Journal of Economic Perspectives—Volume 2, Number 1—Winter 1988—Pages 145–151

Comment on Poirier: The Subjective Perspective of a "Spiritual Bayesian"

John Rust

Professor Dale Poirier has preached a rousing sermon for the Bayesian cause, and in doing so has concisely exposed some of the cracks in the foundations of statistics. One of these cracks is the very interpretation of the concept of probability, a philosophical question that has never been fully resolved and has led to the schism that currently exists in statistics between the Bayesian and classical schools. The paper is valuable for econometricians because it challenges them to think more deeply about the meaning of the statistical procedures they employ. It certainly did that for me.

However a basic question was not addressed in the paper; namely, what is the purpose of statistical inference? I think the answer to this question has at least as much bearing upon whether someone adopts Bayesian or classical methodology as whether one interprets the meaning of probability in subjectivist or frequentist terms. Certainly beliefs about the proper "purpose" of statistical inference diverge every bit as much as do beliefs about the meaning of probability. Many people, especially statisticians, think the purpose is "prediction," a view Poirier sees as favoring the Bayesian school. Others, such as Ed Prescott, think the purpose is "measurement." Under this view, the parameters to be inferred or measured are real physical quantities and the only source of randomness is measurement error. Since physical quantities have an existence independent of our mental representations of them, this view leads one to avoid the subjective Bayesian approach in favor of "objective" classical methods. Still others

■ *John Rust is Associate Professor of Economics, University of Wisconsin, Madison, Wisconsin.*

(myself included), think that the purpose is to "learn." This view treats statistical inference as a process of formulating a sequence of *models* that are abstract but in some sense increasingly accurate representations of certain aspects of reality that we seek to understand. Although difficult to define precisely, learning typically involves three phases: observation and summarization of data, interpretation of the evidence, and revision of previously held beliefs. Bayesian decision theory is perhaps the most well-known formal model of learning that incorporates all three of these phases. Thus, to the extent that we use statistical inference to learn from data, it seems that rational scientists "ought" to be Bayesians.

Limitations of Bayesian Estimation Theory

As Poirier has observed, many economists are somewhat schizophrenic in their application of Bayesian decision theory. The schizophrenia is aptly summarized in a joke Thomas Sargent once told about one of his colleagues: "He models all his agents as knowing Bayesian decision theory even though he doesn't." Actually, the colleague in question is an expert in Bayesian decision theory; he simply chooses not to use the theory to govern the way he does empirical work. Many other applied economists behave this way, often unwittingly. We are perfectly willing (probably even compelled), to model all our agents as hyper-rational Bayesian decision makers, yet very few of us actually use the Bayesian theory when we do our empirical work. How can we explain the apparent hypocrisy?[1] As a general response to the points in Poirier's paper, let me list five reasons why I think many applied economists fail to use formal Bayesian methodology, even though they seem to understand the theory and use it to build their economic models.

First, as many Bayesians such as Poirier and Leamer admit, it is very difficult to quantify our subjective beliefs formally in terms of a prior probability distribution over an abstract parameter space. Although I think it is clear that all economists have a more or less definite system of initial beliefs before they analyze a given data set, the internal, intuitive representation of these beliefs does not appear to correspond well to a subjective probability distribution over an unknown "parameter vector." It may be true that we can be viewed as acting "as if" we had such a prior distribution; however, the process of attempting to elicit these beliefs in terms of a formal prior has proved to be a very difficult and often very arbitrary exercise.

[1]The hypocrisy highlights the distinction between the "normative" vs. "prescriptive" roles of Bayesian decision theory. The hypocritical behavior of econometricians shows that Bayesian decision theory is not a good normative "meta-theory" of how econometricians actually learn, nor is it a good prescriptive theory of how we "ought" to learn. I believe that our real learning process is a far more complex and creative form of behavior than the simple Bayesian theory admits. Economists who model their agents as Bayesian learners do so primarily because of analytical convenience and convention rather than out of a belief that agents "really" learn that way. Thus, our hypocrisy may in fact be a reflection of a sensible learning strategy known as *Occam's Razor*: start with the simplest possible model capable of explaining the data. At the level of abstraction of current models, building more detailed models of learning by economic agents is probably not called for.

Second, even if the applied economist succeeded in eliciting a personal prior and has formally conditioned on the data to derive a posterior, it is not clear why a reader of a journal article should be interested in the author's subjective prior and posterior. Poirier concedes this point, and stresses Leamer's slogan "the mapping is the message." Under this view, an empirical paper should present the relevant information to allow the reader to perform the mapping from prior to posterior, or at least conduct sensitivity analyses showing the impact of changes in the author's subjective prior on the calculated posterior. Unfortunately, there seems to be no general, convenient method for communicating this mapping or conducting general sensitivity analyses in a published journal article. If we follow the Bayesian logic to its limit, empirical papers would consist of a likelihood function, a dataset, and an exhortation to readers to find their own posteriors. At this latter extreme, the information content of published empirical work starts to evaporate; the reader is essentially forced to redo most of the empirical work himself. This defeats one of the basic functions of learning: to summarize complex data.

Third, the statistical models that are of interest to applied economists are rarely the simple textbook linear or Bernoulli models that admit convenient conjugate priors. Instead, interest has focused on increasingly complex dynamic, structural choice and equilibrium models whose functional forms are highly nonlinear in parameters, and for which even a single likelihood function evaluation can be relatively expensive to compute. For these types of models the computational cost of a formal Bayesian analysis (which would be forced to perform direct numerical or Monte Carlo integration over the entire parameter space in order to map the prior into a posterior) appears to be prohibitive. Die-hard Bayesians who insist on limiting the scope of their economic models to the simple textbook functional forms and specific priors which admit convenient posterior calculations have become increasingly susceptible to the criticism that their models are contrived, that their inferences are fragile, and that their models have limited economic content.

Fourth, Bayesian methodology is heavily dependent on the use of parametric functional forms for the likelihood function. However, if you ask most applied economists, they will admit that they are not at all confident about their choice of a particular functional form, especially when it comes to a parametric specification of unobservable variables that are typically included to obtain a nondegenerate statistical model.[2] Over the past ten years there has been substantial progress in the development of flexible, semi-parametric, and nonparametric statistical methods

[2]A statistical model is *degenerate* if a subset of variables is an exact function of remaining variables in the model. In practice, however, we will never be able to predict economic variables perfectly so we introduce *unobservables* ε_t to account for the discrepancies. For example, in the case of dynamic choice models the data $\{d_t, x_t\}$ consists of observations on agents' decisions d_t and states x_t. Blackwell's Theorem implies that under general conditions the solution to the dynamic programming problem takes the form of a *decision rule* $d_t = f(x_t, t)$, for some nonstochastic function f. Thus, if the econometrician observes all of x_t, the resulting choice model is degenerate: knowledge of f would allow us to perfectly predict agents' choices. However if we include a vector ε_t of state variables observed only by the agent, we obtain a nondegenerate model $d_t = f(x_t, \varepsilon_t, t)$. Integrating over the unobserved variables ε_t produces a nondegenerate conditional choice probability $P(d_t|x_t)$ that can be used as a basis for maximum likelihood estimation of the decision rule f.

which in effect allow the researcher to specify infinite-dimensional parameter spaces, avoiding arbitrary choices of parametric functional form. By and large, most of this work has been done in the classical tradition. In principle, one would like to do nonparametric Bayesian analysis, with a prior over an infinite-dimensional parameter space. Poirier concedes this point in his "pragmatic principle of model building 5": never assign prior probability 1 to a particular parametric likelihood function through which you view the data. Unfortunately, there are some very serious problems confronting the development of a formal, nonparametric Bayesian decision theory. To date, researchers have encountered difficulties specifying interesting, non-degenerate priors over infinite-dimensional spaces. Even in cases where well-behaved parameter spaces and non-degenerate priors have actually been formulated, there is an additional problem that the nonparametric Bayesian approach may be *inconsistent*: the posterior distribution need not converge to a unit mass on the "true" parameter vector as the sample size tends to infinity. This is a serious problem, since it implies that the nonparametric Bayesian approach can lead us to the wrong answer; that is, it can lead to incorrect learning even given infinite amounts of time and data. It is well-known that in the case of finite-dimensional parameter spaces Bayes rule is consistent. As early as 1963, however, the Bayesian statistician Freedman produced examples of inconsistency of Bayes rule for infinite-dimensional parameter spaces.[3] In a recent article, Diaconis and Freedman (1986) present an even more disturbing example of an inconsistency in the context of a simple nonparametric model of location. Without going into detail about their results, let me quote from their conclusion (p. 14):

> Often a statistician has prior information about a problem, but does not really have a sharply defined prior probability distribution. Many different distributions would have the right qualitative features, and a Bayesian typically chooses one on the basis of mathematical convenience. In smooth, low dimensional problems, this ought to help, and anyway cannot lead to disaster, because the data will swamp the details of the prior. Unfortunately in high-dimensional problems, arbitrary details of the prior can really matter; indeed, the prior can swamp the data no matter how much data you have. That is what our examples suggest, and that is why we advise against the mechanical use of Bayesian non-parametric techniques.

My fifth and final point is that although Bayesian decision theory is one of the best known formal models of learning, it is not the only model of learning, and

[3]Freedman's (1963) example used an infinite-dimensional parameter space of the set of all probability distributions over the integers. Thus, a prior distribution in Freedman's set-up is a probability distribution μ over probability distributions on the integers. Freedman constructed an example where the true parameter generating the data is a geometric distribution with probability of success equal to 1/4. Freedman constructed a prior μ such that as the number of observations tended to infinity, the posterior converged weakly to a unit mass on a geometric distribution with probability 3/4. Freedman's prior μ was not a "quirk:" he showed that the set of prior probability distributions μ such that the posterior distributions are consistent is a set of the first category in the weak topology on probability distributions; that is, the set of consistent priors is topologically "negligible."

moreover, it is not clear that it is a literally correct model of how human beings actually learn. Any attempt to "mechanize" our learning process through rigid adherence to a formal mathematical model of learning may end up inhibiting the learning process, especially if that model is not a correct model of how we actually learn. My colleague Charles Manski has elsewhere referred to overly rigid adherence to Bayesian techniques as a "straightjacket" that can inhibit the creativity of researchers in their attempts to learn from data. He points out that the very completeness and internal logical consistency of the Bayesian approach can limit the creative freedom of applied researchers. Classical statistical theory, on the other hand, is incomplete, and in some instances even internally inconsistent (especially with regard to the theory of hypothesis testing). But in the hands of intelligent researchers, classical statistical methods can offer a wider and more flexible array of techniques for drawing inferences from data, especially in the case of model selection. The extra freedom and creativity permitted by the very limitations and incompleteness of classical statistical theory may aid the learning process more than it hinders it.

Limitations of Classical Estimation Theory

I do not want my criticisms of Bayesian methodology to be interpreted as acceptance or contentment with classical methodology. Quite the contrary, I feel there are major unsolved problems affecting all aspects of classical statistical theory. The most serious of these problems, I believe, are the issues of pretesting and model selection.[4] Classical distribution theory and hypothesis testing procedures do not allow one to condition on the results of previous estimations and hypothesis tests performed as part of a specification search for the "true" model. As a result, the classical distribution theory for estimators and test statistics is likely to be incorrect once we take this conditioning information into account. A closely related limitation of classical distribution theory is the failure to account for increases in model complexity as sample size increases. Standard asymptotic distribution theory assumes that the dimension of the parameter space is fixed, computing the approximate distribution of an estimator or test statistic as the sample size tends to infinity. However, researchers usually attempt to estimate the most complex, realistic model possible, using as many parameters as their sample size permits. In practice, this leads to a parameter space whose dimension is an increasing function of sample size. For example, in a study of estimated wages equations, Roger Koenker (1985) found that the number of parameters grew roughly proportional to the 1/4 power of the sample size. We need to generalize asymptotic distribution theory to account for such increases in model complexity as the sample size grows. Initial work in this direction has been undertaken by the statistician Stephen Portnoy (1985).

I believe intelligent researchers are aware of these problems and make internal, intuitive corrections to the standard distribution theory in the process of drawing

[4] It is telling that to date the most illuminating analysis of these problems is from the Bayesian perspective, in Edward Leamer's book *Specification Searches*.

conclusions from the data. Obviously it would be much better to attempt to generalize classical distribution theory to account explicitly for pretesting and specification searching. Unfortunately, this task appears to be extremely difficult. Until solutions are found, researchers are faced with a choice: either adopt the apparently complete and logically consistent Bayesian methodology, facing the attendant difficulties already described, or else abandon this approach in favor of an incomplete classical theory that relies heavily on asymptotic distribution theory and whose practical application in applied work involves some serious logical difficulties. The philosophical shortcomings of the latter is probably the biggest reason why I regard myself as a Bayesian, at least "in spirit." However I feel that at least at present, the computational limitations of the Bayesian approach are decisive. Since I am interested in learning from data, I use classical methods because they are computationally feasible for the models I am interested in estimating. However, in interpreting the results I try to make intuitive corrections to the approximate distributions of my estimators and test statistics to account for the information I have obtained through my specification searches. You might say that I act "as if" I were a Bayesian even though I employ classical statistical techniques. I can only hope that intelligent researchers using classical methods in this way will be able to learn correctly from the data.

Are there any guarantees that one can learn correctly using the incomplete and sometimes logically inconsistent classical theory? Recent work on formal models of learning behavior has begun to provide some answers. Bray and Savin (1986) studied the behavior of Bayesian and classical agents in a simple stochastic cobweb model. Each period agents update their estimates of a simple linear regression of the market clearing price on a vector of exogenous variables in order to determine the output quantities to supply to the market. Feedback from this learning behavior produces a linear time-varying coefficients model for the actual market price. The model does not reflect optimal learning behavior because agents operate under the continually falsified assumption that the law of motion for the market price is time invariant. However, Bray and Savin demonstrate that under a stability condition this simple learning process will converge to the unique time-invariant rational expectations equilibrium. Marcet and Sargent (1986) have extended this result to a significantly larger class of environments using a result of Ljung (1977) that relates convergence to the stability of a certain differential equation. This seems to indicate that even unthinking automatons can use classical least-squares methodology to learn the correct model despite the fact that they use misspecified models. If we allow agents to use more sophisticated classical statistical methods including hypothesis testing and time-varying parameter models, agents should be able to learn even faster, a conjecture that Bray and Savin are currently studying. While it is somewhat reassuring to know that relatively simple statistical procedures lead to correct learning in specific environments, at the present time we do not know the class of models for which this kind of stability obtains.

I think the biggest learning problem facing the applied researcher is the process of deciding whether or not a given model represents observed data well, and discovering a model that "best fits" the observed data subject to constraints of computation and identification. Viewed in this way, many practical estimation problems are essentially "specification searches" across a more or less well-defined family of alternative models. Although applied researchers might use flexible statistical techniques searching over a sequence of finitely-parameterized models, ultimately I think we are just using these parametric specifications as stepping stones for solving the real problem, which is best regarded as a nonparametric estimation over the infinite-demensional space of all models. If you accept this viewpoint, then I think even the Bayesian methodology has some serious problems of incompleteness and inconsistency. The incompleteness arises because the traditional Bayesian theory requires the researcher to specify a parametric likelihood function (that is, a model), but the theory does not tell us how to choose the likelihood function or how to modify it in light of the observed data. In practice, this question is precisely the one of interest. If a researcher attempts to make the parameter space explicitly infinite-dimensional and conduct a formal search over alternative models, one immediately encounters the problem of inconsistency of the Bayesian methodology that Freedman and others have discovered. Thus, neither the classical nor Bayesian approaches are adequate formal models of learning processes.[5]

References

Bray, M. M., and N. E. Savin, "Rational Expectations Equilibria, Learning, and Model Specification," *Econometrica*, 1986, *54*, 5, 1129–1160.

Diaconis, P., and D. Freedman, "On the Consistency of Bayes Estimates," *Annals of Statistics*, 1986, *14*, 1, 1–26.

Freedman, D., "On the Asymptotoic Behavior of Bayes Estimates in the Discrete Case I," *Annals of Mathematical Statistics*, 1963, *34*, 1386–1403.

Koenker, R., "Asymptotic Theory and Econometric Practice," Univ. of Illinois at Urbana-Champaign, manuscript, 1985.

Leamer, E., *Specification Searches*. New York: Wiley, 1976.

Ljung, L., "Analysis of Recursive Stochastic Algorithms," *IEEE Transactions on Automatic Control*, 1977, AC-22-4, 551–575.

Marcet, A., and T. J. Sargent, "Convergence of Least Squares Learning Mechanisms in Self Referential Linear Stochastic Models," Carnegie-Mellon Univ. manuscript, 1986.

Portnoy, S., "Asymptotic Behavior of M-estimators of p Regression Parameters When p^2/n is Large II: Normal Approximation," *Annals of Statistics*, 1985, *13*, 4, 1403–1417.

[5] My discussion suggests the need for a more accurate "meta-theory" of the econometric learning process. However in analogy to the celebrated "Gödel Incompleteness Theorem" construction of a complete meta-theory may not be logically possible due to inherent problems of self-reference. Gödel's arguments showed that any attempt to "mechanize" the deductive learning process is bound to fail: there will always exist true statements that cannot be formally deduced from any given set of axioms using a fixed set of rules of inference. Similarly, it may be the case that any attempt to formalize the inductive learning process is also bound to fail. If no antecedent limits can be placed on the inventiveness of mathematicians in devising new rules of proof, it seems even less likely that we could construct an econometric meta-theory that delimits the creativity of econometricians in learning from data.

This article has been cited by:

1. Kevin Lang. 2024. How Credible is the Credibility Revolution?. *Journal of Labor Economics* **18**. . [Crossref]
2. Percy K. Mistry, Michael D. Lee. Violence in the Second Intifada: A Demonstration of Bayesian Generative Cognitive Modeling 65-90. [Crossref]
3. Dale J. Poirier. 2012. What is sensible for your agents should be sensible for yourself. *Journal of Econometrics* **170**:1, 249-250. [Crossref]
4. Harald Uhlig. 1994. What Macroeconomists Should Know about Unit Roots: A Bayesian Perspective. *Econometric Theory* **10**:3-4, 645-671. [Crossref]
5. James H. Stock. 1991. Bayesian approaches to the 'unit root' problem: A comment. *Journal of Applied Econometrics* **6**:4, 403-411. [Crossref]
6. John Rust. 1988. Comments On. *Econometric Reviews* **7**:2, 155-160. [Crossref]