals of Probability*, 12(3):768–793, August 1984.

Jacques H. Drèze. Existence of an exchange equilibrium under price rigidities. *International Economic Review*, 16(2):301–320, June 1975.

Darrell Duffie, John Geanakoplos, Andreu Mas-Collel, and Andrew McLennan. Stationary Markov equilibria. *Econometrica*, 62(4):745–781, July 1994.

Duncan K. Foley. A statistical equilibrium theory of markets. *Journal of Economic Theory*, 62(2):321–345, April 1994. doi:10.1006/jeth.1994.1018.

Duncan K. Foley. Statistical equilibrium in a simple labor market. *Metroeconomica*, 47(2):125–147, 1996.

Duncan K. Foley. Statistical equilibrium in economics: Method, interpretation, and an example. In Fabio Petri and Frank Hahn, editors, *General Equilibrium: Problems and Prospects*, chapter 4. Routledge, London and New York, 2003.

John Geanakoplos. Nash and Walras equilibrium via Brouwer. *Economic Theory*, 21(2-3):585–603, 2003. doi:10.1007/s001990000076.

P. Jean-Jacques Herings. *Static and Dynamic Aspects of General Disequilibrium Theory*, volume 13 of *C: Game Theory, Mathematical Programming and Operations Research*. Kluwer Academic Publishers, Dordrecht, The Netherlands, 1996.

Edwin T. Jaynes. Information theory and statistical mechanics. *Physical Review*, 106(4):620–630, May 1957. doi:10.1103/PhysRev.106.620.

Edwin T. Jaynes. Where do we stand on maximum entropy. In Raphael D. Levine and Myron Tribus, editors, *The Maximum Entropy Formalism*, chapter 1, pages 15–118. MIT Press, Cambridge, MA, 1978.

Edwin T. Jaynes. *Probability Theory: The Logic of Science*. Cambridge University Press, Cambridge, U.K., 2003. Edited by G. Larry Bretthorst.

Daniel Kahneman and Amos Tversky. Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2):263–292, March 1979.

Kevin H. Knuth and John Skilling. Foundations of inference. *Axioms*, 1(1): 38–73, 2012. doi:10.3390/axioms1010038.

Tom Krebs. Statistical equilibrium in one-step forward looking economic models. *Journal of Economic Theory*, 73(2):365–394, April 1997. doi:10.1006/jeth.1996.2231.

Solomon Kullback and Richard A. Leibler. On information and sufficiency. *Annals of Mathematical Statistics*, 22(1):79–86, 1951.

Lars Ljungqvist and Thomas J. Sargent. *Recursive Macroeconomic Theory*. MIT Press, Cambridge, Massachusetts, second edition, 2004.

David G. Luenberger. *Optimization by Vector Space Methods*. John Wiley & Sons, New York, 1969.

John J. McCall. Economics of information and job search. *Quarterly Journal of Economics*, 84(1):113–126, 1970.

Richard D. McKelvey and Thomas R. Palfrey. Quantal response equilibria for normal form games. *Games and Economic Behavior*, 10(1):6–38, July 1995. doi:10.1006/game.1995.1023.

Richard Serfozo. Convergence of Lebesgue integrals with varying measures. *Sankhyā: The Indian Journal of Statistics, Series A*, 44(3):380–402, October 1982.

Claude E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423, 623–656, 1948.

John E. Shore and Rodney W. Johnson. Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy. *IEEE Transactions on Information Theory*, IT-26(1):26–37, January 1980.

Herbert A. Simon. Theories of decision-making in economics and behavioral science. *American Economic Review*, 49(3):253–283, June 1959.

Christopher A. Sims. Implications of rational inattention. *Journal of Monetary Economics*, 50(3):665–690, April 2003. doi:10.1016/S0304-3932(03)00029-1.

Alexis Akira Toda. Existence of a statistical equilibrium for an economy with endogenous offer sets. *Economic Theory*, 45(3):379–415, 2010. doi:10.1007/s00199-009-0493-6.

Harald Uhlig. A law of large numbers for large economies. *Economic Theory*, 8 (1):41–50, 1996. doi:10.1007/BF01212011.

Jan M. Van Campenhout and Thomas M. Cover. Maximum entropy and conditional probability. *IEEE Transactions on Information Theory*, IT-27(4): 483–489, July 1981.

Laura L. Veldkamp. *Information Choice in Macroeconomics and Finance*. Princeton University Press, Princeton, NJ, 2011.