IMF Working Paper

Statistical Inference as a Bargaining Game

Eduardo Ley

INTERNATIONAL MONETARY FUND

IMF Working Paper

IMF Institute

Statistical Inference as a Bargaining Game

Prepared by Eduardo Ley[1]

Authorized for distribution by Reza Vaez-Zadeh

April 2002

Abstract

The views expressed in this Working Paper are those of the author(s) and do not necessarily represent those of the IMF or IMF policy. Working Papers describe research in progress by the author(s) and are published to elicit comments and to further debate.

This paper extends the analogy, previously established by Leamer (1978a), between a Bayesian inference problem and an economics allocation problem to show that posterior modes can be interpreted as optimal outcomes of a bargaining game. This bargaining game, over a parameter value, is played between two players: the researcher (with preferences represented by the prior) and the data (with preferences represented by the likelihood).

Contents   Page

I. INTRODUCTION

The formal parallelism between game theory and statistical decision theory has been fruitfully exploited since Wald (1950) realized that the analysis of two-person zero-sum games provided a useful framework for the theory of minimax statistical decision (Savage, 1972). This paper presents a parallelism between the axiomatic solution for a nonzero-sum bargaining game and the posterior mode. Extending the isomorphism between a standard Bayesian inference problem and an economics allocation problem previously established by Edward Leamer, we show that the posterior mode can be interpreted as the Nash bargaining solution of a bargaining game between the researcher and the data.

Leamer (1978a,b) interprets the level curves of the prior and likelihood as indifference curves: the prior isodensity curves representing the preferences of the researcher and the isolikelihood curves representing the preferences of the data. In this paper we argue that the analogy must be established with a very particular economic situation involving public goods and no resource constraint. We also push the parallelism further and show that the posterior mode can be interpreted as the optimal outcome of a bargaining game between researcher and data over the value of a parameter.

The paper is organized as follows. Section II presents the analogy between the statistical inference problem and a public-goods allocation problem. Section III introduces Leamer's result establishing the posterior mode as the outcome of maximizing a social information function. Section IV frames the inference problem as a bargaining game, shows that the posterior mode is the Nash bargaining solution, and presents some examples. Section V concludes.

II. THE ANALOGY

Suppose that we want to make inferences about a parameter, θ, and assume that $f(\theta)$ summarizes the researcher's prior beliefs. The isodensity curves defined by $f(\theta) = k$, where k is a constant, can be interpreted in an analogous way to indifference curves in economics —all the θ's in the same isodensity curve are equally favored by the researcher. Moreover, if $f(\hat\theta) > f(\tilde\theta)$ then the researcher prefers $\hat\theta$ to $\tilde\theta$. Thus, $f(\theta)$ can be interpreted as a *utility function* representing the researcher's preferences about θ. Similarly, the likelihood function, $g(\theta)$, summarizes the data preferences. Figure 1 presents some contour sets of an hypothetical prior (dashed curves), centered at A, and a likelihood function, centered at B. Inner indifference curves are associated with more preferred θ's.[2]

It should be noted that the analogy between parameter inference and an economics allocation problem must be generally made with a very particular situation involving three elements: *(i)* pure public goods, *(ii)* preferences with satiation points, and *(iii)* no resource constraint.

Pure Public Goods. Take a two-agent two-good economy, and denote the total amount of resources by (X, Y). If both goods are a private goods, each individual ($i = 1, 2$) has preferences over her

---

[2] The plot represents two Normal densities. The prior is centered at $(1, 4)$ and has $\sigma_{11} = 1$, $\sigma_{12} = 0.25$, and $\sigma_{22} = 2$. The likelihood is centered at $(2, 3)$ and has $\sigma_{11} = 1$, $\sigma_{12} = -0.125$, and $\sigma_{22} = 1.2$.

Figure 1. Prior and Likelihood Level Curves.
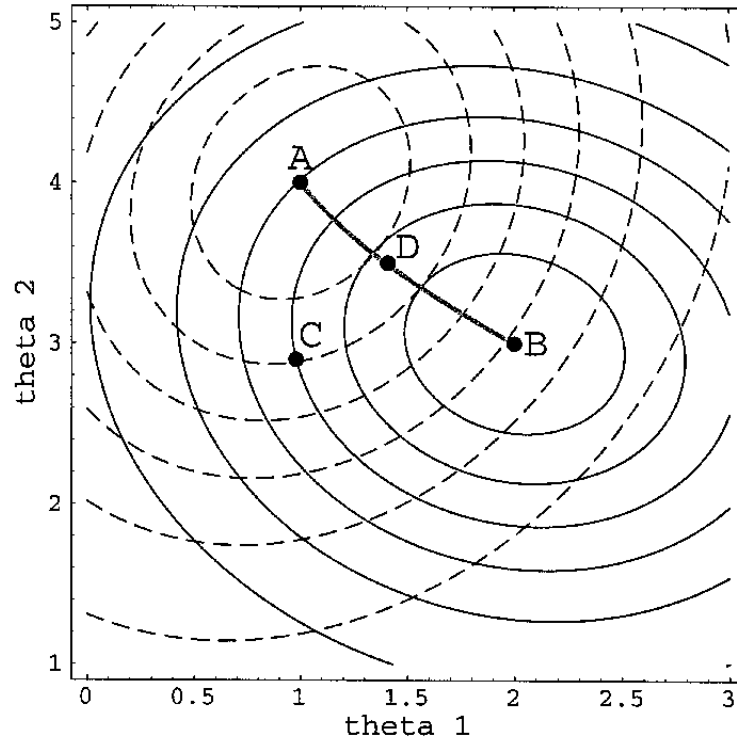
own consumption bundle (x_i, y_i), not on total endowments (X, Y). However, in the parameter-inference problem, matters are different. Individuals have preferences generally defined over all the coordinates of θ. Thus, each coordinate, θ_i, is not divided up among the two agents like in the case of a private good where an agent's consumption rivals other agent's consumption. Indifference maps will then typically look like the ones depicted in Figure 1, often with satiation points.

Lack of a resource constraint. While resource constraints are central to most economic problems, allocation problems involving public goods pose interesting problems where scarcity does not need to play a role at all. Conflict may arise when choosing the level of the goods simply because everyone must consume the same amount. Imagine two roommates, Ana and Bruno, who must choose two public goods, the room temperature (θ_1), and the radio volume (θ_2). Assume further that the landlord pays the energy bills, so cost is not a consideration.[3] The contours in Figure 1 can now be used to represent their preferences in the (θ_1, θ_2) plane. Ana's most preferred allocation is $A = (1, 4)$, elsewhere she is willing to trade along the dashed indifference curves. Bruno's most preferred allocation is given by $B = (2, 3)$, and his preference sets are represented by the solid contours.

The solid gray graph joining both agents' satiation points, A and B, along the tangencies between their indifference curves is the contract curve.[4] Any allocation, C, outside the contract curve

---

[3] This example is a slightly modified version of one in Bergstrom (2001).

[4] There are other tangencies between indifference curves that lie outside the contract curve. Suppose we want fix Ana's utility constant and minimize Bruno's —*i.e.*, we want to optimally hurt him while preserving Ana's utility level. The answer

- 5 -

(Figure 1) is always dominated by some allocation on the contract curve, D, in the sense that at least one agent prefers D to C while the other does not prefer C to D; thus, $f(D) \geq f(C)$, and $g(D) \geq g(C)$, with at least one strict inequality —where $f(\cdot)$ and $g(\cdot)$ now represent the utility functions of Ana and Bruno. As a result, rational agents will exploit mutually advantageous trades and move to allocations situated on the contract curve.

Similarly, the tangencies of the level curves of prior and likelihood define the *information contract curve*.[5] Given a prior and likelihood, we can restrict our attention to outcomes lying in the information contract curve. Any value of θ outside the information contract curve is dominated by some value in the curve which is preferred by both the data and the researcher. The issue remains, however, on how to choose a single θ from the information contract curve.

III. SOCIAL INFORMATION FUNCTION

In economics, maximizing a social welfare function (SWF) provides a centralized solution to picking an allocation, θ. A SWF, $W(\theta)$, is a function that aggregates individual preferences into social preferences, $W(\theta) = \psi(f(\theta), g(\theta))$ (Bergson, 1938; Samuelson, 1951). Similarly, in the inference problem, Bayesian analysis combines prior and data information through the posterior distribution, $h(\theta) \propto f(\theta) g(\theta)$. In this context, the posterior mode, $\hat{\theta}$, can be seen as maximizing a very particular SWF, given by $h(\theta)$, which is proportional to the product of the data and researcher's utility functions:

$$\hat{\theta} \equiv \arg\max_{\theta}\{h(\theta)\} = \arg\max_{\theta}\{f(\theta)\, g(\theta)\}. \tag{1}$$

The SWF given by the posterior, $W(\theta) = h(\theta)$, is called by Leamer (1978b, p. 149) a *social information function* (SIF).

A SWF (or SIF) must not only be consistent with individual valuations, but further assumes that gains and losses for different individuals can be compared —*i.e.*, requires cardinality.[6] In an ordinal preference representation, the utility level assigned to each level curve (the labeling) in Figure 1 is arbitrary, provided that the order of the preferences is preserved. Cardinality, on the other hand, imposes restrictions on the labeling, since now differences in utility levels are meaningful. We shall return to this issue below in Section IV.A.

The choice of a SWF determined by the product of individual utilities is used, among others, by Fair (1971) and Riley (1973). Ray Fair (1971; p. 566) states that this particular SWF "seems to be consistent with commonly held ethical views." John Riley (1973; p. 472) makes such a choice for a SWF "guided by the ethical principle that everyone should have an equal opportunity, and constrained by the mathematical necessity of mathematical simplicity."

---

to this problem is given by the North-West tangencies to Ana's indifference curves.

[5] This curve is called by Dickey (1975) *curve décolletage*; see also Leamer (1978b; page 82). Thurman *et al.* (2001) use the information contract curve to impose regularity on estimated parameter values; see their footnote 1 for references to other works that use this curve.

[6] See Sen (1970) welfare analysis under different degrees of comparability.

IV. THE NASH BARGAINING PROBLEM

The Nash bargaining solution (Nash, 1951) provides yet a better justification for using the product of utilities as the SWF, without recourse to ethical considerations.[7] The Nash bargaining solution is obtained by maximizing the product of cardinal utilities. As Ken Binmore (1994; p. 83) puts it: "When the Nash bargaining solution is used, it is to predict what the result would be, under certain ideal circumstances, if specimens of *homo economicus* were to bargain optimally."

In a Nash bargaining game over the value of θ, researcher and data would, independently and simultaneously, choose utility demands $\hat{f}$ and $\hat{g}$, respectively. If the utility demands are mutually compatible —*i.e.*, if there exists a $\hat{\theta}$, such that $(\hat{f}, \hat{g}) = (f(\hat{\theta}), g(\hat{\theta}))$— then agreement is reached at $\hat{\theta}$ and each player receives his utility demand. If the utility demands are not mutually compatible, the players stay at the *status quo*, or disagreement point. Note that in the inference version of the game presented in this paper, since there is no resource constraint, all θ's are feasible. Nonetheless, mutual compatibility of utility demands is a more restrictive condition which may not obtain. This game has many equilibria (all laying along the contract curve).

Nash (1951,1953) postulates a series of conditions that a plausible solution to this game must satisfy to guarantee its optimality and invariance with respect to certain mathematical transformations. The four axioms are: (i) efficiency, (ii) symmetry, (iii) linear invariance and (iv) independence from irrelevant alternatives.[8] The result is the unique Nash bargaining solution that can be identified by maximizing, with respect to θ, the SWF given by:

$$W(\theta) = (f(\theta) - f(\xi))(g(\theta) - g(\xi)) \tag{2}$$

subject to $f(\theta) \geq f(\xi)$ and $g(\theta) \geq g(\xi)$, where ξ represents the disagreement point. The utility at the disagreement point is the guaranteed utility level that the participants will get in case no agreement is reached. As noted by Crawford (2002): "These axioms generalize the widely accepted principle of sharing the gains from agreement equally to bargaining problems with nonlinear utilities and sets of feasible outcomes, in which the meaning of equal-sharing is not readily apparent."

In the inference problem, the disagreement point has no obvious interpretation. Moreover, an agreement will always be reached, in the sense that prior and data information will always be combined when conducting posterior inference. Under these circumstances, it is natural to place the disagreement point, ξ, in the zero probability region, so that $f(\xi) = g(\xi) = 0$, and completely eliminate ξ from the determination of the Nash bargaining solution. The SIF becomes:

$$W(\theta) = f(\theta)g(\theta) \propto h(\theta), \tag{3}$$

which is maximized at the posterior mode, $\hat{\theta}$. In this sense, *the posterior mode is the optimal agreement point in a bargaining game between data and researcher.*

---

[7] A bargaining problem arises when the status quo is located away from the contract curve, and while the agents may agree to move to an allocation on the contract curve, more than one such allocation is available to them. Nash (1951) used cooperative game theory to study this problem, and obtained the unique NBS characterized by his four axioms. Nash (1953) turned the situation into a noncooperative game, and used the axioms to select the NBS among the many equilibria of the game.

[8] See, *e.g.*, Harsanyi, 1977; p. 144. On the history of the paper by John Nash, see Chapter 9 in Nasar (1998).

How do the axioms fit the statistical problem? Efficiency, symmetry, and independence from irrelevant alternatives should be easy to justify. Linear invariance means that utility functions are unique up to linear transformations —*i.e.*, we are allowed to choose origin and scale. However, likelihoods and priors must integrate to unity, so we cannot generally apply linear transformations and still obtain probability distribution functions. Nonetheless, this axiom is required in the bargaining game to guarantee its generality, *i.e.*, to obtain a solution that does not depend on a particular utility representation. If the utility representation is restricted within the class of von Neumann-Morgestern utilities, the Nash bargaining solution still applies. Consequently, the axioms postulated by Nash can reasonably be applied to the inference problem.

In statistical inference, the posterior mode is an optimal point estimate under a limiting case of a zero-one loss function:

$$L(d, \theta) = \begin{cases} 0 & \text{if } |d - \theta| \leq \varepsilon \\ 1 & \text{if } |d - \theta| > \varepsilon \end{cases} \tag{4}$$

where ε is a constant. When ε is made arbitrarily small, the posterior modal value is the optimal estimate (O'Hagan, 1994). Note that the tails of the posterior distribution are completely ignored in this case.

Zeuthen (1930) developed a solution to the bargaining game which is similar to Nash's. Let agent 1 (the researcher, with preferences $f(\cdot)$), offer θ^f, and agent 2 (the data, with preferences $g(\cdot)$) counteroffer θ^g. Zeuthen then assumes that each agent will make a further concession whenever the opponent's offer is 'closer' to his own than his offer is to his opponent's. Closeness is measured in each agent's utility metric. More precisely, agent 1 will make a further concession if:

$$\frac{f(\theta^f) - f(\theta^g)}{f(\theta^g)} < \frac{g(\theta^g) - g(\theta^f)}{g(\theta^f)}. \tag{5}$$

A similar condition applies to agent 2. Since equation (5) can be rewritten as $f(\theta^f)g(\theta^f) < f(\theta^g)g(\theta^g)$, the dynamics proposed by Zeuthen imply that the party whose offer implies a lower $f \times g$ than his opponent's will always make a further concession. (If the product is identical, both parties make concessions until they both offer the same θ.) This process continues until the product $f \times g$ is maximized at the posterior mode as in the Nash bargaining solution (Harsanyi, 1956).

The Nash bargaining solution is generally presented in utility space. Given a prior and likelihood (and, of course, the data), we can evaluate the pairs $(f(\theta), g(\theta))$ for all possible θ's. The frontier of this set is called the utilities possibility frontier. Figure 2 represents this set for a Gamma prior and Exponential sampling (left and right panels differ only in the assumed number of observations involved in the likelihood).[9] Note that now, in contrast with Figure 1, the axes represent utilities — *i.e.*, values of the probability density functions. Efficient (undominated) allocations must be located in the North-East negatively-sloped portion of the frontier. Along it, higher utility for the researcher implies lower utility for the data and vice versa. Elsewhere there is no need to make a tradeoff

---

[9] Example 1 assumes that we sample from an Exponential distribution with unknown mean $1/\lambda$. The prior for λ is assumed a Gamma with parameters $\alpha = 2$ and $\beta = 3$. The sample mean is taken to be 0.75, and the number of observations, 1 in the left panel and 10 in the right panel. See, *e.g.*, DeGroot (1971).

Figure 2. Effects of increasing sample information: the Prior is downweighted.
(Gamma-Exponential Example: $n = 1$ and $n = 10$.)
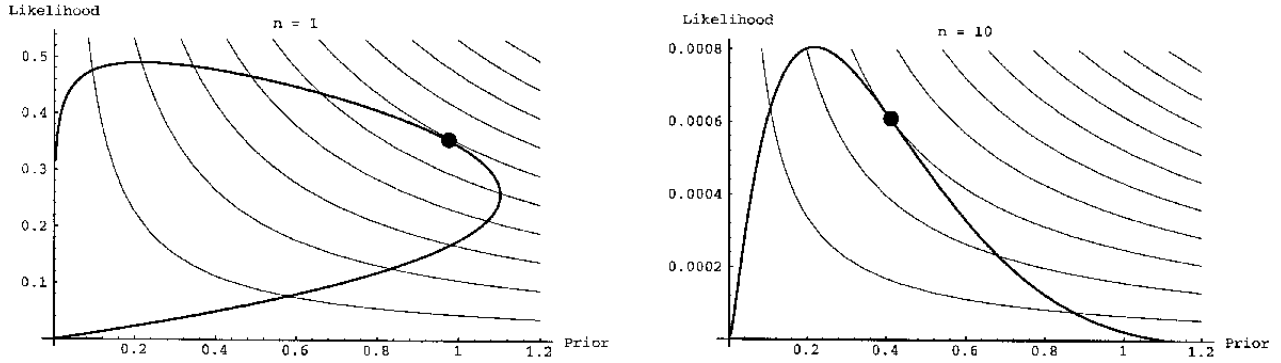
among the two. This negatively-sloped portion of the utilities possibility frontier corresponds to allocations lying in the contract curve.

The family of hyperbolas represent the social indifference curves defined by $W(\theta) = f(\theta)g(\theta) = k$, where k is a constant. The larger k, the higher the social information (welfare) and the more distant from the origin the indifference curve will be. The Nash bargaining solution is determined by the tangency of a social indifference curve with the utilities possibilities frontier, which gives the maximum attainable social indifference curve. This point is, of course, associated with the largest posterior mode.

The effect of increasing the sample information —and consequently, increasingly weighting the likelihood—, is illustrated in Figure 2 by changing the number of observations from $n = 1$ to $n = 10$ while keeping fixed the rest of the parameters of the example. The tangency moves away from a point with an associated high value of the utility of the data, around 1, to the left where the associated value of the prior is just over 0.4. (Note that the scale of the horizontal axis is the same in both graphs, since the prior is the same.)

A. Sensitivity Analysis

As discussed, when the researcher feels confident with the cardinality of his preference representation, the posterior mode is the point chosen by maximizing the SIF represented by the posterior distribution. However, Leamer (1978b) notes that while a researcher may feel comfortable expressing his beliefs through indifference curves such as the ellipsoids represented in Figure 1, the precise labeling of each indifference curve may often be more problematic. In those circumstances, when only an ordinal ranking can be sensibly justified, the whole contract curve becomes relevant. Any outcome in the information contract curve could be achieved as a posterior mode through a suitable labeling function of the prior.

Figure 3 illustrates the effect of trying different priors. The example involves sampling from a Normal distribution with unknown mean, θ and given precision, 1. The sample mean, with $n = 2$, is taken to be 5, and the standard deviation 3. The prior on θ is assumed Normal with mean $\mu = 4$ and precision τ. In the left panel, we have $\tau = 1$ while in the right panel we increase our prior

uncertainty about θ by making the precision 0.25. The posterior mode is now determined almost entirely by the data information.

Figure 3. Effects of reducing prior precision: importance of Likelihood increases.
(Normal-Normal Example: $\tau = 1$ and $\tau = 0.25$.)
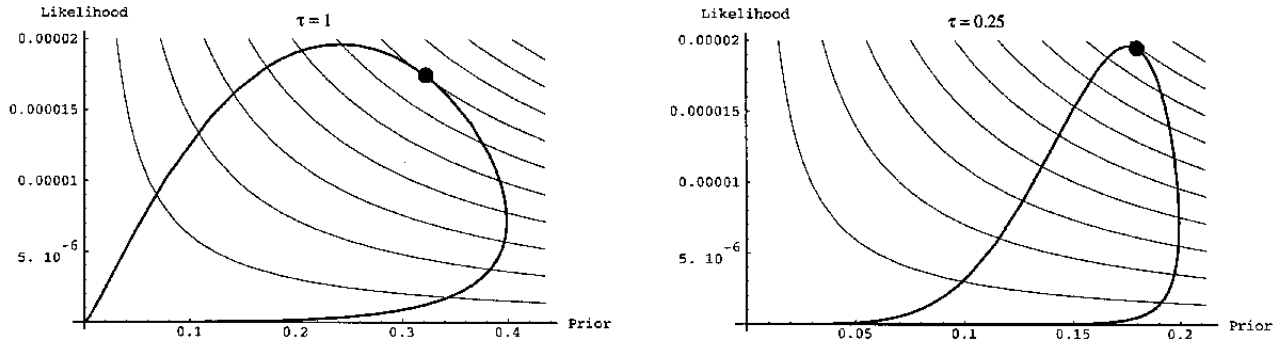
It is worth noting that in both examples while the relabeling of the likelihood or prior causes a deformation in the utilities possibilities frontier, the information contract curve does not change.

B. Nonconvexity

The utilities possibility sets represented here (Figures 2 and 3 below) are all well behaved in the sense that its North-East frontier is convex. However, this need not be the case in inference problems. What makes inference problems different than typical economic problems is that individual preferences may naturally be non-convex since here it need not be the case that 'more is always better.' The possibility of satiation, and even multiple local bliss points (multimodal priors) naturally leads to nonconvex utilities possibilities sets. When the inference problem leads to a nonconvex bargaining game, an extension of the Nash solution leading to the maximization of the Nash product —*i.e.*, corresponding to the posterior mode— can then be chosen and the approach presented here would generally apply.[10]

C. Weighted Nash Bargaining Solution

When the agents involved in a bargaining game may have strategic advantages, an asymmetric solution arises maximizing the geometric weighted average:

$$W(\theta) = [f(\theta)]^\alpha \, [g(\theta)]^\beta \qquad (6)$$

where the non-negative weights α and β reflect the agents' bargaining powers. Zellner (2002) presents various generalized Bayes rules in which a posterior like (6), proportional to a geometric

---

[10] The extension of the Nash bargaining solution to nonconvex problems proposed by Kaneko (1980) does lead to the maximization of the Nash product. Other extensions do not lead to the maximization of the Nash product —*e.g.*, Herrero (1989), and Conley and Wilkie (1996).

weighted mean of prior and likelihood, arises when it may be appropriate to weight prior and sample information differently —for instance, recognizing that their quality may differ.

V. CONCLUSION

This paper presents a parallelism between the axiomatic solution for a bargaining game (Nash, 1951), and the posterior mode. Extending the isomorphism previously established by Leamer (1978a), the standard Bayesian inference problem is framed as a public-goods allocation problem involving the researcher (with preferences represented by a prior distribution) and the data (with preferences represented by the likelihood function). In this context, posterior modes can be interpreted as optimal outcomes (Nash bargaining solutions) of a bargaining game, over the value of a parameter, between researcher (characterized by the prior) and the data (characterized by the likelihood).

REFERENCES

Bergson, Abraham, 1938, "A Reformulation of Certain Aspects of Welfare Economics," *Quarterly Journal of Economics*, Vol. 52, pp. 310–14.

Bergstrom, Theodore C., 2001, "Public Finance," unpublished Graduate Lecture Notes, UC-Santa Barbara. http://www.econ.ucsb.edu/~tedb/econ230b.html.

Binmore, Kenneth, 1994, *Game Theory and The Social Contract I: Playing Fair* (Cambridge Massachusetts: MIT Press).

Conley, J., and S. Wilkie, 1996, "An Extension of the Nash Bargaining Solution to Nonconvex Problems," *Games and Economic Behavior*, Vol. 13, pp. 26–38.

Crawford, Vincent P., 2002, "John Nash and the Analysis of Strategic Behavior," *Economics Letters*, Vol. 75, No. 3, pp. 377–82.

DeGroot, Morris H., 1970, *Optimal Statistical Decisions* (New York: McGraw-Hill).

Dickey, James M., 1975, "Bayesian Alternatives to the F-test and Least-Squares Estimates in the Normal Model," in *Bayesian Studies in Econometrics and Statistics*, edited by S.E. Fienberg and Arnold Zellner (Amsterdam: North-Holland).

Fair, Ray C., 1971, "The Optimal Distribution of Income," *Quarterly Journal of Economics*, Vol. 85, pp. 551–65.

Harsanyi, John C., 1956, "Approaches to the Bargaining Problem Before and After the Theory of Games: A Critical Discussion of Zeuthen's, Hicks', and Nash's Theories," *Econometrica*, Vol. 24, No. 2, pp. 144–57.

Herrero, M.J., 1989, "The Nash Program: Nonconvex Bargaining Problems," *Journal of Economic Theory*, Vol. 49, pp. 266-77.

Kaneko, M., 1980, "An Extension of the Nash Bargaining Problem and the Nash Social Welfare Function," *Theory and Decision*, Vol. 12, pp. 135–48.

Leamer, Edward E., 1978a, "Regression Selection Strategies and Revealed Priors," *Journal of the American Statistical Association*, Vol. 73, 580–87.

————, 1978b, *Specification Searches: Ad Hoc Inference with Nonexperimental Data* (New York: John Wiley).

Nash, John F., 1951, "The Bargaining Problem," *Econometrica*, Vol. 18, pp. 155–62.

————, 1953, "Two-Person Cooperative Games," *Econometrica*, Vol. 21, pp. 128–40.

Nasar, Sylvia, 1998, *A Beautiful Mind: A Biography of John Forbes Nash, Jr.* (New York: Touch-stone).

O'Hagan, Anthony, 1994, *Kendall's Advanced Theory of Statistics, Vol 2B, Bayesian Inference* (New York: John Wiley).

Riley, John G., 1973, "Gammavilie: An Optimal Town," *Journal of Economic Theory*, Vol. 6, pp. 471–82.

Samuelson, Paul A., 1951, "Social Indifference Curves," *Quarterly Journal of Economics*, Vol. 70, pp. 1–22.

Savage, Leonard J., 1972, *The Foundations of Statistics* (New York: Dover, 2nd ed.).

Sen, Amaryta K., 1970, "Interpersonal Aggregation and Partial Comparability," *Econometrica*, Vol. 38, pp. 393–403.

Thurman, Walter N., Tyler J. Fox and Tayler H. Bingham, 2001, "Imposing Smoothness Priors in Applied Welfare Economics: An Application of the Information Contract Curve to Environmental Regulatory Analysis," *Review of Economics and Statistics*, Vol. 83, No. 3, pp. 511–22.

Wald, Abraham, 1950, *Statistical Decision Functions* (New York: John Wiley).

Zellner, Arnold, 2001, "Information Processing and Bayesian Analysis," *Journal of Econometrics*, Vol. 107, pp. 41–50.

Zeuthen, F., 1930, *Problems of Monopoly and Economic Warfare* (London: Routledge and Kegan Paul).