6

VAGUE THEORY AND MODEL UNCERTAINTY IN MACROSOCIOLOGY

*Bruce Western**

Vague macrosociological theory generates uncertainty about statistical models. As a result, statistical inferences in macrosociology—where data sets are small and collinear—may carry considerably more error than is indicated by the conventional estimation and testing methods. I describe a Bayesian approach that propagates uncertainty about statistical models through to the final results. This approach is illustrated in a model of cross-national welfare state development and a pooled cross-sectional analysis of strike activity. These examples show that classical inferences can be seriously misleading when vague theory weakly guides the model specification. These ideas are considered in relation to recent methodological debates in macrosociology.

Macrosociologists are highly uncertain about their statistical models. This uncertainty is rooted in the distance between macrosociological

I gratefully acknowledge Ed Amenta, David Draper, Larry Griffin, Adrian Raftery, Robin Tamarelli, and *Sociological Methodology* reviewers who commented on earlier drafts. I also received helpful comments from seminar participants at Columbia, New York University, and Princeton. An earlier version was presented at the meetings of the American Sociological Association, Los Angeles, August 1994. The pooled cross-sectional data analyzed in the strike example can be obtained from Statlib at http://lib.stat.cmu.edu/datasets/strikes. All variables except union density are available. The union density data were kindly supplied by Jelle Visser at the University of Amsterdam. E-mail: western@lydgate.princeton.edu.

*Princeton University

ideas and their representation as statistical models. The theories that drive quantitative macrosociology provide little guidance beyond suggesting large sets of candidate variables for regression analysis. Consequently, macrosociology is a highly inductive enterprise that engages the data in double duty: first to select a model, and then to estimate parameters. This type of uncertainty, which necessitates a data-driven model search, is not accommodated by conventional statistics. When the same data are used for model selection and parameter estimation, the usual *p*-values and standard errors can be dramatically biased in the direction of positive findings. This bias is particularly acute in the macrosociological setting where samples are small and collinear.

These over-optimistic conclusions result from unreported uncertainty about the model specification. Bayesian statistics provides a way of accounting for model uncertainty. In the Bayesian approach, results are averaged across models, weighted by the support each model receives from the data and prior belief (Leamer 1978). To illustrate this idea, I present analyses of welfare state development and strike activity in 18 industrialized democracies. The welfare state analysis exemplifies the familiar problem of not knowing what independent variables to include in a regression equation. Pooling results from different models in this application revises conclusions in a conservative direction. However, findings are not always diluted by allowing for model uncertainty. In the strike analysis, theory does not say whether to fit a normal or fat-tailed, robust, distribution to the data. The data favor the robust specification and prior uncertainity about the strike model is reduced by Bayesian inference.

By critically examining quantitative macrosociology, this paper adds to recent debates about the value of statistics for studying large-scale social processes. The critics argue that nonprobabilistic methods can remedy the shortcomings of quantitative applications (Ragin 1987; Abbott 1988; Griffin 1993). I add to the critique, arguing that most substantive conclusions in quantitative macrosociology are not justified by the reported statistical analyses. Unlike earlier discussions, I use statistical ideas to resolve a distinctive methodological problem of macrosociology.

This paper also contributes to a line of research which claims that Bayesian inference is useful for studying comparative-historical questions. Previous work investigated the utility of Bayesian subjective probability and prior information (Berk, Western, and Weiss 1995; Western and Jackman 1994). These papers assumed perfect knowledge about the model specification. The current task is to explore the consequences of relaxing that assumption.

1. VAGUE THEORY IN MACROSOCIOLOGY

Quantitative macrosociology typically proceeds by appealing to a small number of theoretical perspectives to explain a set of observations (Ragin 1987, ch. 4). Regression models provide a natural framework for this research. Three decisions arise in choosing a regression model: (1) the selection of variables, (2) the choice of functional forms, and (3) the stochastic specification.

Macrosociological theory plays its largest role in variable selection. I distinguish two situations: (1) where theory is vague about control variables (Leamer 1983), and (2) where theory is vague about variables of key interest. To foreshadow the first application below, say a researcher is interested in the impact of social democratic government and economic development on the welfare state (e.g., Korpi 1989; Wilensky 1975). Other variables like the age of the population, the unemployment rate, or the labor force participation rate might also be considered because these could influence the demand for social spending (Pampel and Williamson 1989), and they may covary with the variables of interest. Theory is often vague about these control variables in the sense that the researcher is unsure precisely which controls to include. For example, Nielsen's (1994) analysis of income inequality divides the independent variables into a "core model" that is included in every specification and various subsets of "additional variables." Similarly, Boswell and Dixon's (1993) study of political violence distinguishes variables representing "the main elements of the theory" and "corollary factors" on the one hand, and a variety of "ancillary terms to better control and specify the relationships" on the other. Theory is too vague in both cases to say which control variables should definitely be included to accurately estimate the effects of interest.

Theories can also be vague about these effects of interest. Theories are generally stated as stylized empirical associations that could be operationally described by a large number of variables. As a result, as Ragin (1987, p. 67) observes, "variable-centered" comparative sociology is dominated by issues of operation, measurement, and model specification. In the welfare state example above, the social democracy variable represents a theory that says welfare state development depends on the power resources of working-class representatives. The theory is vague about which power resources are important. Some researchers prefer social democratic government, others like union organization (Esping-Andersen and van Kersbergen 1992 review the literature). Vague theory generating model uncertainty in this case is not a multiple indicator problem. With multiple indicators, structural variables are known with certainty to be in the model, but they are measured with error. With model uncertainty, researchers are unsure about which variables to include. Although measurement error is commonly and perilously ignored, uncertainty about model specification has distinct implications. In the welfare state example, parties and unions are not indicators of an unobserved construct. The inclusion of one or both variables describes different empirical processes that are not known with certainty.

Additionally, theory is often silent about the functional and stochastic parts of a model. Assuming a normal distribution for the dependent variable involves an empirical claim about how data are distributed, but this assumption is determined by the ease of least squares estimation rather than any substantive understanding (see Freedman 1985, p. 349). By simply providing expectations about the direction of associations, macrosociological theory is often quiet about functional form. The relationship between social spending and working-class power resources is often written as linear, but this is adopted out of convenience. Departures from linearity, where specified, are often used for data-analytic reasons, to obtain symmetric residuals or to pull in outliers. (An important exception are theories that specify non-monotonic relationships; see for example Kuznets 1955).

In short, theory is vague in the sense of failing to generate models that are unambiguously preferred to alternative specifications. Instead of picking out points in the space of possible models, macrosociological theory delineates large and fuzzy regions.

2. WEAK DATA IN MACROSOCIOLOGY

How do researchers choose among the models that theory might suggest? Model choice is empirically guided by a battery of tools including transformations and recodes, goodness-of-fit statistics, stepwise regression, and specification tests. Commonly, significant coefficients with hypothesized signs are retained, while others are discarded or replaced by alternatives.

When data sets are small and collinear as in macrosociology (Western and Jackman 1994), regression estimates can be very unstable in such a data analysis. Partly this is due to the large variance of coefficients estimated with small samples. Instability in coefficients is also aggravated by collinearity. With collinearity, estimates for one variable are sensitive to the other variables in the model. The instability of parameter estimates in an empirical model search is well understood by the critics who skeptically claim that data sufficiently tortured will confess the result desired by the data analyst. Similar intuitions are conveyed in warnings against "capitalizing on chance" or pejoratives like "number crunching," "data mining," and "data dredging" (Leamer 1983).

As the intuitions suggest, vague theory and weak data are a potent combination. Freedman's (1983) Monte Carlo study is a striking illustration. Freedman conducted a regression analysis on a data set consisting of 100 observations and 50 randomly generated predictor variables, all independent of the outcome. The true value of each coefficient was thus known to be 0. The analysis began by screening out independent variables that were not significant at the .25 level. After screening, 15 independent variables remained, of which 8 were significant at the .25 level, and 6 at the .05 level. The R^2 was a substantively tolerable .36, significant according to an F-test at the .001 level. Thus in a regression of "noise on noise," strong evidence for structural relationships in the data was found with the help of a screening procedure. Analytical results showed that calculated t-statistics of 1.96 obtained in the second regression run had a true probability level of 40 percent, in contrast to their nominal level of 5 percent (Freedman 1983, p. 154). Other studies show that data-driven model searches yield inflated R^2s and F-statistics while the lengths of confidence intervals are downwardly biased (Freedman, Navidi, and Peters 1988; Hurvich and Tsai 1990).

These biases are severe precisely in the situation that macrosociologists encounter—when the number of candidate independent variables is large in relation to the number of cases (Freedman et al. 1988). Collinearity compounds the problem by contributing to the sensitivity of regression estimates. Thus empirically guided model searches resulting from vague theory may be ubiquitous in sociology, but the pitfalls for statistical inference are especially severe for macrosociologists because of the weak data they analyze.

These problems are not news. To guard against overfitting, macrosociologists sometimes use statistics like the adjusted R^2 that penalize large models. Less commonly, sensitivity analyses (Leamer 1983) are used to investigate estimates from a range of models. Each strategy addresses a distinct problem. Fit statistics with penalties for parameters provide a way of finding a "best" model that is not simply the largest. Although one model among several might be favored by an adjusted R^2, all the concerns about uncalibrated uncertainty about the specification remain. Classical inference for the "best" model will still be too optimistic. Sensitivity analysis is an informal way of incorporating model uncertainty. Consistent results across models indicate that the data support one substantive conclusion, regardless of model choice. But what if the results are inconsistent across models? Which estimates, if any, should be believed? The Bayesian approach to model uncertainty provides a method for resolving this question.

3. A BAYESIAN APPROACH TO MODEL UNCERTAINTY

The problem with conventional statistics is that only uncertainty of a very special type is permitted. Models are assumed to be true up to unknown parameters. This means that all features of the model are known with certainty, except the numbers to slot in for, say, the coefficients of a regression. This idea contrasts sharply with Ragin's (1987, p. 67) observation, cited above, that quantitative macrosociology is dominated by discussions of model specification. If Ragin is right, then conventional statistics—which admit parametric and not model uncertainty—are unsuitable for the job of inference in macrosociology where the effects of vague theory are likely to be large.

Overconfidence in statistical inference resulting from data mining can be understood in this context. Conventional statistical tests only have their probability interpretation conditional on a true model. When researchers are uncertain about whether their models are true, statistical inferences are too optimistic, failing to consider all the sources of uncertainty on which the inferences should be based.

Bayesian statistics provides a useful vocabulary for discussing statistical inference when theories do not lead to a uniquely preferred model. The Bayesian goal is to describe an unknown quantity, such as a regression coefficient, β, with a probability distribution called a posterior distribution. The posterior distribution results from an application of Bayes's rule that combines a probability distribution for the data (called a likelihood) with a probability distribution for nonsample information (called a prior distribution):

$$p(\beta|\text{data}) \propto p(\text{data}|\beta) \times p(\beta),$$

or, in words, the posterior distribution is proportional to the likelihood times the prior. The proportionality sign ($\propto$) here indicates that the posterior must be standardized to be a proper probability distribution. If sample data dominate the prior information, the Bayesian results for regression coefficients will often be close or identical to the results obtained with maximum likelihood estimation. In a linear regression with normal errors and known error variance, ordinary least squares (OLS) results are the maximum likelihood estimates (MLEs). For diffuse priors, the Bayesian results are close to OLS in the sense that the posterior distribution of coefficients is similar to a normal distribution with means and variances close to the OLS coefficients and squared standard errors.

If vague theory generates several different models, the posterior distribution for a coefficient is simply the weighted average of the posterior distributions for each of the models. The weights are given by the posterior probabilities of the models (Leamer 1978, p. 117). Consider three models sharing a predictor with the regression coefficient β. The other sets of independent variables in the models might be overlapping, disjoint, or nested. If the posterior probabilities for the models are written as ρ_1, ρ_2, and ρ_3, and the posterior distributions of β under each model are $p_1(\beta|\text{data})$, $p_2(\beta|\text{data})$, and $p_3(\beta|\text{data})$, then the unconditional posterior distribution for β that takes account of theoretical vagueness generating the different models is

$$p(\beta|\text{data}) = \rho_1 p_1(\beta|\text{data}) + \rho_2 p_2(\beta|\text{data}) + \rho_3 p_3(\beta|\text{data}). \quad (1)$$

Because it is based on a mixture of results from the three models, this final probability distribution, $p(\beta|\text{data})$, is called the Bayesian model-mixing posterior (Raftery 1993). If only one model has high probability, vague theory poses few problems. Estimates from other models receive little weight and do not influence the final conclusions much. If inferences are based on only one model, M, but others have high probability and give quite different results, the final conclusion based only on M will be misleading.

Note that the quantity of interest, β, should have the same meaning across models. For instance, if β is a logit coefficient in one model but a probit coefficient in another, it makes little sense to pool results because β is measured on two different scales. In the example of equation (1), in each model β expresses the change in the dependent variable for a unit change in the independent variable, all relevant factors held constant. The model-mixing posterior quantifies uncertainty about these "relevant factors."

The unusual ingredient for inference is the posterior model probability. Again using Bayes's rule:

$$\text{Posterior model probability} \propto \text{Model likelihood} \times \text{Model prior}.$$

Posterior model probabilities are discussed in greater detail below, but at this point intuitions about the posterior model probabilities can be sharpened by considering diffuse prior information that allows the sample data to decisively shape the final results. In this situation, the posterior model probability is approximately a function of the likelihood evaluated at the MLEs, sample size, and the number of parameters. The posterior model probability for linear regression then is approximately a function of R^2 and the degrees of freedom.

Note that the Bayesian critique goes beyond the well-rehearsed warnings about data mining. Overfitting is only a symptom, and not a cause, of excessively optimistic statistical inferences. The root cause is unreported uncertainty about the model specification. Imagine a researcher who sits down at the computer, prepared for a long day of data analysis. To her surprise, the estimates come out "right" (conform to her prior expectations) with the first pass through the data. Model uncertainty for the Bayesian remains unaccounted for in this situation. When theory only vaguely guides the choice of model, prior uncertainty about the model is not reflected in classical inferences, even in the absence of data mining.

4. STATISTICAL AND COMPUTATIONAL DETAILS

The problem of model uncertainty reduces to making inferences about an unknown quantity such as a (possibly vector-valued) parameter or future observation, Δ, given sample information, y, and uncertainty about a set of model assumptions, M (Draper 1995). For a given dependent variable, regression assumptions contain information about the independent variables, functional form, and stochastics. This information implies a corresponding set of parameters, θ_M. A statistical model is thus defined by (M, θ_M), denoting some assumptions and their corresponding parameters. The subscript M indicates the dependence of the parameters on the model assumptions. (I omit the subscript when its reference is obvious.)

To obtain an inference about Δ that depends on the data and not on one particular model, we average inferences about Δ across models. Here, θ is treated as a nuisance parameter and is integrated out. The posterior, $p(\Delta|y)$, is the weighted average of the posteriors under each model, $p(\Delta|y, M)$, where the weights are given by the posterior probabilities, $p(M|y)$. For a discrete set of m models, integrating over the model space in equation (2) simplifies to:

$$\begin{aligned} p(\Delta|y) &= \sum_{i=1}^{m} \int p(\Delta|y, M_i, \theta_i) p(M_i, \theta_i|y) d\theta_i \\ &= \sum_{i=1}^{m} p(M_i|y) \int p(\Delta|y, M_i, \theta_i) p(\theta_i|M_i, y) d\theta_i \\ &= \sum_{i=1}^{m} p(M_i|y) p(\Delta|y, M_i). \end{aligned}$$

The posterior model probabilities, $\rho_i = p(M_i|y)$, the model-specific posterior means, $d_i = E(\Delta_i|y, M_i)$, and their variances, $V_i = V(\Delta_i|y, M_i)$, can be used to obtain expressions for the model-mixing means and variances (Leamer 1978, p. 118):

$$E(\Delta|y) = \bar{d} = \sum_{i=1}^{m} \rho_i d_i \qquad (2)$$

and

$$V(\Delta|y) = \sum_{i=1}^{m} \rho_i V_i + \sum_{i=1}^{m} \rho_i (d_i - \bar{d})^2. \quad (3)$$

Equation (3) suggests a "model uncertainty audit" in which the posterior variance can be divided into terms expressing uncertainty within models and uncertainty across models (Draper 1995). Even if Δ is estimated precisely under each model, the variance of the model-mixing posterior will be large if the estimates are widely dispersed.

The posterior probability of the model assumptions, $p(M_i|y)$, is closely related to the Bayes factor, B, a statistic used in Bayesian hypothesis testing (Kass and Raftery 1995). The Bayes factor measures the evidence in favor of one model over another provided by the sample data. Comparing two competing models, M_1 and M_0, the Bayes factor works as a multiplier transforming prior into posterior odds:

$$\frac{p(M_1|y)}{p(M_0|y)} = B_{10} \times \frac{\pi(M_1)}{\pi(M_0)},$$

where $\pi(M_i)$ is the prior model probability of model i ($i = 0, 1$). The Bayes factor is the ratio of the model, or integrated, likelihoods,

$$B_{10} = \frac{p(y|M_1)}{p(y|M_0)},$$

where

$$p(y|M_i) = \int p(y|\theta_i, M_i)\pi(\theta_i|M_i)d\theta_i, \quad (4)$$

$\pi(\theta_i|M_i)$ is the prior density of the parameters, and $p(y|\theta_i, M_i)$ is the density of the data, equivalent to the likelihood for θ_i. If $B_{10} = 1$, both models are equally likely, while the data favor M_1 over M_0 if $B_{10} > 1$. Notice also that if two models are equally likely *a priori*, $\pi(M_1)/\pi(M_0) = 1$, the ratio of the posterior probabilities of the models is given by the Bayes factor.

For a comparison of $m + 1$ models, the Bayes factors $B_{10}, \ldots, B_{m0}$ can be used to obtain the posterior probabilities:

$$p(M_i|y) = \frac{\alpha_i B_{i0}}{\sum_{r=0}^{m} \alpha_r B_{r0}}.$$

The prior odds of M_i against M_0 are given by $\alpha_i = \pi(M_i)/\pi(M_0)$. Setting $\alpha_i = 1$ $(i = 1, \ldots, m)$ corresponds to the judgment that no model is *a priori* more likely than any other.

Finding Bayes factors and the posterior model probabilities from equation (4) thus involves specifying priors, $\pi(\theta_i|M_i)$, for the parameters, evaluating the likelihood, $p(y|\theta_i, M_i)$, and integrating the product of the likelihood and the prior. Prior information—especially from qualitative historical studies—will often be available for many macrosociological problems, so this seems an advantage of the Bayesian approach in this area (Western and Jackman 1994). Owing to their subjectivity, prior distributions will vary across researchers, so it is useful to consider priors which are less subjective in the sense that sample data are allowed to dominate the final results (Leamer 1978, p. 111). Such distributions reflect vague or diffuse prior information about the parameters. The following applications make use of these diffuse priors.

A simple approximation for the Bayes factor is based on the Bayesian Information Criterion (BIC) (Raftery 1995). The BIC for model M_1, denoted BIC_1, is an approximation to $2\log B_{10}$, where B_{10} is the Bayes factor for comparing M_1 with a baseline model M_0. With a proper prior equivalent to a single typical observation, the BIC for a linear regression model, M_k, is approximately equal to

$$\text{BIC}_k = n\log(1 - R_k^2) + p_k \log n, \tag{5}$$

where R_k^2 is the R^2 for M_k and p_k is the number of regression coefficients, excluding the intercept. The baseline model for comparison is the null model that includes only the intercept.

For more complicated problems, approximate or numerical methods are often needed for the integration in equation (4) (see Tanner 1992, ch. 6). The Laplace approximation provides an asymptotic estimator based on the assumption that the posterior density, proportional to $p(y|\theta_i, M_i)\pi(\theta_i|M_i)$, is unimodal and that its logarithm is approximately quadratic. Assuming the sample data dominate the prior, so the posterior mode for the parameters is roughly equal to the MLEs, the Laplace approximation can be adapted as follows:

$$\begin{aligned} I_i &= \int p(y|\theta_i, M_i)\pi(\theta_i|M_i)d\theta_i \\ &\approx (2\pi)^{p_i/2}|\hat{\Sigma}_i|^{1/2} L(y|\hat{\theta}_i, M_i)\pi(\hat{\theta}_i|M_i), \end{aligned}$$

where p_i is the dimensionality of θ_i, $\hat{\theta}_i$ is the MLE of θ_i, $\hat{\Sigma}_i$ is its estimated covariance matrix and $\hat{L}_i = L(y|\hat{\theta}_i, M_i)$ is the maximized value of the likelihood. As sample size grows, the influence of parametric prior information shrinks, and the only nonstandard calculations involve taking the determinant of the covariance matrix and evaluating the likelihood:

$$\hat{I}_i = (2\pi)^{p_i/2}|\hat{\Sigma}_i|^{1/2}\hat{L}_i. \tag{6}$$

The accuracy of this approximation depends in part on how $\hat{\Sigma}$ is calculated. Where the covariance matrix is based on the observed Fisher information, the relative error of $\hat{I}$ is of order n^{-1}. If the expected Fisher information is used, as is common in software for generalized linear models, the error is of order $n^{-1/2}$ (Kass and Raftery 1995).

For nested hypotheses in regression, Raftery (1993) improves the accuracy of $\hat{I}$ by placing a proper but diffuse prior on the parameters, and approximating the posterior mode with a single Newton step from $\hat{\theta}$. This yields

$$2\log B_{10} \approx \Lambda + (E_1 - E_0), \tag{7}$$

where $\Lambda = 2[\hat{L}_1 - \hat{L}_0]$ is the usual log likelihood ratio statistic on $(p_1 - p_0)$ degrees of freedom, and

$$\begin{aligned} E_i = {} & 2\lambda(\hat{\theta}_i) + \\ & l_i'(\hat{\theta}_i)^T(F_i + G_i)^{-1}\{2 - F_i(F_i + G_i)^{-1}\}\, \lambda_i'(\hat{\theta}_i) - \\ & \log|F_i + G_i| + p_i\log(2\pi), \end{aligned}$$

where the prior variance $V(\theta_i|M_i) = W_i = G_i^{-1}$, $F_i = \hat{\Sigma}_i^{-1}$, $\lambda(\theta_i)$ is the log of the prior density $\pi(\theta_i|M_i)$, and $\lambda_i'(\hat{\theta}_i)$ is the p_i-vector of derivatives of $\lambda_i(\theta_i)$ with respect to θ_i. The prior distribution of the regression parameters specifies that in standard form they are *a priori* independent and normally distributed with mean zero and variance ϕ. The prior variance calibrates the influence of prior information on the posterior. In the following application, $\phi = 1$, yielding posterior model probabilities close to those obtained from the approximation in equation (5). Similar prior specifications with $1 < \phi < 5$ provide similar posterior inferences. Raftery (1993) discusses prior distributions for equation (7). Software applying (7) to generalized linear models is available in the S-Plus function, glib, which can be obtained by sending the e-mail message "send glib from S" to statlib@stat.cmu.edu.

Bayesians have variously described the tentative and provisional character of model assumptions as "specification uncertainty" (Leamer 1983), "structural uncertainty" (Draper 1995), and "model uncertainty" (Raftery 1995). Leamer (1978, ch. 4) provides a seminal discussion. Draper's (1995) more general treatment covers some recent developments in Bayesian computations (see also Kass and Raftery 1995). Background material for these sources can be found in Lee (1989, pp. 124–28).

5. EXAMPLES

5.1. *Welfare States in 18 OECD Countries*

A simple analysis of welfare state data illustrates inference under model uncertainty. I examine covariates of decommodification in a large group of industrialized democracies using cross-sectional data for 1980. Decommodification, measured by Esping-Andersen (1990, pp. 49–54), refers to the extent to which citizens' livelihoods are separated from their position in the labor market. Although this cross-sectional design cannot tell us for sure how decommodification would change given changes in the independent variables, data of this kind are common in this research (Esping-Andersen and van Kersbergen 1992, p. 190). Because the decommodification variable measures institutional characteristics of the welfare state, it also shows greater cross-sectional than longitudinal variation. Consequently, the welfare state theories make stronger predictions about comparative differences than changes over time.

This analysis focuses on the impact of social democratic government and per capita gross domestic product (GDP) on the development on welfare state institutions. The impact of left party (social democratic) government was mentioned above in the discussion of vague theory. Per capita GDP measures the wealth of a country. In the development theory of the welfare state, social spending is high in wealthy countries because more funds are available for redistribution, and economic development summarizes the "overall industrial maturation and social modernization of a country" (Esping-Andersen 1990, p. 112). Other variables—the size of the aged population, unemployment, and labor force participation—must be controlled for in order to accurately gauge the effects of theoretical interest.

TABLE 1
Data for Regression Analysis of Welfare State Development in 18 OECD Countries, 1980

	(1)	(2)	(3)	(4)	(5)	(6)
Australia	13.0	1.7	9.6	4.1	10.1	70.2
Austria	31.1	7.6	15.4	1.5	10.3	64.6
Belgium	32.4	1.6	14.4	4.5	11.8	63.0
Canada	22.0	.0	9.5	6.7	10.6	71.8
Denmark	38.1	5.3	14.4	4.2	13.0	80.3
Finland	29.2	3.8	12.0	3.6	10.4	76.4
France	27.5	1.0	13.9	4.0	12.1	68.1
Germany	27.7	5.5	15.5	2.5	13.3	68.5
Ireland	23.3	.4	10.7	6.9	5.2	62.3
Italy	24.1	.7	12.9	6.4	6.9	60.8
Japan	27.1	.0	9.0	1.4	8.9	71.8
Netherlands	32.4	1.3	11.5	4.3	11.9	57.7
New Zealand	17.1	1.7	9.7	.7	7.4	65.6
Norway	38.3	5.3	14.8	1.6	14.0	75.3
Sweden	39.1	6.5	16.1	2.0	14.8	81.0
Switzerland	29.8	1.5	13.7	.4	15.9	74.4
United Kingdom	23.4	6.3	14.9	3.7	9.3	58.2
United States	13.8	.0	11.3	6.2	11.4	71.1

Note: Column headings are as follows: (1) Esping-Andersen's decommodification score, 1990; (2) The cabinet share of labor and social democratic parties from 1965–1980 (Cameron 1984); (3) percent of population aged over 65, 1980 (Organisation for Economic Cooperation and Development [OECD] 1987); (4) average unemployment rate, 1971–1980 (OECD 1990); (5) per capita GDP (US$1,000s), 1980 (OECD 1988, 1990); (6) total labor force as a percentage of the population aged 15 to 64 (OECD 1987).

Here, the familiar distinction is drawn between predictors of interest and a large and uncertain set of controls. Data are collected for a group of 18 OECD countries (Table 1). Lacking a natural metric, the decommodification and left cabinet variables are standardized before regression analysis.

Diagnostics reveal no clear departures from the standard linear model assumptions, so the predictors are written with linear effects and the errors are assumed to be normal. This part of the model is not subject to much uncertainty.

How should prior probabilities for the models be chosen? Following the common strategy, I first include the GDP and left cabinet variables on their own. The model is then augmented by various combinations of the controls. This strategy captures how real data analyses often proceed. With uncertainty about the control variables, no particular combination is preferred over any other and the models are treated as equally likely *a priori*. Strong interest in the left cabinet and GDP variables is formalized by a prior that gives zero probability to the possibility that they have no effect. This prior also ensures that the resulting posterior is smooth. Uncertainty about which variables to control is described by a prior belief that the aged population, unemployment, and labor force participation variables may not belong in the regression. Some prior probability is thus placed at 0 for the control variables.

Conclusions are conditional on this prior. Readers who are not convinced in advance that GDP and left cabinet variables have an effect should, rightly, find the conclusions implausible. Alternative priors on the model space could yield substantively different results. Examination of all possible subsets of predictors, for example, represents a more agnostic prior. A wider class of models is entertained and at least some prior probability is concentrated at zero for all coefficients. Estimation with this type of prior is discussed by Raftery, Madigan, and Hoeting (1993).

Table 2 presents the regressions. Moderately strong results for the left cabinet and GDP effects are found in models 1, 2, 3, and 5. Despite the small sample, both coefficients are at least 1.5 times their standard errors in these specifications. A standard deviation change in the left cabinet variable (about three years of social democratic government) is associated with a .4 of a standard deviation increase in the decommodification score—roughly the difference in welfare state development between the United Kingdom and Germany. A $2,500 increase in per capita GDP yields a similar effect. Estimates in these models are substantively large and robust to the inclusion of the unemployment and labor force participation effects. Adding the aged population variable shrinks the coefficients of interest. (Left cabinet and GDP correlate with aged population at .8 and .6). In the models with aged population, the left cabinet effect decreases by three-quarters or more. The GDP effect is more stable, but posteriors under models 7 and 8 are essentially centered over 0. On balance, both effects are sensitive to the specification, but they retain their positive signs and are estimated with small standard errors in several of the simpler models.

TABLE 2

Regression Results from Analysis of the Welfare State Data

Independent Variables	Possible Models								Bayes Mixture Posterior
	1	2	3	4	5	6	7	8	
Intercept	−1.66	−1.71	−1.93	−3.40	−2.01	−3.33	−5.55	−5.56	−2.67
	(.83)	(1.11)	(2.02)	(1.61)	(2.67)	(1.68)	(3.13)	(3.24)	(2.12)
Left cabinet	.39	.39	.39	.11	.39	.06	.02	−.05	.26
	(.21)	(.23)	(.21)	(.30)	(.24)	(.35)	(.32)	(.38)	(.30)
GDP per capita	.15	.15	.14	.10	.15	.09	.04	.02	.12
	(.07)	(.08)	(.09)	(.08)	(.10)	(.09)	(.11)	(.12)	(.10)
Unemployment	–	.01	–	–	.01	−.04	–	−.05	−.02
		(.11)			(.12)	(.12)		(.12)	(.12, .81)
Participation rate ($\times 10^{-2}$)	–	–	.52	–	.54	–	2.97	3.11	1.82
			(3.44)		(3.58)		(3.70)	(3.84)	(3.68, .78)
Aged population	–	–	–	.18	–	.20	.24	.26	.20
				(.14)		(.16)	(.16)	(.18)	(.15, .59)
R^2	.45	.45	.45	.50	.45	.51	.53	.53	–
Adjusted R^2	.37	.33	.33	.40	.28	.36	.38	.34	–
Posterior probability (5)[a]	.39	.09	.09	.24	.02	.06	.09	.02	–
Posterior probability (7)[b]	.38	.09	.09	.26	.02	.06	.08	.02	–

[a]Model probabilities estimated from equation (5).

[b]Model probabilities estimated from equation (7).

Note: Standard errors for models 1 through 8 are in parentheses. Expectations and standard deviations of the Bayes mixture posterior are given for the intercept, left cabinet, and GDP effects. Coefficient estimates for the control variables are the posterior expectations, given that the coefficient is not zero. Numbers in parentheses for these coefficients are the posterior standard deviation given that the coefficient is not zero, and the posterior probability that the coefficient is zero.

The Bayesian model-mixing posterior is based on model probabilities estimated with equations 5 and 7. Although the two equations specify diffuse priors somewhat differently, results are similar. Unlike R^2, the posterior model probabilities penalize large models. The mixture posterior thus gives little weight to models 7 and 8, the two with the highest R^2s. Compare adjusted R^2s that also impose penalties for parameters. These statistics barely distinguish the simple model 1 from highly parameterized models like 4, 6, and 7. Consistent with its goodness-of-fit and parsimony, more than a third of the posterior probability is given to model 1. Model 4, which featured weak GDP and left cabinet effects, might be discounted in a conventional analysis but is allocated a quarter of the posterior probability.

For the mixture posterior, the variances of the left cabinet and GDP effects are larger than several of the OLS results. The posterior mean for the left cabinet effect is a compromise between the strongest and weakest results. The sign of the coefficient is uncertain, as a 90 percent confidence interval, $[-.24, .74]$, overlaps 0 by a large margin. Similarly, the posterior mean for the GDP effect is 20 percent smaller than the strongest results and nearly equal to its standard error. A 90 percent confidence interval, $[-.04, .28]$, also includes 0. Strong conclusions about the sign of the coefficients are thus unavailable once model uncertainty is taken into account. Reliance on any of the results from models 1, 2, 3, or 5 would signficantly understate uncertainty about the model specification.

It may be tempting to interpret the model-mixing posteriors of the control variables by looking at their means and standard errors, but inference should proceed carefully here because these quantities are poor summaries of the posterior distributions. Posteriors for the left cabinet and GDP coefficients are smooth, being mixtures of normal distributions. In the case of the age, unemployment, and labor force participation effects, however, the distributions are mixtures of normals and a probability mass placed at 0. The posteriors for the control variables have a spike at 0 representing the mixture of certain and uncertain prior knowledge about these coefficients that is implied by the modeling strategy. To obtain model-mixing results for the controls (or posterior marginal distributions for any of the coefficients), posteriors are simulated by weighted sampling from multivariate normal distributions with means and covariances given by the OLS results for models 1 to 8. Sample weights equal the posterior

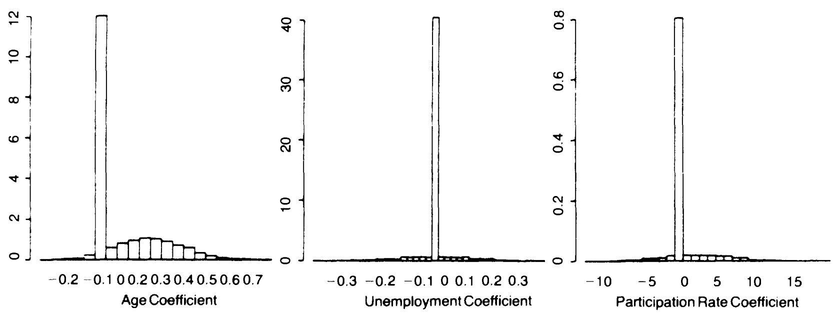

FIGURE 1. Histograms from the simulated marginal posteriors of the age, unemployment coefficients, and labor force participation coefficients.

model probabilities. Simulated coefficients for a control variable were set to 0 in models which omitted that variable. Table 2 summarizes these simulated posterior distributions with three numbers: the expectation and standard deviation of the nonzero part of posterior distribution, and the posterior probability that the coefficient equals zero. These results illustrate the modest support in the data for the age coefficient. While half of the prior probability was placed at zero for the age effect, the data shift an additional 10 percent, giving a 60 percent posterior probability of a zero age coefficient. Corresponding numbers are much larger for the other controls. The irregular shape of the distributions of the control coefficients is indicated in Figure 1, which shows histograms of samples drawn from the mixture posterior.

5.2. *Strikes in the Postwar Period*

The second application reports estimates from several models of strike activity. This analysis of data from 18 OECD countries for the period 1951–1985 makes use of a pooled cross-sectional design that is becoming increasingly common in quantitative comparative work (Esping-Andersen and van Kersbergen 1992, p. 190). Strike activity is defined as the annual number of days lost per 1000 workers through industrial disputes (see Hibbs 1976). Strike volume is written as a linear function of economic and institutional variables. The economic variables include unemployment and inflation. The institutional variables consist of union density, left party parliamentary

TABLE 3
Summary of Strike Volume Data, 18 OECD countries 1951–1985 (N = 625)

Countries	Average Strike Volume	Standard Deviation	Minimum	Maximum
Australia	387	247	109	1252
Austria	26	48	0	277
Belgium[a]	247	260	25	1453
Canada	750	962	28	5918
Denmark	195	448	2	2001
Finland	451	967	11	5568
France	383	1159	41	7000
Germany	44	64	1	260
Ireland	547	399	70	1752
Italy	998	593	257	2875
Japan	174	191	6	1061
Netherlands	26	35	0	142
New Zealand	268	412	29	2462
Norway	86	171	1	946
Sweden	74	199	0	1150
Switzerland	4	8	0	35
United Kingdom	326	316	76	1278
United States	448	274	73	1236

[a]1950–1980.

Note: Standard deviations for each country are calculated for each 1951–1985 time series. Data sources are described in Western (1993).

representation, and a time-invariant measure of union centralization. An intercept for each country captures differences in the average national level of strike activity. For simplicity, I assume this list of independent variables is known with certainty, although the following analysis could be expanded to include various subsets of predictors. This assumption has some realism given a consensus about the important predictors of strike activity (e.g., Hibbs 1976; Korpi and Shalev 1980).

How should randomness in the model be specified? Errors for data with this structure are often treated as autocorrelated and heteroscedastic. Diagnostics show little serial correlation but descriptive statistics in Table 3 indicate that countries with above average strike volume also have high variance. Cross-national variability in the volatility of the strike series can be stabilized with a log transformation. The log transformation also reduces skewness in the data. To remove zeros, I add one to the strike volume variable before applying the transformation. If S_{ij} denotes annual log strike volume in the jth country for year i ($i = 1, \ldots, I$), and covariate information is collected in the vector x_{ij}, expected log strike volume can be written as a linear function:

$$E(S_{ij}) \equiv \mu_{ij} = \alpha_j + x'_{ij}\beta.$$

The logged strike variable could be treated as multivariate normal with different means for each of the 18 OECD countries in the sample:

$$\begin{bmatrix} S_{i1} \\ \vdots \\ S_{I18} \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_{i1}(x'_{i1},\alpha_1,\beta) \\ \vdots \\ \mu_{I18}(x'_{I18},\alpha_{18},\beta) \end{bmatrix}, \begin{bmatrix} \sigma^2 & & \\ & \ddots & \\ & & \sigma^2 \end{bmatrix}\right).$$

Using equation (6), posterior model probabilities can be calculated with the MLEs, their estimated covariance matrix, and the likelihood evaluated at its maximum. The OLS estimates are ML with this specification. The likelihood of a typical observation is given by the usual normal density

$$L_{ij}(\alpha_j,\sigma^2,\beta|S_{ij}) = (2\pi\sigma^2)^{-1/2}\exp\left[-\frac{(S_{ij}-\mu_{ij})^2}{2\sigma^2}\right],$$

and the maximized likelihood equals the product of the likelihood values at the MLEs over all observations.

The log-normal specification may be unrealistic because strike series have large spikes. For example, French strike volume in 1968, or Italian in 1969, far exceeded the average level of industrial disputation in those countries. If extreme values are more common in the strike data than under the normal model, a fat-tailed distribution such as the t_ν (on ν degrees of freedom) may provide a better specification (Lange, Little, and Taylor 1989). The t_ν model is a robust redescending M-estimator of the type recently recommended to comparative sociologists (Dietz, Frey, and Kaloff 1987).The t_ν distribution downweights outliers, allowing estimates to better summarize the majority of the data. The robustness of the t_ν model is controlled by the degrees of freedom, ν. The tails of the likelihood function widen as ν decreases, and outlying cases are increasingly downweighted. As ν grows, the t model resembles the normal. In the robust regression literature, ν is a tuning constant that determines the sensitivity of the estimates to outliners. Because the tuning constant is difficult to specify *a priori,* uncertainty about this aspect of the model specification can also be built into the Bayesian analysis. In the following analysis I experiment with likelihood functions based on $\nu = 1, 1.5, \ldots, 20$.

Applying the log transformation to the raw strike data gives a log-t model that is identical to the normal model except that the normal distribution is replaced by a t distribution:

$$\begin{bmatrix} S_{i1} \\ \vdots \\ S_{I18} \end{bmatrix} \sim t\left(\begin{bmatrix} \mu_{i1}(x'_{i1},\alpha_1,\beta) \\ \vdots \\ \mu_{I18}(x'_{I18},\alpha_{18},\beta) \end{bmatrix}, \begin{bmatrix} \psi & & \\ & \ddots & \\ & & \psi \end{bmatrix}, \nu\right).$$

The contribution of a typical observation to the likelihood is given by

$$L_{ij}(\alpha_j,\psi,\beta|y_{ij}) = \frac{1}{\sqrt{\psi}}\frac{\Gamma\{(\nu+1)/2\}}{\Gamma(1/2)\Gamma(\nu/2)\nu^{1/2}} \times \left[1 + \frac{(S_{ij}-\mu_{ij})^2}{\psi\nu}\right]^{(\nu+1)/2}.$$

An iteratively reweighted least squares algorithm described by Lange et al. (1989) and Rubin (1983) was used for estimation. MLEs are used to form the marginal likelihoods, posterior model probabilities, and the model-mixing results with equation (7). Means and standard deviations of the model-mixing posterior can then be obtained with equations (2) and (3) or by simulation.

The regression results for selected models are reported in Table 4 (country intercepts have been omitted). Log likelihoods indicate the superiority of the t model on 5.5 degrees of freedom. All three models reported provide strong evidence for the positive effects of union density, inflation, and union centralization on strike volume. With the $t_{5.5}$ likelihood, the 10-point decline in union density experienced in the United Kingdom and the United States is estimated to reduce strike volume by nearly a third. A 10-point rise in inflation, characteristic of the Western European countries in the late 1960s, is associated with a 50 percent increase in strike volume. The inflation and union density estimates are consistent with theory and the historical record, which associates the strike waves of the 1960s with rapidly rising prices and labor passivity in the 1980s with shrinking labor organization (Soskice 1978; Shalev 1992). A similar substantive story is provided by the t_1 model, although likelihoods indicate that this model receives little support from the data.

TABLE 4
Regression Analysis of Log Strike Volume in 18 OECD Countries, 1951–1985.

Independent Variables	Error Distribution			Bayes Mixture Posterior
	Normal	t_1	$t_{5.5}$	
Union density	.0265	.0232	.0268	.0269
	(.0070)	(.0037)	(.0057)	(.0061)
Unemployment	−.0212	−.0067	−.0148	−.0164
	(.0202)	(.0102)	(.0165)	(.0176)
Inflation	.0388	.0412	.0387	.0383
	(.0111)	(.0053)	(.0089)	(.0096)
Left parties	.0000	−.0048	.0006	−.0002
	(.0063)	(.0030)	(.0050)	(.0054)
Union centralization	−2.6449	−2.8100	−2.6928	−2.6918
	(.9224)	(.4273)	(.7250)	(.7779)
Log likelihood	−948	−1001	−936	–

Note: Standard errors are in parentheses. The mixture posterior averages over t_ν likelihoods with $\nu = 1, 1.5, \ldots, 20$.

Differences between the robust and normal theory results are reconciled by Bayesian inference. Figure 2 plots the maximized log likelihoods for values of ν. The $t_{5.5}$ model has the highest likelihood, but the likelihood is quite flat in the range $5 \leq \nu \leq 8$. Models in this range account for more than 40 percent of the posterior model probability. The remainder is mostly spread over the thinner-tailed specifications. Models with very thick-tailed error distributions contribute little to the final inference. The expectations of the mixture posteriors are quite close to the $t_{5.5}$ MLEs, but the posterior standard deviations are about 10 percent larger. Still, even when model uncertainty is taken into account, sharp conclusions can be drawn about the direction of all effects except left parliamentary representation and unemployment. Inference based on the normal model alone would have been slightly conservative. In this application, the data give relatively strong support for a small subset of the models and prior uncertainty about the specification is reduced by the data.

6. DISCUSSION

The ideas presented here can be extended in several ways. The Bayesian treatment of model uncertainty suggests several non-

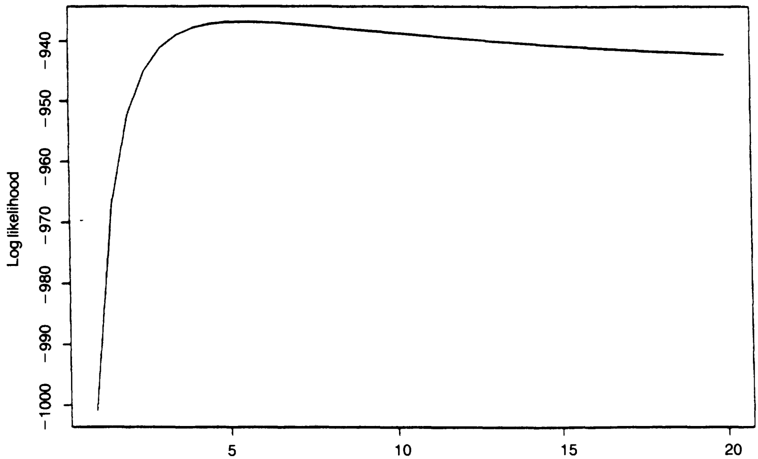

FIGURE 2 Maximized log likelihoods for different values of the degree of freedom parameter, ν, in the log-t model of strike volume.

Bayesian alternatives based upon the introduction of new sample information. This analysis also speaks to recent methodological debates in macrosociology.

One alternative to the model-mixing approach to model uncertainty involves the investigation of new sample data. Cross-validation provides a simple method where, say, half the data are closely analyzed for the purpose of model selection. The remainder are used to estimate parameters. Overfitting in the model selection stage of the data analysis is penalized at the estimation stage (Picard and Berk 1990). Ideally, if two data sets were available, one could be used for model selection and the other for parameter estimation. Neither approach may be feasible in macrosociology because data sets are small and nonstochastic. Data splitting would have little power in the welfare state analysis where multiple regression would be attempted with only nine cases. Because the data are not generated by any repeatable data mechanism (Berk, Western, and Weiss 1995), analyzing new data also seems impractical. The current data are often the only ones available.

Another approach, similar in spirit to cross-validation but less demanding of the data, suggests bootstrapping the entire model selection process (Efron and Gong 1983). A bootstrap sample is drawn from the data, and several models are fit; one is chosen. A distribution is built up from coefficients of the chosen models of many bootstrap samples. This is practical when models are selected with a statistical criterion. In macrosociology, model choice often involves a more complicated mix of substantive and statistical considerations. This is partly because models are fit with one eye to the specific historical stories of certain cases. Bootstrap applications will be less automatic in these settings where substantive criteria play a strong role in model selection. Sometimes the bootstrap will be undesirable on statistical grounds as Monte Carlo experiments show that when the number of candidate variables is close to the sample size, the bootstrap inaccurately measures statistical error due to model uncertainty (Freedman et al. 1988).

Another alternative might informally employ the Bayesian ideas. Discussion of the data analysis would allow readers to judge whether the results were an artifact of the model search. More systematically, sensitivity analyses could report variation in parameters for a large set of models (Leamer 1983). Both strategies informally pool results across models, explicitly considering model uncertainty in the final data summary. Although similar in motivation to the Bayesian approach, these strategies are often guided by poorly specified intuition, and in any event, may be too conservative because the data are not used to assess the likelihoods of the different models.

It might be objected that the problem of model uncertainty is completely internal to quantitative research. A key issue in recent debates on macro methods—the treatment of time and temporality—is ignored, and the models of the welfare state and strike analyses resemble those targeted by the critics of conventional methods. In some ways the problem of model uncertainty issues only an "internal" and not a fundamental "external" challenge to quantitative approaches to comparative and historical questions. However, there are at least three links between this paper and the issues occupying the methodological discussions.

First, one criticism holds that conventional quantitative applications provide an excessively deterministic picture of the social world (Abbott 1988). This partly reflects misunderstandings of critics and researchers alike about the uncertain character of statistical estimates. But it also seems true that excessive determinism results from a mistaken belief in the certainty of statistical findings. If limitations on theoretical knowledge are taken seriously, uncertainty statements will often be revised upward, and researchers will increasingly conclude that the available data do not sustain a confident answer to a research question. New levels of uncertainty about statistical estimates that result from a full accounting of model uncertainty may provide new impressions of contingency in social processes.

Second, by treating assumptions as doubtful, uncertainty could be introduced into deterministic alternatives to statistics, such as narrative or Ragin's (1987) qualitative comparative analysis (QCA). As things stand, narrative methods and QCA are deterministic in the sense that no uncertainty surrounds the data summaries they provide. (Schama 1991 makes this observation for narrative; Ragin 1987, pp. 113–18, discusses this for QCA.) An apparently unique narrative, for example, is sustained by the evidence, and the assumptions behind the analysis (though often inexplicit) are treated as known with certainty. To introduce a stochastic element, researchers could report how narratives or QCA conclusions change as a result of altering assumptions. If several different narratives can be based on the same evidence, these variations could be presented to ensure that the preferred narrative is not atypical. Griffin (1993) offers an analysis of this sort. QCA applications could assess the sensitivity of conclusions to the list of variables included for analysis. In both contexts the objective is to see whether a wide variety of assumptions admits a wide variety of conclusions. Whatever the method, there should be little confidence in conclusions that are highly sensitive to the choice of assumptions.

Third, uncertainty in quantitative macrosociology could be reduced by introducing more prior information, imposing sharper restrictions on the space of possible models. This nonsample information might reasonably come from two sources: (1) interpretation of the historical record or (2) deductive theory. The historical approach dovetails with calls by Tilly (1984) and Isaac and Griffin (1989) for greater historical specificity in quantitative research. In this approach, theoretical content could be developed with the rich historical material that often informs quantitative macrosociology in a more informal way. Prior information with these origins may be too ideographic for those, like Kiser and Hechter (1991), who prefer "general theorizing." For these researchers, prior information with a sound deductive basis may be more plausible. Regardless of the origins of nonsample information, theories that narrow the model space should carry strong prior information about variable selection and stochastic form.

This suggestion, however, is a double-edged sword. Paraphrasing von Clausewitz, methodology is theory by other means. If we claim a lot for our theories, our methods are affected accordingly. Narrowing prior uncertainty about statistical models means that theoretical assumptions—albeit historically grounded or deductively derived—become increasingly influential for research conclusions. In the Bayesian language, prior information increasingly dominates sample information. If these more precisely specified models are reasonable approximations of the processes under study, the increasing influence of prior information is of little concern. On the other hand, if the models are precisely specified but wrong, we would acquire greater certainty at the cost of being wrong. I see no compelling examples of precise theory that are highly informative about variable selection, functional form, and stochastics. If this is a realistic assessment of current knowledge, researchers should routinely examine a large number of models, letting any restrictions on the model space be imposed by the data rather than prior beliefs. Statistical assumptions could then remain as tentative empirical assertions whose uncertainty is recognized with the tools of Bayesian inference.

REFERENCES

Abbott, Andrew. 1988. "Transcending General Linear Reality." *Sociological Theory* 6:169–86.

Berk, Richard A., Bruce Western, and Robert Weiss. 1995. "Bayesian Inference for Apparent Populations." In *Sociological Methodology 1995*, edited by Peter V. Marsden, 421–58. Cambridge, MA: Blackwell Publishers.

Boswell, Terry, and William J. Dixon. 1993. "Marx's Theory of Rebellion: A Cross-National Analysis of Class Exploitation, Economic Development, and Violent Revolt." *American Sociological Review* 58:681–702.

Cameron, David. 1984. "Social Democracy, Corporatism, Labor Quiescence, and the Representation of Interest in Advanced Capitalist Society." In *Order and Conflict in Contemporary Capitalism: Studies in the Political Economy of European Nations*, edited by John Goldthorpe, 143–78. Oxford: Oxford University Press.

Dietz, Thomas R., Scott Frey, and Linda Kaloff. 1987. "Estimation with Cross-National Data: Robust and Nonparametric Approaches." *American Sociological Review* 52:380–90.

Draper, David. 1995. "Assessment and Propagation of Uncertainty" (with discussion). *Journal of the Royal Statistical Society*, Series B 57:45–97.

Efron, Bradley, and Gail Gong. 1983. "A Leisurely Look at the Bootstrap, the Jackknife, and Cross-Validation." *American Statistician* 37:36–48.

Esping-Andersen, Gøsta. 1990. *Three Worlds of Welfare Capitalism.* Madison: University of Wisconsin Press.

Esping-Andersen, Gøsta, and Kees van Kersbergen. 1992. "Contemporary Research on Social Democracy." *Annual Review of Sociology* 18:187–208.

Fox, John. 1984. *Linear Statistical Models and Related Methods with Applications to Social Research.* New York: Wiley.

Freedman, David A. 1983. "A Note on Screening Regression Equations." *American Statistician* 37:152–155.

———. 1985. "Statistics and the Scientific Method." In *Cohort Analysis in Social Research,* edited by William Mason and Stephen Fienberg, 343–66. New York: Springer Verlag.

Freedman, David A., William Navidi, and Steven C. Peters. 1988. "On the Impact of Variable Selection in Fitting Regression Equations." Pp. 1–16 in *On Model Uncertainty and its Statistical Implications,* edited by Theo Dijkstra. New York: Springer-Verlag.

Griffin, Larry J. 1993. "Narrative, Event-Structure Analysis, and Causal Interpretation in Historical Sociology." *American Journal of Sociology* 98:1094–133.

Hibbs, Douglas A. 1976. "Industrial Conflict in Advanced Industrial Societies." *American Political Science Review* 70:1033–58.

Hurvich, Clifford M., and Chich-Ling Tsai. 1990. "The Impact of Model Selection on Inference in Linear Regression." *American Statistician* 44:214–27.

Isaac, Larry and Larry J. Griffin. 1989. "Ahistoricism in Time-Series Analyses of Historical Process: Critique, Redirection, and Illustrations from U.S. Labor History." *American Sociological Review* 54:873–90.

Kass, Robert E., and Raftery, Adrian E. 1995. "Bayes Factors." *Journal of the American Statistical Association* 90:773–95.

Kiser, Edgar, and Michael Hechter. 1991. "The Role of General Theory in Comparative-Historical Sociology." *American Journal of Sociology* 97:1–30.

Korpi, Walter. 1989. "Power Politics and State Autonomy in the Development of Social Citizenship: Social Rights During Sickness in Eighteen OECD Countries Since 1930." *American Sociological Review* 54:309–28.

Korpi, Walter, and Michael Shalev. 1980. "Strikes, Power, and Politics in Western Nations." *Political Power and Social Theory* 1:299–332.

Kuznets, Simon. 1955. "Economic Growth and Income Inequality." *American Economic Review* 45:1–28.

Lange, Kenneth L., Roderick J. A. Little, and Jeremy M. G. Taylor. 1989. "Robust Statistical Modeling Using the *t* Distribution." *Journal of the American Statistical Association* 84:881–895.

Leamer, Edward E. 1978. *Specification Searches: Ad Hoc Inference with Nonexperimental Data.* New York: Wiley.

———. 1983. "Let's Take the Con Out of Econometrics." *American Economic Review* 73:31–44.

Lee, Peter M. 1989. *Bayesian Statistics: An Introduction.* London: Edward Arnold.

Nielsen, François. 1994. "Income Inequality and Industrial Development: Dualism Revisited." *American Sociological Review* 59:654–677.

Organisation for Economic Cooperation and Development. 1987. *Labour Force Statistics*. Paris: OECD.

———. 1988. *National Accounts 1960–1986*. Paris: OECD.

———. 1990. *Main Economic Indicators: Historical Statistics, 1969–1988*. Paris: OECD.

Pampel, Fred C., and John B. Williamson. 1989. *Age, Class, Politics, and the Welfare State*. New York: Cambridge University Press.

Picard, Richard R., and Kenneth N. Berk. 1990. "Data Splitting." *American Statistician* 44:140–147.

Raftery, Adrian E. 1993. "Approximate Bayes Factors and Accounting for Model Uncertainty in Generalized Linear Models." Technical Report No. 255. Department of Statistics. Seattle, WA: University of Washington.

———. 1995. "Bayesian Model Selection in Social Research." In *Sociological Methodology 1995*, edited by Peter V. Marsden, 111–63. Cambridge, MA: Blackwell Publishers.

Raftery, Adrian E., David Madigan, and Jennifer A. Hoeting. 1993. "Model Selection and Accounting for Model Uncertainty in Linear Regression Models." Technical Report No. 262. Department of Statistics. Seattle, WA: University of Washington.

Ragin, Charles C. 1987. *The Comparative Method: Moving Beyond Qualitative and Quantitative Strategies*. Los Angeles: University of California Press.

Rubin, Donald B. 1983. "Iteratively Weighted Least Squares." In *Encyclopedia of the Statistical Sciences*. Vol. 4, 272–75. New York: Wiley.

Schama, Simon. 1991. *Dead Certainties (Unwarranted Speculations)*. New York: Vintage.

Shalev, Michael. 1992. "The Resurgence of Labour Quiescence." In *The Future of Labour Movements*, edited by Marino Regini, 102–32. London: Sage.

Soskice, David W. 1978. "Strike Waves and Wage Explosions, 1968–1970: An Economic Interpretation." In *The Resurgence of Class Conflict in Western Europe Since 1968*, Vol. 2, edited by Colin Crouch and Alessandro Pizzorno, 221–46. New York: Holmes and Meier.

Tanner, Martin A. 1992. *Tools for Statistical Inference: Methods for the Exploration of Posterior Distributions and Likelihood Functions*. New York: Springer-Verlag.

Tilly, Charles. 1984. *Big Structures, Large Processes, Huge Comparisons*. New York: Sage Foundation.

Western, Bruce. 1993. "Postwar Unionization in Eighteen Advanced Capitalist Countries." *American Sociological Review* 58:266–82.

Western, Bruce, and Simon Jackman. 1994. "Bayesian Inference for Comparative Research." *American Political Science Review* 88:412–423.

Wilensky, Harold. 1975. *The Welfare State and Equality: Structural and Ideological Roots of Public Expenditures*. Berkeley: University of California Press.