Concept Paper

Micro-Macro Connected Stochastic Dynamic Economic Behavior Systems

George Judge

ARE, Graduate School and Giannini Foundation, 207 Giannini Hall, University of California Berkeley, Berkeley, CA 94720, USA; gjudge@berkeley.edu

check for updates

Abstract: In this paper, we borrow some of the key concepts of nonequilibrium statistical systems, to develop a framework for analyzing a self-organizing-optimizing system of independent interacting agents, with nonlinear dynamics at the macro level that is based on stochastic individual behavior at the micro level. We demonstrate the use of entropy-divergence methods and micro income data to evaluate and understand the hidden aspects of stochastic dynamics that drives macroeconomic behavior systems and discuss how to empirically represent and evaluate their nonequilibrium nature. Empirical applications of the information theoretic family of power divergence measures-entropic functions, interpreted in a probability context with Markov dynamics, are presented.

Keywords: adaptive behavior; causal entropy maximization; information theoretic methods; minimum power divergence; statistical equilibrium; Markov dynamics

JEL Classification: C1; C10; C2

1. Introduction

In two current papers on the state of macroeconomics, Stiglitz (2018) and Hendry and Muellbauer (2018) (SHM) make clear that reigning paradigms involving dynamic stochastic general equilibrium (DSGE) models, do not provide a basis for understanding the hidden economic behavioral processes and systems and predicting the distributive impacts of economic policies and sharp changes in the direction of the economy. In this paper, as a start to cope with these important problems, we borrow some of the key concepts of entropy methods and nonequilibrium statistical systems, to develop a theoretical economic framework for self-organizing-optimizing economic behavior system of independent interacting agents, that involves nonlinear dynamics at the macro-level and stochastic individual behavior at the micro level. In the context of self-organizing equilibrium seeking stochastic economic behavior systems, we discuss and indicate how to use micro data and information theoretic entropy-divergence methods to represent and evaluate the stochastic dynamic nature of an economic behavior system. We recognize that the macro-level phenomena in such systems has no precise position. Thus, the economy may best be considered in a probabilistic context and its dynamics measured against a statistical equilibrium. Nonlinear probability dynamics represented either by stochastic processes or a random dynamic system, provides a natural language (Jaynes and Bretthorst (2003)) and probabilistic reasoning serves as the scientific logic (Gallager (2014), Bhattacharya and Majumdar (2007)). A dynamic economic behavior system's predictability may then be studied quite naturally with information theoretic entropy divergence measures that reflect a state of knowledge in an inferential context (Judge and Mittelhammer (2012)). The objective is to provide a general principle to capture the possibility of independent agents to self-organizing economic behavioral systems to optimize themselves. This optimizing tendency in behavioral systems is defined by the use of entropy information theoretic recovery methods. The rationale behind this information recovery

framework lies in its ability to combine adaptive intelligent behavior and causal entropy maximization. Such a model has the potential to become a new paradigm for analyzing economic behavior systems.

Currently, as noted by SHM most DSGE models and dynamic economic behavior system theories are expressed in terms of reductionist models and use information recovery methods that seek to understand the behavior of the whole from that of its parts. Their econometric DSGE counterparts appear in a functional form model and use observed data consistent empirical sample moments-constraints such as

$$\mathbf{h}(\mathbf{Y}, \mathbf{X}, \mathbf{Z}; \beta) = n^{-1}[\mathbf{Z}\prime(\mathbf{Y} - \mathbf{X}\beta)] \xrightarrow{P} 0 \tag{1}$$

where $\mathbf{Y}$, $\mathbf{X}$ and $\mathbf{Z}$ are respectively a $n \times 1$, $n \times k$, $n \times m$ vector/matrix of dependent variables, explanatory variables and instruments and β is an unknown and unobservable parameter vector. A solution to the stochastic information recovery problem is usually based on parametric and semiparametric likelihood functions and a range of observed data sampling processes involving a finite set of moments, that are based on indirect noisy observations in the effect domain while our interest is in the causal domain. With uncertainty regarding existing conditions, the observations may not come from a single sampling distribution but many out of equilibrium sampling distributions. Higher order moments are assumed to be well behaved, an assumption often violated in real-world markets where long tail distributions prevail. The resulting econometric formulations do not capture important dynamic nonlinear model components and important microscopic details and leads to a blurred and limited vision at the macro level. Obtaining a meaningful probabilistic basis for a realistic dynamic economic behavior system from this mix of model, data and econometric method, requires a computational task that is neither feasible nor meaningful.

1.1. A Micro Based Information Theoretic Approach

Behavioral systems like physical and biological systems are nondeterministic in nature. Thus, the predictability of these dynamic systems may best be considered in a random context and measured in terms of the statistical equilibrium-disequilibrium of the system. More precisely, the probabilistic dynamics, known either as stochastic processes or random dynamical systems, provide a natural language for this purpose. Since the focus is stochastic in nature, system predictability may then be studied quite naturally by information theoretic functionals and the resulting family of power divergence measures-entropic functions, may be interpreted in a probability context and provide measures of the difference between the relevant statistical distributions. Predictability of a system is then equivalent to the study of the nature of its statistical equilibrium. In this context there are two types of predictions on two different levels. At a single system level, the equilibrium is usually local state with maximum probability. In a stochastic dynamic system, the equilibrium is defined as an entire distribution with fixed probability for each and every possible state of a system. It is equilibrium in this context that the entropy functions provide a quantitative measure. In reality, this means a large amount micro data must be available for constructing a distribution, which then can be compared with an equilibrium distribution.

Perhaps at this point, it is useful to note that economics is concerned with the implications of scarcity and market binary choices that play out as a directed weighted network in a temporal-dynamic dimension. As these temporal restricted choice options manifest themselves in an organizing-optimizing behavioral market context, behavior at the macro level is coupled to micro level dynamics and a statistical representation of a dynamic economic system naturally arises. Information theory and information theoretic methods enter into this discussion as a basis for measuring the unknown, or missing information. In a similar vein, from a behavior system information recovery standpoint, Wissner-Gross and Freer (2013) have proposed a connection between adaptive intelligent behavior and causal entropy maximization (AIB-CEM), that provides an entropy basis for self-organizing in multi agent systems. This causal entropic force connection leads to self-organized equilibrium seeking

behavior in an open dynamic economic system and thus establishes entropy driven statistical dynamics. In this context, behavior-market systems may be viewed as equilibrium-stationary state seeking.

A major remaining question concerns how to represent (i) the possibility of multiple equilibria in such a stochastic dynamics framework and (ii) how to describe an economic system as an ensemble that represent all possible configurations and the probability of reaching each one. If we characterize a macro state by its probability distribution of microstates, entropy then is a function of the probability distribution of the economic system on its ensemble of microstates. This permits us to recast economic-behavioral systems in terms of path microstates where entropy reflects the number of ways a macro state can evolve along a path of possible microstates. As we discuss in a later section, this framework permits us to view markets and behavioral and physical stochastic dynamic systems, in terms of a Markov process and as a directed weighted possibly binary network (see Bargigli et al. (2013) and Judge (2016)).

In contrast to past and current approaches to dynamic economic information recovery, we borrow some key ideas from statistical mechanics-nonequilibrium dynamic systems and develop an entropy-divergence based theoretical stochastic economic framework for analyzing economic phenomena at the macro level that is based on, the behavioral dynamics at the micro-level. In this context the objective is to provide a dynamic statistical description of the evolving macroeconomic world. Moving in this direction, we seek a statistical connection between the micro states-configurations and the evolving macro state. To make this connection we make use of the concept of entropy, which has a long history dating back to Boltzmann in the 1970's, Gibbs (1902); Shannon (1948) and Jaynes (1957a, 1957b, 1978). Under this view, micro state behavior provides a foundation for the macroeconomic state and the time evolution of the dynamic economic behavior system under the theoretical framework of statistical physics and combinatorial stochastic processes. This permits us to study systems made of a large number of interacting micro units by means of a set of macro state variables. To define a macroeconomic state, we may then use the entropy based dynamic process that involves time, the micro economic states and the macroeconomic state (see Rovelli (2018)). In this dynamic micro-macroeconomic behavior system-process, micro information plays a role similar to a molecular process in chemical systems, or energy in a thermodynamic system.

1.2. Looking Ahead

In Section 2, we review the characteristics of a self-organized economic behavior system and discuss the connection between adaptive micro behavior and causal entropy maximization. In Section 3, with a criterion for economic information recovery in mind, we (i) introduce the information theoretic Cressie-Read (CR), multi parametric family of entropy power-divergence measures and (ii) indicate how the CR family of divergence measures may be used in a moment estimating function context and as a Markov process with discrete state and time. In Section 4 we indicate how these divergence measures may be used with sample of micro income data to develop economic income probability density functions-distributions. As an empirical example, in Section 5 we use these divergence measures to develop income probability density functions and time dated entropy measures of income inequality to trace economic performance. In Section 6 we exhibit the implications of using Markov state space processes and a transitional reference distribution-entropic prior, on the predicted economic performance of a country. In Section 7 we discuss networks and agent based models (ACE) as frameworks to study the dynamic collective patterns of economic behavior. In Section 8 we conclude by evaluating the implications of this approach for dynamic economic information recovery and look ahead to possible extensions.

2. Self-Organized Economic Behavior-Entropy Connection

Georgescu-Roegen (1971), appears to be one of the first to note and develop a basis for the connection between entropy and economic processes. More recently, Wissner-Gross and Freer (2013), exhibit a connection between adaptive intelligent behavior and causal entropy maximization, that

provides an entropy basis for self-organization-optimization in multi-agent behavioral economic systems. This causal entropic force connection, that is consistent with the idea that an economic system adapts behavior in line with an optimizing principle (see Judge (2015); Miller and Judge (2015) and Judge (2016)), leads to self-organized equilibrium seeking behavior in an open economic behavioral system and establishes an entropy driven statistical framework. In this context behavior-market systems are equilibrium stationary state seeking and the entropic force quantifies the notion that systems are more likely to move from a low probability to a high probability state. Predictability of a dynamic economic system in terms of an entropy function, or entropy functions, is then equivalent to the study of its statistical-nonequilibrium nature.

By making a connection between entropy and economic processes, we may characterize a macro state by its probability distribution-multiplicity of microstates. Entropy then reflects the number of ways a macro state can evolve along a path of possible microstates. This view permits us to recognize that economic data comes from systems with dynamic adaptive behavior that are non-deterministic in nature, involve uncertainties that can be quantified by the notion of information and are driven toward a certain stationary state associated with a functional and hierarchical structure. As we seek new ways to think about the causal adaptive behavior of large complex and dynamic economic systems, we use entropy as the systems status measure-optimizing criterion. Thus, causal entropy maximization is a link that leads us to believe that an economic-behavioral system with a large number of agents interacting locally and in finite time, is in fact optimizing itself. One of the most general information-theoretical functions that measure uncertainties, or missing information, or discrimination, is a family of divergence-entropy measures.

3. The Cressie-Read Minimum Power Divergence Family

To establish a causal influence-econometric-inferential link to the data and solve the resulting statistical equilibrium problem, information theoretic entropy methods provide a natural basis in questions regarding economic behavioral systems. In this information recovery context, a natural solution is to use estimation and inference methods that are designed to deal with systems that may be fundamentally stochastic in nature, where uncertainty and random behavior are basic to information recovery. To investigate the difference between a statistical and another arbitrary statistical equilibrium we use the Cressie and Read (1984) and Read and Cressie (1988) single parameter CR family of entropic function-minimum power divergence (MPD) measures given by

$$I(\mathbf{p}, \mathbf{q}, \gamma) = \frac{1}{\gamma(\gamma+1)} \sum_{i=1}^{n} p_i \left[\left(\frac{p_i}{q_i} \right)^{\gamma} - 1 \right] \tag{2}$$

In (2), the value of γ is a parameter that indexes members of the CR family, the p_i's represent the subject probability distribution, the q_i's are reference probabilities and $\mathbf{p}$ and $\mathbf{q}$ are $n \times 1$ vectors of p_i's and q_i's, respectively.

The usual probability distribution characteristics of p_i, $q_i \in [0,1]$ $\forall i$, $\sum_{i=1}^{n} p_i = 1$ and $\sum_{i=1}^{n} q_i = 1$ are assumed to hold. In (2), as γ varies, the resulting CR-entropy statistical family of estimators that minimize power divergence, exhibit qualitatively different sampling behavior that includes maximum entropy, and in general a range of additive and correlated systems (see Gorban et al. (2010); Judge and Mittelhammer (2011, 2012)). In identifying the probability space, the CR family of power divergences is defined through a class of additive convex functions that represents a broad family of likelihood functional relationships and test statistics. All well-known divergences-entropies belong to this class of CR functions. In addition, the CR measure exhibits convexity in $\mathbf{p}$, for all values of γ and $\mathbf{q}$ and embodies the required probability system characteristics, such as additivity and invariance with respect to a monotonic transformation of the divergence measures. As γ varies, the optimized value of, $I(\mathbf{p}, \mathbf{q}, \gamma)$, represents a range of data sampling processes, likelihood functions and ensembles and the corresponding estimators that minimize power divergence exhibit qualitatively different

sampling behavior. It is important to note that the family of power divergence statistics defined by (2), is symmetric in the choice of which set of probabilities are considered as the subject and reference distribution arguments of the function. In particular, whether the statistic is designated as $I(\mathbf{p}, \mathbf{q}, \gamma)$, or $I(\mathbf{q}, \mathbf{p}, \gamma)$, the same collection of members of the family of divergence measures are ultimately spanned (Qian (2000)). Three discrete CR alternatives for $I(\mathbf{p}, \mathbf{q}, \gamma)$, where $\gamma \in \{-1, 0, 1\}$, have received the most attention in the literature. For example, in the limit as $\gamma \to 0$, the solution of the first-order condition leads to the maximum entropy and the logistic expression for the conditional probabilities. In the limit $\gamma \to 0$, the Maximum Exponential Empirical Likelihood (MEEL) estimator, is the most likely distribution to be observed from a statistical or combinatorial point of view.

3.1. Reformulated Moment Based Model Econometric Example

In the linear moment-based DSGE model context of (1), if we use (2) as the information recovery criterion, along with moment-estimating function information (1), the estimation problem based on the CR minimum power divergence measure may, for any given choice of γ, be reformulated as the following extremum-type estimator for β:

$$\hat{\beta}(\gamma) = \underset{\beta \in \mathbf{B}}{\operatorname{argmin}} \left[\min_p \left\{ I(\mathbf{p}, \mathbf{q}, \gamma) | \sum_{i=1}^n p_i \mathbf{Z}'_i (\mathbf{Y}_i - \mathbf{X}_i \beta) = 0, \ \sum_{i=1}^n p_i = 0, \ p_i \geq 0 \ \forall i \right\} \right], \tag{3}$$

where $\mathbf{q}$ is often taken as a uniform distribution and $\mathbf{B}$ denotes the appropriate parameter space for β. Note in (3) that $\mathbf{X}_i$ and $\mathbf{Z}_i$ denote the *i*th rows of $\mathbf{X}$ and $\mathbf{Z}$, respectively. A solution to the stochastic inverse problem, based on the optimized value of $I(\mathbf{p}, \mathbf{q}, \gamma)$, is one basis for representing a range of data sampling processes, likelihood functions and ensembles. This class of estimation procedures is referred to as *Minimum Power Divergence* (*MPD*) estimation and additional details of the solution to this stochastic inverse problem are provided in Judge and Mittelhammer (2011, 2012). The MPD optimization problem may be represented as a two-step process. In particular, one can first optimize with respect to the choice of the sample probabilities, $\mathbf{p}$ and then optimize with respect to the structural parameters β, for any choice of the CR family of divergence measures identified by the choice of γ, given $\mathbf{q}$. To establish a statistical equilibrium we need to define a probability measure over possible configurations of the system and the probability of reaching each one. To obtain the ensemble distribution that represents all configurations of the system, one possibility is to use of the sample unbiased estimating functions-constraints (1), in the context of the information recovery rule (2). This permits estimation of the fluctuations around the average moment conditions-constraints and makes apparent the relationship between an econometric model and the corresponding statistical ensemble. Solving (3) yields an estimate of the statistical equilibrium consistent with the moment constraints that provide a link to the data. The data and linear moment based problems and conclusions relating to (1) and DSGEs in Section 1, remain.

3.2. A Markov Process with Discrete State and Time

Given the CR-minimum power divergence information recovery base, we now discuss a useful attribute of the divergence functions of the micro state information-distributions. Assume we are observing an economic system on a relatively long, time scale. Then the current state of the system depends largely on the closest previous state and the history over longer terms does not have much influence and a Markov process with finite discrete state and discrete time is obtained. If we denote this Markov process with the Markov matrix P, then we can represent changes in the CR entropy family with Markov dynamics, where *p(t)* and *q(t)* are two distributions following the Markov chain, with different initial distribution *p(0)* and *q(0)* (Qian and Judge (2014)). To show this, a key step is the re-arrangement

$$\frac{p_i(t+1)}{q_i(t+1)} = \frac{\sum_{j=1} p_j(t)P_{ji}}{q_i(t+1)} = \frac{1}{q_i(t+1)} \sum_{\{j=1\}}^{n} \frac{p_j(t)q_j(t)P_{ji}}{q_j(t)}$$

$$= \frac{1}{\sum_{j=1}^{n} q_j(t)P_{ji}} \sum_{j=1}^{n} q_j(t)P_{ji}\left(\frac{p_j(t)}{q_j(t)}\right).$$

Then with a convex function $\varphi(x)$:

$$q_i(t+1)\varphi\left(\frac{p_i(t+1)}{q_i(t+1)}\right) \le \frac{q_i(t+1)}{\sum_{j=1}^{n} q_j(t)P_{ji}} \sum_{j}^{n} q\_j(t)P_{ji}\varphi\left(\frac{p_j(t)}{q_j(t)}\right)$$

$$= \sum_{j=1}^{n} q_j(t)P_{ji}\varphi\left(\frac{p_j(t)}{q_j(t)}\right)$$

The inequality is reversed if φ is concave.

The function $\varphi(x) = x^{\gamma+1}$ is convex when $\gamma > 0$. Therefore,

$$I(\{p_i(t+1)\}, \{q_i(t+1)\}, \gamma) = \frac{1}{\gamma}\left\{\sum_{i=1}^{n} q_i(t+1)\left(\frac{p_i(t+1)}{q_i(t+1)}\right)^{\gamma+1}\right\}$$

$$\le \frac{1}{\gamma}\left\{\sum_{i=1}^{n}\sum_{j=1}^{n} q_j(t)P_{ji}\left(\frac{p_j(t)}{q_j(t)}\right)^{\gamma+1} - 1\right\} \tag{4}$$

$$= \frac{1}{\gamma}\left\{\sum_{j=1}^{n} q_j(t)\left(\frac{p_j(t)}{q_j(t)}\right)^{\gamma+1} - 1\right\} = I(\{p_i(t)\}, \{q_i(t)\}, \gamma)$$

where the inequality follows from the convexity of $(\cdot)^{\gamma+1}$.

In many cases, we are only interested in a periodic irreducible Markov processes, because the probability of a state being strictly periodic is practically zero. Thus, if the entire system can be reduced into subsystems, we can study the subsystems individually. For these cases, the CR entropy approaches zero as p, q approaches the unique stationary distribution. Finally, in the limit of $I(\mathbf{p}, \mathbf{q}, \gamma)$, when $\gamma \to 0$,

$$\lim_{\gamma \to 0} I(\{p_i(t)\}, \{q_i(t)\}, \gamma) = \sum_{i=1}^{n} p_i \ln\left(\frac{p_i}{q_i}\right), \tag{5}$$

which is the Kullback-Leibler (KL) $\mathbf{p}/\mathbf{q}$ divergence measure in the discrete case and the $\gamma \to 0$ member of the CR family (Kullback and Leibler (1951); Kullback (1959) and Good (1963)). This criterion leads to a natural measure of the deviation of the distribution of probabilities $\mathbf{p}$ $\mathbf{q}$ and permits us determine what new information is contained in p relative to that contained in $\mathbf{q}$. Under the principle of minimum discriminability, the difference $I(\mathbf{p}, \mathbf{q})$ is minimized. The entropy measure gives a monotonicity relation to the Markov process and in contrast to moment-based estimation problems, where matching the leading moments does not guarantee convergence in probability measures, KL-divergence upper bounds the large deviation rate function and thus controls moments of all orders. Thus, the divergence functions over probability distributions provide a strong and rigorous macro-economic index for systems with various behaviors. As noted earlier, whether the statistic is designated as $I(\mathbf{p}, \mathbf{q}, \gamma)$ or $I(\mathbf{q}, \mathbf{p}, \gamma)$, the same collection of members of the family of divergence measures are ultimately spanned.

4. Entropy Based Micro State Income Distributions

To develop the idea of an entropy micro base connected to macroeconomic behavior, we to use country based micro income data. In this micro foundation approach to a statistical description at the macro level, markets provide a microscopic basis for processing information on the level of economic activity and determining the value of the components in the portfolio. For example, at the economic unit-country level, the micro income probability density function-distribution and corresponding measure of inequality, contains macroeconomic information on (i) how the market-economic network-system is functioning, (ii) how the allocation and distribution system is performing and (iii) how the economic system and income inequality has changed and is changing over

time. The micro data distribution and corresponding measures, provide a macro observable and a micro constraint that may be used to measure the consequences of fluctuations in the time dated performance of a dynamic economic behavioral system. Samples of micro data and information theoretic entropy-divergence measures of nonequilibrium provide a constrained optimization framework to characterize and predict economic performance and the statistical equilibrium of the system.

To describe the micro information in an economic system, we use the CR-$I(\mathbf{p}, \mathbf{q}, \gamma)$, $\gamma \to 0$ divergence-distance measure and discretize within a histogram graphic structure, a micro sample of income data by a finite number of nonoverlapping intervals, to establish a data-based link to recover the micro income probability distribution. The histograms represent the micro states that are the subject of inference and the number of micro states-micro configurations, called a multiplicity, yields a probabilistic vision of the macro state. Recovering the income probability density function-distribution from a sample of income data, is represented by histograms that spans the income sample space. Samples from the income partitions, yield histogram outcomes of the discrete random income variable d_j, for $j = 1, 2, \ldots, n$ and under repeated observation, one of the micro state-configurations associated with the macro state income is observed with probability p_j. After a large number of trials, a first-moment sample information can be recovered in the form of the mean value of $\sum_{j=1}^{K} d_j p_j = \bar{d}$. Under this specification, when in the limit $\gamma \to 0$, the CR $I(\mathbf{p}, \mathbf{q}, \gamma)$ converges to an estimation criterion equivalent to the maximum exponential empirical likelihood (MEEL) metric $H(p) = -\sum_{j=1}^{K} p_j \ln(p_j)$. Recovery of the extremum problem micro state-entropy function may then be formulated as

$$\max_{p} \left[-\sum_{j=1}^{K} p_j \ln p_j \,\middle|\, \sum_{j=1}^{K} p_j d_j = \bar{d}, \sum_{j=1}^{K} p_j = 1, p_j > 0, \text{ for all } j \right]. \tag{6}$$

The corresponding Lagrange function-extremum recovery problem is

$$L(\boldsymbol{p}, \eta, \boldsymbol{\lambda}) \equiv -\sum_{j=1}^{K} p_j \ln(p_j) + \boldsymbol{\lambda} \left(d - \sum_{j=1}^{K} p_j d_j \right) + \eta \left(1 - \sum_{j=1}^{K} p_j \right). \tag{7}$$

Solving the first-order conditions yields the exponential result

$$\hat{p}_j = \frac{exp\left(-d_j \hat{\boldsymbol{\lambda}}\right)}{\sum_{j=1}^{K} exp\left(-d_j \hat{\boldsymbol{\lambda}}\right)} \tag{8}$$

for the *j*th income outcome and the mean-related micro state income distribution. Using the CR ($\gamma \to 0$) entropy functional, the joint histograms and a country's income data, the resulting micro states-income distribution and the entropy inequality income measure $\sum_{j=1}^{K} p_j \ln(p_j)$ is recovered. In Equation (8), $\hat{p}_j$ is a function of $\hat{\boldsymbol{\lambda}}$, the Lagrange multiplier for constraint (7) and is obtained from the constrained maximization estimation procedure. Given that $\hat{p}_j$, is an implicit function of $\hat{\boldsymbol{\lambda}}$, the maximum entropy distribution does not have a closed form solution and numerical optimization techniques must be used to compute the probabilities.

5. An Empirical Example

As an example of the income information recovery method and inequality measures discussed in the previous sections, we make use of samples of country based micro income data provided by Eurostat, a Directorate-General of the European Commission (see Villas-Boas et al. (2018)). First, we focus on Sweden (SE) and use over 67,000 micro income sample observations, to develop income distributions and entropy measures of income inequality time path for the decade 2004–2013. To discretize the micro income data, we partition the sample into 12 income bins-histograms in order to define and derive the 2004 and 2013 MEEL income distributions for Sweden. As noted from the

time dated 2004 to 2013 income distributions in Figure 1, there has been an important shift toward income equality in Sweden. The comparison of the 2004 and 2013 micro state-distribution in Figure 1, illustrates the *importance* of looking at the *entire income distribution function*, when drawing static and time related inferences concerning income inequality.
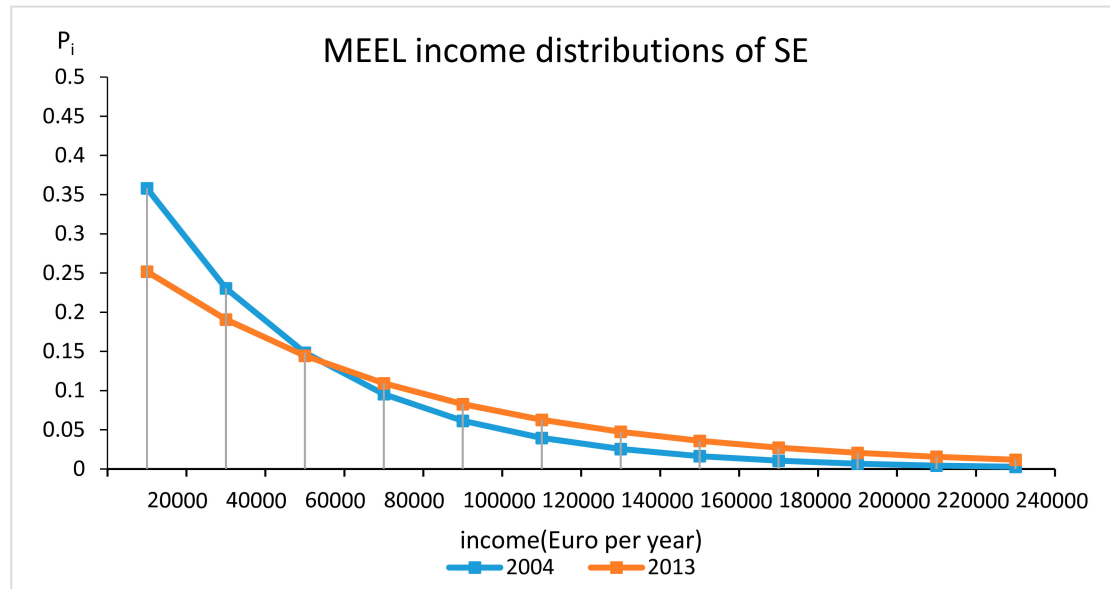

Figure 1. 2004 and 2013 information theoretic micro state-income distributions for Sweden.

The time implication of this change for the complete set of SE micro states-income functions for the years 2004–2013 is displayed in Figure 2, for the yearly entropy income inequality measure $\sum_{j=1}^{12} p_j \ln (p_j)$. Except for the financial crash years of 2009 and 2010, the entropy income inequality measure values for the period 2004–2013 indicate that there has been a steady change over time toward a higher entropy measure-lower inequality-more equal income distribution in Sweden and a higher level of economic activity.
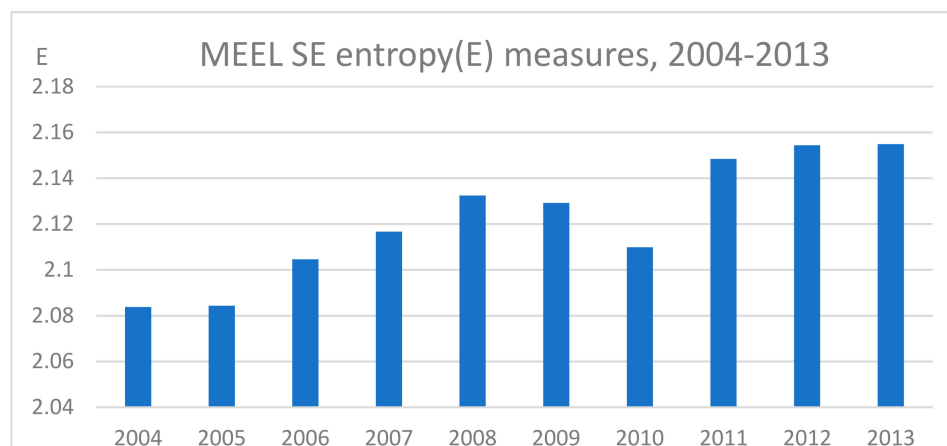

Figure 2. Entropy inequality income measures by year for Sweden.

Taken together, the entropy-based micro states-income distribution information theoretic recovery tools present a basis for measuring and identifying the dynamic nature of economic performance and income inequality. This may, as we demonstrate in the section ahead, provide a basis for predicting the impact of country wise of changes in economic policy.

The Eurostat micro income data may also be used to contrast income distributions over countries that result from different economic and social systems. For example, using data for the period 2008–2013 results in the following Figure 3 MEEL income distributions for SWEDEN, Germany and Greece and corresponding entropy inequality income inequality measures for SWEDEN = 1.987, GERMANY = 1.769 and GREECE = 1.211.
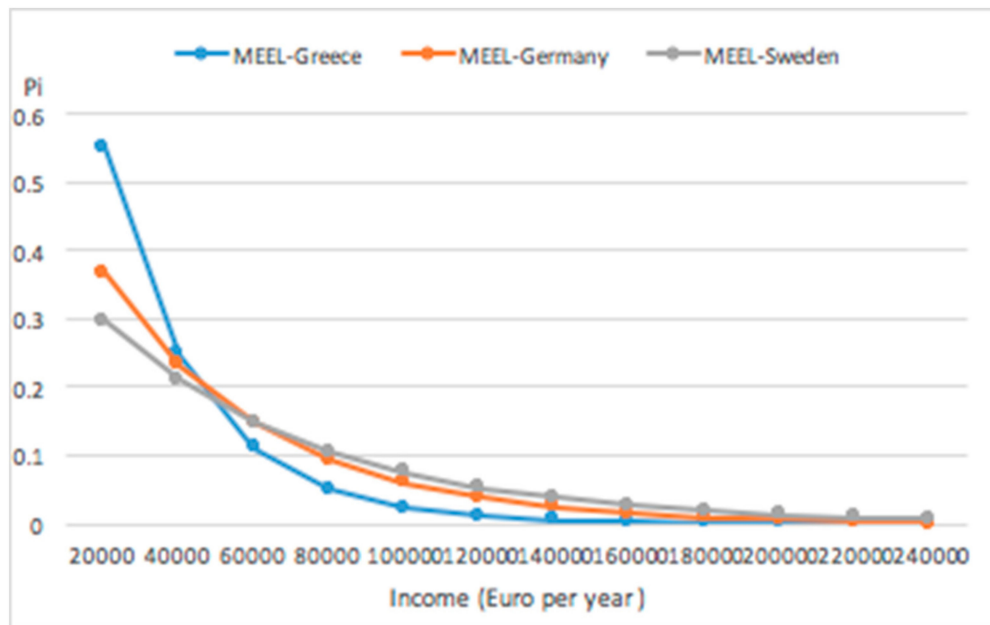

Figure 3. 2008–2013 information theoretic micro state-income. distributions for Sweden, Germany and Greece.

The maximum entropy principle used in obtaining the income configurations in Figure 3, is given the data, a method for obtaining the most likely distribution of observables from the underlying unknown statistical system. In other words, given the micro income data, out of all the distributions that could have happened, these are the ones that could have happened in the greatest number of ways.

6. An Entropic Prior

The micro states-income distributions in Section 5 were obtained, in the context of the CR divergence measure (2), using a uniform reference distribution **q**. In practice when measuring income distributions, we may have non-sample or pre-sample information about the system and the unknown income probabilities $\mathbf{p} = (p_1, p_2, \cdots, p_k)'$, in the form of an entropic prior distribution of probabilities $\mathbf{q} = (q_1, q_2, \cdots, q_k)'$. That is, in addition to the sample of micro data that enters in the consistency relations (7) and (8), there may be some non-uniform reference distribution **q** that is appropriate as a prior micro states-reference distribution. When such prior non-uniform reference information-knowledge exists, we may incorporate this information in the form of (4) and the principle of minimum cross-relative entropy or Kullback-Leibler (KL) directed divergence measure (5). The KL estimation objective function is a particular case of the more general CR family of entropy minimum discrepancy measures (MPD)-distance measures (2) introduced in Section 3. This minimal discriminability principle implies one would choose, subject to the information defined by the moment and other behavioral constraints on the system, the estimate of **p** that can be discriminated, from the non-uniform reference distribution **q** with a minimum of difference. When $\gamma \to 0$, recovering the transition probability density, leads to the minimum divergence entropy **p** and **q**,

$$I(\mathbf{p}, \mathbf{q}) = \sum_{k=1}^{K} p_k \ln\left(p_k/q_k\right) = \sum_k p_k \ln\left(p_k\right) - \sum_k p_k \ln\left(q_k\right) \tag{9}$$

that is nonnegative, equal to zero when **p** = **q** and a convex function of **p**. This criterion is a natural measure of the deviation of the distribution of probabilities **p** and **q**, where **q** is the prior in maximum entropy inferences. Under the principle of minimum discriminability, the difference $I(\mathbf{p}, \mathbf{q})$ is minimized. To take account of both the prior non-uniform reference distribution and the micro data sample information, the minimum cross-entropy solution may be obtained from the minimization problem (9), subject to the moment consistency constraints and the adding up-normalization constraint in (6). In contrast to moment-based estimation problems, where matching the leading moments does not guarantee convergence in probability measures, KL-divergence upper bounds the large deviation rate function and thus controls moments of all orders.

It is important to note again that the *family* of power divergence statistics defined by (2), is symmetric in the choice of which set of probabilities are considered as the subject and reference distribution arguments of the function. As noted in Section 3, whether the divergence statistic is designated as $I(\mathbf{p}, \mathbf{q}, \gamma)$ or $I(\mathbf{q}, \mathbf{p}, \gamma)$, the same collection of members of the family of divergence measures are ultimately spanned and, is a function of the information contained in both the data and the prior. The CR MPD alternative, is a more general measure of dis-similarity between distribution p_i's and q_i's, than the widely used Kullback-Leibler (KL) divergence (5). As we have noted, under $I(\mathbf{p}, \mathbf{q}, \gamma)$ when gamma limits to zero, the information divergence the two statistical equilibriums outcome distributions may be defined as KL divergence. When γ limits to -1, the empirical likelihood member of the family of divergences $I(\mathbf{p}, \mathbf{q}, \gamma)$ is spanned. Therefore, the divergence functions over probability distributions provide a strong and rigorous macro-economic index for systems with various behaviors.

A Transitional Prior Application

Sweden, one of the more egalitarian countries has managed remain to remain efficient in a production context, while keeping its inequality low. The Kullback-Liebler directed divergence principle makes it possible to use the prior information from one country to gauge the impact on the income distribution of another country that may have a higher or lower entropy measure. In this regard in Figure 4 we show the impact of using Sweden's income distribution information and entropy measure of 1.987 as a prior for the income distribution for Germany, which has an entropy measure of 1.769. As a result, the transitional BMEEL income distribution for Germany moves in the direction of more income equality and the entropy measure changes to1.897. In another sense, the transitional income distribution may serve as prediction for either Germany or Sweden and provide a basis for reflecting income distribution dynamics.

In another potential transitional income distribution context, Greece has been facing economic problems and these difficulties have been reflected in terms of its entropy measure and the inequality nature of its income distribution (see Figure 3). Recently the European union imposed a list of economic restrictions on Greece in the hopes of improving its economy and decreasing the level of inequality. To gauge the possible economic impact on Greece of these economic conditions, one possibility is to use the cross entropy-KL divergence principle, with the income distribution of Germany as transitional prior information for Greece. This provides one way of predicting changes in country income distributions and indicating something about the underlying dynamics and statistical equilibrium.
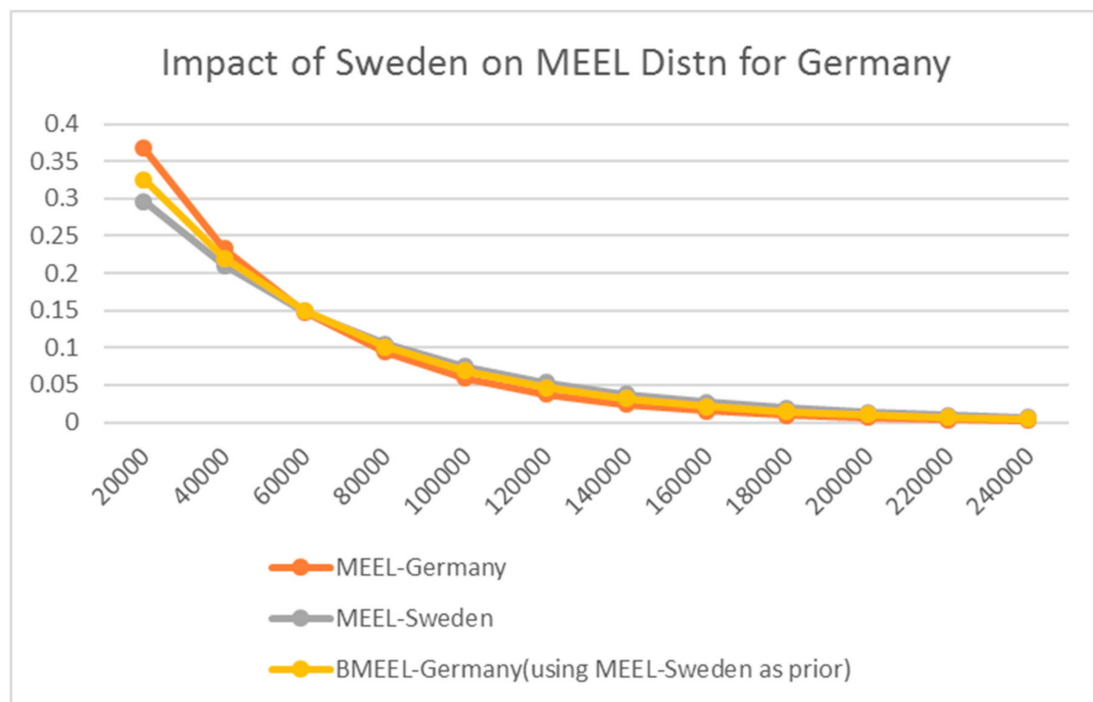
Figure 4. The predicted income distribution for Germany.

7. Related Literature Review

In this paper we have used the concept of entropy to indicate how stochastic individual behavior at the micro level may generate emergent macro regularities. This general topic has received much attention in the economic literature in the last few years. So, for completeness, in the following subsections we recognize the area of network information recovery and the agent-based bottoms up approach to macroeconomics.

7.1. Network Information Recovery

Networks provide an additional useful conceptual framework to study the stochastic dynamic collective patterns of economic behavior. As an extension of the information theoretic approach of the Section 3, we use the concept of a network to provide a basis for analyzing the micro foundations of aggregate-macroeconomics behavior systems. As we seek a systematic way to generate networks-graphs that display desired properties, consider a finite discrete linear version of a network problem

$$\mathbf{y} = A\,\mathbf{p}, \tag{10}$$

where we are given data in the form of a vector $\mathbf{y}$ and the network linear operator A matrix and we wish to determine the unknown and unobservable network ensemble frequency $\mathbf{p} = (p_1, p_2, \dots, p_j)$. Thus, out of all the probability distributions that satisfy (10) and fulfill the non- negative and adding up probability conditions, we wish to recover the unknown pathway probabilities that cannot be justified by just random interaction. For a discussion of a statistical equilibrium representation of an economic market system as a directed complex network, see Bargigli et al. (2013) and Judge (2016).

To indicate the nature and application of information theoretic entropy inference methods to information recovery in behavioral networks, we use the CR family (6) of entropy functionals $I(\mathbf{p}, \mathbf{q}, \gamma)$ network protocols in the form of a constraining adjacency matrix A_{ij} and the unknown ensemble of pathway probabilities, p_j. This results in a nonlinear system of equations where the corresponding Lagrange parameters, provide the solution for the unknown pathway probabilities p_j. If in the context of (10), we indicate by $\mathbf{y}$ the vector of observed flows and by $\mathbf{p}$ the intermediate pathway measures,

the activity of an origin-destination network can be written as the corresponding finite discrete linear inverse problem as

$$y_i = \sum_{j=1}^{N} A_{ij}\, p_j, \tag{11}$$

where y_i and A_{ij} are known, p_j is unknown and $\sum_{j=1}^{n} p_j = 1$, we may use the CR family of entropic divergence measures (6) and write static and dynamic problems and define path entropy in the following constrained conditional optimization form.

$$L = I(\mathbf{p}, \mathbf{q}, \gamma)\,\Big|\, \sum_{i=1}^{n} \lambda_i \Big(\sum_{j=1}^{N} (A_{ij} p_j - y_i) \Big),\ \sum_{j=1}^{n} p_j = 1, p_j \ge 0 \tag{12}$$

Taking a derivative with respect to (12) yields a solution for the path probabilities, p_j, in terms of the Lagrange multipliers and the underlying probability distribution. Since we work with random networks, the observables conditioned by constraints are formulated in terms of statistics computed over the network ensemble. In the solution process we associate the ensemble average with an empirical estimate of such an average. Since the observables depend on network realizations, we equate the ensemble average against the probability of observing such an average. The resulting constrained statistical ensemble acts as a null model for the observed network data. Thus, this type of network inference and monitoring problem has a direct connection to the inverse-regularization problem discussed in Section 4.

In an economic-behavioral network the efficiency of information flow is predicated on discovering or designing protocols that efficiently route information. To carry the above behavioral routing information flow analogy a bit farther, consider the problem of determining least-time point-to-point flows between sub networks, when only *the average origin-destination rate flow volume is* known. Given information about the network protocol in the form of a matrix composed of binary elements A_{ij} that encode the information about the possible connections, the problem is to recover the origin-destination routing flows from the noisy network data. In this origin-destination micro-macro connection-flow context for our problem, we have cars and micro income data, routes and histograms, flow probabilities, destinations, entropy measure heterogeneity of the network and income inequality. The preferred probability distribution of dynamical pathways is the one obtained by maximizing route-flow entropy. In general, the solution for this fundamentally underdetermined problem does not have a closed-form expression and the optimum values of the unknown parameters must be numerically determined (see Cho and Judge (2014) and Squartini et al. (2015)).

7.2. Agent Based Models

Finally, we note a closely related program of research where economic relations are not variants of mechanical DSGE models and researchers have been exploring Agent-based Computational Economics (ACE) models. These models are concerned with the study of open-ended dynamic systems in which the repeated interactions of agents over time can result in sustained out-of-equilibrium dynamics. There is a substantive literature relating to the ACE models that has led to a paradigm shift reflecting how individual behavior at the micro level may generate emergent macro regularities. Tesfatsion and Judd (2006); Bruun (2003); Gatti et al. (2010); Wagner (2012) and Gomes (2014), are some of the theorists who are responsible for the emergence of this new paradigm. The ACE approach treats equilibrium as a testable hypothesis rather than as an imposed constraint. Complex networks provide a useful setting in which to study ACE patterns of collective behavior (see Bargigli and Tedeschi (2014)). The study of agent-based interacting micro units and associated microstate ensembles and the complex networks associated with them, provides an alternative to representative agent approaches and a complement to other applied and theoretical approaches to the study of dynamic behavioral economics. Micro sample

data and agents and Markov processes are an obvious connections between the information recovery methods proposed in this paper and ACE.

8. Summing Up and Looking Ahead

In this paper we have used information recovery methods to mitigate the failure of current paradigms to provide a basis for understanding the hidden dynamics generated from complex macroeconomic behavioral processes and systems. We use samples of household micro income data and borrow some of the key concepts of entropy based minimum divergence methods and physical systems dynamics, to develop an entropy-information theoretic framework for self-organizing-optimizing macroeconomic behavior systems, that is based on nonlinear stochastic dynamics of individual behavior at the micro-level. We focus on statistical economic systems that can be characterized by a macro state for which many configurations exist, that are compatible with it. In a stochastic Markov dynamic context, a transitional reference distribution is used to measure the impact of the use of agent country based prior information on the macro state of another country. Finally, in Section 8 we use the concept of a network to provide a basis for analyzing the micro foundations of aggregate-macroeconomics behavior problems.

Looking ahead we plan to develop a basis for making use of samples of micro data-information to make time dated predictions of the macro variable(s). The use of a transitional reference distribution introduced in Section 6 of this paper and the behavioral model with a Markov process in Section 3, is a start in this direction. Moving from histograms, where we are essentially discretizing the income space, we plan to look at the micro dynamics of a continuous Brownian motion model. The information recovery methods used should provide a basis for combining the information from traditional econometric methods and data based methods such as AI.

Conflicts of Interest: The author declares no conflicts of interest.

References

Bargigli, Leonardo, and Gabriele Tedeschi. 2014. Interaction in Agent Based Economics: A survey of the Network Approach. *Physica A* 399: 1–15. [CrossRef]

Bargigli, Leonardo, Andrea Lionetto, and Stefano Viaggiu. 2013. A Statistical Equilibrium Representation of Markets as Complex Networks. *Journal of Statistical Physics* 165: 351–70. [CrossRef]

Bhattacharya, Rabi, and Mukul Majumdar. 2007. *Random Dynamical Systems: Theory and Applications*. Cambridge: Cambridge University Press.

Bruun, Charlotte. 2003. The Economy as an Agent-based Whole-Simulating Schumpeterian Dynamics. *Industry and Innovation* 10: 475–91. [CrossRef]

Cho, Wendy, and George Judge. 2014. An Information Theoretic Approach to Network Tomography. *Applied Econometric Letters* 97: 201–7. [CrossRef]

Cressie, Noel, and Timothy Read. 1984. Multinomial Goodness of Fit tests. *Journal of the Royal Statistical Society Series B* 4: 440–64.

Gallager, Robert G. 2014. *Stochastic Processes: Theory for Applications*. Cambridge: Cambridge University Press.

Gatti, Domenico Delli, Edoardo Gaffeo, and Mauro Gallegati. 2010. Complex Agent-Based Macroeconomics: A Manifesto for a New Paradigm. *Journal of Economic Interaction and Coordination* 5: 111–35. [CrossRef]

Georgescu-Roegen, Nicholas. 1971. *The Entropy Law and the Economic Process*. Harvard: Harvard University Press.

Gibbs, J. Willard. 1902. *Elementary Principles in Statistical Mechanics*. New York: Charles Scribner.

Gomes, Orlando. 2014. Complex Networks in Macroeconomics. In *Applied and Computational Mathematics*. vol. 3, p. e138. [CrossRef]

Good, Irving. 1963. Maximum Entropy for Hypothesis Formulation, Especially for Multidimensional Contingency Tables. *The Annals of Mathematical Statistics* 34: 911–34. [CrossRef]

Gorban, Alexander N., Pavel A. Gorban, and George Judge. 2010. Entropy: The Markov Ordering Approach. *Entropy* 12: 1145–93. [CrossRef]

Hendry, David, and John Muellbauer. 2018. The Future of Macroeconomics: Macro Theory and Models at the Bank of England. *Oxford Review of Economic Policy* 34: 287–328. [CrossRef]

Jaynes, Edwin T. 1957a. Information Theory and Statistical Mechanics. *Physics Review* 106: 620–30. [CrossRef]

Jaynes, Edwin T. 1957b. Information Theory and Statistical Mechanics. *Physics Review* 106: 171–90. [CrossRef]

Jaynes, Edwin T. 1978. Where Do We Stand on Maximum Entropy? In *The Maximum Entropy Formalism*. Edited by Raphael D. Levine and Myron Tribus. Cambridge: MIT Press, pp. 15–118.

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

Judge, George. 2015. Entropy Maximization as a Basis for Information Recovery in Dynamic Economic Behavioral Systems. *Econometrics* 3: 91–100. [CrossRef]

Judge, George. 2016. Econometric Information Recovery in Behavioral Networks. *Econometrics* 4: 38. [CrossRef]

Judge, George G., and Ron C. Mittelhammer. 2011. *An Information Theoretic Approach to Econometrics*. Cambridge: Cambridge University Press.

Judge, George G., and Ron C. Mittelhammer. 2012. Implications of the Cressie-Read Family of Additive Divergences for Information Recovery. *Entropy* 14: 2427–38. [CrossRef]

Kullback, Solomon. 1959. *Information Theory and Statistics*. New York: John Wiley.

Kullback, Solomon, and Richard Leibler. 1951. On Information and Sufficiency. *The Annals of Mathematical Statistics* 22: 79–86. [CrossRef]

Miller, Douglas J., and George Judge. 2015. Information Recovery in a Dynamic Statistical Markov Model. *Econometrics* 3: 187–98. [CrossRef]

Qian, Hong. 2000. Relative entropy: Free energy associated with equilibrium fluctuations and nonequilibrium deviations. *Physical Review E* 63: 042103. [CrossRef]

Qian, Hong, and George Judge. 2014. *Econometric Information Recovery in Stochastic Dynamic Systems*. Working Paper. Berkeley: University of California.

Read, Timothy R., and Noel A. Cressie. 1988. *Goodness of Fit Statistics for Discrete Multivariate Data*. New York: Springer.

Rovelli, Carlo. 2018. *The Order of Time*. New York: Riverhead Books.

Shannon, Claude Elwood. 1948. A Mathematical Theory of Communications. *Bell Systems Technical Journal* 27: 379–423. [CrossRef]

Squartini, Tiziano, Enrico Ser-Giacomi, Diego Garlaschelli, and George Judge. 2015. Information Recovery in Behavioral Networks. *PLoS ONE* 10: e01277. [CrossRef]

Stiglitz, Joseph. 2018. Where Modern Macroeconomics Went Wrong. *Oxford Review of Economic Policy* 34: 70–106.

Tesfatsion, Leigh, and Kenneth Judd. 2006. *Agent Based Computational Economics: A Constructive Approach to Economic Theory*. Amsterdam: Elsevier.

Villas-Boas, Sofia B., Qiuzi Fu, and George Judge. 2018. Entropy Based European Income Distributions and Inequality Measures. *Physica A* 514: 686–98. [CrossRef]

Wagner, Richard. 2012. A Macro Economy as an Ecology of Plans. *Journal of Economic Behavior and Organization* 82: 433–44. [CrossRef]

Wissner-Gross, Alexander D., and Cameron E. Freer. 2013. Causal Entropic Forces. *Physical Review Letters* 110: 167802. [CrossRef] [PubMed]