

Judge, George

Article

Entropy maximization as a basis for information recovery in dynamic economic behavioral systems

Econometrics

Provided in Cooperation with:

MDPI – Multidisciplinary Digital Publishing Institute, Basel

Suggested Citation: Judge, George (2015) : Entropy maximization as a basis for information recovery in dynamic economic behavioral systems, *Econometrics*, ISSN 2225-1146, MDPI, Basel, Vol. 3, Iss. 1, pp. 91-100,
<https://doi.org/10.3390/econometrics3010091>

This Version is available at:

<https://hdl.handle.net/10419/171818>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<http://creativecommons.org/licenses/by/4.0/>

Article

Entropy Maximization as a Basis for Information Recovery in Dynamic Economic Behavioral Systems

George Judge

Graduate School, 207 Giannini Hall, University of California Berkeley, Berkeley, CA 94720, USA;
E-Mail: gjudge@berkeley.edu; Tel.: +1-510-642-0791

Academic Editor: Kerry Patterson

Received: 16 December 2014 / Accepted: 5 February 2015 / Published: 16 February 2015

Abstract: As a basis for information recovery in open dynamic microeconomic systems, we emphasize the connection between adaptive intelligent behavior, causal entropy maximization and self-organized equilibrium seeking behavior. This entropy-based causal adaptive behavior framework permits the use of information-theoretic methods as a solution basis for the resulting pure and stochastic inverse economic-econometric problems. We cast the information recovery problem in the form of a binary network and suggest information-theoretic methods to recover estimates of the unknown binary behavioral parameters without explicitly sampling the configuration-arrangement of the sample space.

Keywords: information-theoretic methods; adaptive behavior; causal entropy maximization; pure and stochastic inverse problems; binary network; dynamic economic systems

JEL Classification: C10; C24; C40

1. Introduction

A central problem in economics is to predict or make inferences concerning behavioral choices. We refer to the process of modeling, estimation, and inference as information recovery. In economics, information recovery is often based on observed and quasi-experimental behavioral data. The resulting mix of microeconomic models and data usually requires the solution of a pure or stochastic ill-posed inverse problem. Given the constraints of the traditional theoretical micro behavioral model and data mix, the objective of this paper is to provide a basis for recovering the expected value of the unknown behavioral parameters without explicitly sampling the sample space. As a status-optimizing criterion measure, we emphasize the connection between adaptive intelligent behavior, causal entropy

maximization, and self-organized equilibrium seeking behavior in an open dynamic economic system. Information-theoretic methods are suggested as a solution to the inverse information recovery problem. As an example of this type of information recovery problem, we consider a binary dynamic network problem and use an adaptive intelligent behavior causal entropy maximization (AIB-CEM) basis for recovering the unknown binary behavioral parameters.

In contrast to a standard deterministic equilibrium solution of the competitive microeconomic model, in this paper, we recognize that economic data comes from dynamic adaptive behavior systems that are non-deterministic in nature, involve information and uncertainty, and are driven toward a certain optimal stationary state associated with a functional and hierarchical structure [1–3]. The resulting dynamic economic system involves interdependent micro components and gives rise to an instantaneous feedback adaptive behavior world that is seldom, if ever, in equilibrium. As we seek new ways to think about the causal adaptive behavior of large, complex, and dynamic micro economic systems, in the sections ahead we use entropy as the systems status measure. 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 and the more diverse the number of microstates, the larger the causal path entropy. A uniform-unstructured distribution of the microstates corresponds to a macro state with maximum entropy and minimum information.

From an economic information recovery standpoint, we follow Wissner-Gross and Freer [4] and recognize the connection between adaptive intelligent behavior, causal entropy maximization (AIB-CEM) and self-organized equilibrium seeking behavior in an open dynamic economic system. In this context, economic systems are equilibrium-stationary state seeking, but may not be in equilibrium. Although there is only one stationary state consistent with an economic system in equilibrium, there are a large number of ways an economic path-dependent competitive interacting processes-system may be out of equilibrium. Thus in the behavioral area, 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. In this setting, information, a commodity, economic value, optimal resource allocation, and causal path entropy represent essentially the same thing and data outcomes are behavior-related in the same sense that prices do not behave, people behave. In a similar vein, in the computer science area, maximum causal inference has been used with dynamically-sequential information revealed from interacting processes (see [5,6]).

The connection between causal adaptive behavior and entropy maximization, based on a causal generalization of entropic forces, is consistent with the idea that economic social systems do not evolve in a deterministic or a random way, but tend to adapt behavior in line with an optimizing principle. This is a natural process in an effective working system. One reason for seeking an entropy-based adaptive behavior causal framework is that it permits the interpretation of adaptive economic behavior in terms of entropic functions and, thereby, the use of information-theoretic methods. This consistency of the economic and econometric models, the data, and the information recovery-estimation and inference processes, has potential for turning economics from a descriptive science to a predictive or at least a comprehensive and behavior related quantitative one.

In the sections ahead, with information recovery in mind, we recognize that measuring evidence and making inferences for this type of behavior-flow problem, requires the solution of a pure or stochastic inverse problem. As a basis for solving this type of problem we use the adaptive intelligent behavior

causal entropy maximization (AIB-CEM) connection and suggest an information-theoretic family of entropic functions, as a basis for linking the data and the unknown and unobservable behavioral parameters. As an example, we suggest an information-theoretic framework as a basis for recovering the unknown optimum pathway probabilities of a general binary network based on aggregate behavioral flow data.

2. The Information Recovery Base

In the behavioral sciences, there has been a growing interest in developing economic-econometric information-theoretic formulations that will aid in drawing conclusions and making inferences about causal relationships/influences in complex dynamic systems. The recovery of causal information is of course a basic objective and central to all branches of science. In this context, our ability to measure evidence and make inferences is directly linked to the economic-econometric model, the available sample of data and the appropriate information recovery method.

In attempts to recover causal information, it is important to recognize the fragile nature of behavioral economic-econometric models. Although behavioral models add the component structure, incorrect constraints may close the system and lead to the identification of incorrect system stationary states and the distribution of the underlying statistical noise. In the behavioral sciences, the data usually consists of indirect-noisy effects observations that come from an uncontrolled observational sampling process that often contain a variety of systematic errors. This type of data makes it impossible to distinguish between mutual influence and causal influence, and thus does not contain location or directional information. Even introducing a lag in the mutual observations fails to distinguish information that is actually exchanged from shared information and does not support time causality. Confounders of various sorts are usually present and abundant. These concerns taken together or individually may dominate the statistical variability of the item of interest and obscure the information that one hopes to measure. These types of specification and data problems lead to biases and incorrect inferences and, in reality, do not present a reliable basis for developing dynamic microeconomic theory and making causal inferences of a supposed treatment in observational and quasi experimental settings (see for example [7–9]).

Given imperfect reductionist and often toy economic-econometric models and indirect noisy effects data, a final question concerns the choice of a stochastic basis for causal information recovery, measuring evidence and making defensible inferences. At this juncture, it is important to realize that the indirect noisy observations used as a basis for identifying the underlying adaptive behavior of dynamic microeconomic systems and to measure causal influence, usually requires the solution of a pure or stochastic inverse problem. The data are in the effects domain, and our interest lies in the causal domain. The number of measurements-data points, are often smaller than the number of unknown parameters to be estimated, and, thus, the stochastic inverse problem is in addition ill-posed. Without a large number of assumptions, tuning parameters, pseudo likelihoods, kernel distributions, and regularization methods, the resulting stochastic ill-posed underdetermined inverse problem cannot be solved by traditional information recovery methods discussed by Hastie *et al.* [10] and applied by Smith *et al.* [11]. Since stochastic inverse problems in behavioral economic-econometrics appear to be

the rule rather than the exception, in the next section, we discuss information-theoretic methods designed for this type of information recovery problem.

3. Information Recovery Framework

In Section 2, we noted the connection between adaptive intelligent behavior and causal entropy maximization. This connection suggests a basis for establishing a causal influence-econometric model link to the data. With this behavioral-entropy connection, a natural solution is to make use of information-theoretic estimation and inference methods that are designed to deal with the nature of economic-econometric models and data, and the resulting pure and stochastic inverse problems. In developing a basis for the use of information-theoretic (IT) methods, we focus on a stochastic ill-posed inverse problem of which the pure-without noise inverse problem is just a special case. In this context, the Cressie and Read [12] Read and Cressie [13] (CR) family of entropic functions, provide a basis for linking the data and the unknown and unobservable behavioral model parameters. These functions permit the researcher to exploit the statistical machinery of information theory to gain insights about the causal behavior of a dynamic process from a system that may not be in equilibrium. Thus, in developing an information-theoretic econometric approach to estimation and inference, the CR parameter family represents a way to link the entropic behavior informational functions with the underlying sample of data. Information-entropic functions of this type have an intuitive interpretation that reflects uncertainty as it relates to a model of the adaptive behavior of micro economic processes.

In identifying estimation and inference measures that may be used as a basis for characterizing the data sampling process for indirect-noisy observed data outcomes, we begin with the CR multi parametric convex family of entropic functional-power divergence measures:

$$I(\mathbf{p}, \mathbf{q}, \gamma) = \frac{1}{\gamma(\gamma + 1)} \sum_{i=1}^n p_i \left[\left(\frac{p_i}{q_i} \right)^\gamma - 1 \right]. \quad (3.1)$$

In Equation (3.1), γ is a parameter that indexes members of the CR family, p_i 's represent the subject probabilities and the q_i 's, are interpreted as reference probabilities. Being probabilities, 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 Equation (3.1), as γ varies, the resulting CR family of estimators that minimize power divergence, exhibit qualitatively different sampling behavior that includes Shannon's entropy, the Kullback-Leibler measure, and, in a binary context, the logistic distribution-divergence.

The CR family of power divergence is defined through a class of additive convex functions that encompass a broad family of test statistics, and represents, within a moments-based estimation context, a broad family of likelihood functional relationships. In addition, the CR measure exhibits proper 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. In the context of extremum metrics, the general CR family of power divergence statistics represents a flexible family of pseudo-distance measures from which to derive empirical probabilities and encompasses a wide array of empirical goodness-of-fit and information recovery criteria.

As γ varies, the resulting estimators that minimize power divergence exhibit qualitatively different sampling behavior.

3.1. Traditional Economic-Econometric Behavioral-Choice Models

As a first example, consider a stochastic economic-econometric model of behavioral equations that involve endogenous and exogenous variables. Data consistent with the economic-econometric model may be reflected in terms of empirical sample moments-constraints such as $\mathbf{h}(\mathbf{Y}, \mathbf{X}, \mathbf{Z}; \boldsymbol{\beta}) = n^{-1} [\mathbf{Z}'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})] \xrightarrow{p} \mathbf{0}$, where \mathbf{Y} , \mathbf{X} and \mathbf{Z} are respectively a $n \times 1$, $n \times k$, $n \times m$ vector/matrix of explanatory variables and instruments, with parameter vector $\boldsymbol{\beta}$ the objective of information recovery. A solution to the stochastic inverse problem, in the context of Equation (3.1) and based on the optimized value of $I(\mathbf{p}, \mathbf{q}, \gamma)$, is one basis for representing a range of data sampling processes and likelihood-entropy functions. As γ varies, the resulting rules that minimize power divergence exhibit qualitatively different sampling behavior. Using empirical sample moments, a solution to the stochastic inverse problem, for any given choice of the γ parameter, may be formulated as the following extremum-type information recovery basis for $\boldsymbol{\beta}$

$$\hat{\boldsymbol{\beta}}(\gamma) = \arg \min_{\boldsymbol{\beta} \in \mathbf{B}} \left[\min_{\mathbf{p}} \left\{ I(\mathbf{p}, \mathbf{q}, \gamma) \mid \sum_{i=1}^n p_i \mathbf{z}'_i (y_i - \mathbf{x}_i \boldsymbol{\beta}) = \mathbf{0}, \sum_{i=1}^n p_i = 1, p_i \geq 0 \forall i \right\} \right], \quad (3.2)$$

Unless out of sample information is available, \mathbf{q} is usually taken as a uniform-non-informative distribution. For a discussion of these Minimum Power Divergence (MPD) information recovery methods see [14–16].

In connection with the MPD information-theoretic methods, it is important to mention the analysis of binary response data-models (BRMs), that include discrete choice econometric models (see for example [17]). With these models, the objective is to predict probabilities that are unobserved and unobservable, from indirect noisy observations. Traditionally the estimation and inference methods, used in empirical analyses of binary response models, converts this fundamentally ill-posed stochastic inverse problem into a well-posed one that can be analyzed via conventional parametric statistical methods. This is accomplished by imposing a parametric functional form on the underlying data generating distribution. Seeking to minimize the use of unknown information concerning model components, we characterize the $n \times 1$ vector of Bernoulli random variables, \mathbf{Y} , by the universally applicable stochastic representation

$$\mathbf{Y} = \mathbf{p} + \boldsymbol{\varepsilon}, \text{ where } E(\boldsymbol{\varepsilon}) = \mathbf{0} \text{ and } \mathbf{p} \in \prod_{i=1}^n (0,1) \quad (3.3)$$

The specification in Equation (3.3) implies only that the expectation of the random vector \mathbf{Y} is some mean vector of Bernoulli probabilities \mathbf{p} , and that the outcomes of \mathbf{Y} are decomposed into means and noise terms. Given sampled binary outcomes from Equation (3.3), if the Bernoulli probabilities in Equation (3.3) are allowed to depend on the values of explanatory variables \mathbf{x} , we may use $k < n$ empirical moment representations of the orthogonality conditions, $n^{-1} \mathbf{x}'(\mathbf{y} - \mathbf{p}) = \mathbf{0}$, to connect the data space to the unknown-unobservable probabilities. It is straightforward to extend the univariate distribution formulations to their multivariate counterparts. For example, one such extension when

$\gamma = 0$, subsumes the multivariate logistic distribution as a special case and results in a multinomial specification of the minimum power divergence estimation problem in Lagrange form as

$$L(\mathbf{p}, \boldsymbol{\lambda}) = \sum_{i=1}^n \sum_{i=1}^n \left(\frac{1}{\gamma(\gamma+1)} \sum_{j=1}^m p_{ij} \left[\left(\frac{p_{ij}}{q_{ij}} \right)^\gamma - 1 \right] \right) + \sum_{j=1}^m \boldsymbol{\lambda}'_j \mathbf{x}'_j (y_j - p_j) + \sum_{i=1}^n \eta_i \left(\sum_{j=1}^m p_{ij} - 1 \right) \quad (3.4)$$

Solving first order conditions with respect to the p_{ij} 's leads to the standard multivariate logistic distribution, when the reference distributions are uniform.

3.2. Convex Entropic Divergences

In choosing a member of the CR family of likelihood-divergence functions, one might follow Gorban and Karlin [18] and consider a bounded parametric family of convex information divergences which satisfy additivity and trace conditions. Convex combinations of $\gamma = 0$ and $\gamma = -1$ span an important part of the probability space and produce a remarkable family of distributions. This parametric family of divergence measures is essentially the linear convex combination of the cases where $\gamma = 0$ and $\gamma = -1$. This family is tractable analytically and provides a basis for joining (combining) statistically independent subsystems. When the base measure of the reference distribution \mathbf{q} is taken to be a uniform non-informative probability density function, we arrive at a one-parameter family of additive convex dynamic functions. From the standpoint of extremum-minimization with respect to \mathbf{p} , the generalized divergence family, under uniform \mathbf{q} , reduces to

$$S_\alpha^* = \sum_{i=1}^n \left(-(1-\alpha) p_i \ln(p_i) + \alpha \ln(p_i) \right), \quad 0 \leq \alpha \leq 1 \quad (3.5)$$

In the limit, as $\alpha \rightarrow 0$, the minimum I divergence of the probability mass function \mathbf{p} , with respect to \mathbf{q} , is recovered. As $\alpha \rightarrow 1$, the maximum empirical likelihood (MEL) solution is recovered. This generalized family of divergence measures permits a broadening of the canonical distribution functions and provides a framework for developing a quadratic loss-minimizing estimation rule.

4. Binary Network Problem

Given an information recovery framework, in order to go beyond traditional overly simplified modeling and mathematical anomalies, consider a new network based paradigm that is developing under the name of Network Science (for example, see [19,20], and the references contained therein). It is based on observed adaptive behavior data sets that are indirect, incomplete and noisy. This representation of markets arises quite naturally from microeconomic theory. In fact, in many ways, markets and binary linked networks are equivalent (see [21]). There are several things that make this approach attractive for information recovery in economics and the other social sciences. In the economic-behavioral sciences, everything seems to depend on everything else and this fits right in to the interconnectedness of the nonlinear dynamic network paradigm. There is also a close link between evolving network structures and the equilibrium or disequilibrium of economic-behavioral systems and entropy maximization. Finally, in terms of a methodology, network problems are consistent with the information-theoretic approach to information recovery.

In general, the representation of a market as a network presents the consideration of a market in terms of a micro canonical ensemble. Thus, if B is a binary network, then its links only take on binary values and may be represented by a matrix A with binary values. This leads to a binary network with N vertices that is specified by an $N \times N$ matrix A , with entries $A_{ij} = 1$, if the vertices i and j are connected, and $A_{ij} = 0$ otherwise. Analytically, we seek an expression for the probabilities that are connected in the random-statistical ensemble of pathways. Our objective is to recover expected values across the ensemble that can be computed analytically, without explicitly sampling the configuration space.

Given information about the network routing protocol in the form of a matrix A , with entries A_{ij} , the unknown P_{ij} , pathway probabilities often must be estimated from aggregate data that may be noisy in nature. In addition the number of unknown pathway parameters of the protocol matrix A , are much larger than the number of measured aggregate origin-destination data points. Thus the components of matrix A cannot be observed directly. As a result, indirect and possibly noisy observable data must be used to recover information on these unobserved and unobservable model components. This means that, although the observed data are considered to be directly influenced by the values of model components, the observations are not themselves the direct values of these components and only indirectly reflect the influence of the components. The relationship characterizing the effect of unobservable components on the observed data must be somehow inverted to recover information concerning the unobservable model components from the indirect observations. Thus, the analyst must use indirect noisy observations to recover information on the unobserved vector of parameters and unobserved and unobservable random components. This means that this type of ill-posed pure or stochastic inverse regularization problem cannot be solved by traditional information recovery-econometric methods, without making use of regularization schemes such as noted in Section 2. As a solution basis, entropy pathway maximization problems of this type may be formulated as a problem of maximizing the entropy over the pathways, subject to constraints. The result provides an exact expression for the occurrence of the unknown probabilities over the ensemble of pathways and yields the preferred probability distribution (see [22]).

A Network Behavior Recovery

To indicate the applicability of the information-theoretic approach in the binary network area, an example may be useful. In an economic-behavioral network, the efficiency of information flow is predicated on discovering or designing protocols that efficiently scale free patterns. In many ways this is like a transportation network where the emphasis is on design and efficiency in routing the traffic flows (for example, see [23] and the references therein). To carry this information flow analogy a bit farther, consider the problem of determining least-time point-to-point traffic flows between sub networks, when only aggregate origin-destination volumes are known. Given information about the network protocol in the form of a matrix A_{ij} composed of binary elements, traffic flows may be estimated from the noisy aggregate traffic data. If the amount of unknown origin to destination routes is much larger than the amount of origin-destination data, then we have an ill-posed linear inverse problem of the type first introduced in Section 2. If we write the inverse problem as

$$y_i = \sum_{j=1}^N A_{ij} p_j, \quad (4.1)$$

where, y_i and A_{ij} are known, p_j is unknown and $\sum_{j=1}^N p_j = 1$, we may make use of the CR family of entropic divergence measures Equation (3.1) and write the problem as the following constrained conditionally optimization problem:

$$I(\mathbf{p}, \mathbf{q}, \gamma) \mid \sum_{i=1}^n \lambda_i \left(\sum_{j=1}^N (A_{ij} p_j - y_i) \right), \sum_{j=1}^n p_j = 1, p_j \geq 0 \quad (4.2)$$

This is just the solution to a standard problem when a function must be inferred from insufficient sample-data information. Thus network inference and monitoring problems have a strong resemblance to an inverse problem in which key aspects of a system are not directly observable. Details of the application of information-theoretic entropic methods to this type of network information flow problem are discussed in Cho and Judge (2015, 2007) [24,25] and Ziebart, Bagnell, and Dey [5,6]. Finally, it is worth emphasizing that in actual networks, the flows from one node to another will themselves affect node-to-node capacities that may impact deterministic or statistical predictions [26]. Finally, it is interesting that network theory presents a model for producing scale free patterns that are manifestations in the physics world of least-time-free energy consumption. In other words if economic systems did not consume energy in the least time, these patterns would not be present.

5. Conclusions

In this paper we have:

- (i) Exhibited a connection between adaptive economic behavior and causal entropy maximization in self organizing equilibrium seeking dynamic economic systems,
- (ii) Used a broad family of entropic functionals to provide an information-theoretic solution for ill-posed pure and stochastic inverse problems,
- (iii) Used a binary network to illustrate the applicability of information-theoretic methods,
- (iv) Demonstrated that networks are a useful way to model micro systems models and can be adapted to serve various purposes, and
- (v) Demonstrated the general applicability of the adaptive-optimizing behavior information theoretic concept in the context of ill-posed inverse economic settings. Given the importance of recovering dynamic economic behavioral information, a natural question arises as to the continued use of traditional regularization methods as a solution basis for traditional pure and stochastic inverse type problems.

Finally, we noted the statistical implications of using imperfect economic-econometric models and data in Section 2, and, for solution purposes, we argued for the need to solve a pure or stochastic inverse problem. As a start toward mitigating these problems, in Section 3, we suggested an adaptive intelligent behavior causal entropy maximization connection and a corresponding information-theoretic recovery framework. In contrast, many traditional observational, experimental, and game-theoretic data based economic and econometric models and methods are disconnected from the underlying

nature of the dynamic behavioral process. Consequently, a natural question arises as to the frequent use of ad hoc traditional status measures/criteria as a solution basis. The connection between adaptive dynamic economic behavior and causal entropy maximization appears to offer one way to move economics in the direction of a behavior-related predictive quantitative science.

Acknowledgments

I would like to acknowledge the helpful comments of A. Annala; W. Cho; D. Giles; W. Griffiths; D. Hammer; K. Jack; J. Lee; R. Mittelhammer; J. Perloff; S. Villas-Boas; A. Wissner-Gross; A. Zaman and B. Ziebart.

Conflicts of Interest

The author declares no conflict of interest.

References

1. Georgescu-Roegen, N. *The Entropy Law and the Economic Process*; Harvard University Press: Cambridge, MA, USA, 1971.
2. Raine, A.; Foster, J.; Potts, J. The new entropy law and the economic process. *Ecol. Complex.* **2006**, *3*, 354–360.
3. Annala, A.; Salthe, S. Economies evolve by energy dispersal. *Entropy* **2009**, *11*, 606–633.
4. Wissner-Gross, A.; Freer, C. Causal entropic forces. *Phys. Rev. Lett.* **2013**, *110*, 168702, doi:10.1103/PhysRevLett.110.168702.
5. Ziebart, B.; Bagnell, J.; Dey, A. Modeling interaction via principle of maximum entropy. In Proceedings of the International Conference on Machine Learning, Haifa, Israel, 21–24 June 2010.
6. Ziebart, B.; Bagnell, J.; Dey, A. The principle of maximum causal entropy for estimating interacting processes. *IEEE Trans. Inf. Theory* **2013**, *59*, 1966–1980.
7. DiPrete, T.; Gangl, M. Assessing bias in the estimation of causal effects: Rosenbaum bounds on matching estimators and instrumental variables estimation with imperfect instruments. *Sociol. Methodol.* **2004**, *34*, 271–310.
8. Angrist, J.D.; Krueger, A.B. Instrumental variables and the search for identification: From supply and demand to natural experiments. *J. Econ. Perspect.* **2001**, *15*, 69–85.
9. Smith, V. Experimental methods in economics. In *The New Palgrave Dictionary of Economics*, 2nd ed.; Palgrave Macmillan: Cambridge, UK, 2008.
10. Hastie, T.; Friedman, J.; Tibshirani, R. *The Elements of Statistical Learning*; Springer: New York, NY, USA, 2009; Volume 2.
11. Smith, A.; Bernheim, B.D.; Camerer, C.; Rangel, A. Neural activity reveals preferences without choices. *Am. Econ. J. Microecon.* **2014**, *6*, 1–3.
12. Cressie, N.; Read, T. Multinomial goodness-of-fit tests. *J. R. Stat. Soc. Ser. B (Methodol.)* **1984**, *46*, 440–464.
13. Read, T.; Cressie, N. *Goodness-of-Fit Statistics for Discrete Multivariate Data*; Springer Verlag: New York, NY, USA, 1988.

14. Gorban, A.; Gorban, P.; Judge, G. Entropy: The Markov ordering approach. *Entropy* **2010**, *12*, 1145–1193.
15. Judge, G.; Mittelhammer, R. *An Information Theoretic Approach to Econometrics*; Cambridge University Press: Cambridge, UK, 2011.
16. Judge, G.; Mittelhammer, R. Implications of the Cressie-Read family of additive divergences for information recovery. *Entropy* **2012**, *14*, 2427–2438.
17. Mittelhammer, R.; Judge, G. A family of empirical likelihood functions and estimators for the binary response model. *J. Econom.* **2011**, *164*, 207–217.
18. Gorban, A.; Karlin, I. Family of additive entropy functions out of thermodynamic limit. *Phys. Rev. E* **2003**, *67*, 016104, doi:10.1103/PhysRevE.67.016104.
19. Willinger, W.; Doyle, J. Mathematics and the Internet: A source of enormous confusion and great potential. *J. Am. Math. Soc.* **2009**, *56*, 586–599.
20. Barabasi, A. The network takeover. *Nat. Phys.* **2012**, *8*, 14–16.
21. Bargigli, L.; Lionetto, A.; Viaggiu, S. A Statistical Equilibrium Representation of Markets as Complex Networks. Available online: <http://arxiv.org/abs/1307.0817> (accessed on 22 July 2013).
22. Presse, S.; Kingship, G.; Lee, J.; Dill, K. Principles of maximum entropy and maximum caliber in statistical physics. *Rev. Mod. Phys.* **2013**, *85*, 1115, doi:10.1103/RevModPhys.85.1115.
23. Castro, R.; Mark, C.; Gang, L.; Robert, N.; Yu, B. Network tomography: Recent developments. *Stat. Sci.* **2004**, *19*, 499–517.
24. Cho, W.T.; Judge, G. An information theoretic approach to network tomography. *Appl. Econ. Lett.* **2015**, *22*, 1–6
25. Cho, W.T.; Judge, G. Information theoretic solutions for correlated bivariate processes. *Econ. Lett.* **2007**, *97*, 201–207.
26. Hartonen, T.; Annala, A. Natural networks as thermodynamic systems. *Complexity* **2012**, *18*, 53–62.

© 2015 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).