In estimating contrasts for ANOVA situations, the BP will be much smaller. Consider the simplest case of testing for equality of location parameters in a balanced two-sample situation with $n/2$ observations in each sample. Any two-sample test statistic constructed as the difference of location estimates for each of the samples will break down as soon as either of the two location estimates breaks down. The best we can hope for is to achieve the HBP of 0.5 *for each of the samples*. But this results in a BP of 0.25 relative to the entire sample size of n. The situation will quite clearly be worse for the higher way ANOVA situations of major interest. The basic problem in ANOVA is that typically there are only a small number of observations per parameter, and thus a small number of wild points can spoil the various contrasts of interest. Thus, there will be a premium on maintaining the highest possible BP in ANOVA situations. From this point of view, the BP of 0.5 for each parameter obtained with median-based methods (e.g., median polish where applicable) seems preferable for example to the Hodges-Lehmann estimate BP of 0.29.

Returning to the M-estimates versus R-estimates issue for a moment, it is important to note that the latter have the attractive property that they do not require estimation of a nuisance scale parameter. On the other hand, M-estimates based on the popular psi functions, e.g., Huber's favorite or Tukey's bisquare, do require an auxiliary scale estimate. For the small sample sizes per parameter that occur in many ANOVA situations, this may result in loss of robustness in level and power relative to R methods. This issue needs to be studied in detail.

One other robustness consideration is worth noting. For those situations where one really needs to estimate an effect rather than a true contrast, it will be important to control the maximum bias due to asymmetric contamination. In this regard it may be useful to consider *min-max bias robust* estimates. For example, Huber (1964) showed that among all translation equivariant estimates of location, the median solves the problem for ε contamination models. Hence, in that case the HBP and min-max bias properties coincide, but this is not always the case. Recently Yohai, Zamar and I have been working on min-max bias robust estimates of scale (Martin and Zamar, 1987) and regression (Martin, Yohai and Zamar, 1987). In the solutions found to date, the breakdown point $\mathrm{BP} = \mathrm{BP}(\varepsilon)$ will be relatively close to 0.5 for all but very small ε in an ε-contaminated model, and furthermore $\mathrm{BP}(\varepsilon) \to 0.5$ as $\varepsilon \to 0.5$. Perhaps the min-max bias robust approach will be of some utility in ANOVA problems.

ADDITIONAL REFERENCES

CARROLL, R. J. (1980). Robust methods for factorial experiments with outliers. *Appl. Statist.* **29** 246–251.

DONOHO, D. L. and HUBER, P. J. (1983). The notion of breakdown point. In *A Festschrift for Erich L. Lehmann* (P. J. Bickel, K. A. Doksum and J. L. Hodges, Jr., eds.) 157–184. Wadsworth, Belmont, Calif.

FELLNER, W. H. (1986). Robust estimation of variance components. *Technometrics* **28** 51–60.

FIELD, C. A. and RONCHETTI, E. (1985). A tail area influence function and its application to testing. *Sequential Anal.* **4** 19–41.

HAMPEL, F. R. (1974). The influence curve and its role in robust estimation. *J. Amer. Statist. Assoc.* **69** 383–393.

HAMPEL, F. R., RONCHETTI, E. M., ROUSSEEUW, P. J. and STAHEL, W. A. (1986). *Robust Statistics: The Approach Based on Influence Functions*. Wiley, New York.

HUBER, P. J. (1964). Robust estimation of a location parameter. *Ann. Math. Statist.* **35** 73–101.

LAMBERT, D. (1981). Influence functions for testing. *J. Amer. Statist. Assoc.* **76** 649–657.

MARTIN, R. D. and YOHAI, V. J. (1985). Robustness in time series and estimating ARMA models. In *Handbook of Statistics* **5** (E. J. Hannan, P. R. Krishnaiah and M. M. Rao, eds.) 119–155. North-Holland, Amsterdam.

MARTIN, R. D., YOHAI, V. J. and ZAMAR, R. H. (1987). Min-max bias robust regression. *Ann. Statist.* To appear.

MARTIN, R. D. and ZAMAR, R. H. (1987). Min-max bias robust estimates of scale. Technical Report 72, Dept. Statistics, Univ. Washington.

YLVISAKER, D. (1977). Test resistance. *J. Amer. Statist. Assoc.* **72** 551–556.

Rejoinder

David Draper

Let me open the rejoinder by thanking the discussants for their insightful and kind comments. They have not given me much to disagree with, but (since it is at least as much the job of the rejoinderer to be contentious as it is for the discussants themselves) I'll see what I can do. I begin with some remarks about the influence of robustness work on actual practice to date; continue with some comments on the relevance of expert systems research to the comparison of modeling strategies; and finally devote most of my

attention to some differences between the frequentist and Bayesian approaches to robustness and their comparative strengths and drawbacks, along the way noting some basic distinctions between inference and prediction. By way of placing these remarks in the context of the main paper under discussion, I should say that in the years since I completed the research on which the paper is based I have become considerably more sympathetic to the Bayesian viewpoint.

1. ROBUSTNESS AND REALITY

Morrie DeGroot has invited distinguished representatives from most of the various frequentist robustness schools—*L*, *M*, *R* (and *S*, as Roger Koenker and Stephen Portnoy remind us)—to comment, and each has spoken up for the merits of one approach at the expense of the others: Alan Welsh and Koenker and Portnoy for *L* procedures, Doug Martin for *M*-estimates, Tom Hettmansperger and Jay Aubuchon for the *R* school, and Peter Bickel as a comparative agnostic. This has given the discussion the flavor of a horse race (an alternative metaphor that springs to mind comes from a conversation I had with Raj Bahadur some years ago after attending my first Joint Statistical Meetings. I commented on the bewildering array of methods and viewpoints being advocated by the presenters in talks and poster sessions, and Raj said that the meetings always reminded him of market day at the central bazaar in Calcutta, with people shouting "Buy from me!" in all directions). So the question becomes, "From whom should, and will, people buy?"

The plain truth is that a betting person, asked to wager on which school would win the race toward making a practical difference in science and decision making, would be tempted on the basis of the thirty-year track record of robustness work to date to bet on "none of the above." On the positive side, there have been some nice mathematical results, some clarity has been reached on some useful ideas about stability in inference and the like, and some interesting and potentially useful methods have emerged that have had wide use descriptively and in exploratory settings. But there are so few examples of robust methods actually helping people to solve real inferential problems that it seems fair to say that the impact on practice so far has been small relative to the effort expended. Martin takes me to task for not emphasizing practical examples in this paper, a criticism that is right on the mark. In my defense I can only say that an already long paper would have been twice as long and that I intend to put out another paper contrasting robust, data-analytic and model expansion approaches to inference in a series of real-world case studies. The trouble is that we have apparently *all* been meaning to write such a paper, and it hardly ever seems to get written. For robustness work to make a bigger difference, there has to be a much wider availability of relevant software, as Martin notes, and this phantom paper has to get written by a number of people, each using a different set of real examples, so that we may build up a core of knowledge about the comparative strengths of the numerous approaches to inference now available in theory to practitioners.

2. CALIBRATION AND EXPERT SYSTEMS

Welsh touches on the good, and related, point that as a general matter the overall frequency properties of the data-analytic strategy are not well understood. Frequentists are inherently curious about such questions, but Bayesians should be, too. Everybody ought to be interested in whether they got the right answer or not (or how close they came), and the ultimate test of statisticians who are actually trying to help people solve problems is to look back over their lifetimes with an eye to the validity and efficiency of their separation of signal from noise, and the quality of their predictions, across the range of problems they tackled. (I sometimes imagine St. Peter greeting statisticians at the pearly gates (those who end up there, anyway) with the news of how they did: "Do you remember back in 1959 when you said that drug A was better than drug B? Well, you were wrong," and so on.) Thus the calibration process, which it would seem is, or should be, of universal interest, is inherently frequentist in nature. This is a special case of a general position that currently makes good sense to me on what Jerzy Neyman used to refer to as "Bayesian-frequentist controversy": It is useful to think like a Bayesian in formulating one's inferences and predictions, but it is also useful to think like a frequentist in evaluating their quality. Box (1980) expressed similar views.

The point of all this is that whether you are a Bayesian or a frequentist it seems interesting to ask: What are the operating characteristics of John Tukey? (Substitute some other good data analyst if you wish.) A related question that might actually be answerable is, What are the operating characteristics of a computer program that embodies some significant portion of his data-analytic wisdom? Current and future expert systems work (such as Gale, 1986a,b; Adams, 1988) can help in this direction. The idea is to write programs that embody various data-analytic strategies, run them on many different data sets and look at the validity and efficiency of the raw scale back-transformed findings. It is hard to think of another way to make the data-analytic approach objective enough so that systematic comparisons with the robust and model expansion strategies become possible.

3. BAYESIAN VERSUS FREQUENTIST ROBUSTNESS: INFERENCE VERSUS PREDICTION

Welsh also notes correctly that nonparametric regression is not part of robustness work, at least as he defines it, but it is clearly part of the more general underlying problem: quantifying the information content of modeling assumptions, understanding the sensitivity of final conclusions to model specification—in short, assessing and appropriately dealing with the uncertainty that arises in the modeling process. In this regard there is a worthwhile contrast to be drawn between Bayesian and frequentist robustness, which brings me to my final topic. In making this comparison it helps to consider inference and prediction separately, and to contrast the consequences of modeling uncertainty in each. I will use I. J. Good's convenient device of referring to a person making modeling judgments as You.

Inference

When inference is the goal, as is often the case in science (for example) interest typically focuses on a particular quantity like the mass of the electron or a regression coefficient in a biological model, and other features of the model specification process (heteroscedasticity, the underlying error distribution and so on) are like nuisance parameters. Pulling a single standard model off the shelf, as in Section 1's "do-nothing" strategy, is like plugging in known values for these nuisance parameters, and You may get it wrong. Frequentist robustness work has mainly concerned itself to date with inference, and in that context it involves a search for methods that walk a fine line in relation to the "modeling nuisance parameters"—methods that, within a certain *neighborhood* characterized by certain departures from some target model, are valid (in some consistency sense) and never do badly in precision of estimation no matter what the "right" model might be (within the neighborhood), but which accomplish this by and large without formally trying to adapt to the data at hand by estimating the "model nuisance parameters." The result, when this fine line is walked successfully, is often an *increase* in apparent precision (smaller standard errors) when compared with the conclusions drawn from the off-the-shelf model, and when this occurs it is because the data are able to argue fairly convincingly that some other model in the neighborhood fits better than the off-the-shelf choice. The Bayesian approach to robustness in inference typically arrives at results similar to those of the frequentist approach, but by a different route; the two approaches will be exemplified and contrasted below.

Prediction

When prediction is the goal, as it is so often in decision making (for instance) interest typically focuses on questions like "What will the world average spot price of oil be two years from now if such-and-such a policy is put into place?" The statistical task in such cases is to say what You think would happen in the future under given sets of conditions and attach believable measures of uncertainty to Your predictions, so that an appropriate amount of hedging may be built into the actions taken. Overall predictive uncertainly may be partitioned into three basic sources (Hodges, 1987; Draper, Hodges, Leamer, Morris and Rubin, 1987): *structural or modeling* uncertainty (How will the available predictor variables behave in the future? How are the outcomes of interest linked to the predictors?); *estimation* uncertainty (The model linking outcomes to predictors typically involves unknown parameters; what are reasonable values for these parameters, conditional on the model?); and *prediction* uncertainty (since unexplained stochastic fluctuation is built into most predictive methods even if there is no structural or estimation uncertainty).

Typical practice is to make a single structural choice and assess estimation and prediction uncertainty conditional on that choice, as if the chosen model were "right." A frequent outcome of predictive exercises is the unpleasant realization that Your prediction errors are larger than You expected them to be. Retrospective investigation, when conducted, often reveals that the observations in the previous two sentences are causally connected; uncertainty in the modeling step on the analysis has not been accurately assessed and propagated through to the overall measure of predictive uncertainty. To improve on this outcome it is evidently necessary to remove in some way the conditioning on a single model that has led to the understatement of uncertainty. The result will then often be a *decrease* in apparent precision (larger predictive variances) when compared with the conclusions drawn from a single model, and when this occurs it is because substantial uncertainty exists as to which model is "right" and conditioning on a single model would underpropagate that uncertainty.

Propagation of Model Uncertainty

Two frequentist ways to approach this removal of conditioning on the model in predictive settings have been suggested (Draper, Hodges, Leamer, Morris and Rubin, 1987): *mixture likelihoods*, which will be discussed below, and *bootstrapping the modeling process*, as in Efron and Gong (1983). The approach taken in the latter is to generate a number of bootstrap

replicates of the full data set and to carry out the modeling process separately on each bootstrap data set (as independently as possible; this is another place where expert systems can help). This permits both within-model and between-model uncertainty to be assessed and then combined. Attention will be restricted here to a Bayesian approach to improved propagation of model uncertainty (de Finetti, 1974/1975; Leamer, 1978; Hodges, 1987; Draper, 1987) that, in the taxonomy of Section 1 of the main paper, can be thought of as both a Bayesian form of the "model expansion" strategy and as a version of Bayesian robustness (Smith, 1983). The idea is as follows.

On the predictive scale a model is just a joint probability distribution for the observables, so it becomes possible to conceive of the space of all possible models as the collection of all such distributions. This space may seem hopelessly complicated, and indeed in real life problems it is exceedingly rich, but it is useful to think about no matter how complex it is, and in simple cases one can even visualize it quite clearly and work with it mathematically. When there are no covariates and the outcomes are binary, for example, the set of all probability distributions on an n vector of 0's and 1's is the 2^n-dimensional simplex (Fienberg, 1968; Diaconis, 1977; Meeden, 1986). What people usually refer to as a "model" with unknown parameters is a low-dimensional curve in the space of all possible models, indexed by the parameters (the word "model" in quotes below will refer to such a subspace). Choosing a single "model" corresponds in a Bayesian sense to putting a prior distribution on model space that concentrates all its mass on such a curve. To improve on this You have only to entertain a richer prior distribution on model space and perform an additional integration over this new source of uncertainty. In practice the priors on model space needed for realism in the modeling process often take the form of enlargements of, or *neighborhoods* around, one or more standard "models." Examples of this approach will be given below.

Thus we have two situations, inference and prediction, in which uncertainty in the modeling process arises, and procedures based on conditioning on a single model can be improved on. In inference one goal of robustness work is to diminish the needlessly large "give or takes" associated with estimates arising from a single partly misspecified model; in prediction the goal is the production of more realistic predictive uncertainty assessments, which will often be larger than those implicit in a single modeling choice. In both cases the robustness work undertaken in search of this improvement concerns itself with neighborhoods in model space around certain "models." How is the uncertainty in the modeling process implied by the specification of these neighborhoods treated in the frequentist and Bayesian approaches? How are these neighborhoods specified in practice?

Optimization Versus Integration

In a familiar echo of other Bayesian-frequentist comparisons, the operation performed in frequentist robustness is typically an *optimization* of some kind over neighborhoods in model space, whereas in Bayesian robustness an *integration* over the modeling uncertainty captured by the choice of neighborhood is performed, quite literally treating model specification as a (possibly high dimensional) nuisance parameter. To take two examples of the frequentist optimization process in estimation settings, the idea that led Peter Huber to M-estimates (Huber, 1964, 1981) was to minimize the asymptotic variance of an estimator over the relevant neighborhood subject to Fisher consistency and a bound on asymptotic bias. Hampel's (1974) influence function approach, which often leads to results similar to Huber's (Hampel, Ronchetti, Rousseeuw and Stahel, 1986), involves an "optimal" compromise between efficiency and resistance in the neighborhood, based on a simultaneous minimization of asymptotic variance and asymptotic bias.

Two examples of the Bayesian approach based on integration over modeling uncertainty are worth examining in contrast. First, in perhaps the earliest paper on Bayesian robustness, Box and Tiao (1962) considered the one-sample location problem with unknown mean μ and variance σ^2, and expanded the usual normal theory model $p(x \mid \mu, \sigma^2)$ by adding a third unknown parameter α for kurtosis in such a way that $\alpha = 0$ corresponded to the original, unenlarged model: $p(x \mid \mu, \sigma^2, 0) = p(x \mid \mu, \sigma^2)$. They then integrated out the model uncertainty captured by α to obtain a marginal posterior for μ and σ^2,

$$(1) \quad p(\mu, \sigma^2 \mid x) = \int p(\mu, \sigma^2 \mid x, \alpha) p(\alpha \mid x)\, d\alpha.$$

This approach moves away from a prior distribution on model space that puts all its mass on a single "model" by indexing a somewhat richer subspace of the space of all models parametrically with the one-dimensionally quantity α. As a second example in a predictive setting that is considerably more complicated, Smith and West (1983) used the state-space, time series, multiprocess Kalman filter methodology of Harrison and Stevens (1971, 1976) to build a system that monitors serum creatinine levels in kidney transplant patients to detect changes in the level of functioning of the transplanted organ. When state-space methods are used in a short-term forecasting setting such as this, model uncertainty takes the form of lack of knowledge about which of several states the time

series is in: steady state, single-period outlier, sudden change in level and sudden change in slope, for example. The posterior predictive distribution (ppd) characterizing the next period forecast is obtained as a weighted average of the ppd's conditional on each "model"/state, using as mixing weights the posterior distribution on the "models":

$$p(\text{future} \mid \text{data}) = \int p(\text{future} \mid \text{data}, \text{"model"}) \cdot p(\text{"model"} \mid \text{data})\, d\,\text{"model"}. \tag{2}$$

Equation (2) is the predictive analogue of equation (1)'s integration over model uncertainty (α) in an estimation setting. This setup differs from Box and Tiao's in the form that the prior distribution on model space takes: instead of expanding a single "model" smoothly with a continuous parameter, Smith and West put a discrete prior on model space with mass on the four "models" corresponding to the four states above, so that the integration in (2) is really a summation here. Considerable medical interest focuses in this example on how the posterior distribution on model space $p(\text{"model"} \mid \text{data})$ unfolds over time, because the states correspond directly to clinical stages of the patient's progress; this distribution arises as a natural part of the Bayesian updating calculations.

Whence the Neighborhoods?

One issue ought always to be uppermost in the specification of the neighborhoods in model space (as it was in the Smith and West example)—realism—but in practice two other matters often constrain: tractability and identifiability. In frequentist robustness the optimization problem has been sufficiently difficult that the neighborhoods have been restricted to date by tractability considerations mainly to {point mass on iid models, varying only the underlying error distribution F} (although there are some recent exceptions: see Martin and Yohai (1986) in a time series context, for instance). Embracing the Bayesian approach in all its generality leads to potentially horrendous integration problems over model space, to say nothing of the prospect, at first sight daunting, of putting a prior distribution on the space of all possible models. The first of these obstacles seems to be yielding to steady recent progress with such methods as Laplace approximations (Tierney and Kadane, 1986) and reparameterized Gaussian quadrature (Smith, Skene, Shaw, Naylor and Dransfield, 1985); for some thoughts on why the second task is not nearly as daunting as it might seem, see Draper, Hodges, Leamer, Morris and Rubin (1987). As for identifiability, in inferential settings considerable care is required in specifying a neighborhood around some target "model" so that the parameters of interest retain their meaning in all "models" over which uncertainty is entertained; in his comments Welsh puts the same point somewhat differently: "$\cdots$ without the specification of a precise underlying model, the whole concept of a deviation from that model loses its foundation." But note that identifiability restrictions are completely absent in predictive settings: the common denominator for all "models" is the next observation on the outcome scale of interest, so all models are fair game, and the prior distributions on model space in the Bayesian predictive approach can be very rich. This is a key distinction between inference and prediction.

Which Is Better?

Under what conditions do the Bayesian and frequentist approaches to dealing with model uncertainty lead to similar results, and under what conditions does one approach produce better results than the other, where "good" means both feasible in practice and yielding inferences and predictions that stand up well to out-of-sample validation? The question is similar to others like it in frequentist-Bayesian comparisons (When to maximize a likelihood, when to integrate it?), but more complex because it is not just a few parameters that are at stake but uncertainty in the entire modeling process. Not much seems to be known in general about the answer to this question. An example of a simple inferential setting in which the two approaches are competitive is given by Smith (1983) and Spiegelhalter (1981). Smith considers the familiar one-sample location problem $y_i = \mu + \sigma e_i$, $i = 1, \cdots, n$, but incorporates uncertainty about the underlying distribution of the errors e_i by making them iid from one (and only one) of m known distribution functions $\mathbf{F} = \{F_1, \cdots, F_m\}$ symmetric about 0 and having $\text{Var}(e_i \mid F_j) = 1$. Starting from a prior distribution $p(\mu, \sigma, \mathbf{F})$ and likelihoods $p(y \mid \mu, \sigma, F_j)$, it is straightforward to update to a marginal posterior $p(\mu \mid y)$. Spiegelhalter takes $m = 3$, $\mathbf{F} =$ {normal, double exponential, uniform}, and a "neutral" prior specified by $p(F_j) = \frac{1}{3}$ and $p(\mu, \sigma \mid F_j)$ proportional to $1/\sigma$ to construct a "Bayesian adaptive robust estimator of location," namely the posterior mean arising from this prior, that performs well in comparison with M-estimates and adaptive trimmed means. An alternative frequentist formulation, not investigated by these authors, that more closely mimics the modeling story told by the Bayesian setup is to employ a sort of random-effects two-stage sampling model. Nature chooses a distribution at random from $\{F_1, \cdots, F_m\}$ according to $p(F_j) = p_j$; then, conditional on F_j, generates n iid values e_i; and finally lets the statistician observe $y_i = \mu + \sigma e_i$. This induces a *mixture likelihood*

$$p(y \mid \mu, \sigma) = \sum_{j=1}^{m} p_j p(y \mid \mu, \sigma, F_j). \tag{3}$$

One can now maximize this in μ and σ and do sensitivity analysis on $(p_1, \cdots, p_m)$. Here this should just correspond to the Bayesian inference with flat priors on μ and σ, although I am not aware of any comparative work confirming or denying this.

In the vacuum created by the absence of comparative studies there is plenty of room for speculation. Based on the examples I have looked at, it would seem (1) that the Bayesian and frequentist approaches produce similar results in the fairly simple inferential settings in which the frequentist optimization problem can be solved; (2) that in really complicated inferential settings the optimization problem may be too hard; (3) that the Bayesian integration problem in complicated settings is not to be glossed over, either, but that it seems to be yielding to the recent intensive efforts by Bayesians interested in practical applications; and (4) that, in complicated predictive problems like the oil price forecasting example mentioned at the beginning of this section or the kidney transplant example of Smith and West above, the Bayesian approach seems well suited to the kinds of regions in model space over which uncertainty needs to be propagated to achieve realism in uncertainty assessment, whereas the available frequentist approaches seem much more problematic. (More detail on this position may be found in Draper, Hodges, Leamer, Morris and Rubin, 1987.) Thus these days I favor the Bayesian approach, on grounds of flexibility and comprehensiveness, acknowledging the computational difficulties, but for lack of enough evidence the jury may still be said to be out in the most important arena of all: actual practice. This can be remedied, but it requires that we concentrate on real problems and write case studies for each other on what works and what doesn't, two things that sometimes seem more at a premium than they should in our ostensibly empirical science.

Acknowledgment

I am indebted to John Rolph for remarks that improved the clarity of presentation of these ideas.

ADDITIONAL REFERENCES

ADAMS, J. (1988). Evaluation of regression strategies. Ph.D. thesis (in preparation), Dept. Applied Statistics, Univ. Minnesota.

BOX, G. E. P. (1980). Sampling and Bayes inference in scientific modeling and robustness (with discussion). *J. Roy. Statist. Soc. Ser. A* **143** 383–430.

DE FINETTI, B. (1974/1975). *The Theory of Probability* **1, 2**. Wiley, New York.

DIACONIS, P. (1977). Finite forms of de Finetti's theorem on exchangeability. *Synthese* **36** 271–281.

DRAPER, D. (1987). On exchangeability judgments in predictive modeling and the role of data in statistical research. Comment on "Prediction of future observations in growth curve models," by C. R. Rao. *Statist. Sci.* **2** 454–461.

DRAPER, D., HODGES, J. S., LEAMER, E. E., MORRIS, C. N. and RUBIN, D. B. (1987). A research agenda for assessment and propagation of model uncertainty. The RAND Corporation, N-2683-RC, November 1987.

EFRON, B. and GONG, G. (1983). A leisurely look at the bootstrap, the jackknife, and cross-validation. *Amer. Statist.* **37** 36–48.

FIENBERG, S. E. (1968). The geometry of an $r \times c$ contingency table. *Ann. Math. Statist.* **39** 1186–1190.

GALE, W. A. (1986a). REX review. In *Artificial Intelligence and Statistics* (W. A. Gale, ed.). Addison-Wesley, Reading, Mass.

GALE, W. A. (1986b). Student phase I: A report on work in progress. In *Artificial Intelligence and Statistics* (W. A. Gale, ed.). Addison-Wesley, Reading, Mass.

HAMPEL, F. R. (1974). The influence curve and its role in robust estimation. *J. Amer. Statist. Assoc.* **69** 383–393.

HAMPEL, F. R., RONCHETTI, E. M., ROUSSEEUW, P. J. and STAHEL, W. A. (1986). *Robust Statistics: The Approach Based on Influence Functions*. Wiley, New York.

HARRISON, P. J. and STEVENS, C. F. (1971). A Bayesian approach to short-term forecasting. *Oper. Res. Q.* **22** 341–362.

HARRISON, P. J. and STEVENS, C. F. (1976). Bayesian forecasting (with discussion). *J. Roy. Statist. Soc. Ser. B* **38** 205–247.

HODGES, J. S. (1987). Uncertainty, policy analysis, and statistics (with discussion). *Statist. Sci.* **2** 259–291.

HUBER, P. J. (1964). Robust estimation of a location parameter. *Ann. Math. Statist.* **35** 73–101.

LEAMER, E. E. (1978). *Specification Searches: Ad Hoc Inference with Nonexperimental Data*. Wiley, New York.

MARTIN, R. D. and YOHAI, V. J. (1986). Influence functionals for time series. *Ann. Statist.* **14** 781–818.

MEEDEN, G. (1986). Sufficiency and partitions of the class of all possible discrete distributions. *Amer. Statist.* **40** 42–44.

SMITH, A. F. M. (1983). Bayesian approaches to outliers and robustness. In *Specifying Statistical Models: From Parametric to Non-parametric, Using Bayesian or Non-Bayesian Approaches* (J. P. Florens, M. Mouchart, J. P. Raoult, L. Simar and A. F. M. Smith, eds.). Springer, New York.

SMITH, A. F. M., SKENE, A. M., SHAW, J. E. H., NAYLOR, J. C. and DRANSFIELD, M. (1985). The implementation of the Bayesian paradigm. *Comm. Statist. A—Theory Methods* **14** 1079–1102.

SMITH, A. F. M. and WEST, M. (1983). Monitoring renal transplants: An application of the multiprocess Kalman filter. *Biometrics* **39** 867–878.

SPIEGELHALTER, D. J. (1981). Sampling properties of a finite mixture model. Unpublished manuscript, Dept. Mathematics, Univ. Nottingham.

TIERNEY, L. and KADANE, J. B. (1986). Accurate approximations for posterior moments and marginal densities. *J. Amer. Statist. Assoc.* **81** 82–86.