The Statistician (2002)
51, *Part* 1, *pp.* 1–20

Confessions of a pragmatic statistician

Chris Chatfield
University of Bath, UK

[Received May 2001. Revised November 2001]

Summary. The paper reflects on the author's experience and discusses how statistical theory, sound judgment and knowledge of the context can work together to best advantage when tackling the wide range of statistical problems that can arise in practice. The phrase 'pragmatic statistical inference' is introduced.

Keywords: Bayesian approach; Context; Frequentist approach; Model building; Problem solving; Statistical inference

1. Introduction

After about 35 years in statistics, it may be worthwhile to set down my approach to the subject, both to encourage and guide practitioners and also to counterbalance a literature that can be overly concerned with theoretical matters far removed from the day-to-day concerns of many working statisticians.

I claim no originality for most of what is written, except perhaps in the explicit formalization of what I call *pragmatic statistical inference*. As well as problem solving in general, I also discuss model building and inference from a practical point of view. Some general themes are the importance of context in statistical practice, the key role of data analytic descriptive methods and recognition of the iterative nature of much statistical work. I also emphasize the underlying unity of different statistical approaches, because I subscribe to the growing consensus that different forms of inference are appropriate in different practical situations, and that the frequentist and Bayesian perspectives are complementary in applied statistical work. I begin with an example—my most recent piece of consulting at the time of writing.

2. Prologue example

A large manufacturing company sought my advice on the best way of forecasting future values of a particular time series. A meeting was arranged with the person who was responsible for this task. It transpired that the company did not employ a statistician, even though they employed several thousand people, collected large amounts of data and had numerous quality control, reliability, experimental design and planning problems. My client said that this was not for want of trying, but because they had been unable to recruit a suitable person. Whether or not this was true, it is sadly typical that many large companies do not employ a statistician and rely on engineers who 'know a bit of statistics', or on external consultants who may not have the requisite inside knowledge of the company's workings.

Address for correspondence: Chris Chatfield, Department of Mathematical Sciences, University of Bath, Claverton Down, Bath, BA2 7AY, UK.
E-mail: cc@maths.bath.ac.uk

 0039–0526/02/51001

My client showed me the data, which consisted of just 11 quarterly observations—a short series. The observed variable was described as the 'number of services and part orders in one quarter' for a particular model of a type of electrical appliance. A 'part order' was explained as involving the supply of one or more spare part replacements for a single machine. The data have been coded to preserve confidentiality and the scaled data are 1579, 1924, 1868, 2068, 2165, 1754, 2265, 2288, 2526, 2568 and 2722, starting in 1998, quarter 1. The series is plotted in Fig. 1 and shows an upward trend, which is approximately linear. There is no apparent seasonality. The value in 1999, quarter 2, seems unusually low, but my client could provide no explanation for this. The series is rather short to use any sort of sophisticated forecasting procedure. The company had heard of exponential smoothing and wondered whether that would be an appropriate method to use.

My first impression was that the series was so short that any simple univariate projection method would be reasonable, provided that it allowed for trend. However, before committing myself, I asked some questions to obtain further background information to understand the context better. In particular, I needed to find out why a forecast was required and what would be done with it. I soon ascertained the key information that production of the particular brand was shortly to be stopped so that a new model could be introduced. Forecasts were desired to plan the run-down of work on the brand. Sales through shops would continue for a while but would be phased out within 3–5 months. Thus the upward trend in the data would soon cease. Then, almost by chance, I discovered the second crucial item of information, namely that the observed variable was not as originally described but was actually the 'number of services and part orders *for machines less than 2 years old*'. This is because the company is only directly involved in servicing and supplying spare parts when a machine is covered by warranty. The

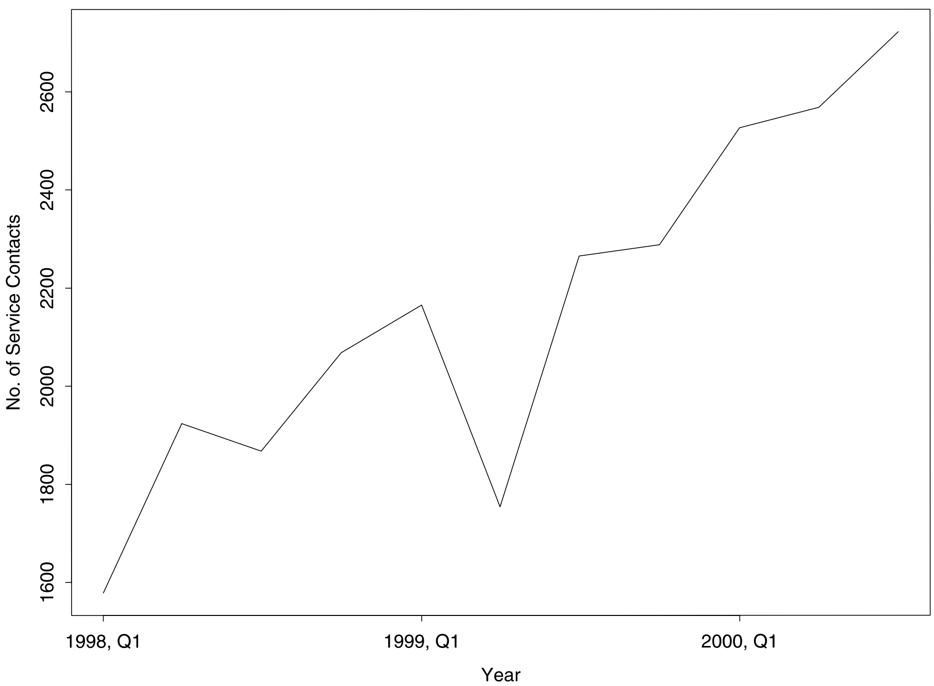

Fig. 1. Number of services required for a particular type of consumer durable in successive quarters from 1998, quarter 1, to 2000, quarter 3

latter expires when a machine is 2 years old. This means that the observed series will soon start decreasing and should die out within $2\frac{1}{2}$ years!

Using Fig. 1 by itself, it seems reasonable to use something like Holt's linear trend version of exponential smoothing. However, we now know that this would be completely inappropriate! The contextual information tells us that the value $2\frac{1}{2}$ years ahead will be 0, and so we can forecast with certainty! In the interim, I suggested that some sort of damping procedure, taking account of the known numbers of past sales, would be appropriate. This would best be carried out in house by people who were familiar with the problem. Simple extrapolation is totally inappropriate, and my expertise as a time series 'expert' appears wasted! In fact, it was my more general skills as a problem solver that proved useful.

3. What is statistics?

I do not claim that the prologue example is a 'typical statistical problem'. There is (fortunately?) no such thing. However, as often happens at the end of a piece of consultancy, I wondered whether this example could teach us anything about problem solving in general. I would say that some elements of the above problem feature in many statistical tasks—the importance of context, the messiness of the data and the lack of clarity in objectives.

Before we go any further, it may be helpful to attempt to specify what is meant by the subject entitled 'statistics', though this is clearly difficult to do in a way that will please everyone. A typical dictionary definition is 'the science of collecting, classifying and using numerical facts'. I would vary this somewhat to the 'science of collecting, analysing and interpreting data to describe phenomena and to answer specific questions'. A radically different definition is given by Lindley (2000) as 'the study of uncertainty', and this enabled Lindley to concentrate on the measurement of probability as the prime activity of statisticians. However, although I might accept an alternative definition of statistics as 'the pursuit of knowledge in the presence of uncertainty', I cannot accept an approach that removes data from centre stage or which makes the subject primarily deductive, rather than inductive.

In practice, the scope of statistical problem solving is so wide that any definition can look suspect. It involves interaction with the real world and hence with many different kinds of uncertainty—certainly a longer list than that considered by Lindley (2000). There may, for example, be uncertainty about what the question is, uncertainty about the quality and meaning of the data and uncertainty about the form of relationships between variables, not to mention the usual well-documented sources of statistical uncertainty such as random variation and measurement errors. Thus my paper is not concerned with idealized inference but, like Rubin (1984), with the

> 'activities of applied statisticians in the real world who are subject to constraints of finite resources, many problems to examine, and mixed expertise of consumers'.

4. Tackling real life statistical problems

This section briefly revisits some key aspects of problem solving. *Good statistical practice* (e.g. Deming (1965), Preece (1987) and Chatfield (1995a)) is an essential prerequisite to any application of theory, and any competent statistician, whatever his or her philosophy, will start by asking questions, clarifying objectives, obtaining background information to tackle the problem 'in context', collecting 'good' relevant data, assessing the quality and structure of the data, deciding what to do about outliers and missing observations, and so on.

4.1. Context

Context is recognized as crucial in scientific investigations. For example, a doctor treating a patient with a high temperature and shivering might diagnose influenza in Europe but malaria in West Africa. Context is equally crucial in statistics but is rarely recognized in theoretical work. In particular, there have been few attempts to provide systematic guidelines on how to *formulate the problem* (but see Hand (1994)) even though this is the essential first step in any statistical investigation. I spend an increasing proportion of my time in problem formulation as opposed to data analysis and note the importance of asking (many) questions, of finding out about any prior knowledge and looking at any prior data, and ensuring that any necessary background information is understood. Some tips, and further references, on consulting and collaboration are given by Chatfield (1995a), chapter 10, and Derr (2000). From bitter experience, I particularly advise against consulting by telephone or electronic mail, where one cannot see the data.

In statistical inference, the use of contextual information raises rather different issues. The Bayesian may be able to incorporate some contextual information via priors, whereas the frequentist may, for example, be able to incorporate external constraints, such as requiring a particular coefficient to be positive. However, such matters tend to become wrapped up in the inferential apparatus and are unrelated to more important questions, such as what is the objective and what is a sensible model given the context? Many examples could be given (e.g. the prologue example and Piepel and Redgate (1998)) where contextual issues need to be centre stage, and hence be an explicit ingredient of statistical inference.

4.2. Collecting the data

There is a large literature on collecting 'good' data, some of which recognizes the many practical difficulties involved. It is customary to try to collect *random* samples and statistical inference often proceeds by pretending that the data are random even when they are clearly not! The fact that some data sets are not completely random may, or may not, matter. Randomization is the means, but representativeness is the goal.

The term *quasirandom* may be used to describe data that are as near to random as the practical conditions allow. In the non-experimental sciences, genuine random samples are rare but are generally analysed as if they are random. Sample surveys and observational studies can suffer greatly from non-response (as well as misrecording). Even scientific experiments are usually less random than they may at first sight appear. For example the scientist who takes a 'random' sample in a particular laboratory on a particular day using a particular batch of material may have difficulty specifying the population from which this is a random sample. Here, Deming's distinction between enumerative and analytic samples (e.g. Hahn and Meeker (1993)) is important.

The statistical literature is rather sparse on the question of collecting confirmatory samples. Statisticians tend to be involved with analysing one-off samples, which can have serious limitations. All scientific findings need to be checked, e.g. by repeating an experiment in a different laboratory on a different day. Thus statisticians, if they wish to be regarded as good scientists (Box, 1990), need to give more emphasis to *collecting more than one data set* wherever possible, as that is the route to scientifically valid and generalizable results. However, there is little advice in the literature on how to collect follow-up samples and traditional inference has rather little to say when there are a series of data sets, rather than a single set of data. There has been some recent work on *meta-analysis* that looks at between-study heterogeneity, as well as the findings from single studies, while Ehrenberg's (1975) work on seeking empirical generalizations provides an alternative route. More generally the question whether 'statistical method' is, or should be, the same thing as 'scientific method' is not a topic that I can fruitfully pursue here—see for example Nelder (1986, 1999) and MacKay and Oldford (2000).

4.3. Analysing the data

Having collected some 'good' data, the next stage is to analyse them. The first step is usually to carry out an *initial data analysis* (IDA)—see Chatfield (1985) and Chatfield (1995a), especially chapter 6. This includes assessing the quality and structure of the data, calculating summary statistics and plotting appropriate graphs. In my experience, deciding what to do about peculiarities of the data, such as outliers and missing observations, is often more important than anything else. Indeed many readers will know that I am an ardent proponent of a careful IDA, which is vital to understand the data and is sometimes all that is needed if, for example, the results are very clear cut or reveal such poor data quality that a more sophisticated model-based analysis cannot be justified. Where appropriate, IDA is also vital in building a model—see Section 5—and choosing an appropriate method of analysis.

An instructive example of the importance of IDA is the reanalysis of some data from a field trial comparing 10 varieties of barley by Cleveland (1993), chapters 1 and 6. This shows fairly conclusively that the 1931 and 1932 data for a variety called velvet at a site called Grand Rapids were erroneously interchanged. Sadly, many previous analyses failed to spot this.

4.4. Avoiding trouble

A key skill for the practitioner is the knack of avoiding trouble, as I (just) managed to do in the prologue example. It is far more important to avoid trouble than to achieve some sort of notional optimality. Some ways of doing this, some things to look out for and a variety of examples are given by Chatfield (1991). I have made many mistakes over the years but have (fortunately) managed to keep most of them quiet or to correct them before much harm was done. Like others, I naturally avoid broadcasting such errors, even though we probably learn more from our failures than from our successes.

4.5. Communicating the results

At the University of Bath, we send many of our undergraduates out for a year on industrial placement as part of their degree course. In my experience, the most important skill that they learn during this time is the ability to communicate with their work colleagues, both statisticians and non-statisticians. Report writing is a key skill, while giving good oral presentations is also vital—see Chatfield (1995a), chapter 11. I particularly recommend asking someone to comment on a draft report, while overhead transparencies should (obviously) be legible at the back of a lecture room. I find it amazing that so many experienced lecturers still ignore the latter guideline.

5. Model building

Most statistical problems require some sort of *model* to be constructed. This involves formulating a sensible model, fitting it to the data, checking the residuals and revising the model if that proves necessary. The phrase 'statistical inference' is sometimes used synonymously with the *estimation* of model parameters (i.e. fitting the model). However, modern computer packages enable the analyst to fit most classes of model very easily, and so the real problem is deciding what model to fit. I have argued elsewhere (Chatfield, 1995b) that modern inference should include the whole model-building process, including model formulation and model criticism.

The statistician needs to 'match the model to the data' (Nelder, 1986) rather than just to fit it, and it is the model formulation stage that is usually the most difficult.

Although some general guidance can be given about model formulation, it is often more of an art than a science and is heavily context dependent. The practising statistician needs to formulate a model based on background theory, on previous empirical evidence and/or on a preliminary look at the data. Exactly how this is done will vary widely and will usually involve subjective judgment, but there is, for example, a world of difference between a plausible guess (which all statisticians have to make from time to time) and a model based on substantial prior empirical knowledge. Given that model formulation is a key aspect of problem solving, it is unfortunate that it is often skated over when considering inferential questions. For example a theorist might pose the problem 'Let $X_1, X_2, \ldots, X_n$ be independently and identically distributed $N(\mu, \sigma^2)$: what is the best way to estimate μ?' without saying why the data are thought to be independent normal random variables with a constant variance. A man in daily muddy contact with field experiments will not have much faith in a model that assumes *independent* normal errors (Box, 1976).

The word 'model' normally invokes ideas of simplification and idealization (Cox, 1995) but there is always some tension between aiming for a parsimonious model with few variables and wanting to include all conceivably relevant variables. Related to this is the distinction between a model seen as a useful approximation and a model seen as the 'truth'. However much data we have, and however careful the modelling procedure, we can never be certain that the true model has been found. Indeed most statisticians would agree that there is no such thing as a 'true' model (e.g. Durbin (1987), page 178). *Model uncertainty* (Draper, 1995a; Chatfield, 1995b) is a fact of life and arguably the most serious component of uncertainty. However, it is also the component about which least is known and which is typically ignored in statistical inference. Yet inference is crucially dependent on the model assumed, and Tsay (1993) goes as far as saying that 'Since all statistical models are wrong, the maximum likelihood principle does not apply'!

Taking time series analysis as an example, the statistician who decides (rightly or wrongly) to fit a model from the autoregressive integrated moving average (ARIMA) class will typically try a range of plausible models (after looking at diagnostic tools such as the correlogram), pick the best fitting model and then make inferences and forecasts on the assumption that the model selected is true. Such inferences fail to take account of uncertainty about the structure of the model. A Bayesian may prefer to use a dynamic linear model (e.g. Pole *et al.* (1994)) as it arguably provides a more natural way of quantifying prior information about physically meaningful quantities such as trend. Moreover, estimates of model parameters can be readily updated by using the Kalman filter. However, these advantages should not be allowed to obscure the many assumptions that have to be made in setting up the model. Model uncertainty is still present.

A key component of much model building is the use of IDA—see Section 4.3. Although data analytic in character, IDA is a vital preliminary to model building and inference. However, the extent to which the analyst can and should look at the data before formulating and fitting a model is controversial. For the Bayesian, there is a real danger in looking at the data before choosing the 'prior' since one ends up looking at the data twice. But if this is not done, and the data contain some unexpected feature that has not been allowed for in the prior, then a zero prior leads to a zero posterior regardless of what the data are saying (sometimes called Cromwell's law).

The frequentist also faces problems in deciding whether to incorporate IDA when formulating a model. When a model is entertained *after* looking at the data, it can be shown (Chatfield, 1995b) that least squares theory no longer applies, estimates are biased and diagnostic checks, using the *same* data previously used to formulate a model, are tainted. In particular, my experience in time series forecasting suggests that estimates of uncertainty, based on assuming that the best-fit model is true, are woefully inadequate in that the out-of-sample forecast accuracy is typically twice as bad as expected from the within-sample fit. This emphasizes the importance of checking any model with new data to guard against overoptimism resulting from using a best-fit model, and it explains my preference for simple models. The use of more complicated models may lead to a better (within-sample) fit but may give worse out-of-sample forecasts.

It is clear from the above that statisticians do not usually fit a single model, but rather try several, either all at the same time or in a sequential way. The first guess at a model will be modified in response to diagnostic checks or to the collection of more data and will typically be improved in an iterative–interactive way. The circular (closed loop) process for real life model building contrasts with the (open loop) process of traditional inference where a model is assumed *a priori*. The merits of iteration in statistical work are propounded, for example, by the work of Deming (e.g. Deming (1986)) and Box (e.g. Box (1994)). Sadly, the iterative nature of modelling is often suppressed in published work, thereby giving a misleading impression.

6. Inference

To use a statistical model, the analyst will estimate model parameters and then make appropriate inferences. Many experienced statisticians adopt a general approach that can best be described as 'pragmatic', and this is a way ahead that I also recommend. A 'pragmatist' is someone who chooses a sensible, practicable course of action, and the relevant dictionary definition of the adjective pragmatic is 'advocating behaviour that is dictated more by practical consequences than by theory'. (The school of philosophy called the pragmatic method (e.g. James (1975)) is concerned with studying ideas by tracing their respective practical consequences.) This adjective has already been used to describe what statisticians do (e.g. Bartholomew (1995) and Spiegelhalter *et al*. (1994), page 378). Such an approach is flexible, undogmatic and eclectic in character. Although implicit in some published work, I want to formalize it as far as is possible (which is not very far) and to give it a distinctive name. First, we briefly revisit traditional inference.

6.1. *Traditional statistical inference*

Traditional inference has two main ingredients:

(a) a *model*—$\mathcal{M}$—a *prespecified* family of parameter-indexed probability models;
(b) a *set of data*—$\mathcal{D}$—assumed to be a random sample from some known population where $\mathcal{M}$ applies.

The problem is to estimate and/or test hypotheses about the parameters of $\mathcal{M}$ using $\mathcal{D}$. Various approaches may be taken, and the Bayesian approach further assumes prior knowledge about the model parameters or about particular hypotheses.

Although of theoretical interest, traditional inference forms a rather small part of what statisticians actually do. The theory is primarily intended to be applied to (smallish) samples collected from well-defined populations by using a proper randomized survey or experimental design when there is a particular model in mind. This theory is 'nice' to teach for the more mathematically inclined and can be formalized in such a way that it sounds 'objective'. Unfortunately, it diverts attention from the wealth of alternatives that statisticians may face in practice.

Much statistical work involves rather little inference (Smith, 1986). As well as IDA, non-inferential techniques include data analytic methods, such as cluster analysis and multidimensional scaling, which need not depend on a probability model and are designed to reveal structure in data. When used as a precursor to inference, they are typically ignored in any subsequent theoretical calculations. Bartholomew (1995) distinguished four types of statistical procedures, of which three do not readily fit into a traditional inference scenario, namely

(a) the analysis of census and large sample data,
(b) the analysis of complex dynamic stochastic systems (rather than static populations) and
(c) the assessment of uncertainty in wider areas of public concern.

Some statistical problems (e.g. regression with observational data and time series analysis) are typically tackled by using traditional inferential apparatus, despite doubts about whether the sample really is random and despite not knowing the model *a priori*.

Modern inference is generally interpreted more broadly than traditional inference with more emphasis on *model selection* and *model criticism*. However, it still fails to cover many of the activities of working statisticians.

6.2. Pragmatic statistical inference

I suggest that pragmatic inference may be thought of as having three main ingredients:

(a) the *context*—$\mathcal{C}$—including objectives and known background information;
(b) the *model*—$\mathcal{M}$—(or a series of models, $\mathcal{M}_1, \mathcal{M}_2, \ldots$) entertained *before*, *during or after* looking at the data;
(c) the *data*—$\mathcal{D}$—(or a series of data sets $\mathcal{D}_1, \mathcal{D}_2, \ldots$) are collected in a random or *quasi-random* way from some known population (and the populations will not be exactly the same for each $\mathcal{D}_i$ when a series of data sets is collected). The adjective quasirandom was introduced in Section 4.2.

This differs from traditional inference in its explicit inclusion of the context, and hence of the problem to be solved, and the recognition that there may be a series of models and data sets, perhaps examined in an iterative way. The problem may involve model selection as well as parameter estimation and hypothesis testing, but the last two are usually a means to an end, and not the end itself. Rather the statistician will want to use the fitted model, e.g. to make predictions, to make decisions or choices, or simply to provide a succinct description. Note that no restriction is placed on the inferential approach to be used, and the analyst is free to use ideas from more than one approach even within the same problem.

I realize that this approach is already used implicitly by many or most statisticians. Even the most dyed in the wool frequentist or Bayesian will recognize the importance of context and, if honest, will admit to 'peeping' at the data before formulating a model. Thus in one sense the message of the paper is not new. However, I want to go further than merely suggesting that the statistician selects the most appropriate inferential approach from a list of alternatives by explicitly recognizing

(a) the paramount importance of context,
(b) the key role that is played by pre-inferential descriptive and data analytic methods,
(c) the iterative–interactive nature of statistical modelling,
(d) that statisticians may use ideas from different schools of inference, not only in different problems but perhaps even within the same problem, and
(e) that statisticians need to be flexible and to make good use of that hard-to-define, but priceless, commodity called 'common sense'.

The formalism given above for pragmatic inference emphasizes the difficulties that are involved in trying to develop any corresponding theory. As Mallows (1998) says, it is not clear that there could be a 'theory of applied statistics', since 'by definition, context and theory are different things'. The context is impossible to formalize in a general way, and it may even be the case, for example, that the objectives change iteratively as the study progresses. The model may also develop iteratively as more data are collected, and this also makes a general theory elusive. There have been several attempts to formalize statistical practice (e.g. Gale (1986) and Oldford (1990)), but real statistical analysis includes components that cannot be completely formalized, given the complex process by which an experienced analyst examines a data set and formulates a model.

Of course, one attraction of traditional inferential procedures is that we can develop certain 'road test' properties so that a procedure is known to be 'good' under given assumptions. More specifically, we would like to be sure that a procedure is not systematically misleading when applied repeatedly (Cox, 1997). Unfortunately, it looks to be impossible to develop notions of 'goodness' to the same extent for pragmatic inference. However, it may be possible to define 'best' in alternative ways or at least to draw on traditional inference results in regard to *part* of some pragmatic procedure. It should in any case be remembered that optimality results for traditional inference may depend on dubious assumptions. What matters in practice is not whether a model or procedure is 'best' but whether it is 'sufficiently good' in the given context.

One important method for checking a model that does not depend on traditional theory is the *predictive validation of out-of-sample results*. This is routine in time series forecasting but can readily be used in many other applications. The use of calibration is an area where Bayesian and frequentist ideas can both be used within the same problem. For example, Rubin (1984) recommended the 'calibration of Bayesian probabilities to the frequencies of actual events', while predictive calibration, with its inherent frequency character, can be used to validate or falsify a model formulated in a Bayesian way (e.g. Draper (1995a) and Draper (1995b), page 142).

6.3. Why an eclectic approach?

This subsection amplifies my wish to promote further the virtues of a flexible approach to inference. The word 'eclectic' means to be willing 'to select from each school of thought such doctrines as please him' (*Concise Oxford Dictionary*), and this description has already been applied to inference (Cox, 1978, 1995; Durbin, 1987; Moore, 1998). Some approaches to traditional statistical inference have rather restricted application (e.g. decision theory) or are not widely understood (e.g. fiducial inference). Thus I concentrate my remarks on seeking common ground between the frequentist and Bayesian approaches, both of which are widely used.

Historically there has been a wide gulf between the two schools of thought. The sharp exchanges in the discussion of Efron (1986) provide an example from fairly recent years. Some frequentists continue to be highly sceptical of methods involving subjective beliefs (e.g. 'Bayesianism may seal the doom of applied statistics as a respected profession' (Hamaker, 1977)), while avowed Bayesians may still claim, for example, that their approach is the only one to be 'coherent' (e.g. 'Every statistician would be a Bayesian if he took the trouble to read the literature thoroughly' (Lindley, 1986)). Coherency aims to ensure that the statistician's actions are internally consistent, but at the price of requiring that the analyst is able and willing to express consistent preferences about different courses of action, even when there are different kinds of uncertainty, perhaps imprecisely known. Whether such superrational beings exist is open to doubt! Despite this, there is no doubt that Bayesian ideas are useful. For example Pocock (1994) says that

> 'frequentists are prone to become Bayesians when faced with practical research questions, especially in clinical trials, which do not readily lend themselves to frequentist solutions',

and goes on to suggest the title 'closet Bayesian' for statisticians

> 'who adopt strategies in study design and data interpretation which include concepts of prior belief, but who do not express them in a formal Bayesian framework'.

Likewise Cox (1986) says that he has rarely, if ever, found it feasible to incorporate prior knowledge via the Bayesian formalism, but that 'it seems crucial that such prior knowledge is used, if extremes of empiricism are to be avoided'.

Many statisticians have begun to question whether such disagreements are as real as the more extreme protagonists appear to suggest, or whether practising statisticians can, in fact, gain much from both schools of thought. In particular, there are many common features to the different inferential approaches, including the importance of clarifying the problem, the importance of collecting good data, the desirability of carrying out a preliminary data examination before starting a more formal statistical analysis and the use of the likelihood function. We agree about far more than we disagree.

Given the wide variety of practical problems that require solution, it seems unnecessarily limiting to restrict attention to one particular inferential approach. By being flexible, the statistician will not 'fall into the rigid fundamentalisms which are so destructive and, ultimately, self-defeating' (Bartholomew, 1995). Views similar to my own have already been expressed by many others such as Smith (1994) who said (page 17) that 'there is no single right method of inference', and that 'different types of inference are relevant for different problems', by Lindsey (1999), page 16, who said that 'we should not attempt to apply one mode of inference in all contexts', and by Box (1983) and Box (1994), page 217, who used the word 'ecumenism' in his earlier paper, deriving from its use in the search for worldwide (Christian) unity. Barnett (1982), page 285, summarized the approach of George Barnard as being

> 'a catholic one—stressing the fact that no single concept or attitude is sufficient to cover the range of different needs in statistical inference',

and Barnard (1996) himself said that we need to be familiar with different approaches so that 'an appropriate choice ... can be made to problems as they arise'. Cox (1986) said 'A single approach is not viable', whereas Rice (1995), page 597, remarked

> 'the subject matter under investigation and the role that statistics plays in the investigation should effectively determine whether it is more appropriate for the user of statistics to take a Bayesian, decision-theoretic, frequentist or purely data-analytic point of view, or some combination of these'.

While agreeing with Smith (1994), page 18, that we should 'enjoy the diversity of our subject', I also think that more effort is needed to unify the discipline. (Lindley (1997) is unusual is saying that attempts at reconciliation should be resisted, but the rejoinder of Berger *et al.* (1997), page 156, is persuasive.) As Efron (1998) said, 'the world of applied statistics seems to need an effective compromise between Bayesian and frequentist ideas'. There have been several formal attempts at this (e.g. Good (1992)), and I want to take this one stage further with the explicit recognition of *pragmatic inference*. One consequence is that there is no need for statisticians to label themselves as 'Bayesian' or 'frequentist' or 'unsure'. Rather we can all simply use the label 'statistician'.

Morris (1986) referred to a 'frequency-Bayes compromise' over 15 years ago, whereas Spiegelhalter *et al.* (1994) referred to *Bayesian–frequentist rules*. It is now not uncommon to see both words in the title of a paper (e.g. Berger *et al.* (1994, 1997) and Bernard (1996)). Further evidence of the coming together of the different approaches is provided by several discussants of

Spiegelhalter *et al.* (1994) including A. P. Grieve, J. A. Lewis, D. G. Altman and F. E. Harrell, Jr. (However, A. P. Dempster, in a rare note of discord, recommends the removal of 'frequentist crutches' and relying solely on Bayes, thereby demonstrating that there is still life in old controversies!)

It is inappropriate here to reiterate the perceived advantages and disadvantages of different inferential approaches (see, for example, Cox (1978) for an early summary, Barnett (1982) for a detailed exposition and Pocock (1994)), especially those of a mathematical nature, as we wish to focus on the complementary qualities of different approaches as applied to real life problem solving. Many problems lead to very similar solutions using different inferential approaches, and it is rather rare for different approaches to give qualitatively different results. However, when prior information is sharp and non-controversial, then frequentists methods could be seriously disadvantaged. Moreover the different approaches will generally *appear* to be different, especially in the way that probability is interpreted.

In the notation of Section 6.2, the frequentist approach enables the calculation of quantities of the form prob($\mathcal{D}|\mathcal{M}$) but not of prob($\mathcal{M}|\mathcal{D}$). If the latter is what is really wanted, then it can be obtained with a Bayesian approach, but only by specifying prior beliefs about the model. That may be difficult, especially if the appropriate model was not even thought of before looking at the data! (Note that the theory for the traditional frequentist approach is also strictly inapplicable when a model is entertained after looking at the data.) Moreover Bayesian computations can be demanding even in this computer age. Thus there is little doubt in my mind that we can never find an inferential approach that covers all practical situations.

The pragmatic statistician, who is unconstrained by philosophical dogma, will generally be able to choose an appropriate approach for a given problem, given the context and objectives. To take just two examples, consider the Dirichlet model of consumer purchasing behaviour (e.g. Goodhardt *et al.* (1984)), which has been built up over many years based on many data sets. The presence of much prior information might appear to suggest a Bayesian approach, but when a new data set is collected, perhaps on a new product or under new marketing conditions, the question is not 'Is the Dirichlet model true' but 'do the new data conform to the well-established model?'. Here prob($\mathcal{D}|\mathcal{M}$) may be of interest, suggesting a frequentist approach. Alternatively the analyst may simply make a judgment (neither frequentist nor Bayesian) about whether the deviations from the model are of any practical importance. Suppose instead that some econometric model has been suggested by economic theory, and that we are interested in the truth or otherwise of some particular hypothesis—$\mathcal{H}$—perhaps specifying a subclass of a more general econometric model. Now the quantity prob($\mathcal{H}|\mathcal{D}$) seems of prime interest. This suggests a Bayesian approach, provided that widely acceptable priors can be agreed.

6.4. Point estimation

The need for an eclectic approach is nowhere more important than when computing point estimates of model parameters. Proponents of a particular approach like to posit problems where alternative approaches fail. Although there is much to learn from such problems, it is not that method A is 'good' and method B 'bad'; it is that different methods suit different problems. As one example, Samaniego and Reneau (1994) present a paper that, according to the title, is aimed at reconciling the Bayesian and frequentist approaches. In reality, it is primarily a contest between the two approaches for a particular empirical situation, namely estimating the proportion of first words, containing six or more letters, on the 758 pages of a particular book. This shows (surprise, surprise!) that the Bayesian approach is better when good prior information is supplied, but that the frequentist approach is better when the analyst is ignorant, or worse misinformed, about the prior. Samaniego and Reneau (1994) concluded that

> 'every statistician ... should sometimes be a Bayesian and sometimes be a frequentist. Knowing when to be one or the other remains a tricky question.'

The only clear message is that the context is crucial. To clarify the choice of a suitable approach, the analyst should ask such questions as the following.

(a) Why is a parameter estimate required? How will it be used to solve the problem?
(b) What subject-matter knowledge is available? What prior information, if any, is available about the parameter?
(c) How confident are we that the model entertained is a (very) good approximation?
(d) How were the data collected? Are they a genuine random sample from the target population?
(e) Do we still believe the model after seeing the data? Does conditioning on the observed data make some inferences silly?

6.5. More on modelling

Parameter estimation, as discussed above, is one highly specialized form of statistical activity where theory generally assumes knowledge of a model. In practice, it is the construction of the latter that is usually the key step, and issues of model formulation are usually more important than a discussion of different modes of inference (Cox, 1995). However, model formulation is not easy—see Section 5—and all statisticians need to develop skills for doing this. The Bayesian needs an additional skill for *eliciting sensible priors*—a rather neglected and difficult topic. For example, Pocock (1994) said

> 'it is difficult to specify a meaningful prior that will be accepted by all interested parties. Incorporating clinical opinion is ... heavily dependent on who is asked.'

And Evans (1994) said that clinicians' beliefs are often unsupported by evidence. This being so, the use of a client's subjective prior may be dangerous given that one aim of a statistical analysis is to protect the client from seeing merely what he would like to see. However, Bayesians will say that some clients welcome the opportunity to try to quantify their prior beliefs in regard to physically meaningful quantities.

This raises the much asked question whether the model-building process can be made 'objective'. Some scientists prefer frequentist procedures because they think that they are 'objective' but then apply them in a completely mindless way (Wang (1993), epilogue) that ignores the context and may make a nonsense of them. In contrast, Bayesian inference is sometimes criticized for being 'too subjective'. The implication is that the frequentist approach is more objective because it separates factual evidence from prior opinions (or prejudices) so that two analysts (using the same model) should derive the same answer. In practice the frequentist approach, and indeed any statistical problem solving, involves the application of subjective judgment, both to clarify the objectives, to assess a suitable cost function and to formulate a model. The pragmatic Bayesian will 'scratch down' a 'prior' in much the same way that a pragmatic frequentist will scratch down a model. (Note that the description 'objective prior' is sometimes used for a uniform or ignorant prior (Jeffreys, 1961), but this prior is anything but objective.) Neither should be afraid to go back to modify their assumptions if subsequent inferences appear qualitatively unsound or counter-intuitive.

The point is that statistical analysis is never undertaken in a knowledge-free situation. Rather the ideas that led to the study being undertaken will influence the questions that are asked, the formulation of the model and the form of analysis. The tools used, and the interpretations made, rely strongly on the good sense, experience and subjective judgment of the analyst, which rely in turn on the answers to questions such as whether randomization has been used and whether outliers are present. For example, if randomization has not been used, then it is unwise to try to build a model and a simple descriptive analysis is indicated. Alternatively if randomization *has* been used, then the IDA will suggest what model assumptions are reasonable and whether a parametric or nonparametric test is appropriate. Although inference may appear to have a low profile, a further consideration suggests that the ideas of pragmatic inference actually pervade all that the statistician does. What is clear is that a statistical analysis can never be entirely objective, and so I suggest that a subjective–objective dichotomy is unhelpful.

6.6. Probability

The concept of *probability* plays a central role in statistical inference and I suggest that the pragmatic statistician also needs an eclectic attitude here. To relate probability to real life problems, we must take a view on how it should be understood, and establish a mapping between the real world and the mathematical construct. The three main ways of interpreting probability are

(a) the classical equally likely case,
(b) the frequentist approach for phenomena that are inherently repeatable under idealized identical conditions and
(c) the personal (or subjective or Bayesian) approach that relates probability to the odds at which you would be willing to bet that a particular outcome will arise.

These interpretations are complementary to each other and appropriate to different situations. Thus it would be silly to try to apply frequentist ideas to an inherently unrepeatable event like 'The UK Labour Party will be re-elected with a majority at the next election' (written in April 2001), and it may be unwise to try to attach personal probabilities to events associated with a physical mechanism such as a particular form of radioactive decay.

Even apparently simple situations can give rise to discussion. Suppose that a fair penny has been tossed but you are not allowed to see the result. Is it sensible to say that the probability of a head is $\frac{1}{2}$, even though the event in question has already occurred and there is no random variable in sight? I would be happy to regard this as a subjective probability with a long run frequency justification, in much the same way that I interpret confidence intervals. Consumers of statistical results will typically interpret confidence intervals, rightly or wrongly, as subjective probability statements about the likely values of unknown parameters (Rubin (1984), section 2.3). Thus, for a 95% confidence interval (say), 95% expresses the odds at which the analyst is willing to bet on what will happen in the particular case. Christensen (1995) disliked the long run frequentist interpretation but also harshly described the subjective interpretation as a 'blatant attempt at making Bayesian omelettes without breaking Bayesian eggs or even admitting to making omelettes'—a lovely expression! Unfortunately, his proposed alternative, namely 'the collection of parameter values that are consistent with the data as determined by an α level test', seems even less persuasive. Thus the pragmatic statistician may find it helpful to use both (frequentist) confidence intervals and Bayesian posterior intervals according to the context, while making clear the different assumptions and uncertainty logics on which they are based.

More generally, many statisticians, including me, are happy to apply different types of probability even within the same problem, e.g. by using frequentist ideas to check Bayesian predictions. This needs to be clearly stated, and evaluated.

7. The effect of computational advances

One other important influence, which has been largely ignored in theoretical development, is the effect of the astonishing computational advances that have been made. Even a computing dinosaur like me can, for example, readily fit ARIMA models to time series data. At the start of my career, I was reliant on punched cards or even paper tape to fit one model at a time with much difficulty. Although frustrating, this scenario did make the statistician think very carefully about what should be done. Nowadays, there is an understandable tendency to fit models without much thought. The inevitable outcome is that most statisticians routinely fit many models to the same data. This is sometimes called *data dredging* or (statistical) *data mining* (DM). A better term might be *data-driven inference*. The effect that this has on modelling is barely understood (Chatfield, 1995b). When we use a best-fit model, having looked at dozens of candidate models, what will be the properties of the resulting inferences and forecasts? Most of us still behave as if the best-fit model is the one that we thought of before looking at the data.

Another important area is that of DM (e.g. Hand (1998)), where the term is used in its computing science sense to mean the extraction of information from very large data sets, which may be 'messy'. An alternative term is *knowledge discovery in databases*. It is 'open to debate whether Statistics as a field should embrace DM as a subdiscipline or leave it to Computer Scientists' (Friedman, 2001). My limited experience suggests that we should not leave DM entirely to computer scientists and that there are plenty of interesting questions to tackle. Modern technology now produces huge quantities of automatically recorded data, which are usually 'secondary', in that they were collected for some other purpose or for no particular purpose. Their enormous size poses handling problems and they are often stored in database management systems. A standard DM package includes many techniques, such as neural networks and pattern recognition, which are on the fringes of statistics, but is unlikely to offer 'bread-and-butter' statistical techniques like analysis of variance. Much of DM is outside the experience of most statisticians. But should it be? Statisticians need to rethink the way that we tackle very large data sets. In most cases, statistical inference is usually neither necessary nor desirable. Instead, descriptive and data analytic methods combined with subject-matter insight may be the best way to assess, explore and understand the data. They are usually too large to perform much in the way of an IDA, and yet it may be the peculiar observations that are of most interest. It may not be possible to find a model to describe all the data, and it may be necessary to build different models for different parts of the data, or to use algorithms, rather than models, to make decisions. And what is meant by a 'significant' effect or relationship, given that we may have the complete set of data and we may use a computer to try out numerous potential effects? While the hype about DM has been rather overdone, the achievements of DM have often been fairly basic in a statistical sense, as, for example, cleaning up a long list of names and addresses to save duplicated mailing shots or spotting that beer sells well when placed next to nappies in a supermarket.

In the area of inference, the computer has enabled the Bayesian analyst to tackle problems that have no analytic solution and has also allowed the analyst more generally to make more realistic assumptions without worrying that this may make a problem intractable. There has also been increased activity in other computationally intensive methods, such as bootstrapping, and in the last few years I have sat through many seminars featuring Markov chain Monte Carlo (MCMC) methods, Gibbs sampling, the Metropolis algorithm and other trendy topics. Sometimes I have been impressed by what has been achieved, but on other occasions I have been disturbed by the need to rely on intuition to validate the results. This seems unsatisfactory. I am also aware that practitioners naturally suppress any results which go clearly wrong. (I admire the honesty of one software package that has a preamble saying 'Beware—MCMC sampling can be dangerous. The package might just crash, which is not very good, but it might just carry on and produce answers that are wrong, which is even worse.') But how do we tell? The increasing (over)reliance on these techniques gives me some cause for anxiety, given unsolved convergence issues and the difficulties in checking what others have done.

Computational advances have also enabled the practitioner to tackle various other problems in the broad area of inference such as estimating functions (rather than parameters), estimating models with large numbers of parameters and estimating models with time-varying parameters. I suspect that the cumulative effect of these computational advances will be to reveal, even more clearly, the artificial nature of any barriers that exist between the different approaches to inference.

8. Examples

This section presents two examples to illustrate the pragmatic statistician at work, with the emphasis on model building in its broadest sense. Scanning my recent work drew attention to the wide variety of problems, the high proportion of descriptive (as opposed to inferential) analyses, the iterative nature of any modelling and, when consulting, the importance of asking questions and preventing 'clients' from doing 'silly' things. I also realized that it would be difficult to choose representative examples to illustrate the many points made in this paper. Moreover I recognize that most experienced statisticians would actually do the sort of things portrayed here, even though they might not write about them. Nevertheless I hope that these examples will be helpful, especially to the less experienced reader.

8.1. Example 1: is laser therapy effective for rheumatoid arthritis?

About 10 years ago, a physiotherapist from our local hospital called me to arrange a consultancy meeting to discuss a proposed trial on the treatment of rheumatoid arthritis (RA). What follows demonstrates some of the difficulties that are involved in consulting.

Reports in the literature suggested that low level laser therapy (LLLT) may be effective for a wide variety of disorders, including wound healing and rheumatic complaints, and the method was being used to treat RA even though there was no direct evidence of its efficacy. It was agreed that a randomized, double-blind and placebo-controlled study would be carried out to assess the efficacy of LLLT on chronic RA patients with arthritic finger joints. My client showed a good grasp of the importance of collecting good data with a proper randomized trial.

About 3 years later, I was invited to visit the hospital to see the results and to discuss their analysis. In particular they asked me to test the hypothesis that there was no difference between the treatment and control groups. First, I began, as always, by asking questions to remind myself about the background, and to find out exactly how the data were collected. In particular, I checked that randomization really was used, and that there had been no major problems in collecting the data. I also asked whether there was any prior information from previous studies that could be used, but the answer was effectively 'no'. Some patients had dropped out of the study, primarily because they had received additional alternative therapies, such as a cortisone injection, and we had to discuss the effect of these missing data on the results. We also discussed whether a null hypothesis of no difference was really sensible, and whether a one- or two-sided alternative was appropriate.

When we came to look at the data, I found that a fairly simple IDA was all that was necessary. My clients had already computed the means and standard deviations of the observations in the two groups. I also looked briefly at the distributions of the individual observations and checked that there were no gross outliers. I saw immediately that the differences between the groups were (very) 'small', as compared with the standard deviations, and this was really the only feature of note. Even if the treatment did have some small effect, it was unlikely to be sufficiently large to be 'interesting'. This being so, there is little point in doing a test as a 'significant' difference is of no interest if it is not sufficiently large to be of practical importance. Nevertheless, to keep them happy, I carried out the test and showed that it was nowhere near being significant. Of course, with missing values, the *P*-values are suspect anyway, but in this case they were so large that there was clearly no evidence that LLLT was effective, and they helped to convince my clients that there really was no effect. Normally I would have recommended a confidence interval as an alternative to, or in addition to, a *P*-value, but there seemed little point here. (A *P*-value is often not what is really needed, even if it is what scientists think they want.)

At first my clients were devastated that the results were not significant. The therapy was already being used widely and was expected to have a beneficial effect. Initially, they asked me about the power of the test. Could the non-significant result arise from using too small a sample? Given the very small observed effect, I thought not and I eventually managed to persuade them that a non-significant result need not be regarded as a failure. Rather it should be seen in a positive light as indicating that this particular form of electrotherapy was not worth using for this particular condition. More generally, the suppression of non-significant results is a serious problem leading to what is usually called publication bias (Begg and Berlin, 1988). My clients eventually agreed that they should write the work up to spread the message that the treatment was ineffective.

This study had a salutary twist. About 3 years later, I received a copy of an offprint (Hall *et al.*, 1994) 'out of the blue' reporting these results and noting that LLLT was ineffective in this particular context. In the acknowledgements, I found myself thanked for statistical advice even though I had not seen the paper in either draft or final form. I was therefore not pleased to find at least one terrible statistical feature in the paper, namely some inappropriate graphs, incorrectly described as histograms, which purported to show differences between group means. Presumably the refereeing process also bypassed statistical advice, possibly because a statistician appeared in the acknowledgements!

8.2. *Example 2: forecasting in the presence of an outlier*

There is a vast literature on the treatment of outliers (e.g. Barnett and Lewis (1994)) which concentrates on tests for discordancy and on robust procedures for accommodating outliers. However, in my experience, the treatment of outliers is so heavily context dependent that these procedures are used rather rarely in practice (except perhaps for multivariate data where outliers may not be 'obvious'). Instead one useful option is to repeat the analysis with, and without, the outlier and to see whether qualitatively different results are obtained. Alternatively the use of 'native wit' (Barnett and Lewis (1994), page 8) may be called for, and the pragmatic treatment of outliers fits much more naturally into the ambit of pragmatic statistical inference.

As part of some consulting work, I was given 4 years of monthly sales data and asked to produce point and interval forecasts for up to 1 year ahead. The problem related to an insurance claim and a large amount of money was involved. To preserve confidentiality, the data have been scaled before plotting them in Fig. 2, and so the units are not given. There was no prior information about a suitable time series model other than that the data were known to be seasonal. No covariates were provided so only univariate forecasts were considered.

As in any time series analysis, I began by looking at the time plot to assess trend, seasonality and the possible presence of discontinuities and/or outliers. In this particular study, the time plot

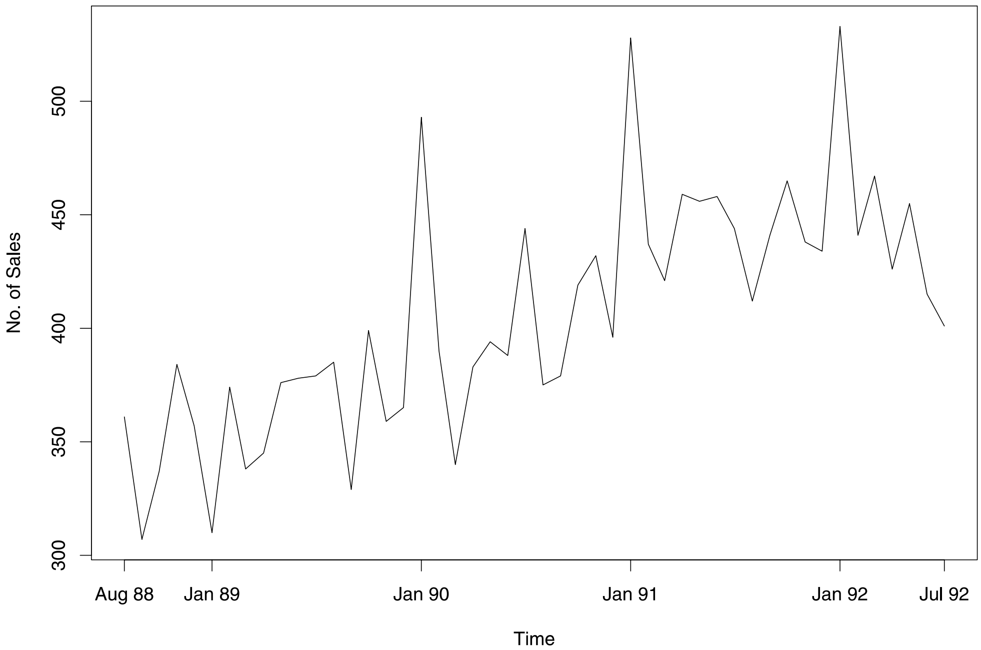

Fig. 2. Coded monthly sales for 4 years starting in August 1988

showed some upward trend over the first couple of years, which flattens out towards the end of the series. There is also some seasonality, with three January figures being comparatively high. However, the first January observation is unusually low compared with other January results, and the treatment of this outlier took centre stage in the modelling and inference process.

There seems little point in testing for discordance, given that the sixth observation is so clearly out of line. Rather, the company was asked why such an outlying value might have occurred. Asking such questions is usually more important than the choice of forecasting method. The simple answer was that no-one seemed to know, presumably because the observation occurred several years ago. There was speculation that it was due to changes in the accounting procedure that resulted in some sales being carried over into neighbouring months. The question was what to do about it. We could have imputed the value, or used some sort of robust modelling procedure, but, given the proximity to the start of the series, it seemed more sensible to omit the first 6 months of data (which are the least relevant to forecasts). The drawback is that we then had to work with just $3\frac{1}{2}$ years of data. This is shorter than one would like and suggests using a simple approach. The client preferred a model-based approach, rather than a smoothing method like Holt–Winters exponential smoothing, and so seasonal ARIMA modelling was used. After looking at the usual diagnostic tools and trying several models with different forms of differencing, it was found that seasonal differences were close to being a random walk. The details need not be given here. Forecasts were then produced using this model, together with prediction intervals, calculated in the usual way, conditionally on the best-fit model. The latter are likely to be too narrow in not allowing for model uncertainty (Chatfield, 1996). However, this analysis seemed the best that could be done here and the client appeared satisfied (though that is not always a conclusive indicator!).

The twist in this story is that the statisticians working for the ‘other side’ in this insurance battle not only produced alternative forecasts, which were qualitatively different but also made disparaging remarks about my approach. It took some effort to defend myself from what I thought was unfair criticism. The statistician must be prepared to defend his or her approach in what is often a hostile climate. In the event, I understand that the lawyers involved eventually agreed to take the average of the two sets of forecasts—a pragmatic, and not unreasonable, outcome.

9. Closing comments

This paper has attempted to give some impression of my experience in tackling statistical problems. I maintain that the driving force in statistics should be a desire to *solve problems*, whether they be in science, technology, management or social and public affairs, and not to worry excessively about the underlying philosophical foundations of our subject, though we do need to guard against the overly simplistic and to beware of the attitude that ‘If it feels good, it’s OK’ (Copas, 1993).

The tension between the different schools of inference can be creative or destructive, and I urge that it be the former. Likewise the tension between *theory* and *applications* can be constructive or destructive (Cox, 1995), but the vast majority of statisticians will surely see them as being complementary. What does worry me is that the literature seems overly concerned with the former. Ever more complicated techniques are developed, leading to what Cox (1997) has called ‘methodological overkill’. Of course, new methods can be good, but they can also lead to a fragmentation of the subject, while the absence of complementary help on general strategy, detailing when and how a new method should be used, is a drawback to the practitioner. My response has been to focus on the practical aspects of problem solving and to make a modest start on describing explicitly a general pragmatic inclusive approach to statistical inference.

In summary, the pragmatic statistician realizes that the really important actions during a statistical study include

(a) exploring the *context*—obtaining sufficient background information to formulate the problem carefully,
(b) collecting the necessary *data* in a valid way,
(c) carrying out a preliminary examination of the data,
(d) formulating an appropriate *model* and being willing to revise it,
(e) checking the predictive accuracy of the model by using out-of-sample results wherever possible,
(f) taking active steps to avoid trouble and
(g) communicating the results clearly.

Acknowledgements

I am very grateful for constructive comments on much earlier drafts especially from David Draper, Andrew Ehrenberg, David Hand and Jim Zidek. The opinions that remain are my own.

References

Barnard, G. A. (1996) Fragments of a statistical autobiography. *Student*, **1**, 257–268.
Barnett, V. (1982) *Comparative Statistical Inference*, 2nd edn. Chichester: Wiley.
Barnett, V. and Lewis, T. (1994) *Outliers in Statistics*, 3rd edn. New York: Wiley.
Bartholomew, D. J. (1995) What is statistics? *J. R. Statist. Soc.* A, **158**, 1–20.
Begg, C. B. and Berlin, J. A. (1988) Publication bias: a problem in interpreting medical data (with discussion). *J. R. Statist. Soc.* A, **151**, 419–463.

Berger, J. O., Boukai, B. and Wang, Y. (1997) Unified frequentist and Bayesian testing of a precise hypothesis (with discussion). *Statist. Sci.*, **12**, 133–160.
Berger, J. O., Brown, L. D. and Wolpert, R. L. (1994) A unified conditional frequentist and Bayesian test for fixed and sequential hypothesis testing. *Ann. Statist.*, **22**, 1787–1807.
Bernard, J.-M. (1996) Bayesian interpretation of frequentist procedures for a Bernoulli process. *Am. Statistn*, **50**, 7–13.
Box, G. E. P. (1976) Science and statistics. *J. Am. Statist. Ass.*, **71**, 791–799.
———(1983) An apology for ecumenism in statistics. In *Scientific Inference, Data Analysis and Robustness* (eds G. E. P. Box, T. Leonard and C. F. Wu). New York: Academic Press.
———(1990) Commentary on a paper by Hoadley and Kettenring. *Technometrics*, **32**, 251–252.
———(1994) Statistics and quality improvement. *J. R. Statist. Soc.* A, **157**, 209–229.
Chatfield, C. (1985) The initial examination of data (with discussion). *J. R. Statist. Soc.* A, **148**, 214–253.
———(1991) Avoiding statistical pitfalls (with discussion). *Statist. Sci.*, **6**, 240–268.
———(1995a) *Problem Solving: a Statistician's Guide*, 2nd edn. London: Chapman and Hall.
———(1995b) Model uncertainty, data mining and statistical inference (with discussion). *J. R. Statist. Soc.* A, **158**, 419–466.
———(1996) Model uncertainty and forecast accuracy. *J. Forecast.*, **15**, 495–508.
Christensen, R. (1995) Letter to the Editor. *Am. Statistn*, **49**, 400.
Cleveland, W. S. (1993) *Visualizing Data.* Summit: Hobart Press.
Copas, J. B. (1993) On some important statistical problems. *Statist. Comput.*, **3**, 185–187.
Cox, D. R. (1978) Foundations of statistical inference: the case for eclecticism. *Aust. J. Statist.*, **20**, 43–59.
———(1986) Some general aspects of the theory of statistics. *Int. Statist. Rev.*, **54**, 117–126.
———(1995) The relation between theory and application in statistics (with discussion). *Test*, **4**, 207–261.
———(1997) The current position of statistics: a personal view (with discussion). *Int. Statist. Rev.*, **65**, 261–290.
Deming, W. E. (1965) Principles of professional statistical practice. *Ann. Math. Statist.*, **36**, 1883–1900.
———(1986) *Out of the Crisis.* Cambridge: Cambridge University Press.
Derr, J. (2000) *Statistical Consulting: a Guide to Effective Communication.* Pacific Grove: Duxbury, Thomson Learning.
Draper, D. (1995a) Assessment and propagation of model uncertainty (with discussion). *J. R. Statist. Soc.* B, **57**, 45–97.
———(1995b) Inference and hierarchical modeling in the social sciences. *J. Educ. Behav. Statist.*, **20**, 115–147.
Durbin, J. (1987) Statistics and statistical science. *J. R. Statist. Soc.* A, **150**, 177–191.
Efron, B. (1986) Why isn't everyone a Bayesian (with discussion)? *Am. Statistn*, **40**, 1–11.
———(1998) R. A. Fisher in the 21st century (with discussion). *Statist. Sci.*, **13**, 95–122.
Ehrenberg, A. S. C. (1975) *Data Reduction.* London: Wiley.
Evans, S. J. W. (1994) Discussion on 'Bayesian approaches to randomized trials' (by D. J. Spiegelhalter, L. S. Freedman and M. K. B. Parmar). *J. R. Statist. Soc.* A, **157**, 395.
Friedman, J. H. (2001) The role of statistics in the data revolution? *Int. Statist. Rev.*, **69**, 5–10.
Gale, W. A. (ed.) (1986) *Artificial Intelligence and Statistics.* Reading: Addison-Wesley.
Good, I. J. (1992) The Bayes/non-Bayes compromise: a brief review. *J. Am. Statist. Ass.*, **87**, 597–606.
Goodhardt, G. J., Ehrenberg, A. S. C. and Chatfield, C. (1984) The Dirichlet: a comprehensive model of buying behaviour (with discussion). *J. R. Statist. Soc.* A, **147**, 621–655.
Hahn, G. J. and Meeker, W. Q. (1993) Assumptions for statistical inference. *Am. Statistn*, **47**, 1–11.
Hall, J., Clarke, A. K., Elvins, D. M. and Ring, E. F. J. (1994) Low level laser therapy is ineffective in the management of rheumatoid arthritic finger joints. *Br. J. Rheum.*, **33**, 142–147.
Hamaker, H. C. (1977) Bayesianism; a threat to the statistical profession? *Int. Statist. Rev.*, **45**, 111–115.
Hand, D. J. (1994) Deconstructing statistical questions (with discussion). *J. R. Statist. Soc.* A, **157**, 317–356.
———(1998) Data mining: statistics and more? *Am. Statistn*, **52**, 112–118.
James, W. (1975) *Pragmatism.* Cambridge: Harvard University Press.
Jeffreys, H. (1961) *Theory of Probability*, 3rd edn. Oxford: Clarendon.
Lindley, D. V. (1986) Comment on a paper by Efron. *Am. Statistn*, **40**, 6–7.
———(1997) Comment on a paper by Berger *et al. Statist. Sci.*, **12**, 149–152.
———(2000) The philosophy of statistics (with comments). *Statistician*, **49**, 293–337.
Lindsey, J. K. (1999) Some statistical heresies (with discussion). *Statistician*, **48**, 1–40.
MacKay, R. J. and Oldford, R. W. (2000) Scientific method, statistical method and the speed of light. *Statist. Sci.*, **15**, 254–278.
Mallows, C. (1998) The zeroth problem. *Am. Statistn*, **52**, 1–9.
Moore, D. S. (1998) Statistics among the liberal arts. *J. Am. Statist. Ass.*, **93**, 1253–1259.
Morris, C. N. (1986) Comment on a paper by Efron. *Am. Statistn*, **40**, 7–8.
Nelder, J. A. (1986) Statistics, science and technology. *J. R. Statist. Soc.* A, **149**, 109–121.
———(1999) From statistics to statistical science (with comment). *Statistician*, **48**, 257–269.
Oldford, R. W. (1990) Software abstraction of elements of statistical strategy. *Ann. Math. Artif. Intell.*, **2**, 291–307.
Piepel, G. and Redgate, T. (1998) A mixture experiment analysis of the Hald cement data. *Am. Statistn*, **52**, 23–30.

Pocock, S. (1994) Discussion on 'Bayesian approaches to randomized trials' (by D. J. Spiegelhalter, L. S. Freedman and M. K. B. Parmar). *J. R. Statist. Soc.* A, **157**, 388–390.
Pole, A., West, M. and Harrison, J. (1994) *Applied Bayesian Forecasting and Time Series Analysis*. New York: Chapman and Hall.
Preece, D. A. (1987) Good statistical practice. *Statistician*, **36**, 397–408.
Rice, J. A. (1995) *Mathematical Statistics and Data Analysis*, 2nd edn. Belmont: Duxbury.
Rubin, D. B. (1984) Bayesianly justifiable and relevant frequency calculations for the applied statistician. *Ann. Statist.*, **12**, 1151–1172.
Samaniego, F. J. and Reneau, D. M. (1994) Towards a reconciliation of the Bayesian and frequentist approaches to point estimation. *J. Am. Statist. Ass.*, **89**, 947–957.
Smith, A. F. M. (1986) Comment on a paper by Efron. *Am. Statistn*, **40**, 10.
Smith, T. M. F. (1994) Sample surveys 1975-1990; an age of reconciliation (with discussion)? *Int. Statist. Rev.*, **62**, 5–34.
Spiegelhalter, D. J., Freedman, L. S. and Parmar, M. K. B. (1994) Bayesian approaches to randomized trials (with discussion). *J. R. Statist. Soc.* A, **157**, 357–416.
Tsay, R. S. (1993) Comment on a paper by Chatfield. *J. Bus. Econ. Statist.*, **11**, 140–142.
Wang, C. (1993) *Sense and Nonsense of Statistical Inference: Controversy, Misuse and Subtlety*. New York: Dekker.