JOURNAL OF ECONOMIC SURVEYS

doi: 10.1111/joes.12226

INFORMATION THEORETIC APPROACHES IN ECONOMICS

Jangho Yang*

The New School for Social Research

Abstract. Economics has seen a recent rise in interest in information theory as an alternative framework to the conventional notion of equilibrium as a fixed state, such as Walrasian market-clearing general equilibrium. The information theoretic approach is predicated on the notion of statistical equilibrium (SE) that takes a distribution over all possible states as an equilibrium, and therefore predicts the endogenous fluctuations of the system along with its central tendency simultaneously. For this reason, SE approaches can explain the observed data without relying on arbitrary assumptions about random noise and provide useful insights for many interesting economic problems that conventional methods have not been able to satisfactorily deal with. In this paper, we review the key elements of information theory focusing on the notions and applications of entropy and SE in economics, particularly paying attention to how entropy concepts open up a new frontline of economic research.

Keywords. Information theory; Statistical equilibrium; Maximum entropy principle; Entropy-constrained model

1. Introduction

There have been important attempts to introduce information theoretic approaches in economics in an effort to propose an alternative framework to the conventional notion of equilibrium as a fixed state of the system, such as a market-clearing Walrasian equilibrium or Sraffian price of production model (Farjoun and Machover, 1983; Foley, 1994; Stutzer, 1995; Sims, 2003; Smith and Foley, 2008; Toda, 2015). The information theoretic approach is predicated on the notion of statistical equilibrium (SE) pioneered by physicists Maxwell, Boltzmann, and Gibbs, and understands equilibrium as a probability distribution over all possible states. While the conventional equilibrium framework assumes that the fluctuations are often induced by some exogenous shocks, as in the Real Business Cycle models stemming from Kydland and Prescott (1982), the SE approach can predict the endogenous fluctuations of the system along with its central tendencies simultaneously. Furthermore, what makes the SE approach relatively powerful compared to many other post–general equilibrium theories, such as behavior and information economics, is the fact that it does not have to assume biases of economic agents in the model to predict the fluctuations. Even without any behavioral or informational biases, the SE model can explain heterogeneity in the observed behavior, suggesting it can serve as an overarching alternative to the conventional equilibrium models in economics.

However, despite its importance, there have been few attempts to discuss the methodological and theoretical advantages of using the information theoretic methods for economic problems, nor have many papers been published to assess the current development of information theory in economics. A very few exceptions are survey articles by Lux (2009), Rongxi Zhou and Tong (2013), and Rosser (2016). However, Lux (2009) reviews a broad econophysics literature without particularly focusing on

*Corresponding author contact email: yangj994@newschool.edu; Tel: 347-276-4512.

Journal of Economic Surveys (2018) Vol. 32, No. 3, pp. 940–960

the information theoretic approaches. Rongxi Zhou and Tong (2013) survey wide applications of entropy concepts but only addresses finance literature. Rosser (2016) gives an overview of a few economic and financial applications of SE, but does not discuss the SE models of individual decision makers, such as rational inattention (RI) literature, one of the most influential economics applications of information theory.

This paper addresses deficiencies of previous attempts and provides an extensive and critical survey on information theoretic approaches in economics. The paper understands SE methods as an alternative framework for studying complex economic systems with central tendencies and endogenous fluctuations. Two different but closely connected lines of SE models in economics are (1) SE models of economic interactions and (2) SE models of individual decision makers. The first line of models are based on the maximization of the entropy of a target economic system subject to macroeconomic constraints. In contrast, the second line of models maximizes the expected payoff function of individual economic agents subject to information constraints. The two classes of models have distinctive interpretations: the former considers the complex interactions among economic agents given a macroeconomic structure as the central piece of the model while the latter puts emphasis on the bounded-rationality of nonoptimizing agent due to the positive degree of uncertainty. However, as we will demonstrate in the following sections, these two distinctive classes of models are mathematically equivalent.

The paper begins with a brief overview of the key elements of information theory that are relevant to economic models covered in this paper. Special attention will be paid to the notion of entropy and the maximum entropy principle (MEP). After the overview of information theory, Sections 3 and 4 provide detailed discussions of two classes of SE models. To study the SE models of economic interactions, the paper covers four topics of central economic concern: market transaction, the rate of profit, asset pricing and Tobin's q in finance, and income distribution. For the survey of SE models of individual economic agents, the paper reviews the mutual-information constrained model and Shannon entropy-constrained model. Sim's RI model is discussed in this section. In Section 5, the paper addresses some controversies over the interpretation and justification of SE methods in economics, centering around the legitimacy of applying the physical concepts to social system.

2. Overview of Information Theory

We begin with a general introduction of the information theoretic concepts that are relevant to economic applications covered in this paper. The main concepts to be introduced include the Shannon entropy, Kullback–Leibler divergence, channel capacity, and SE. These are the fundamental concepts on which the information theoretic economic models are constructed. Because the present paper is an economic survey paper, it does not provide comprehensive discussions of these concepts. Instead, we will limit ourselves to outlining their core logic to the extent it helps to understand information theoretic economic models reviewed in this paper. For detailed discussion on information theory, see Jaynes (2003), MacKay (2005), and Cover and Thomas (2006).

2.1 *Entropy as a Measure of Uncertainty*

2.1.1 *Shannon's Entropy*

In information theory, the *Shannon entropy*, H, of a probability measure $p \geq 0$ on a finite set $X = \{x_1, x_2, \ldots, x_n\}$ is defined as the expected value of *information content* $I(p) = \log[1/p]$:[1]

$$H(p) = -\sum_{i=1}^{n} p(x_i) \log\left[p(x_i)\right], \tag{1}$$

where $\sum_{i=1}^{n} p(x_i) = 1$. By convention, $0 \log[0] = 0$. The continuous version is

$$H(P) = -\int_{-\infty}^{+\infty} P(x) \log [P(x)]\, dx, \tag{2}$$

where $P(x)$ is the density function of x so that $\int_{-\infty}^{+\infty} P(x)dx = 1$. It is worthwhile to emphasize that entropy depends only on the probability distribution $p(x)$ or $P(x)$, not on the random variable x.[2]

The information content of an outcome x, $I(p) = \log[1/p]$ is monotonically decreasing in p and therefore reflects the important intuition that the occurrence of a highly likely outcome does not provide much information, whereas a highly unlikely outcome provides a great deal of information to an observer. This leads to the interpretation of the Shannon entropy, which is the average information content of the probability distribution, as a *degree of uncertainty* of a system composed of x outcomes. The more uncertain we are about the system, the more information we can expect from observing the actual occurrence of the set of outcomes, and thus, the higher entropy H is. For example, if the probability distribution is represented by the Dirac Delta function in which only one outcome is possible, then we are perfectly certain which event will occur. The entropy of this probability function is predictably zero because $H[P] = -(1 \log[1] + 0 \log[0] + \cdots + 0 \log[0]) = 0$. In contrast, if the probability distribution is uniform, in which $p = 1/n$, we are maximally uncertain about which event will occur. In this case, the entropy of system is $\log[n]$, which is actually the maximum entropy of this system.[3] The Shannon entropy as the degree of uncertainty is the basis of the information theoretic models of individual decision makers, which we will review later in Section 4.

2.1.2 *Kullback–Leibler Divergence*

A close look at the Shannon entropy reveals that this measure represents the average information content of a probability distribution when there is no other reference information as to the actual occurrence of the events (or that the reference information is a uniform distribution, and thus is uninformative). Hence, an extension of this logic includes a situation where we want to obtain the information contents of the occurrence of the events of which we have prior information represented by a probability distribution $q(x_i)$. A relative entropy, known as Kullback–Leibler divergence (KL divergence), measures the average informational gain from p when the prior probability is q (Kullback and Leibler, 1951). The KL divergence of the discrete distributions of p and q, $D_{\mathrm{KL}}(p\|q)$, is defined as follows:

$$D_{\mathrm{KL}}(p\|q) = \sum_i p(x_i) \log \frac{p(x_i)}{q(x_i)}. \tag{3}$$

The continuous case replaces the summation with the integral over the probability density functions of $p(x)$ and $q(x)$. A lower KL divergence means that $p(x)$ and $q(x)$ are similar, and therefore, there is a small informational gain (or a small surprise) from the distribution of the outcomes given our prior distribution on those outcomes.[4]

2.1.3 *Mutual Information and Channel Capacity*

The mutual information of two random variables x and y, $I(x; y)$ is defined as the change in our information after observing y given the prior information on x, which is equivalent to the KL divergence of a joint distribution x and y relative to the independent prior $p(x)$ and $p(y)$:

$$I(x; y) = \sum_{i=1}^{n} \sum_{j=1}^{m} p(x_i, y_j) \log \left[\frac{p(x_i, y_j)}{p(x_i)p(y_j)} \right],$$

which leads to the following relation (MacKay, 2005; Cover and Thomas, 2006):

$$I(x, y) = H(x) - H(x|y) \geq 0. \tag{4}$$

where $H(x|y)$ is the conditional entropy defined as $\sum p(x_i, y_i) \log[p(x_i)/p(x_i, y_i)]$.

This result shows that the change in our information on the inputs can be expressed by the difference between the entropy of the prior and the posterior distribution of inputs. A high mutual information of x and y implies that the amount of information about the variable x obtained from the variable y is significant. In contrast, low mutual information implies that the output variable is not helpful to obtain information about the input variable. When the mutual information is zero, therefore, there is no informational gain from the output about the input, implying that the two variables are independent.

The channel capacity C is defined as the maximum mutual information of transmission, that is, the maximum amount of information that can be transmitted through the channel given the prior distribution on x (MacKay, 2005; Cover and Thomas, 2006):

$$C = \max I(x, y). \tag{5}$$

The channel capacity plays a critical role in the RI literature, which we will review later in Section 4.

2.2 *SE and the MEP*

2.2.1 *Entropy as a Measure of the Number of States in the System*

So far, we have discussed some key concepts in information theory, which are centered around the notion of entropy as a measure of uncertainty. In this section, we will introduce another way to interpret entropy based on the result from statistical mechanics. Over the course of our discussion, we will introduce a new concept, the *MEP* and *SE*.

In statistical mechanics, entropy represents a measure of the number of microstates of constituent components of a system consistent with a particular macrostate described by the distribution in question (Jaynes, 1957, 1978, 2003). Given n number of constituent components and m different macrostate, the total number of possible microstates with $n_1, n_2, \ldots, n_m$ components in each state m can be given by the multinomial coefficient:

$$\binom{n}{n_1, n_2, \ldots, n_m} = \frac{n!}{n_1!n_2!\ldots n_m!} = \frac{n!}{(np_1)!(np_2)!\ldots(np_m)!}, \tag{6}$$

where $n = n_1 + n_2 + \cdots + n_m$ and $n_i/n = p_i$. Since the multinomial coefficient is a fast growing function with its argument and cannot be easily approximated by a Taylor expansion, it is mathematically more convenient to take the logarithm of this multiplicity.[5] Taking the logarithm of the multinomial coefficient and using the Stirling approximation $\log(n!) \approx n\log[n] - n$, we have

$$\log\left[\frac{n!}{(np_1)!(np_2)!\ldots(np_m)!}\right] \approx -n\sum_{i=1}^{m} \log[p_i]p_i. \tag{7}$$

This logarithm of the multinomial coefficient is entropy, nH, in statistical mechanics (Jaynes, 1957, 1978, 2003). The entropy per outcome is $H = -\sum_{i=1}^{m} \log(p_i)p_i$ which is the same as the Shannon entropy.

As the derivation shows, entropy is proportional to the total number of possible microstates of the system. The higher the entropy is, the more microstates the system has, and therefore is more likely to happen. Finding the maximum entropy of a system is, therefore, equivalent to obtaining the most likely configuration in the system expressed as a probability distribution of a target variable. The most likely probability distribution of the variable is called *SE* (Jaynes, 1957, 1978, 2003).[6]

2.2.2 *Maximum Entropy Principle*

Finding SE of the system through maximizing entropy gives the most likely state of the system, from which the macroscopic properties of interest can be induced. Following Jaynes (1957, 1978, 1982, 2003), we call this general inferential framework the MEP.[7]

From the standpoint of information theory, the maximum entropy state is maximally uninformative because the uncertainty of the system is maximized.[8] Therefore, the MEP as an inferential method, suggests that the investigator takes only given information into account and makes the maximally noncommittal guess about the problem (Jaynes, 1957, 1978). The MEP can incorporate the constraints in the form of linear moment functions. Then, it finds the most likely state of the system that is consistent with these constraints.[9]

The MEP formalism can be expressed as a constrained optimization problem. In the case of discrete variables x_i, the maximum entropy program with h constraints on the expectations of the functions $f_h(x_i)$ can be expressed as follows:

$$\max \quad -\sum p(x)\log[p(x)], \tag{8}$$

$$\begin{aligned} \text{s.t} \quad & \sum p(x) = 1 \\ & \sum_{i=1}^{n} p(x_i) f_h(x_i) = F_h \qquad h = 1, \ldots, m. \end{aligned} \tag{9}$$

The standard method to solve the constrained maximum entropy program is to use the dual constrained program as follows (Borwein and Lewis, 1991; Cover and Thomas, 2006):

$$\min \quad \log\left[\sum_{i=1}^{n} \exp\left(\sum_{h=1}^{m} \lambda_h f_h(x_i)\right)\right] - \sum_{h=1}^{m} \lambda_h F_h, \tag{10}$$

where λ is a vector of Lagrangian multiplier. The solution implies the following maximum entropy frequency distribution (SE):[10]

$$p(x_i) = \frac{1}{Z(\lambda_1, \ldots, \lambda_h)} \exp\left[\lambda_1 f_1(x_i) + \cdots + \lambda_h f_h(x_i)\right], \tag{11}$$

where $Z(\lambda_1, \ldots, \lambda_h) = \sum_{i=1}^{n} \exp[\lambda_1 f_1(x_i) + \cdots + \lambda_h f_h(x_i)]$.

The maximum entropy distribution suggests that the equilibrium state of the system is not a single state but is inherently a probability distribution of all possible states. This SE approach is the bedrock of the application of information theoretic approaches in economics as we will review in the following sections.

3. SE Models of Economic Interactions

This section reviews prominent economic models of the SE. These models study the economic process in a complex reality in which a great number of market agents interact with one another subject to some possible economic constraints. For this reason, we call them *SE models of economic interactions*. Four topics to be reviewed are the SE model of market transactions, profit rate, asset pricing and Tobin's q in finance, and income and wealth distribution. The resulting SE in the models is qualitatively different from the Walrasian and Sraffian equilibrium models in that the equilibrium states are characterized by endogenously generated fluctuations.

3.1 *SE Model of Market Transaction: An Alternative to Walrasian Equilibrium Models*

Foley (1994) is the first attempt in economics to theorize a market transaction in the SE framework. It starts with the fundamental question concerning how the complex interactions between economic agents in a decentralized market lead to remarkable regularities in market transactions. In an economy with m number of commodities whose transactions are represented by points in $R^m : x = (x_1, \ldots, x_m)$, we assume that there is a probability distribution of potential transactions $p(h[x])$. To find the SE of the market transaction, Foley proposes the maximum entropy program for the transaction distribution $p(h[x])$ for n number of agents with r different types of agents with the constraint that the net transaction is zero:

$$\max \ -\sum_{k=1}^{r} w^k \sum_{x \subseteq R^m} p(h^k[x]) \log[p(h^k[x])], \tag{12}$$

$$\text{s.t} \ \sum h^k[x] = 1$$

$$\sum_{k=1}^{r} w^k \sum_{x \subseteq R^m} p(h^k[x]) = 0 \qquad k = 1, \ldots, m, \tag{13}$$

where w^k is the weight of the k type agent in the economy. The solution to this program, that is, the (unique) most likely market transaction distribution is the Gibbs canonical distribution with entropy prices π (a vector of shadow prices for each traded commodity) among many possible configurations of mutually advantageous transactions:

$$h^k[x] = \frac{\exp[-\pi x]}{Z^k[\pi]}, \tag{14}$$

where $Z^k[\pi] = \sum_x \exp[-\pi x]$. This implies that the probability of a particular trade occurring is proportional to the value of that trade given the market-clearing entropy price.

This SE theory of market provides a powerful alternative to the Walrasian theory of competitive equilibrium, the fundamental basis of the neoclassical economics. First, in line with other SE approaches, Foley's theory does not aim to pick a particular market transaction as an equilibrium but finds the most likely distribution of all feasible (Pareto-improving) transactions. By doing so, it removes the existence of the auctioneer from the equilibrium theory. Second, having more than one (market-clearing) price ratios and non–single-point offer sets, Foley's theory predicts inequality in the final consumption bundle even with identical agents with the same preferences, endowment, and production possibilities. This is possible because transactions can take place even when the equilibrium prices are not determined, unlike the tâtonnement case. As Foley (1994) points out, this aspect of the theory has a particularly important policy implication because if we believe that transactions do not occur at equilibrium prices but occur at different price ratios, price control can prevent disadvantageous transactions. Finally, Foley's theory is more general than Walras' theory and has the Walrasian equilibrium price and Pareto-efficiency as a special case of the model when the initial endowments are already Pareto-efficient so that no production or transactions are necessary. This methodological advantage renders Foley's theory more applicable to many other economic problems.

In a subsequent paper, Foley (1996) shows how his general theory of statistical market equilibrium can be applied. In the labor market model, the general pure exchange model is reduced to a two commodity case with labor power and wages as particular substances of exchange. One of the most important features of the derived SE of the labor market is that, unlike Marshallian or Walrasian equilibrium, the mean wage can be higher than the reservation wage in the SE, implying involuntary unemployment. Unemployment does not have a perfectly (or instantly) effective counteracting impact on wages, so that, the labor market in this model cannot fully exhaust all the opportunities for mutually advantageous transactions.

Toda (2010) further extends Foley's SE model of pure exchange by introducing production economies with endogenous offer sets. In Foley's original model, transactions depend on the utility function with the exogenoulsly given preference, endowment, technology, and expectation. Toda points out that, in production economies, firms' expectations are endogenously determined by the expectation of price, because their demand function, in pursuing profit maximization, ultimately depends on price. Using the fixed-point theorem, he proves the existence of the SE of this general model of market exchange with the endogenous offer set. In Toda (2015), he further generalizes his model and formally proves that the SE model of market exchange is compatible with a classical Walrasian equilibrium depending on the types of offer set in relation to the upper contour set. This implies that the Walrasian equilibrium could also arise from a complex economy where a great number of nonoptimizing agents interact with one another.

Foley's pioneering works and Toda's important generalization raise intriguing questions for the further theoretical extensions of their models. Two fundamental questions deserve our attention. First, as Foley (2003) points out, his examples on a pure exchange do not actually spell out how the SE of market transactions actually takes place. The market in this model is a black box where the maximum entropy over the constraint on market-clearing prices lead to a collection of heterogeneous market exchanges, but exact dynamics as to how such a collection is achieved are not self-explanatory in the model. This apparent drawback, however, actually highlights the importance of using the SE approach as the basis of all non-Walrasian microeconomics, such as behavior or information economics, because it offers the most general framework for modeling macro results of micro behavior, through which other behavioral or informational assumptions can be systematically incorporated. Second, this class of SE model inevitably raises a question about operationalization, primarily because SE, like other equilibrium approaches, is inherently static and cannot fit into transient process. However, if we believe that a static approach to dynamic systems can still help us to improve our knowledge, this issue reduces to a practical question as to what is a right time span for operationalization of the equilibrium analysis to avoid the situation where the underlying parameters move faster than the market processes that lead to the equilibrium (Foley, 2003).

Despite Foley and Toda's breakthrough in SE reasoning for economic problems, the SE models of market exchange remain as a theoretical entity that needs to be empirically tested. If there are disagreements between the model prediction and the actual observations, this suggests that investigators should come up with other constraints of the model that might affect the economic process (Jaynes, 1978).

3.2 *SE Model of the Profit Rate: An Alternative to Classical Deterministic Model*

3.2.1 *Farjoun and Machover's Seminal Work on the Gamma Distribution of the Profit Rate*

Farjoun and Machover's book *Laws of Chaos: A Probabilistic Approach to Political Economy* (1983) is an original contribution that opened up the use of the SE approach in the study of the profit rate. They start the book with the critique of the deterministic view adopted in much of economics literature, such as the classical political economy on the theory of price, arguing that it cannot fully capture rich aspects of the economic process in a complex reality where millions of market agents interact with one another over a long period of time. The deterministic view is the tendency in economics to take the aggregate or average of a particular economic variable as the representative of the system and derive its equilibrium state in relation to other variables whose quantity is also based on the average. The main methodological break of their work from the conventional approaches in political economy is that it adopts a probabilistic approach and treats economic variables as random variables. From this perspective, the average is just one of many states of the random variable, and does not tell more than the general location of the distribution. Instead of reducing entire states to a single quantity, Farjoun and Machover try to find the equilibrium of distribution itself, which is the SE . Based on an analogy with a confined gas, they propose the gamma distribution as the SE of the rate of profit (Farjoun and Machover, 1983).[11]

Their gamma conjecture of the SE of the profit rate, however, has been criticized by the succeeding works. Above all, the gamma distribution is constrained in the positive domain, and therefore, it does not explain the observed phenomenon of negative profit rates. It comes from the fundamental problem that Farjoun and Machover do not correctly define the state–space of the profit rate and therefore assign unjustified zero probability to the negative values. This is not only against empirical observation (negative profit rates are actually very common in observed data), but also against the spirit of SE approach. The latter is the case because the observed distribution of the profit rate is not a result from interactions only among those firms with a positive profit rate but among all firms with different levels of profit rates including negative ones.

3.2.2 *Subbotin Distribution of the Profit Rate*

With regard to the empirical evidence of the SE of the rate of profit, Alfarano *et al.* (2012) find that the distributions of firms' profit rates in the United States from 1980 to 2006 are well described by the non-Gaussian exponential power (Subbotin) distribution. Based on their previous work on the firm's growth rate (Alfarano and Milaković, 2008), they suggest a maximum entropy program as follows:

$$\max \; -\int_{-\infty}^{+\infty} p(x)\log p(x)dx, \tag{15}$$

$$\text{s.t} \; -\int_{-\infty}^{+\infty} p(x)dx = 1,$$
$$-\int_{-\infty}^{+\infty} p(x)\left|\frac{x-m}{\sigma}\right|^{\alpha} dx = 1, \tag{16}$$

whose solution is the Subbotin distribution:

$$f(x;m,\sigma,\alpha) = \frac{1}{2\sigma\alpha^{1/\alpha}\Gamma(1+1/\alpha)}\exp\left(-\frac{1}{\alpha}\left|\frac{x-m}{\sigma}\right|^{\alpha}\right), \tag{17}$$

where m is a location parameter, σ is the a scale parameter, and α is a shape parameter. Alfarano *et al.* (2012) proposes a drift-diffusion model and provides a detailed discussion on the shape parameter in which α is interpreted as the degree of diffusion. Scharfenaker and Semieniuk (2017), using the Compustat database, also find that the U.S. firm's profit rates are well approximated by the (asymmetric) Laplace distribution, which is a special case of the (asymmetric) Subottin distribution. In the same fashion with the Alfarano and Milaković 's approach, they derive the Laplace distribution by maximizing entropy over two moment constraints: the mean and the expected absolute distance from the mean. They use their evidence to refute the claim by Farjoun and Machover that the profit rate data are described by the gamma distribution.

It is a great achievement of these works to find empirical regularities for the rate of profit. They also try to demonstrate that the maximum entropy approach can be a parsimonious way to explain the observed data. However, it is questionable whether the SE method they have employed and the economic theories are coherently tied up together. All of these applications of the SE require deliberate discussions of the constraints of the model because those constraints actually determine the features of the resulting SE. Despite the flawless mathematical derivation of each SE distribution given a particular set of constraints, it is not always easy to see plausible economic interpretations of those constraints in the above works. The dispersion of profit rates as a constraint in Alfarano *et al.* (2012) is conveniently assumed for a mathematical derivation of the Subbotin distribution. Their diffusion-drift theory (Alfarano *et al.*, 2012) is one way of deriving the Subbotin distribution, but is not directly related to the maximum entropy approach. The mean and expected distance constraints in Scharfenaker and Semieniuk's model are not based on any relevant theory of the firm's behavior either.

3.2.3 *Discrete Choice (Quantal Response) Social Interaction Model of the Profit Rate*

We will discuss Scharfenaker and Foley (2017) on the social interaction model of the profit rate as an example that shows how to utilize an economic theory as a constraint of the maximum entropy problem in the study of the profit rate. Their model is represented by an action variable $A = (a = entry, \bar{a} = exit)$ of the entry–exit decision of firms in each subsector and the aggregate profit rate, x. They set up a maximum entropy program to find the SE of the joint distribution of x and A, $p(x, A)$, from which the marginal distribution of the profit rate, $p(x)$ can be derived. In doing so, they suggest one constraint based on Adam Smith's competition theory that the firms' decision to enter or exit has a negative feedback on the observed profit rate. For example, if a firm enters a sector seeking to maximize her profit rate, it will lower the profit rate of that sector through changes in supply and consequently the price of the commodity in the sector. The degree of negative feedback (competitive pressure) δ is represented by the difference between the two conditional expectations of x weighted by the marginal probability of the action variable as follows:

$$E(x|a)p(a) - E(x|\bar{a})p(\bar{a}) \leq \delta. \tag{18}$$

They also assume that the firm has a positive uncertainty in processing the market signal regarding the profit rate, which is expressed by the quantal response constraint on the firm's action, specifying $p(a|x)$ as $1/(1 + e^{\frac{x-\mu}{T}})$. μ determines the indifference point in which exit and entry have the same probability and T measures the responsiveness of the action frequency in response to changes in the profit rate. The maximum entropy program with these two constraints implies the following marginal distribution of the profit rate:

$$\hat{p}(x) = \frac{e^{H_{x,\mu,T}} e^{-\beta \mathrm{Tanh}[\frac{x-\mu}{2T}]x} e^{-\gamma x}}{\sum_x e^{H_{x,\mu,T}} e^{-\beta \mathrm{Tanh}[\frac{x-\mu}{2T}]x} e^{-\gamma x}}, \tag{19}$$

where $H_{x,\mu,T} = H(\frac{1}{e^{-\frac{x-\mu}{T}}+1}, \frac{1}{e^{\frac{x-\mu}{T}}+1})$[12] and $\mathrm{Tanh}[\alpha]$ is the hyperbolic tangent function, written as $\frac{e^{2\alpha}-1}{e^{2\alpha}+1}$. This distribution turns out to fit the observed data on the rate of profit extremely well. It tells us that the quantal action of a great number of profit-maximizing capitalists to enter and exit the sector and its effect on the profit rate of the sector produce a highly organized pattern of profit rate as an unintended consequence.[13] Unlike Alfarano *et al.* (2012) and Scharfenaker and Semieniuk (2017), the constraints of the maximum entropy program of the profit rate in Scharfenaker and Foley (2017) are formulated based on an economic theory, the Smithian competition theory, and the quantal response of the economic agent.

One minor technical problem with their approach is that the actual recovery of SE of the joint distribution of x and A requires discretizing x (coarse-grained bins). However, it is often difficult to coherently prove that the result is not sensitive to bin sizes.

3.3 *SE Model of Financial Market: Asset Pricing and Tobin's q*

Financial economics has seen remarkable progress in the applications of entropy concepts, especially in portfolio selection and asset pricing. These applications are not limited to the Shannon entropy but utilize different entropy measures, such as the Tsallis entropy, Rényi entropy, and the *f*-divergence. As Rongxi Zhou and Tong (2013) document, the different entropy concepts have been used as a measure of risk, capital increment, and portfolio diversification in portfolio selection theory. These applications do not directly rely on the maximum entropy approach, and therefore, are not covered in the present paper.

In contrast, the asset pricing theory in finance has seen a wide use of SE approach, originating in the seminal works by Stutzer (1995, 1996, 2010) and Buchen and Kelly (1996).

Notably, Stutzer (1995, 1996) develops an SE model of asset pricing (or more concretely derivative security valuation) on the basis of the KL divergence . In this work, the question of interest is to recover the probability distribution of the future value of the asset price P at any time T, which is represented by the local volatility function, $\sigma(P, t)$. An analyst only observes the current value of the asset price but does not know $\sigma(P, t)$ itself. Stutzer proposes *canonical valuation*, which is essentially a method to find the most likely risk neutral density of asset price π^* with the prior distribution of the empirical risk neutral density, $\hat{\pi}$, computed using time series of past gross asset returns R, dividend yields D, and riskless interest rates r. This inverse problem can be solved by minimizing the KL divergence of π^* to $\hat{\pi}$ subject to the risk-neutral measure of the asset price, which is the second constrain in the following program:[14]

$$\min \quad I(\pi^*, \hat{\pi}) = \sum_{h=1}^{H-T} \pi^*(h) \frac{\pi^*(h)}{\hat{\pi}(h)} \tag{20}$$

$$\text{s.t} \quad -\sum_{-\infty}^{+\infty} p(x) = 1,$$

$$\sum_{h=1}^{H-T} \frac{R(-h)}{r^T} \frac{\pi^*(h)}{\hat{\pi}(h)} = 1, \tag{21}$$

where $R(-h)$ is the rolling historical time series of T-period returns obtained by computing $P(-h)/P(-h-T)$, $h = 1, 2, \ldots, H-T$ and $P(t)$ is a historical time series of the previous asset price with $t = -1, -2, \ldots, -H$. The resulting SE is the well-established *Gibbs canonical distribution*:

$$\hat{\pi}^* = \frac{\exp\left(\gamma^* \frac{R(-h)}{r^T}\right)}{\sum_h \exp\left(\gamma^* \frac{R(-h)}{r^T}\right)}, \qquad h = 1, \ldots, H-T, \tag{22}$$

where γ^* is a Lagrangian multiplier or the shadow price.[15] Stutzer uses the resulting distribution of the SE and successfully replicates the previous tests on Black–Scholes call option prices with far less data.

Similarly, Buchen and Kelly (1996) develop a simple maximum entropy model of the nonstochastic option price. The key constraint is the finance model of dividend-free option price x in an equilibrium market. For a call option, the constraint is given by

$$c_i(x) = D(T)(x - K_i)^+ = D(T)\max(0, x - K_i), \tag{23}$$

and for a put option

$$c_i(x) = D(T)(K_i - x)^+ = D(T)\max(0, K_i - x), \tag{24}$$

where $D(T)$ is the nonstochastic present-value discount factor to time T and can be chosen to be e^{-rT}, and K_i is the strike price.

Thus, the maximum entropy program is given by

$$\max \quad -\int_0^{\infty} p(x) \log p(x) dx, \tag{25}$$

$$\text{s.t} \quad \int_0^{\infty} p(x) c_i(x) dx = d_i,$$

$$-\int_0^{\infty} p(x) dx = 1, \tag{26}$$

where d_i for $i = 1, \ldots, m$, is the observed data on the option price. The solution to this problem is $\exp(\sum_{i=1}^{m} \lambda_i c_i(x))/\int_0^\infty \exp(\sum_{i=1}^{m} \lambda_i c_i(x))$. Buchen and Kelly note that the numerical solution to the resulting maximum entropy distribution has a very good fit to the exact distribution for option prices given at different simulated strikes.

These works by Stutzer, and Buchen and Kelly demonstrate that SE approaches can provide a parsimonious method to study the dynamics of asset prices, which are inherently generated by numerous interactions among the agents in financial market.

3.3.1 *Tobin's q*

As another example of the application of the SE approach to the study of financial market, we will review Scharfenaker and dos Santos (2015) on Tobin's q. They report the empirical frequency distributions of Tobin's q for the private, nonfinancial corporations traded in the U.S. equity market over 50 years. The recorded Tobin's q distributions have a asymmetric unimodal shape with a highly peaked mode. They conjecture that the asymmetric Laplace distribution may be understood as a good representation of the empirical data. To derive the asymmetric Laplace distribution from the MEP, they set up two constraints, the expected values of the infra mode M_L and the supra-mode M_H with the mode c, $-\int_{-\infty}^{c} p(x)(c - x)dx = M_L$ and $-\int_{c}^{+\infty} p(x)(c - x)dx = M_H$, respectively.

The resulting maximum entropy distribution has an explicit analytical solution known as the asymmetric Laplace distribution:

$$f(x; c, M_L, M_H) = \frac{1}{\left(\sqrt{M_H} + \sqrt{M_L}\right)^2} \times \begin{cases} \exp\left(\frac{(x-c)}{\sqrt{M_H M_L + M_L}}\right) & \text{if } x \le c \\ \exp\left(-\frac{(x-c)}{\sqrt{M_H M_L + M_H}}\right) & \text{if } x > c. \end{cases} \tag{27}$$

The derived asymmetric Laplace distribution fits the data extremely well with 98% information of the data recovered by the distribution according to the Soofi informational index (Scharfenaker and dos Santos, 2015).

An immediate criticism arises of their approach in so far as they did not spell out any economic theory to justify the constraints, as we pointed out in our discussion in the SE models of the profit rate. dos Santos and Scharfenaker (2016) try to provide tentative economic explanations of the constraints. The mean constraint is formulated based on the scaling principle, by which systemic interdependences between individual measures of Tobin's q is defined, while the modal deviation constraint is interpreted as a result of the imperfect capacity of a typical agent to process the market signal on capital valuation. Regardless of how convincing this interpretation of constraints is, their approach seems to be predicated on somewhat weak methodological stance that does not fully utilize the strength of SE reasoning. This is because they attach an economic interpretations of the constraints a posteriori to match the mathematical constraints of the maximum entropy program for the (asymmetric) Laplace distribution. In other words, the two moment constraints for the program have been assumed by the investigators even before they formulate a proper theory of them. The methodological problem of this approach is that one can actually propose many different interpretations of the constraints and there is no way to quantitatively differentiate them.

3.4 *SE Model of Income and Wealth Distribution*

The distribution of income and wealth has also been one of the most intensively studied topics in the econophysics literature. Among many important contributions, this section focuses on more recent works in econophysics literature, a collective scholarly work by physicists and economists that introduces elements of physics into economics. The econophysics literature is broad and is not confined to SE methods. Therefore, we will be quite selective in this section and pay attention only to those works that explicitly utilize the notion of SE, represented by Dragulescu and Yakovenko (2000, 2001).[16]

The seminal study of income and wealth distribution from the statistical equilibrium perspective is due to the works by Dragulescu and Yakovenko (2000, 2001). They find that the income distribution in the United States and the wealth distribution in the United Kingdom are well described by the mixture of the exponential and the power law distributions. Yakovenko and Rosser (2009) provide SE explanations for the lower part of the empirical distribution by proposing a maximum entropy program with a mean constraint that reflects the idea that the total income and total stock of assets are conserved.[17] With the mean constraint, the resulting SE of income, x, is the famous Gibbs distribution $c\exp(-x/T_x)$ where T_x is the "income temperature" and c is a constant. Dregulescu and Yakovenko (2001) show that, even with rapidly increasing income inequality, the lower part of the income distribution has persistently exhibited the exponential distribution in the U.S. during the 80s and 90s. Combined with the fat tail behavior characterized by the Pareto distribution, they argue that the coexistence of two different distributions in the income distribution reveals the two-class structure of the U.S. economy. Shaikh *et al.* (2014) expands Dragulescu and Yakovenko's findings by showing that the mixture of the exponential and the Pareto distribution is also found in the income distribution by race and gender in the United States. A recent study by Toda (2012, 2011) shows that not only the upper tail but also the lower tail of the (detrended) income distribution are well described by the power-law distribution and therefore the distribution follows the double Pareto distribution. Also, Schneider (2015) argues that the mixture of exponential and the power-law might not satisfactorily explain the observed data. He proposes the mixture of the exponential and the log-normal distribution for a better fit of the data.

One caveat of the SE models of income and wealth distribution stemming from Dragulescu and Yakovenko (2000, 2001) is that the assumption about the conservation principle is not properly explained. In the income distribution case, the conservation principle implies a fixed amount of total wages available in the economy for a given period. Therefore, someone's increase in wages is automatically offset by others' decrease in wages in the economy, which, without a further elaboration, inevitably invokes the once discredited wage fund theory. This criticism in line with the previous examples on the profit rate and Tobin's q shows some potential pitfalls of applying the SE method to economic problems when the constraints of the maximum entropy program are not supported by relevant economic theories.

4. SE Models of Individual Decision Makers

Another prominent strand of economics literature that utilizes the concepts of Information Theory is the information-constrained behavior model. This line of models is based on optimizing the payoff function of an individual economic agents under the assumption that they are information-constrained. Therefore, we call this class of model *SE models of individual decision makers*. Two versions of this model is the mutual information-constrained model under the name of *RI* (Sims, 2003, 2006; Woodford, 2009; Luo and Young, 2009, 2013; Yang, 2015) and the Shannon entropy constrained model under the name of *quantal response model* (Scharfenaker and Foley, 2017). In these models, individual decision makers are assumed to have a limited capacity to process market signals or they are exposed to a positive degree of uncertainty about the system. Under this constraint, the optimization of the payoff function will lead to a non-degenerate distribution of heterogeneous behaviors.

We will show that the information-constrained behavior model is dual to the maximum entropy model (SE model of the economic interactions), where the entropy of the system, not the individual payoff function, is maximized given the economic theory as constraints.

4.1 *Mutual Information Constrained Model: Sims' RI Program*

The RI program, initiated by Christopher Sims (2003, 2006), is a prominent example of the information-constrained behavior model. In the RI models, economic behavior is modeled as information flow through human cognitive channels. Optimizing agents make economic decisions, a, (savings, wage bargaining, stock market investment, etc.) by processing external random signals x about relevant economic variables (interest rate, expected inflation, market report on company performance, etc.). RI models introduce the idea that these economic agents have limited capacity to process market signals. In other words, agents are not able to be fully attentive to information and therefore fail to translate all the available market signals into actions.

Mathematically, the limited capacity to process market signals is expressed by a finite Shannon channel capacity defined as the maximum of the mutual information between the input and the output in Equation (4). In RI models, the input is the relevant market signals, x, while the output is the processed signal that leads to the corresponding economic action, a. Therefore, the finite channel capacity of economic agents is expressed as the following:

$$I(A, X) = \int \log[p(a, x)]p(a, x)dadx - \int \log\left[\int p(a, x')dx'\right] p(a, x)dadx \\ - \int \log\left[g(x)\right] g(x)dx \leq C,$$

where $g(x)$ is the marginal distribution of the input and C is the maximum information flow between a and x, the channel capacity. The limits of integrals are determined by the domain of random variables a and x of the question of interest and the assumptions about the joint distributions of the random variables. For example, if two random vectors are finite subsequences drawn from a joint normal distribution, a rate per observation of the information flow is approximated by the diagonalization of the variance and covariance matrices using Fourier transform matrices, whose integral is defined between $-\pi$ and π (Sims, 2003). Recalling the definition of the mutual information, when I between x and a is zero, it implies that x and a are independent and x does not give any information about a. As I becomes larger, x has more useful information that is translated into the action of a. A finite constraint on I, therefore, implies that the translation from x to a is not perfect but limited.

The optimizing agent in RI models maximizes her payoff through a mixed strategy over possible actions a and market signals x, $U(a, x)$, subject to a finite channel capacity C, which boils down to the following maximization problem:

$$\begin{aligned} \max \ & \int U(a, x)p(a, x)dadx, \quad \text{s.t} \int p(a, x)da = g(x) \\ & \int \log[p(a, x)]p(a, x)dadx - \int \log\left[\int p(a, x')dx'\right] p(a, x)dadx \\ & - \int \log[g(x)]g(x)dx \leq C, \end{aligned} \tag{28}$$

where $p(a, x) \geq 0$. Depending on the choice of the functional form of the payoff function $U(.)$ and the prior joint probability of a and x, this maximization program can provide the conditional distribution of $p(x|a)$, the possible configuration of the signal x given the action a. In the original paper, Sims (2003, 2006) considers a simple example where the prior joint probability of a and x is a joint normal distribution. In this case, the posterior distribution of $p(x|a)$ turns out to be a normal distribution with the mean and the variance as a function of a, the parameters of the quadratic form of $U(a, x) = -(\omega_{aa}a^2 + 2\omega_{ax}ax + \omega_{xx}x^2)$, and the Lagrange multiplier on the information constraint λ:

$$p(x|a) \sim N\left(\frac{\omega_{aa}}{\omega_{ax}a}, \frac{\omega_{aa}}{\omega_{ax}^2}\lambda\right). \tag{29}$$

It is quite interesting to see that, under a fairly weak assumption of a quadratic payoff function, just maximizing expected payoff subject to a mutual information constraint leads to a normal posterior distribution, which represents the most common stochastic model of economic variables.

As a natural implication of information theoretic models, RI models predict endogenous fluctuations and the inherent heterogeneity of economic behavior. Put in the context of standard macroeconomics, RI models provide an alternative view to the rational expectation hypothesis. This is because RI models predict the sluggish and erratic responses of economic agents, which are not observed in the standard rational expectation models. Those apparent abnormalities are endogenously induced while the rational expectation models tend to treat them as induced from external shocks. In the context of policy evaluation, RI models are more flexible and can predict that economic agents could respond differently to economic fluctuations with different degrees, while the rational expectation models tend to have fixed responsive coefficients. This comes from the fact that optimizing agents in RI models will allocate more attention to more volatile fluctuations (erratic response) and less attention to small variations (delayed response) (Sims, 2003, 2006).

Sims' RI program has motivated a number of applications. Maćkowiak and Wiederholt (2015) and Luo and Young (2009) have applied the RI hypothesis to the standard general equilibrium model, Matêjka and McKay (2015) and Woodford (2009) have proposed the RI discrete choice models, and Yang (2015) has proposed the rational inattention game theory model.

4.2 *Shannon Entropy-Constrained Model: Quantal Response Model*

Information constrained models can also be formulated by directly constraining the Shannon entropy of action instead of constraining the mutual information of action and signal variables. The entropy constraint has a direct implication that economic agents have a positive degree of uncertainty over their actions. The *bounded rationality models*, particularly, the *quantal response models* (McFadden, 1976) can be succinctly derived using the minimum entropy constraint.

To see the formal expression of the general entropy constrained model with action variables, let us first start with a simple mixed strategy problem in which the typical agent chooses a mixed strategy $p(a)$ given $U(a)$ from a convex set:

$$\begin{aligned} \max \quad & \int U(a)p(a)da, \\ \text{s.t} \quad & \int p(a)da = 1. \end{aligned} \tag{30}$$

With no further constraint, the solution to this problem is the Dirac Delta function:

$$p(a) = \text{DiracDelta}[a - \hat{a}[u]], \tag{31}$$

where $\hat{a}[u]$ is the payoff maximizing behavior. The Dirac Delta function has the following property

$$\text{DiracDelta}\,(a) = \begin{cases} 1, & a = 0 \\ 0, & a \neq 0 \end{cases} \qquad \text{and} \qquad \int_{-\infty}^{\infty} \text{DiracDelta}\,(a)\,da = 1, \tag{32}$$

so that the resulting frequency distribution of p puts unit weight on the payoff maximizing action while putting zero weight on the others.

The same maximization program with a minimum entropy constraint can be formulated as follows:

$$\begin{aligned} \max \quad & \int U(a)p(a)da, \\ \text{s.t} \quad & \int p(a)da = 1, \\ & -\int p(a)\log[p(a)]da \geq \text{H}_{min}, \end{aligned} \tag{33}$$

where H_{min} represents the minimum entropy (uncertainty) of the typical agent about the target economic variable. A Lagrangian function of this maximization problem is as follows:

$$\mathrm{L} = \int p(a)U(a) - \mu\left(\int p(a) - 1\right) + T\left(-\int p(a)\log[p(a)] - \mathrm{H}_{min}\right), \tag{34}$$

whose resulting frequencies of a at a behavior temperature T is

$$p(a) = Z(U, T, a)^{-1} e^{\frac{U(a)}{T}}, \tag{35}$$

where $Z(.)$ is a normalizing partition function, $\int e^{\frac{U(a)}{T}} da$. Equation (35) suggests that the payoff maximizing distribution of p with a minimum entropy constraint is the Gibbs canonical distribution with the temperature T.

Based on this basic framework, Scharfenaker and Foley (2017) derive the quantal response function $p(a|x)$ with a discrete action variable a and outcome variables x. This model, therefore, assumes that the economic agent has a positive degree of uncertainty of her action when she sees a particular state of aggregate economic variable. This can be mathematically expressed as follows:

$$\begin{aligned} \max \quad & \sum U(a, x)p(a|x), \\ \text{s.t} \quad & \sum p(a|x) = 1, \\ & -\sum p(a|x)\log[p(a|x)] \geq \mathrm{H}_{min}. \end{aligned} \tag{36}$$

The resulting frequencies of $p(a|x)$ at a behavior temperature T is

$$p(a|x) = Z(U, T, A)^{-1} e^{\frac{U(a,x)}{T}} = \frac{e^{\frac{U(a,x)}{T}}}{\sum e^{\frac{U(a,x)}{T}}}. \tag{37}$$

Since a is a binary (quantal) variable in their model, $p(a|x)$ can be written as follows:

$$p(a = 0|x) = \frac{e^{\frac{U(a=0|x)}{T}}}{e^{\frac{U(a=0|x)}{T}} + e^{\frac{U(a=1|x)}{T}}},$$

$$p(a = 1|x) = \frac{e^{\frac{U(a=1|x)}{T}}}{e^{\frac{U(a=0|x)}{T}} + e^{\frac{U(a=1|x)}{T}}}.$$

Depending on the specification of the payoff function, the recovered link function $p(a|x)$ represents the probability of a particular action given observed economic variable x. Figure 1 shows the logit quantal response function with the behavior temperature T.

This figure graphically shows that the higher T is the more uncertain the quantal decision is. Except the unattainable case when $T = 0$, the function generally predicts a gradual increase in the frequency of the action in response to the aggregate variable. When T is sufficiently large, the quantal response is almost uniform between different actions. When the behavior temperature T is close to zero, the function becomes a step function. Except for the latter case, the link function is not degenerate and assigns positive probabilities to heterogeneous responses, reflecting the agents' uncertainty.

4.3 *Duality between Maximum Entropy Models and Entropy Constrained Models*

There is a duality between the SE models of economic interaction (maximum entropy models) and the SE of individual decision makers (entropy constrained models). This can be proved by showing that both problems have the same associated Lagrangian multiplier. Let us suppose a simple discrete one-variable

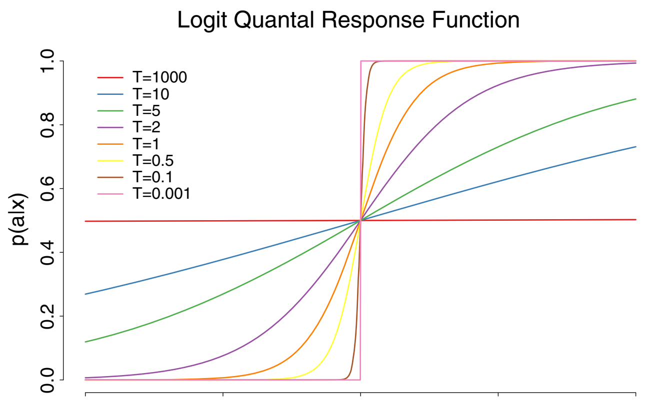

Figure 1. A Logit Quantal Response Function with the Behavior Temperature T. The Horizontal Axis Represents the Social Variable, Such as the Profit Rate, while the Vertical Axis Represents the Frequency of the Action in Response to the Aggregate Variable. The Higher the Behavior Temperature Is the More Uncertain the Quantal Decision Is. [Colour figure can be viewed at wileyonlinelibrary.com]

case. The maximum entropy approach with a given expected value of the payoff is formulated as follows:

$$\begin{aligned} \max \quad & -\sum p(a)\log[p(a)], \\ \text{s.t} \quad & \sum p(a) = 1, \\ & \sum U(a)p(a) = \hat{U}(a), \end{aligned} \tag{38}$$

whose Lagrangian is:

$$\mathrm{L} = -\sum p(a)\log[p(a)] - \mu\left(\sum p(a) - 1\right) - \beta\left(\sum U(a)p(a) - \hat{U}(a)\right). \tag{39}$$

In contrast, as we have derived above in Equation (34), the entropy constrained model with a given payoff function has the Lagrangian as follows:

$$\mathrm{L} = \sum p(a)U(a) - \mu\left(\sum p(a) - 1\right) + T\left(-\sum p(a)\log[p(a)] - \mathrm{H}_{min}\right).$$

Both Lagrangian functions lead to the same first-order condition with respect to $p(a)$ when $T = 1/\beta$, implying the duality between the two different classes of model.

From this important result, we can see that the payoff function in the entropy constraint model actually implies the moment constraints of maximum entropy models. For example, if the payoff function is a linear function (the Taylor expansion until the first term), it amounts to putting a mean constraint in the maximum entropy problem. When the payoff function is a linear quadratic function (the Taylor expansion until the second term), this implies constraints on the mean and the variance. As the payoff function reflects a particular economic theory, so too do the constraints of the maximum entropy models.

5. Discussion

We have reviewed two prominent classes of SE models in economics. This section addresses some controversies over the interpretation and justification of SE methods in economics, centering around the legitimacy of applying physical concepts to social system.

5.1 *SE and the Purposeful Behavior of Agents*

One fundamental difference between social systems and physical systems is that human behavior is "purposeful" unlike the movements of particles in a gas. For example, while conscious human beings purposefully keep track of the balance sheets to plan future expenditures, the movement of the particles suspended in a gas is the consequence of their random collision with the fast-moving molecules. The purposefulness in human behavior raises the issue of the legitimacy of using the physical concept of SE to economic problems. Our survey of information theoretic approaches in economics in this paper reveals two promising approaches to reconcile the SE methods and economic analysis.

First, the SE models of individual decision makers is fundamentally the von Neumann–Morgentstern model (von Neumann and Morgenstern, 1944), in which the economic agent purposefully chooses a mixed strategy to maximize the expected payoff, with the additional constraint that the agent is exposed to a positive degree of uncertainty in processing the market signals. In this model, the payoff function can transparently express purposeful human behaviors that might differ in various institutional settings. The entropy-constrained model incorporates other conventional microeconomic theories with complete information, including general equilibrium theory and the fundamental welfare theorems, as a special case because it predicts the central tendency of human behavior along with its endogenous fluctuation.

Second, the purposeful human behavior can be directly incorporated into the maximum entropy model as a constraint. For example, Foley and Scharfenaker's model of profit rate (Scharfenaker and Foley, 2017) introduces the pursuit of higher profit rates by capitalists in terms of the quantal response function as a constraint of the maximum entropy program. This approach has some advantages over other maximum entropy models in that it can simultaneously express the constraint on the purposeful human behavior and the other macroeconomic constraints in the same maximum entropy program.

5.2 *Maximum Entropy Approach as a General Inferential Tool in Economics*

Another way to justify the SE reasoning for the economic problems is to understand the maximum entropy approach as a general inferential procedure. All inference problems in science tries to explain data as a combination of regularities (signal) and residual variation (noise). As we discussed in the overview of information theory, the maximum entropy approach is one that maximizes the randomness of the variation of our knowledge state. Therefore, from the observer's standpoint, the maximum entropy state is maximally uninformative so that it provides the most noncommittal description of the system. Regardless of the physical, chemical, or social systems, the maximum entropy reasoning can be used as long as one can express observer's knowledge state in term of probability distribution.

In this regard, we note that one of the greatest insights of the SE reasoning is that data and theory are closely intertwined. As Stutzer (1995, 1996) shows, data and theory both constraint the optimization problem simultaneously. In Stutzer's model, for example, historical data on the derivative price comes as the prior distribution of the minimum KL divergence, yielding the maximally uninformative distribution of derivative price under the risk-neutral density constraint. Since the maximum entropy procedure is constrained both by theory and data, this procedure is greatly helpful for economics in which theories are abound but relevant data are often limited. With the help of economic theories, the resulting maximum entropy distribution will always lead to the maximally noncommittal inference regardless of how limited data we have (Golan *et al.*, 1996; Golan, 2002).

6. Conclusion

Information theory and statistical mechanics have inspired many economists in pursuing more parsimonious economic models. As we have shown in this paper, SE provides a strong alternative to the conventional notion of economic equilibrium as a fixed state because it gives a distribution over all possible states as an equilibrium. In doing so, it predicts not only the central tendency but also the fluctuations around the tendency as well. For this reason, SE approaches can explain the observed data without relying on arbitrary assumptions about the random noise or external shocks. From the standpoint of SE, the fluctuations or variability around the central tendency is not noise but crucial information about the system because they reflect the outcome of constant interactions among economic agents under the constraints investigators need to figure out.

As a general inferential method, the information theoretic approach can be used for economic problems with limited data. This is because the MEP will lead to the most likely state of the system regardless of data availability. Information theoretic approaches firmly based on economic theories can provide useful insights for many interesting economic problems that the conventional methods have not been able to satisfactorily deal with.

Acknowledgments

This paper benefited from extensive discussions with Duncan Foley and Paulo dos Santos. I am grateful for their helpful comments on the previous versions of this draft. I also owe my colleague Luis Daniel Torres Gonzales for detailed comment on this paper. I am grateful to two anonymous referees for helpful comments and suggestions that significantly improved the paper.

Notes

1. For the discussions on the derivation of entropy, see Shannon (1948), Cover and Thomas (2006), Caticha and Giffin (2006), Chakrabarti and Chakrabarty (2005), Csiszár. (2008), and Knuth and Skilling (2012).
2. The continuous version of entropy (often called the differential entropy), in fact, is not a correct limit of the discrete entropy, because the differential entropy is not invariant to changes in units to measure the random variable (Jaynes, 1957). Two differential entropies of the same quantity measured in different units, for example, oz and liter, will have different value, which is not the case in the discrete entropy. Therefore, a correct way to formulate the different entropy is to add a reference measure for the entropy to be dimensionless in measurement, which leads to $H(P) = \int P(x) \log \frac{P(x)}{m(x)}\, dx$, where $m(x)$ is an invariant measure factor. When $m(x)$ is a probability function, the differential entropy actually is relative entropy.
3. This point can be proven using the Gibbs inequality. See Cover and Thomas (2006).
4. Using the Bayes theorem, the KL divergence indicates the informational gain on average when one replaces the prior distribution $q(x)$ with the posterior distribution $p(x)$.
5. The reason as to the necessity of the logarithm is explained slightly differently in information theory. The logarithm is necessary because the logarithm of n states (or messages) correspond to information contents. For example, if there are two codes to send n messages, $\log_2(n)$ bits of information is required, since $n = 2^{\log_2(n)}$.
6. The basic logic of the derivation, taking the entropy as a number of microstates resulting in a particular macrostate, holds for nonmultinomial systems as well. However, the actual derivation takes different forms depending on different systems. For detailed discussion on the derivation entropy for nonmultinomial systems, see Hanel *et al.* (2014).

7. There is an axiomatic derivation of MEP by Shore and Johnson (1980) that proves that the MEP is the uniquely correct method for inference when information is given in the form of moment constraints.
8. Maximally uninformative, maximally uncertain, and maximally random are exchangeable.
9. In the example of the moving particles in statistical mechanics, one prominent constraint is that the total energy of microstates of particles is constant, which leads to the famous Boltzman–Gibbs distribution of energy level with the temperature T, $p_i \propto e^{-\frac{E_i}{kT}}$, where i is a certain state with an energy E_i and k is Boltzmann's constant.
10. The Shannon entropy is a concave function, so that the First Order Condition (FOC) are enough to secure a global maximum.
11. The gamma distribution is a two-parameter family of continuous probability distributions whose support lies on the positive real line. The functional form of the gamma distribution is the following:

$$f(x; k, \theta) = \frac{x^{k-1} e^{-\frac{x}{\theta}}}{\theta^k \Gamma(k)} \quad \text{for } x > 0 \text{ and } k, \theta > 0.$$

    where k and θ are the shape and the scale parameter.
12. $H(X, Y)$ is the joint entropy of x and y and is defined as $-\sum_x \sum_y P(x, y) \log[P(x, y)]$.
13. Foley and Scharfenaker's approach to the underdetermined problem in which the action variable is unobservable is an example that shows how the principle of maximum entropy is particularly useful for economic problems with limited data. Economists encounter many kinds of underdetermined problems primarily due to a lack of data. The maximum entropy approach provides a systematic way to make a consistent and a robust inference out of ill posed problems. Foley and Scharfenaker's exercise successfully proves this point by applying this method to the profit rate analysis.
14. It is important to note that minimization of KL divergence generalizes the MEP since the latter is the special case of the former when the prior distribution is assumed to be the uniform distribution. For detailed discussion on this issue, see Shore and Johnson (1980), van Campenhout and Cover (1981), Caticha and Giffin (2006), and Knuth and Skilling (2012).
15. Using the same logic, Stutzer (2010) derives the generalized Black–Scholes model of asset prices based on the minimization of relative entropy of conditional risk neutral density.
16. Notably, Mantegna and Stanley (2000) and Sornette (2003) heavily utilize models in physics in analyzing financial market but do not explicitly use SE reasoning.
17. In the same paper, Yakovenko and Rosser (2009) discuss that the maximum entropy distribution of money is also the canonical Gibbs distribution in a closed economy. Unfortunately, it is hard to empirically corroborate their argument due to the lack of data. More importantly, it is not entirely clear if it makes sense to assume that the total amount of money is conserved, considering that the monetary institutions change rapidly and therefore monetary variables, such as the interest rest and the velocity of money, change rapidly as well.

References

Alfarano, S. and Milaković, M. (2008) Does classical competition explain the statistical features of firm growth? *Economics Letters* 101(3): 272–274.

Alfarano, S., Milaković, M., Irle, A., and Kauschke, J. (2012) A statistical equilibrium model of competitive firms. *Journal of Economic Dynamics and Control* 36(1): 136–149.

Borwein, J.M. and Lewis, A.S. (1991) On the convergence of moment problems. *Transactions of the American Mathematical Society* 325(1): 249–271.

Buchen, P. and Kelly, M. (1996) The maximum entropy distribution of an asset inferred from option prices. *The Journal of Financial and Quantitative Analysis* 31: 143–159.

Caticha, A. and Giffin, A. (2006) Updating probabilities. *AIP Conference Proceedings* 872(1): 31–42.

Chakrabarti, C.G. and Chakrabarty, I. (2005) Shannon entropy: Axiomatic characterization and application. *International Journal of Mathematics and Mathematical Sciences* 2005(17): 2847–2854.

Cover, T.M. and Thomas, J.A. (2006) *Elements of Information Theory* (2nd Edition). Hoboken, New Jersey: John Wiley & Sons, Inc.

Csiszár., I., (2008) Axiomatic characterizations of information measures. *Entropy* 10(3): 261–273.

dos Santos, P.L. and Scharfenaker, E. (2016) Capital-market competition, informational performance, and the distribution of tobin's q. NSSR Working Paper Series, 06/2016.

Dragulescu, A. and Yakovenko, V.M. (2000) Statistical mechanics of money. *The European Physical Journal B* 17: 723–729.

Dragulescu, A. and Yakovenko, V.M. (2001) Exponential and power-law probability distributions of wealth and income in the United Kingdom and the United States. *Physica A* 299: 213–221.

Farjoun, E. and Machover, M. (1983) *Laws of Chaos: Probabilistic Approach to Political Economy*. London: Verso Books.

Foley, D.K. (1994) A statistical equilibrium theory of market. *Journal of Economic Theory* 62(2): 321–345.

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

Foley, D.K. (2003) Statistical equilibrium in economics: method, interpretation, and an example. In Frank Hahn, F.P. (ed.), *General Equilibrium: Problems and Prospects* (pp. 95–116). London: Routledge.

Golan, A. (2002) Information and entropy econometrics: Editor's view. *Journal of Econometrics* 107(1,2): 1–15.

Golan, A., Judge, G.G., and Miller, D. (1996) *Maximum Entropy Econometrics: Robust Estimation with Limited Data*. Chichester, New York: Wiley.

Hanel, R., Thurner, S., and Gell-Mann, M. (2014) How multiplicity determines entropy and the derivation of the maximum entropy principle for complex systems. *Proceedings of the National Academy of Sciences* 111(19): 6959–6910.

Jaynes, E.T. (1957) Information theory and statistical mechanics. *Physical Review* 106(4): 620–630.

Jaynes, E.T. (1978) Where do we stand on maximum entropy? In Levine, R.D. and Tribus, M. (eds.), *The Maximum Entropy Formalism* (pp. 15–118). Cambridge, MA: MIT Press.

Jaynes, E.T. (1982) On the rationale of maximum-entropy methods. *Proceedings of the IEEE* 70(9): 939–952.

Jaynes, E.T. (2003) *Probability Theory: The Logic of Science*. Cambridge University Press.

Knuth, K.H. and Skilling, J. (2012) Foundations of inference. *Axioms* 1(1): 38–73.

Kullback, S. and Leibler, R. (1951) On information and sufficiency. *The Annals of Mathematical Statistics* 22: 79–86.

Kydland, F.E. and Prescott, E.C. (1982) Time to build and aggregate fluctuations. *Econometrica* 50(6): 1345–1370.

Luo, Y. and Young, E.R. (2009) Rational inattention and aggregate fluctuations. *The B.E. Journal of Macroeconomics* 9(1), https://www.degruyter.com/view/j/bejm.2009.9.1/bejm.2009.9.1.1700/bejm.2009.9.1.1700.xml.

Luo, Y. and Young, E.R. (2013) Rational inattention in macroeconomics: A survey. Munich Personal RePEc Archive.

Lux, T. (2009) Applications of statistical physics in finance and economics. In Rosser, J.B. (ed.), *Handbook on Complexity Research* (pp. 213–258). Cheltenham: Edward Elgar.

MacKay, D.J. (2005) *Information Theory, Inference, and Learning Algorithms*. Cambridge, England: Cambridge University Press.

Maćkowiak, B. and Wiederholt, M. (2015) Business cycle dynamics under rational inattention. *The Review of Economic Studies* 82(4): 1502–1532.

Mantegna, R.N. and Stanley, H.E. (2000) *Introduction to Econophysics: Correlations and Complexity in Finance*. Cambridge, England: Cambridge University Press.

Matêjka, F. and McKay, A. (2015) Rational inattention to discrete choices: A new foundation for the multinomial logit model. *American Economic Review* 105(1): 272–98.

McFadden, D.L. (1976) Quantal choice analysis: A survey. *Annals of Economic and Social Measurement* 5(4): 363–390.

Rongxi Zhou, R.C. and Tong, G. (2013) Entropy and econophysics. *Entropy* 15: 4909–4931.

Rosser, J.B. (2016) Entropy and econophysics. *The European Physical Journal Special Topics* 225: 3091–3104.
Scharfenaker, E. and dos Santos, P.L. (2015) The distribution and regulation of Tobin's q. *Economics Letters* 137: 191–194.
Scharfenaker, E. and Foley, D.K. (2017) Quantal response statistical equilibrium in economic interactions: Theory and estimation. *Entropy*, 19(9).
Scharfenaker, E. and Semieniuk, G. (2017) A statistical equilibrium approach to the distribution of profit rates. *Metroeconomica* 68(3): 465–499.
Schneider, M.P.A. (2015) Revisiting the thermal and superthermal two-class distribution of incomes. *The European Physical Journal B* 88(1): 5.
Shaikh, A., Papanikolaou, N., and Wiener, N. (2014) Race, gender and the econophysics of income distribution in the usa. *Physica A: Statistical Mechanics and its Applications* 415: 54–60.
Shannon, C. (1948) A mathematical theory of communication. *Bell System Technical Journal* 27: 379–423.
Shore, J.E. and Johnson, R.W. (1980) Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy. *IEEE Transactions on Information Theory* 26(1): 26–37.
Sims, C.A. (2003) Implications of rational inattention. *Journal of Monetary Economics* 50(3).
Sims, C.A. (2006) Rational inattention: A research agenda. Available at http://sims.princeton.edu/yftp/RIplus/RatInattPlus.pdf
Smith, E. and Foley, D.K. (2008) Classical thermodynamics and economic general equilibrium theory. *Journal of Economic Dynamics and Control* 32(1): 7–65.
Stutzer, M.J. (1995) A Bayesian approach to diagnostic of asset pricing models. *Journal of Econometrics* 68: 369–397.
Stutzer, M.J. (1996) A simple nonparametric approach to derivative security valuation. *The Journal of Finance* 51(5): 1633–1652.
Stutzer, M.J. (2010) Simple entropic derivation of a generalized Black-Scholes option pricing model. *Entropy* 2: 70–77.
Toda, A.A. (2010) Existence of a statistical equilibrium for an economy with endogenous offer sets. *Economic Theory* 45(3): 379–415.
Toda, A.A. (2011) Income dynamics with a stationary double Pareto distribution. *Physical Review E* 83: 1–4.
Toda, A.A. (2012) The double power law in income distribution: Explanations and evidence. *Journal of Economic Behavior and Organization* 84: 364–381.
Toda, A.A. (2015) Bayesian general equilibrium. *Economic Theory* 58(2): 375–411.
van Campenhout, J. and Cover, T. (1981) Maximum entropy and conditional probability. *IEEE Transactions on Information Theory* 27(4): 483–489.
von Neumann and Morgenstern (1944) *Theory of Games and Economic Behavior*. Princeton, New Jersey: Princeton University Press.
Woodford, M. (2009) Information-constrained state-dependent pricing. *Journal of Monetary Economics* 56: S100–S124.
Yakovenko, V.M. and Rosser, J.B. (2009) Colloquium: Statistical mechanics of money, wealth, and income. *Reviews of Modern Physics* 81: 1703–1725.
Yang, M. (2015) Coordination with flexible information acquisition. *Journal of Economic Theory* 158: 721–738.