



Statistical COMPUTING & GRAPHICS

A WORD FROM OUR CHAIRS

Statistical Computing



Sallie Keller-McNulty is the 1996 Chair of the Statistical Computing Section. She welcomes your feedback and comments on any issue of interest to your section.

You are probably receiving this Newsletter shortly after returning from the 1996 Joint Statistical Meetings. As you probably noticed, the Statistical Computing Section had a full slate of activities at the Meetings. From our invited, contributed paper and poster sessions to our roundtable lunches and continuing educations courses, the Computing Section had a lot to offer. As usual, our Joint Mixer and Business Meeting was a big success. If you didn't stop by, you missed a great evening of food,

CONTINUED ON PAGE 2

Statistical Graphics



William DuMouchel is the 1996 Chair of the Graphics Section. He has recently joined AT&T Labs — Research.

Having just returned from the Joint Statistical Meetings in Chicago, I can report that the activities there sponsored by the Section on Statistical Graphics went very well. On the first day of the meetings, the short course, "Tools for Discovering Patterns in Data", taught by John Elder and David Scott, attracted 45 students and was well received. Thank you, John and David, for a terrific job! By the way, it is already time to plan for whatever Continuing Education events we want to

CONTINUED ON PAGE 3

SPECIAL ISSUE ON BAYESIAN FUNCTION ESTIMATION

In this issue, we are very excited to be able to bring to you three articles from leading researchers in the field of Bayesian statistics. The common theme of the three pieces is function estimation. For more information, see the Editorial on Page 2.

Bayesian Strategies for Wavelet Analysis

By Merlise Clyde, Giovanni Parmigiani and
Brani Vidakovic

1. Introduction

Wavelets are a family of basis functions that can be used to approximate other functions. They combine powerful properties such as orthonormality, compact support, different degrees of smoothness, localization in time and scale (frequency), and fast implementation. An advantage of wavelet transformations is that they have the "whitening" property in that dependent data, with a dependence structure hard to describe, can become almost independent in the wavelet domain. Discrete wavelet transformations can be used to represent data very concisely. In signal processing terms, the energy is the total sum of squares of the data and the energy components are the squared observations. Wavelet transformations, being orthogonal, or close to orthogonal, will preserve the total energy of the data. However, essentially all of the energy in the wavelet domain could be contained in only a few energy components.

2. Wavelet Models

Suppose that the function f is sampled at N data points and observed with an additive noise ϵ , where Y is the vector of noisy observations (Y_1, \dots, Y_N) , so that the

CONTINUED ON PAGE 4

Special Issue on Bayesian Function Estimation

Our good intentions to have the newsletter mailed to you by the end of July remained just that, intentions, and you will receive this issue after the end of the Chicago Meetings. Production deadlines force us to defer detailed coverage of the Meetings to the December issue. There are, however, two important items that were discussed in the Executive Committee meeting of the Graphics Section and that we wish to bring to your attention. The first is a call for papers on "Fresh Ideas in Statistical Graphics." The selected papers will be presented in a special contributed paper session at the 1997 Joint Meetings in Anaheim (refer to the article by Dianne Cook on page 30 for submission guidelines). The second is the Graphics Section sponsorship of the 1997 Data Exposition at the Anaheim Meetings. All papers contributed to the the Data Expo must analyze a common data set. Colin Goodall describes this year's data and provides information on how to participate in the Expo in his article on page 30.

As usual, the newsletter blends news pieces and announcements of interest to the members of our sections with substantive articles on hot research topics. This issue is very strong in the latter department. We are very excited to be able to bring to you three excellent contributions from leading researchers in the field of Bayesian statistics. The common theme of the three pieces is function estimation. In an article beginning on the page 1, Merlise Clyde, Giovanni Parmigiani and Brani Vidakovic review some recent applications of Bayesian methods to wavelet analysis. Hugh Chipman, Edward George and Robert McCulloch address the issue of Bayesian approaches to CART modeling in their article on page 9. A general Bayesian methodology for nonparametric regression modeling and estimation is presented in the article by Michael Smith and Robert Kohn starting on page 14. An additional common theme likely to be of interest to many of our readers is that all three articles supply examples of sophisticated applications of MCMC estimation methods.

We are also extremely happy to be able to continue publishing two of our regular columns, "Topics in Information Visualization" and "Unix Computing." Among the authors of the article on visualization that appears on page 19 we recognize the familiar name of Daniel Carr. This time, Dan teams up with Richard Valliant and Daniel Rope of the Bureau of Labor Statistics to tell us

about plot interpretation and metadata access in information webs. Our Unix expert, Phil Spector, brings us an article on the Perl programming language. Why should statisticians know about this language? Turn to page 27 to find out.

The most recent issues of the newsletter are now available on-line at <http://cm.bell-labs.com/cm/ms/who/cocteau/newsletter/index.html>. We would like you to take a look at the site and send us your comments and suggestions for improvements. There may come a day when the newsletter will only be published electronically. Would you be in favor of that happening or would you rather continue receiving the printed version? Would the savings in printing costs be enough to offset the lost pleasure of reading the newsletter sprawled out on your living room couch? Again, let us know what your thoughts are.

In what has become a tradition of sorts, we conclude our editorial with a cry for help! We have lost some of our regular contributors along the way, due to their busy schedules, and we would like somebody else to take their place. If you would like to be that somebody, please let us know.

Thank you and enjoy this great issue.

Mark Hansen
Editor, Statistical Computing Section
Bell Laboratories
cocteau@bell-labs.com

Mario Peruggia
Editor, Statistical Graphics Section
The Ohio State University
peruggia@stat.ohio-state.edu



FROM OUR CHAIRS (Cont.) . . .

Statistical Computing

CONTINUED FROM PAGE 1

fun, and fabulous door prizes! I would like to emphasize the fact that it is you, the members, that make the Statistical Computing Section Program such a success at the annual meetings. This is through your hard work in the preparation of your presentations and your energetic attendance at our Section's sessions and activities.

The Statistical Computing Executive Committee had

its annual meeting at the Joint Statistical Meetings as well. On the agenda were plans to discuss ways we can better provide leadership and direction to the statistical computing community, and also ways we might better serve our members. The minutes will appear in a later newsletter. In the remainder of my column I will be pointing out some things I would like to discuss with all of our members.

During my rotation as the National Science Foundation's Program Director of Statistics I have become aware of many of the stresses and needs of the academic community. Three items that come up over and over are the ability and means to travel, the lack of current computing equipment and software, and the need for technical support. There is a crying need for travel funds for statistical researchers in all stages of their careers (from undergraduates to emeritus faculty) and in all types of departments. Even though we are very electronically connected as a community, travel to workshops, conferences, and for short-term and long-term visits to other departments and places is still critical activity for the success and growth of our field. This helps people broaden, learn new things from experts, and build new collaborations. The availability of funds from most departments for travel is traditionally very limited, and the opportunities of getting funds for travel from external sources are diminishing quickly.

For those working in the area of statistical computing and graphics, ranging from the development of computationally intensive methods to software and visualization implementations, there is a need for equipment and technical assistance. In part due to the 1986 IMS workshop and report (Eddy 1986) that argued statistics should be considered a laboratory science with computers as our tool (Eddy 1986), the vast majority of our departments across the country are now providing faculty start-up packages for equipment. But, these funds are still usually not enough to purchase the level of equipment that will carry these researchers for more than a couple of years. After the initial purchase, the Universities expect the researchers to find funding "off-campus" for their equipment needs. This is presenting a serious problem for our community. There is also an ever increasing need for technical assistance. It is no longer (as if it ever was!) cost effective to expend valuable faculty and graduate assistant time doing system administration, computer hook-ups, porting software from one system to another, routine computation, etc. Where should the funds come for increased technical support?

Is there anything we can do as a community to encourage the Universities to make the needs of the statistics

community a priority and to take some responsibility in the directions discussed above? Where can we suggest members of our community go to get some of these financial needs met? What role should/could the ASA play? The funds available from NSF do not even put a small dent in meeting these needs across the broad range of our community. Please share with me any suggestions you have along these lines.

This then brings me to the area where I feel rather uninformed. What are the stresses and concerns of our members outside of the academic community? Are they the same? In particular, what is the situation for the non-Ph.D. segment of our community? Again, I would like to hear from you.

Reference

Eddy, William F. (1986), "Computers in Statistical Research," *Statistical Science*, 419-437.

Sallie Keller-McNulty
Kansas State University
sallie@cecil.stat.ksu.edu



FROM OUR CHAIRS (Cont.) . . .

Statistical Graphics

CONTINUED FROM PAGE 1

sponsor for next year's meetings in Anaheim, California. If you have any suggestions for courses with a graphics theme, send them to me or Karen Kafadar, our CE committee person, or any member of the Executive Committee.

During the regular meetings we had eleven other sessions, counting invited, contributed or poster sessions, the roundtable luncheon and the undergraduate data analysis contest. Many thanks to our Program Chair this year, Stephen Eick, for putting in a lot of work to make it all happen. I was not able to attend every session, but those I did make were excellent, well attended, and generated enthusiastic responses. One thing that impressed me was how many of the talks discussed the World Wide Web and its potential for aiding both Statistics teaching and research. What a difference a year makes! Diane Cook, our Program Chair for 1997, reports that she has over half of next year's invited program lined up, and she welcomes any suggestions for sessions to be sponsored or co-sponsored by the Graphics Section. Congratulations are also due to our newly elected officers:

Mike Meyer as Chair-Elect (Chair in 1998), Ed Wegman as Program Chair-Elect, Bob Newcomb continuing as Secretary/Treasurer (and volunteer Webmaster), Mike Minnotte as Publications Officer, and Roy Welsch as a Representative to the Council of Sections.

And of course our Annual Meeting and Joint Statistical Computing and Statistical Graphics Mixer was a great success as usual! We were again fortunate to have a great turnout, aided by the prospect of winning one of the door prizes donated by many of the exhibitors. If you want to check out who won what, donated by whom, it's on our Web page. Finally, there are two other projects discussed at the Joint Meetings that I would like to bring to your attention. First, we are responsible for running the Data Analysis Contest for 1997, and we have tentatively decided to use a very interesting data set being prepared by Colin Goodall for use in "hospital report cards." For each of 1,088 US hospitals in ten different states, data are given on 30 patient study populations, including those treated in Obstetrics, Pediatrics, Cardiology, etc. Variables comprise number of cases, outcome rate, complication rate, mortality, average length of stay, average cost, and average charge for each procedure or diagnosis presented. Colin points out that such hospital "report cards" are a hot topic, with versions published by *US News and World Report* as well as many other State and nongovernmental organizations. Watch for further announcements of data availability and, with a little creative data analysis, you could make the news too!

The other project, discussed at our Executive Board meeting, is our intention to sponsor a prize for the outstanding paper on statistical graphics published each year in the *Journal of Computational and Graphical Statistics*. What didn't get settled was just what ground rules should apply. Should we define just what constitutes a paper on statistical graphics? How do we do the judging and how large a committee of judges is needed? Should we define criteria for excellence? How large should the prize be? Stephen Eick has volunteered to draft a proposed set of rules in the next few months to get us going. If you would like to give him some suggestions, send them to eick@bell-labs.com. What do you think constitutes a "great paper on statistical graphics"?

Bill DuMouchel
 AT&T Labs — Research
dumouchel@research.att.com



SPECIAL ISSUE ON BAYESIAN FUNCTION ESTIMATION: FIRST FEATURE ARTICLE

Bayesian Strategies for Wavelet Analysis

CONTINUED FROM PAGE 1

model is

$$Y_i = f_i + \epsilon_i.$$

Estimation of f by wavelet shrinkage involves three steps: transforming Y by a discrete wavelet transformation (DWT); shrinking wavelet coefficients towards 0; and transforming shrunk wavelet coefficients back to the time domain (IDWT) to give an estimate of f .

For discrete data, there is an orthogonal wavelet transformation, so that the vector of observations and function values can be represented in a wavelet regression form

$$Y = W\beta + \epsilon \quad (1)$$

where W is an orthonormal matrix, and $W'f = \beta$ (see Daubechies (1988, 1992) and Bruce and Gao (1994) for software). The least squares estimator of β is $\hat{\beta} = W'Y$, and is, in fact, the DWT of Y . Under the one-to-one reparameterization of (1), one can also consider directly the model in the wavelet coefficient space,

$$\hat{\beta} = \beta + \epsilon^* \quad (2)$$

where $\beta = W'f$ and $\epsilon^* = W'\epsilon$.

Several thresholding and shrinkage policies have been applied in the wavelet domain. In a series of recent papers Donoho and Johnstone (1994, 1995) show that in estimating β simple shrinkage rules have asymptotic optimality properties for a rich class of function spaces. In most areas of application, there is the need for shrinkage and thresholding strategies that adapt to data, and also for formal assessment of the uncertainty associated with the shrinkage and subset selection process. For these reasons, statistical analysis using wavelets offers a very natural terrain for the application of Bayesian estimation, prediction and model selection techniques. We will review some of the recent applications of Bayesian methods to wavelets.

3. Bayesian modeling in wavelets

Vidakovic (1994) considers models for (2) given by the following independent distributions:

$$\begin{aligned} \hat{\beta}_{jk} | \beta_{jk}, \sigma &\sim N(\beta_{jk}, \sigma) \\ \sigma &\sim \text{Exponential}(\mu) \\ \beta_{jk} &\sim t_n(0, \tau) \end{aligned} \quad (3)$$

(the indices j and k in subscripts jk correspond to scales and shifts in the wavelet domain). The posterior mean has the shape of standard Donoho-Johnstone soft thresholding rules, but requires numerical integration or approximations to calculate it in most cases. Monte Carlo integration to calculate the estimator, however, only takes a few seconds of computing time with $N = 1024$. Shrinkage is determined by the data and the prior hyperparameters.

The posterior means determined by model (3) are not hard thresholding rules in the sense of Donoho-Johnstone. To achieve hard thresholding Vidakovic considered Bayesian testing of precise hypothesis on β_{jk} . The coefficient β_{jk} is removed from the model if the hypothesis $H_0(\beta_{jk} = 0)$ is accepted, otherwise it is estimated as $\hat{\beta}_{jk}$ with no shrinkage. The prior used for testing is a mixture of a point mass at 0 and the proper student t distribution used in (3). Posterior probabilities of hypotheses are obtained in a finite form, but require numerical integration as in the calculation of posterior means for this model. The paper of Ruggeri and Vidakovic (1995) gives other strategies for model choice and prior specification, yielding Bayes rules with attractive properties.

Vidakovic and Müller (1995) relax the assumption of prior independence of β_{jk} , which may be simplistic in real-life applications, in favor of dependence of β_{jk} within the same level j . An independent normal model on the vector $\hat{\beta}$, as in (3), and a conjugate normal-inverse-gamma prior on (β, σ^2) produce a normal-inverse-gamma posterior. The resulting posterior mean of β is a linear transformation of $\hat{\beta}$,

$$E(\beta|Y) = (I_N + \Sigma^{-1})^{-1} \hat{\beta}$$

where Σ is the prior covariance matrix of β . The choice of Σ can lead to differential shrinkage depending on the level. In addition to the function estimation problem outlined here, they also consider density estimation.

More powerful and flexible Bayesian models are obtained using a hierarchical structure, assigning a probability distribution to the space of subsets of basis elements (the columns of W). The resulting estimators provide flexible nonlinear shrinkage, such as in Vidakovic (1994), but also allow for a coherent approach to hard-thresholding like rules through multiple shrinkage.

4. Hierarchical Models and Multiple Shrinkage

The multiple shrinkage estimators are obtained as the posterior mean of a hierarchical model. The highest stage of the hierarchy is the probability distribution over subset selections; the middle stage is the distribution of

wavelet coefficients given a subset; and the lowest stage is the distribution of the data given the subset and the coefficients. The coefficient estimates and predictions are generated by integrating over the subset selection (model averaging). Regression coefficients are shrunk towards their respective prior mean as a result of uncertainty about their value, and are shrunk towards 0 as a result of uncertainty about whether the corresponding basis elements should be included. Thus they can be viewed as multiple shrinkage estimators. George (1986) discusses desirable Bayes and minimax properties of multiple shrinkage estimators. In the context of wavelets, the appeal of multiple shrinkage is both in the flexibility of the shrinkage patterns that can be generated, and in the fact that shrinkage patterns adapt to the data at hand. Two proposals along this lines have been put forward to date; one by Chipman, Kolaczyk and McCulloch (1995) and one by the authors. Both use the same sampling model in (1) based on the assumption that

$$Y|\beta, \sigma^2 \sim N_N(W\beta, \sigma^2 I_N), \quad (4)$$

but differ in their assumptions about the prior distributions. The approach used by Chipman et al. (1995) is based on the stochastic search variable selection (SSVS) model proposed by George and McCulloch, but with the assumption that σ is known. The approach used by Clyde et al. (1995) is based on a limiting form of the conjugate SSVS prior in George and McCulloch (1994). (See George and McCulloch (1994) for a discussion of these models and related literature.) Both Chipman et al. (1995) and Clyde et al. (1995) use the posterior mean under the resulting mixture models as an estimate of β , and then estimate f by applying the inverse discrete wavelet transformation (IWDWT) to the posterior mean. We will discuss the approach of Clyde et al. (1995) here.

4.1. The SSVS Model

In the linear regression context with orthogonal variables, Clyde, DeSimone, and Parmigiani (1996) consider a limiting case of the conjugate SSVS prior for β where the distribution is a mixture of a point mass at 0 if the variable is excluded from the regression and a normal distribution if it is included,

$$\beta_{jk}|\gamma_{jk}, \sigma^2 \sim N(0, 0(1 - \gamma_{jk}) + c_{jk}\sigma^2\gamma_{jk}).$$

The γ_{jk} are indicator variables that specify which basis element or column of W should be selected. The set of all possible γ 's will be referred to as the subset space. The prior distribution for σ^2 is inverse χ^2 (i.e. $\lambda\nu/\sigma^2 \sim \chi_\nu^2$), and the γ_{jk} 's are independently distributed as Bernoulli(θ_{jk}) random variables.

The posterior mean of β conditional on γ is

$$E(\beta|Y, \gamma) = \tilde{\beta}_\gamma = \Gamma(I_N + C^{-1})^{-1} \hat{\beta} \quad (5)$$

where Γ is a diagonal matrix with γ_{jk} on the diagonal and 0 elsewhere; C is also a diagonal matrix with c_{jk} on the diagonal and 0 elsewhere. For a particular subset given by γ , this corresponds to hard thresholding with linear shrinkage.

The posterior mean is obtained by averaging over all models. Model averaging leads to multiple shrinkage estimators of β :

$$E(\beta|Y) = \sum_{\gamma} \pi(\gamma|Y) \Gamma (I_N + C^{-1})^{-1} \hat{\beta}$$

based on additional nonlinear shrinkage of the coefficients to 0 resulting from the uncertainty about which subsets should be selected. Here, $\pi(\gamma|Y)$ is the posterior probability of the particular subset γ .

Calculating the posterior probabilities of γ and the mixture estimates for the posterior mean of β above involves sums over all 2^N values of γ . For any realistic problem, such enumeration is prohibitive and either we must use approximations or turn to stochastic methods for selecting subsets with high posterior probability to estimate subset probabilities. It is desirable to have fast algorithms for calculating the posterior mean without having to resort to MCMC sampling, as discussed by Chipman et al. (1995). When variables are orthogonal, Clyde, DeSimone, and Parmigiani (1996) obtain an approximation to the posterior probability of γ . Clyde, Parmigiani, and Vidakovic (1995) adapt this approximation to the wavelet setting.

4.2. Approximations

The approximation of Clyde et al. (1995) results in independence of the elements of γ . The approximate model probabilities are a function of the data through the regression sum of squares and are given by:

$$\pi(\gamma|y) \approx \tilde{\pi}(\gamma|y) = \prod_{j,k} p_{jk}^{\gamma_{jk}} (1 - p_{jk})^{1-\gamma_{jk}} \quad (6)$$

$$p_{jk} = \frac{a_{jk}(Y)}{1 + a_{jk}(Y)}$$

where

$$a_{jk}(Y) = \frac{\theta_{jk}}{1 - \theta_{jk}} (1 + c_{jk})^{-1/2} \cdot \exp \left\{ \frac{1}{2} (N + \nu) L \sum_{l=1}^m \frac{S_{jk}^{2l}}{l(\nu\lambda + Y'Y)^l} \right\}$$

$$S_{jk}^2 = \hat{\beta}_{jk}^2 / (1 + c_{jk}^{-1}),$$

and L is a calibration constant.

The p_{jk} can be used to obtain a direct approximation to the multiple shrinkage Bayes rule. Based on the independence form of the approximation, the posterior mean

is approximately

$$p_{jk}(1 + c_{jk}^{-1})^{-1} \hat{\beta}_{jk}. \quad (7)$$

Equation (7) can be viewed as a level dependent wavelet shrinkage rule, generating a variety of nonlinear rules. Depending on the choice of prior hyperparameters shrinkage may be monotonic, if there are no level dependent hyperparameters, or non-monotonic.

The posterior distribution of γ given σ also exists in closed form where γ is distributed as a product of independent Bernoulli random variables with $P(\gamma_{jk} = 1|\sigma, Y) = p_{jk}(\sigma)$ and

$$p_{jk}(\sigma) = \frac{a_{jk}(Y, \sigma)}{1 + a_{jk}(Y, \sigma)} \quad (8)$$

$$a_{jk}(Y, \sigma) = (1 + c_{jk})^{-1/2} \left\{ \frac{\theta_{jk}}{1 - \theta_{jk}} \right\} \exp \left\{ \frac{1}{2} \frac{S_{jk}^2}{\sigma^2} \right\}.$$

If σ is known, or if it is unknown and a robust estimate of it as in Chipman et al. (1995) is used, then the $p_{jk}(\sigma)$ can be used in place of p_{jk} in (7) to approximate the posterior mean of β under model averaging. (Note: if σ were known the result is exact.) The results in Chipman et al. (1995) can also be adapted to obtain approximate variances, that, in turn, can be used to calculate approximate uncertainty intervals about the estimate of the function f .

4.3. Sampling

If there is substantial uncertainty about σ or concern about using the approximations, then one can use importance sampling or MCMC methods to reduce any bias induced by the approximations. Using (6), we can generate a model γ by generating each γ_{jk} as an independent Bernoulli random variable with $P(\gamma_{jk} = 1) = p_{jk}$. An estimate of $\pi(\gamma|Y)$ from the importance sampler is then used in place of $\pi(\gamma|Y)$ in the multiple shrinkage estimator. All sums are over the set of sampled models. See Clyde et al. (1995) for more details on the importance sampler or Gibbs sampler. Müller and Vidakovic (1995), in the context of density estimation with wavelets, also use a hierarchical model that generalizes the SSVS models used here and in Chipman et al. (1995), but the extra complexity requires that inference is based on Markov chain Monte Carlo methods.

4.4. Prior Hyperparameters

The distribution of σ , by specifying the amount of noise that can be expected in the data, has impact on the overall smoothness of the curve and parsimony of the model.

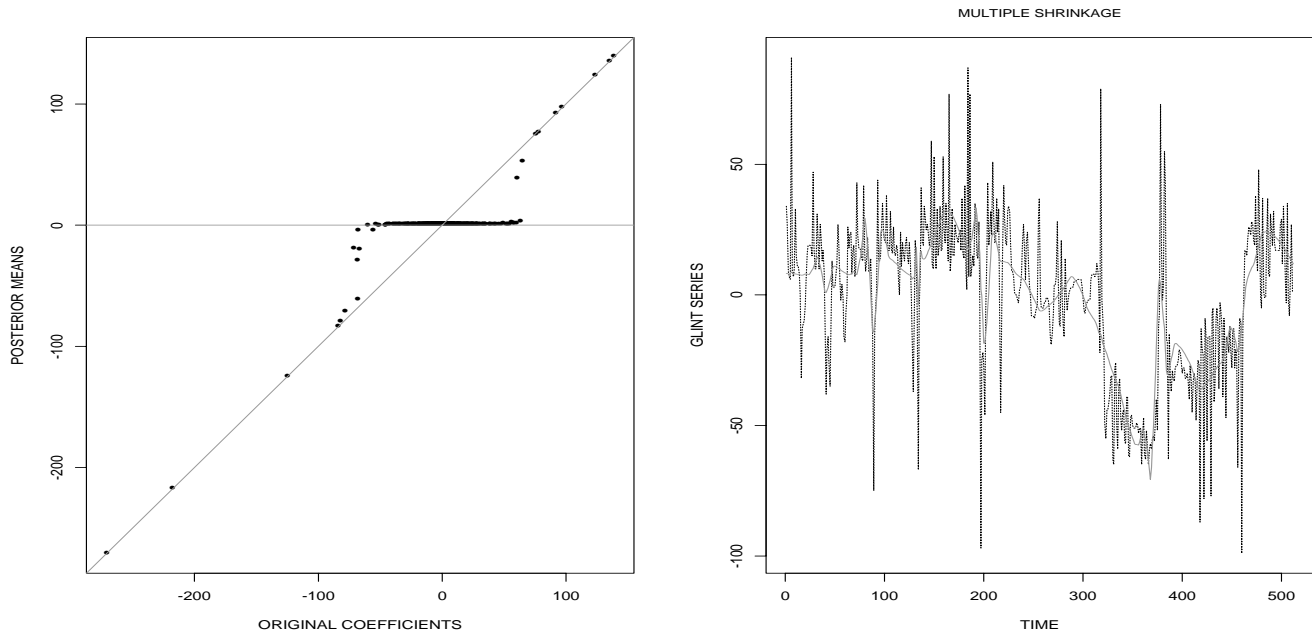


Figure 1. Glint Data: shrinkage curve (left) and fit (right) resulting from approximation to the posterior model probabilities.

The hyperparameters λ and ν can be selected to reflect this. The values of c_{jk} and θ_{jk} can be used to shrink coefficients differently. It is often appealing to assign values of θ_{jk} by level. A simple option used is based on uniform values of θ_{jk} within level j . Assigning $\theta_{jk} < .5$ imposes a penalty for adding the basis element; in this way, increasing penalties for inclusion can be imposed to higher levels of detail. We implement this by using a prior for θ_{jk} decaying geometrically with the level of detail,

$$\theta_{jk} = a^{-1}b^{j-J}. \quad (9)$$

This choice may lead to an overall non-monotone shrinkage patterns, since shrinkage depends on the data through the subset probabilities, but shrinkage within a level will be monotone and nonlinear, if $c_{jk} = c_j$.

One approach to choosing C is to let $C = cI_N$. Foster et al. (1995) propose a method for calibrating c so that, conditional on σ , model selection based on Bayes factors corresponds to classical model selection methods, such as AIC, BIC or RIC. While some aspects of the models in Chipman et al. (1995) are different, their suggestions for elicitation of the hyperparameters may be helpful in the present context.

We conclude by looking at an example.

5. Glint Example

To illustrate features of the multiple shrinkage approach described here we used the glint data presented by Bruce and Gao (1994). The data series includes 512 equally

spaced observations. The true signal is a low frequency oscillation about zero, resulting from rotating an airplane model. As the model rotates, the center of mass will shift slightly. The measurements, given in angles, are subject to large errors, and can be off the truth by as much as 100 degrees.

Throughout, we used the least asymmetric Daubechies 8-tap mother wavelet “s8” from the S+WAVELETS package. This basis provides a sensible balance between the compactness of support and smoothness of the wavelet. The default fit resulting from universal thresholding is not satisfactory in this application, where the true signal is believed to be smoother. The fit using the largest 12 coefficients, while ad hoc, gives a more satisfactory answer. Moving to multiple shrinkage, we begin by defining the prior distribution used. The prior hyperparameters for σ are given by $\lambda = 1000$ and $\nu = 22$, reflecting the belief that errors can be large. The choice of c is based on the Foster et al. (1995) RIC rule, leading to $c = 262130.5$. The prior on the subset space is based on independence and the exponential decay expression in (9), with $b = 2$ and $a = 2$. Under this specification, basis elements at level J are included with probability $1/2$.

Figure 1 shows the corresponding shrinkage curve and the resulting fit using the approximation (7). The shrinkage curve displays both nonlinearities and nonmonotonicities. Nonlinearities arise from multiple shrinkage, nonmonotonicities from the specific prior chosen, which

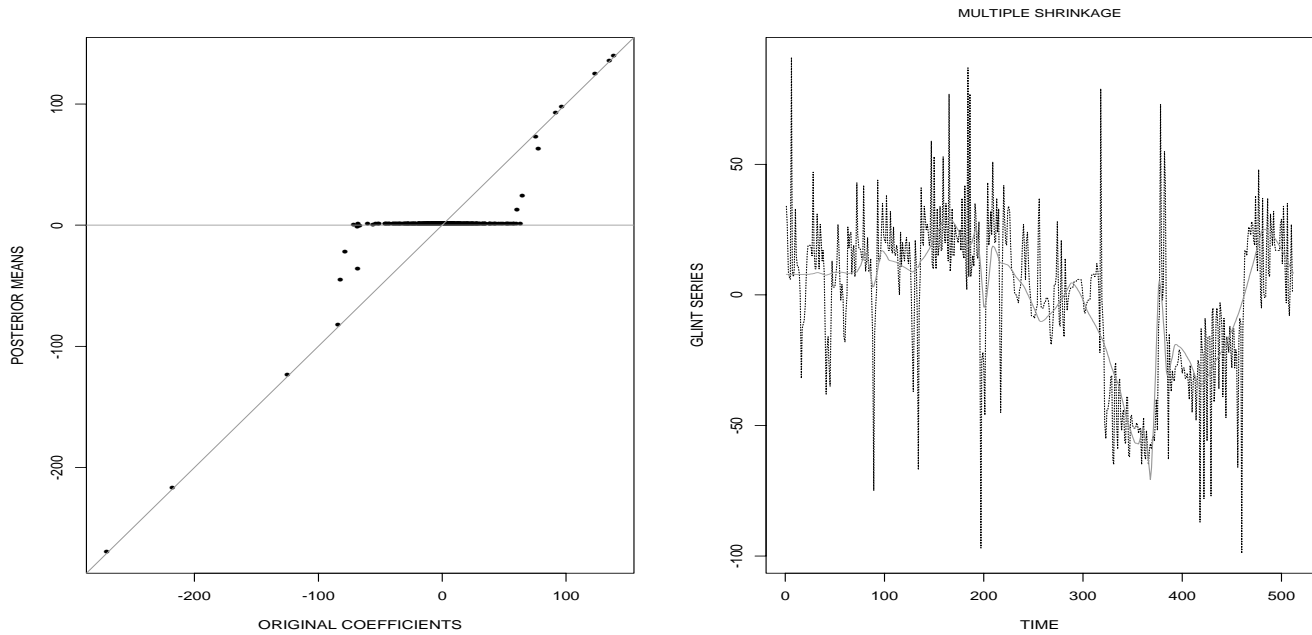


Figure 2. Glint Data: shrinkage curve (left) and fit (right) using importance sampling.

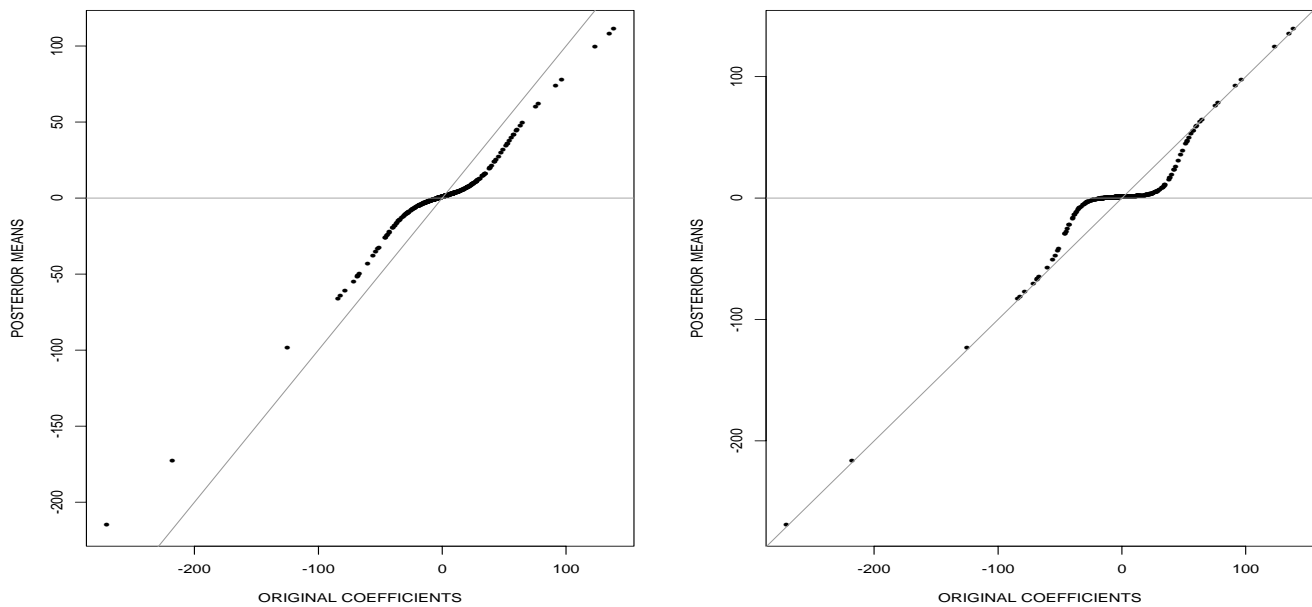


Figure 3. Shrinkage for the AIC prior (left) and the BIC prior (right).

allows a stronger shrinkage at higher level of detail. However, shrinkage within a level is monotone. These approximate estimates can be rapidly computed analytically. In high dimensional problems they can provide a valid alternative to performing the simulation necessary to evaluate the posterior distribution and the posterior mean.

We also show the results based on evaluating the posterior means of the coefficients by sampling. We illustrate

the results of the importance sampler. Figure 2 shows the shrinkage curve and the fit. The importance sampler uses the analytic approximation as an importance function, and then reweights the selected models proportionally to the exact posterior probability. Similar results are obtained using a Gibbs sampler (Clyde et al. 1995).

Alternative prior specifications, however, can have a greater impact on the shrinkage. Figure 3 illustrates the

results of specifying C via the AIC and BIC criteria. This leads to $c = 504.72$ and $c = 3.92$ respectively. The prior distribution on σ is the same as before. The prior on the subset space is given by $b = 1$ and $a = 2$, meaning that each element is, a priori, included with probability $1/2$. The shrinkage is not as strong as in the previous example, primarily as a result of the change in the prior on the subset space. Under the AIC prior, all of the coefficients undergo some shrinkage; the difference between the AIC and BIC prior is due to the smaller value of c . Note that the shrinkage in these two cases is monotone in the original coefficients, since there are no level dependent hyperparameters as in the first prior specification.

References

- Bruce, A., and Gao, H-Y. (1994), *S+Wavelets, Users Manual*, Seattle, WA.: StatSci.
- Chipman, H., Kolaczyk, E., and McCulloch, R. (1995), "Adaptive Bayesian Wavelet Shrinkage", TR 4, The University of Chicago, Econometrics and Statistics Group.
- Clyde, M., DeSimone, H., and Parmigiani, G. (in press), "Prediction via orthogonalized model mixing," *Journal of the American Statistical Association*.
- Clyde, M., Parmigiani, G., and Vidakovic, B. (1995), "Multiple Shrinkage and Subset Selection in Wavelets," Discussion Paper 95-37, ISDS, Duke University.
- Daubechies, I. (1988), "Orthonormal bases of compactly supported wavelets," *Communications on Pure Applied Mathematics*, 41, 909-996.
- Daubechies, I. (1992), *Ten Lectures on Wavelets*, Society for Industrial and Applied Mathematics.
- Donoho, D., and Johnstone, I. (1994), "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, 81, 425-455.
- Donoho, D., and Johnstone, I. (1995), "Adapting to unknown smoothness via wavelet shrinkage," *Journal of the American Statistical Association*, 90, 1200-1224.
- Foster, D., and George, E. (1994), "The risk inflation criterion for multiple regression," *Annals of Statistics*, 22, 1947-1975.
- Foster, D., George, E., and McCulloch, R. (1995), "Calibrating Bayesian variable selection procedures," Manuscript.
- George, E.I., and McCulloch, R. (1993), "Variable Selection via Gibbs Sampling," *Journal of the American Statistical Association*, 88, 881-889.
- George, E.I., and McCulloch, R. (1994), "Approaches to Bayesian Variable Selection," Technical Report, Graduate School of Business, University of Chicago.
- Müller, P., and Vidakovic, B. (1995), "Bayesian Inference with Wavelets: Density Estimation," Discussion Paper 94-33, ISDS, Duke University.
- Raftery, A.E., Madigan, D.M., and Hoeting, J. (1993), "Model selection and accounting for model uncertainty in linear regression models," TR 262, Department of Statistics, University of Washington.
- Vidakovic, B. (1994), "Nonlinear wavelet shrinkage with Bayes rules and Bayes Factors," Discussion Paper 94-24, ISDS, Duke University.
- Vidakovic, B., and Müller, P. (1995), "Wavelet Shrinkage with Affine Bayes Rules With Applications," Discussion Paper 95-34, ISDS, Duke University.

Merlise Clyde
Duke University
clyde@stat.duke.edu

Giovanni Parmigiani
Duke University
gp@stat.duke.edu

Brani Vidakovic
Duke University
brani@stat.duke.edu



SECOND FEATURE ARTICLE

Trees à la Bayes

By Hugh Chipman, Edward I. George and Robert E. McCulloch

1. Introduction

One of the reasons tree based models are so popular is that they are so simple to explain, and they make it possible to graphically depict the model even with many predictors. It seems that perhaps the less statistics you know, the more attractive these models may be. Since their introduction (Sonquist and Morgan 1964) these models have remained popular and been a source of continuing research. In this article, we discuss current Bayesian research in this area, and how it relates to other current work.

Overviews of the Classification And Regression Tree (CART) model are given in Breiman et al. (1984), and

Clark and Pregibon (1992). A short summary follows. The CART model consists of two components: A recursive partition of the predictors (X), and a parametric model for the response (Y) within each partition. The partition is defined by a binary tree with splitting rules based on the predictors. An example of a regression tree is given in Figure 1. In this case, there are only two predictors (X_1, X_2), so the rectangular partitions may also be plotted. Each bottom node of the tree corresponds to a rectangular partition, and within each partition, Y is assumed to have a fixed parametric distribution. Here, the distribution is normal, with constant variance and means that vary across partitions. If Y were a categorical variable, say with a multinomial distribution, such a tree would be called a classification tree.

CART models are traditionally fit by a greedy algorithm. The algorithm starts with all observations in a single group. It identifies the variable and splitting rule which provides the best fit for the two resultant nodes. In Figure 1, the first split variable is X_1 , and the split is at $X_1 = 0.6$. The procedure is then repeated. For every bottom node, the optimal splitting rule is found, and the

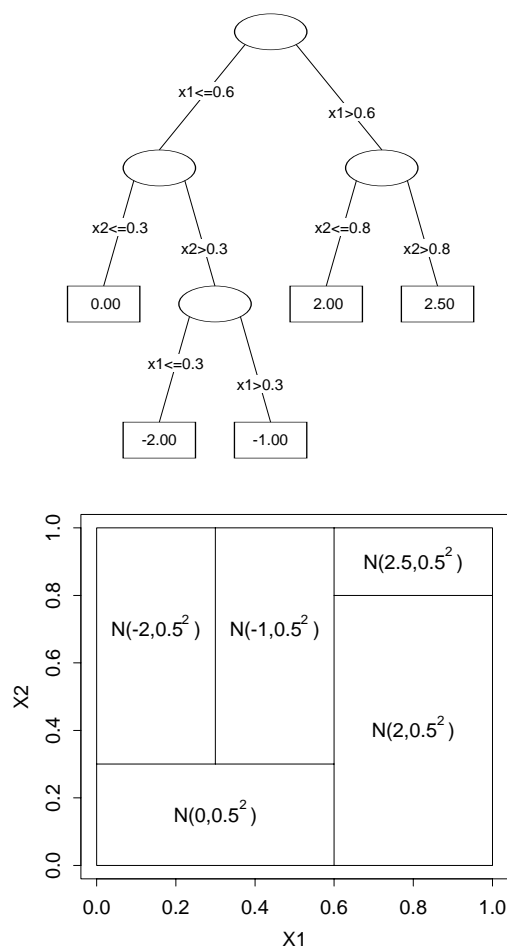


Figure 1. Two perspectives of a CART model.

node with the best improvement is chosen to split. Since the form of the splitting rule depends on the rules further up the tree, a greedy procedure will not usually identify the best fitting tree of a given size. This procedure is similar to forward stepwise regression and like that method, it cannot guarantee a global optimum. A large tree is usually grown, and then reduced by either pruning or shrinking (see next section).

We'd like to summarize some of the main issues in current research and relate these to Bayesian CART (Section 2) and describe aspects of the Bayesian approach to this problem (Section 3). We then present an illustrative example (Section 4).

2. Current Research Issues

Recent work on CART has been extensive. Here we focus on ideas related to Bayesian approaches. Bayesian alternatives are developed in more detail in Section 3.

CART models fit by greedy methods are unlikely to provide the optimal fit for a given number of nodes. They are also unstable, in the sense that small changes in the data can produce substantial changes in the fitted model. Bootstrapping capitalizes on this fact, in the **bootstrap umbrella** of **parameters** (“bumping”) approach of Tibshirani and Knight (1995), and the **bootstrap aggregation** (“bagging”) method of Breiman (1994). In both cases, multiple data sets are produced by sampling with replacement from the original data. For each resampled data set, a greedy tree is grown. These trees often provide better fit (to the original data), and when averaged together, can provide more stable predictions. Bayesian analogs of these two concepts arise naturally. A stochastic search based on a Metropolis algorithm finds better models than the greedy methods, and a mixture over multiple trees (weighted by their posterior probabilities) gives more stable predictions.

Both Bayesian and non-Bayesian CART models are very flexible, and unless care is exercised, they can overfit the data. In non-Bayesian situations, trees are either pruned or shrunk. In the former, pairs of bottom nodes are removed until fit degrades substantially. In the latter the large tree is maintained, but the fitted values in a pair of nodes are shrunk towards the fitted values of their parent (Hastie and Pregibon 1990). Bayesian priors can incorporate both of these structures. Priors on the tree structure may enforce parsimony analogous to pruning, while shrinkage may be achieved through priors on the parametric model for the response at the end nodes.

A consequence of both Bootstrap and Bayesian methods is that they grow a forest of good-fitting trees. These

trees may have little in common, unlike the nested sequence of trees produced by pruning, or the single tree structure that forms the basis of a shrunk tree. Methods for exploring and summarizing trees are needed, since examination of more than a few trees is not practical. A metric on trees would aid in this exploration, allowing similar trees to be grouped.

3. Bayesian Approaches to Cart

Recently Chipman, George, and McCulloch (1996) and Denison, Mallick, and Smith (1996) have developed Bayesian approaches to CART modeling. A Bayesian approach requires a specification of a prior and a method for exploring the resulting posterior distribution on the set of models. While neither of these tasks is straightforward, both offer ways of addressing some of the basic issues in the analysis of CART models.

Markov chains designed to explore the posterior distribution are powerful methods for searching the complex space of tree models for models which describe the data. These chains provide a rich class of alternatives to search methods such as bumping. The posterior can also be used to gauge our uncertainty about which tree structure best fits the data and to smooth our tree based model by using a combination of trees weighted by their posterior likelihoods. This Bayesian smoothing provides an alternative to methods such as bagging.

Simple models are easiest to understand and often perform best in out of sample prediction. This simplicity is achieved by either growing a small tree (or pruning), or shrinking the parametric response model. By choosing priors which favor smaller trees (simpler models) the Bayesian approach can naturally restrain the tendency to overfit the data. Trees may be shrunk by choosing priors on the parametric model for the response at bottom nodes. Nodes which are close in the predictor space would (a priori) be more likely to have similar responses.

3.1. Exploring the Posterior

Both Chipman et al. (1996) and Denison et al. (1996) use a similar approach to finding trees which have high posterior probability. The Metropolis-Hastings algorithm is used to step through the tree space by iterating a Markov Chain. At each iteration a candidate new tree is generated by a modification of the current tree. The algorithm either rejects the candidate tree (keeps the current tree) or moves to the candidate probabilistically, where the probability of accepting the candidate is large when the posterior probability of the candidate is large relative to the current tree. At the heart of the approach is the choice of a simple set of possible ways to modify

the current tree to obtain a candidate tree to step to. The most basic types of modification (or step) correspond to growing and pruning the tree. In a growth step, a bottom node is split into two new bottom nodes and in a pruning step two bottom nodes are combined. Other types of steps alter nodes by changing the corresponding splitting variable or rule.

As the Markov chain is run, we wander through the space of trees, growing, pruning, and modifying the tree in a constant search for trees with high posterior probability. Many choices are possible for both the type of step and generation of a new candidate with a given step type so that a wide variety of possible stochastic searches in the tree space can be constructed from the basic Metropolis approach. Additional experience and experimentation are needed to find the chains which are most effective.

Besides the basic issue of the specification of the chain, there are other important aspects of the approach which need further work. Perhaps the most fundamental problem is that we have an embarrassment of riches. The chains generate large numbers of trees. What do we do with them? How do we communicate what “kinds” of trees have been found and how likely they are? Trees may be viewed as a high-dimensional objects, and so a natural solution is to consider a metric on the space of trees. One such metric is explored in the next section.

Another basic issue is that the Markov chains tend to get “stuck” in local modes. As Denison et al. (1996) note, once a large tree is grown “proposed changes in split variables or rules near the root node are extremely unlikely to be accepted”. Thus, although our experience indicates that the chains can be used as effective search mechanisms it is not clear that actual frequencies of trees can be used in the usual Monte Carlo method to approximate posterior probabilities. Chipman et al. (1996) discuss an approach based on restarting the chain essentially in the spirit of Gelman and Rubin (1992). Again it seems that much work is needed to learn which strategies are most effective although experience to date indicates that even naive use of the methods can be very powerful.

3.2. Choosing the Prior

How is the prior specified for a CART model? In thinking about the prior, it is useful to decompose the “parameter” into two parts. One part is the tree structure which describes the partitioning of the predictor space with each bottom node of the tree corresponding to a partition. Within each partition we usually assume that the distribution of the response belongs to some para-

metric family of distributions so that each bottom node is associated with a parameter value θ . The second part of our parameter is the set of parameter values corresponding to the bottom nodes. If we let T denote the tree and Θ denote the set of values for θ (one for each bottom node) a CART node may be represented by the parameter (T, Θ) where the dimension of Θ depends on T . We then specify a joint prior $p(T, \Theta)$ by choosing the marginal $p(T)$ and conditional $p(\Theta|T)$.

Denison et al. (1996) choose an appealingly simple form for $P(T)$. Let B denote the number of bottom nodes in a tree. Then we can specify $p(T)$ in the form $p(T) = p(B)p(T|B)$. A (truncated) Poisson prior is chosen for $p(B)$ and $p(T|B)$ is uniform so that all trees having the same number of bottom nodes are equally likely. This simple prior specification allows us to express our basic preference for parsimony in model choice by placing prior weight on small values of B . They then choose a “noninformative” uniform prior for $p(\Theta|T)$.

Chipman et al. (1996) consider priors which allow for a richer expression of preferences at the price of some additional complexity. The prior $p(T)$ is specified by defining a process on the space of trees T . The process describes how a tree is grown by a series of growth steps analogous to those used in the computation of the posterior. By making the growth steps less and less likely as the tree becomes larger, prior preference for smaller trees is again obtained. In principle just about any type of prior belief can be captured by this approach. For example we may have a preference for certain splitting variables or splitting rules. It is even possible to capture preferences for the type of tree. Bushy trees, in which there are many short branches, may be preferred to trees with a few long branches. In the case where the response is normal given the node, Chipman et al. (1996) also discuss useful prior forms for $p(\Theta|T)$. A hierarchical prior based on the tree structure captures the idea that the mean value should be more similar for “nearby” nodes than for those that are far apart. This yields a version of tree shrinking.

3.3. Object Oriented Programming of Bayesian CART

Chipman et al. (1996) have implemented their approach in C++ using the object oriented features of the language. The basic idea is that it is convenient to think of a tree as a set of nodes and a set of relationships between the nodes. Each node is an object which knows who its parent and children are. A tree is then identified by its top node, since all other nodes can be found recursively. A simplified version of the class definition is:

```
class Node {
public:
    int status[3];
        //0:top 1:bot 2:grand (1=yes 0=no)

    Node *parent; //pointer to parent node
    Node *c_le; //pointer to left child node
    Node *c_gt; //pointer to right child node

    int var; // index of split variable
    int split; // index of split rule

    Parameter theta; // (for bottom nodes)

    void findnode(double *x,Node **n);
        // find node corresponding to predictor x
    void describe(); // describe node
    int num_nodes(); // number of nodes in tree
    void bottoms(List **list);
        //list of bottom nodes
    double logpri_prob(); // log prior of tree
    void draw_prior(); // grow tree from prior
    double node_logprob();
        // loglikelihood of tree
    Node();
    ~Node(){ }
};
```

The Node * pointers tell how the node relates to the rest of the tree by identifying its parent and children. The construction of a tree out of node objects allows for simple addition or removal of nodes. Since a tree is identified uniquely by its top node, lists of trees are simply lists of top nodes.

Some of the tree functions are also included. For example num_nodes counts the number of nodes in the tree below the given node. Using the object structure such of function can be written very simply:

```
int Node::num_nodes()
{
    if(status[1]==1)
        return 1;
    else
        return (c_le->num_nodes() +
                c_gt->num_nodes());
}
```

In words, the function returns 1 if the node is a bottom node, and otherwise returns the sum of the number of nodes possessed by its children. Many functions used are of this simple recursive nature.

4. An Example: Mileage Data

To illustrate some of the issues raised in this article, we present the analysis of a car mileage data set. It

contains mileage and various other characteristics of 392 cars (after missing values are removed). The data are available on the world wide web at the address <http://lib.stat.cmu.edu/datasets/cars.data>.

The response is mileage, in miles per gallon. In this example, we use five predictors: number of cylinders, displacement, horsepower, model year, and weight. Here, the predictors have 4, 10, 10, 10, and 12 equally spaced split points, respectively.

Our prior on trees puts most mass on models with 10 nodes or less. The prior on split variable and split location are uniform, and a relatively uninformative conjugate prior is placed on the mean and variance of the response within each bottom node. The Metropolis algorithm was started 30,000 times, and ran 3,000 cycles for each start. The 100 trees with largest posteriors were stored, and are considered here.

The tree with the most posterior mass has 11 nodes. When the posteriors for the 100 trees are renormalized, the modal tree accounted for 25% of the posterior mass. The top three trees accounted for 50% of the mass, and were the only trees with individual posterior probabilities of greater than 10%.

The greedy algorithm was used to grow a 23 node tree, which was then pruned back to produce an 11 node tree. For comparison, the greedy algorithm was allowed to choose from the same set of splits used in the Metropolis algorithm. The modal and greedy trees turned out to be quite different. For example the greedy tree never used the predictor displacement to split, while the modal tree used it four times. The greedy tree was slightly less “bushy,” in the sense that it had a wider range of terminal node depths. The trees found by our procedure were more bushy because the prior penalizes growth at a deeper level more heavily.

The trees identified as having high posterior probability are compared to a pruned sequence of greedy trees in Figure 2. The vertical axis is the residual sum of squares for each tree, and the horizontal the number of bottom nodes. The line represents the nested sequence of greedy trees, obtained by starting with the largest tree, and pruning the node that reduces the RSS the least. The dots represent the RSS for the trees in our posterior. Notice that all 100 trees in our posterior beat the greedy tree of the same size.

The large number of trees that appear better than the greedy tree suggest that this search technique can be far more effective than the greedy method. Some of these trees are quite different from each other, with different first split points. This suggests that not only can the

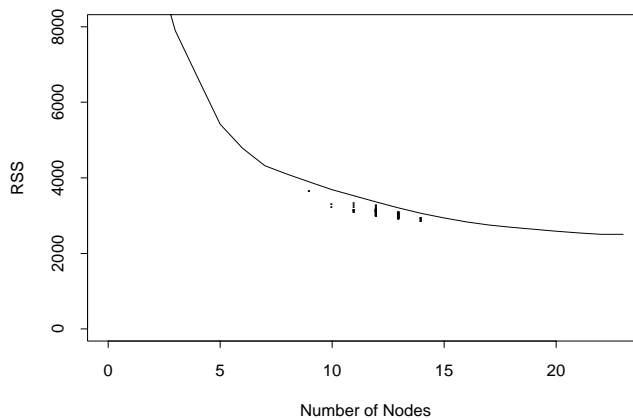


Figure 2. Residual sums of squares for trees of different sizes. Dots represent the trees in the posterior, and the line represents a nested sequence of greedy trees.

greedy tree be beat, but a number of alternate models provide a better fit to the data.

To explore these trees, we use a metric based on predictions. It is defined as the average absolute difference between predicted responses for all observations in the dataset. This metric is presented in Figure 3 in the form of a distance matrix between 10 posterior trees and the greedy tree with 11 nodes. The posterior trees are ordered from bottom to top of the plot by decreasing probability. Each shaded rectangle gives the average absolute difference in predictions for a pair of trees, with dark squares representing greater differences. The range of distances is between 0 and 1.57 mpg. It’s quite evident that the greedy tree is most different from the other trees, although some differences exist between the

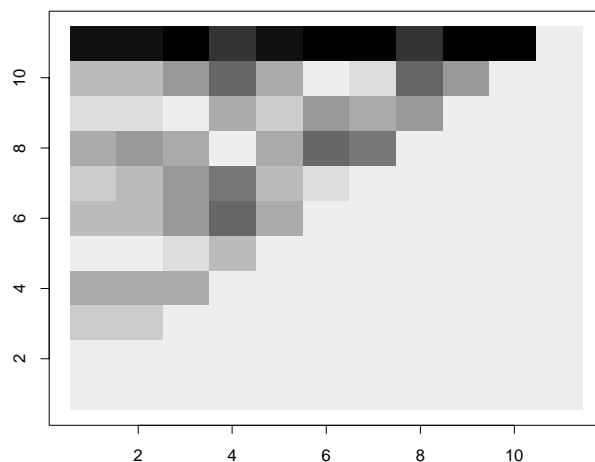


Figure 3. Distance matrix for ten most probable trees and greedy tree (top row). The ten trees are ordered from bottom row to top by decreasing posterior probability.

10 most probable trees. The two most probable trees (1 and 2) are evidently very similar; inspection of the trees indicates that they differ only at a pair of end nodes. The third tree differs more substantially from the first two.

References

Breiman, L. (1994), "Bagging Predictors," Technical Report, University of California – Berkely, Department of Statistics.

Breiman, L., Friedman, J., Olshen, R., and Stone, C. (1984), *Classification and Regression Trees*, Wadsworth.

Chipman, H., George, E.I., and McCulloch, R.E. (1996), "Bayesian CART," Technical Report, University of Chicago.

Clark, L., and Pregibon, D. (1992), "Tree-Based Models," in *Statistical models in S*, J. Chambers and T. Hastie, Eds., Wadsworth.

Denison, D.G.T., Mallick, B.K., and Smith, A.F.M. March (1996) "Bayesian CART," Imperial College, London.

Gelman, A., and Rubin, D. (1992), "Inference from Iterative Simulation Using Multiple Sequences," *Statistical Science*, 7, 457–511.

Hastie, T., and Pregibon, L. (1990), "Shrinking Trees," AT&T Bell Laboratories Technical Report.

Sonquist, J., and Morgan, J. (1964), *The Detection of Interaction Effects.*, Monograph 35, Survey Research Center, Institute for Social Research, University of Michigan.

Tibshirani, R., and Knight, K. (1995), "Model Search and Inference by Bootstrap 'Bumping' ," University of Toronto Technical Report.

Hugh Chipman
University of Chicago
hugh.chipman@gsb.uchicago.edu

Edward George
University of Texas
egeorge@mail.utexas.edu

Robert McCulloch
University of Chicago
rem@gsbrem.uchicago.edu



THIRD FEATURE ARTICLE

Nonparametric regression: a general methodology

By Michael Smith and Robert Kohn

1. Introduction

This paper outlines a general methodology for estimating various nonparametric regression models. Functions that require estimation are approximated by regression splines with many knots. The process is rendered nonparametric by accounting for the distribution of the significant subsets of these knots using a stochastic Bayesian approach. This subset selection scheme, described in Section 2, uses a Markov chain Monte Carlo algorithm that is both faster and has better empirical convergence properties than earlier algorithms found in the literature (see Smith and Kohn, 1996a, for details).

This makes the resulting nonparametric estimator a reliable data driven procedure that is practical to implement. Section 3 discusses this nonparametric estimator in the univariate independent normal error case. It provides a small simulation study against two high performance contemporary estimators that is intended to illustrate the competitive nature of the approach. More comprehensive benchmarking can be found in Smith and Kohn (1996a) and Smith, Sheather and Kohn (1996).

The subset selection procedure can also be extended to more complex regression models, such as where an outlier or autocorrelation process exists in the error structure. Section 4 gives a brief discussion of how this is achieved and outlines how nonparametric estimators that account for such error distributions can therefore be obtained. This section also demonstrates how the approach can be extended to additive nonparametric models with normal, outlier or autocorrelation error processes.

2. Subset selection in linear regression

2.1. Description of the hierarchical model and prior assumptions

This section reviews subset selection in linear regression as it forms the basis of the suggested approach to nonparametric regression. Consider the linear model

$$y = X\beta + e \quad (1)$$

where y is the $n \times 1$ vector of observations, X is the $n \times r$ design matrix, $e \sim N(0, \sigma^2 I_n)$ is the error vector

and $\beta = (\beta_1, \dots, \beta_r)'$ is the $r \times 1$ vector of regression coefficients. Let γ be the $r \times 1$ vector of indicator variables with the i th element γ_i , such that $\gamma_i = 0$ iff $\beta_i = 0$ and $\gamma_i = 1$ iff $\beta_i \neq 0$. Given γ , let β_γ consist of all the nonzero elements of β and let X_γ be the columns of X corresponding to those elements of γ that are equal to one. Therefore, the linear model can be rewritten conditional on γ as

$$y = X_\gamma \beta_\gamma + e.$$

The following prior assumptions are made on the model parameters:

- 1: The prior $\beta_\gamma | \gamma, \sigma^2 \sim N(0, c\sigma^2(X'_\gamma X_\gamma)^{-1})$, where c is a positive scale factor specified by the user. As the conditional prior variance is proportional to the variance of the least squares estimate of β_γ , a large enough value of c ensures that the conditional prior $\beta_\gamma | \gamma, \sigma^2$ contains little information about the location of β_γ compared to the likelihood. Therefore, we are motivated to choose $c = 100$, a value which we use throughout all the empirical work presented in this paper.
- 2: The prior of σ^2 is taken *a priori* independent of γ so that $p(\sigma^2 | \gamma) \propto 1/\sigma^2$. This is a commonly used prior for σ^2 as it makes $\log(\sigma^2)$ uniform.
- 3: The γ_i are assumed to be *a priori* independent with $p(\gamma_i = 1) = \pi_i$, $0 \leq \pi_i \leq 1$, for $i = 1, \dots, r$. In our applications we take the $\pi_i = \frac{1}{2}$ which represents no prior knowledge about whether a variable is included or excluded. This results in the prior probability of any model γ being $p(\gamma) = 2^{-r}$.

2.2. Gibbs sampling from the model

For a given γ , if $q_\gamma = \sum_{i=1}^r \gamma_i$ is the number of nonzero elements of β and

$$S(\gamma) = y'y - \frac{c}{1+c} y'X_\gamma(X'_\gamma X_\gamma)^{-1} X'_\gamma y, \quad (2)$$

then the likelihood

$$\begin{aligned} p(y|\gamma) &\propto \int_{\sigma^2} \int_{\beta_\gamma} p(y, \beta_\gamma, \sigma^2 | \gamma) d\beta_\gamma d\sigma^2 \\ &\propto \int_{\sigma^2} \left\{ \int_{\beta_\gamma} p(y|\beta_\gamma, \sigma^2) p(\beta_\gamma | \sigma^2) d\beta_\gamma \right\} p(\sigma^2) d\sigma^2 \\ &\propto (1+c)^{-q_\gamma/2} S(\gamma)^{-n/2} \end{aligned} \quad (3)$$

Here, β_γ is integrated out as a normal integral and σ^2 is integrated out as an inverse gamma integral. Therefore, the posterior distribution of any possible subset γ is

$$\begin{aligned} p(\gamma|y) &\propto p(y|\gamma)p(\gamma) \propto (1+c)^{-q_\gamma/2} S(\gamma)^{-n/2} \\ &\times \left\{ \prod_{i=1}^r \pi_i^{\gamma_i} (1-\pi_i)^{1-\gamma_i} \right\} \end{aligned} \quad (4)$$

Because γ has support on 2^r distinct values, it is impractical to obtain its posterior by direct enumeration unless r is small. In our applications r is large and we use the following Gibbs sampler (Gelfand and Smith, 1990) to traverse the parameter space in a stochastic manner.

Gibbs sampler for subset selection.

- (i) Choose an initial value $\gamma^{[0]} = (\gamma_1^{[0]}, \dots, \gamma_r^{[0]})$
- (ii) Successively generate from $p(\gamma_i | y, \gamma_{j \neq i})$, $i = 1, \dots, r$.

Step (ii) is iterated many times and in two consecutive stages. The first is a warmup period at the end of which it is assumed that the sampler has converged to the joint distribution of $p(\gamma|y)$. The second stage is a Monte Carlo sampling period producing a sequence of iterates $\gamma^{[1]}, \gamma^{[2]}, \dots, \gamma^{[K]}$ from which parameter inference can be determined (see Section 3.2).

From (3) the conditional probability of γ_i is

$$\begin{aligned} p(\gamma_i | y, \gamma_{j \neq i}) &\propto p(y|\gamma)p(\gamma_i) \\ &\propto \{ \pi_i^{\gamma_i} (1-\pi_i)^{1-\gamma_i} \} (1+c)^{-q_\gamma/2} S(\gamma)^{-n/2} \end{aligned} \quad (5)$$

Because γ_i is binary, the conditional probability $p(\gamma_i | y, \gamma_{j \neq i})$ is obtained by evaluating (5) for $\gamma_i = 0$ and $\gamma_i = 1$ and normalizing.

Notice that due to the use of the point prior formation (where β_i exactly equals 0 if $\gamma_i = 0$) in Section 2.1, calculation of $S(\gamma)$ requires the factorization of a $q_\gamma \times q_\gamma$ matrix, rather than the $r \times r$ matrix that would arise from a non-point prior. As the problems examined here usually have r large and q_γ small (see Section 3), the resulting improvement in speed is often very significant. For implementation details, comparison with previous Bayesian subset selection procedures, such as that by George and McCulloch (1993) and Mitchell and Beauchamp (1988), and an investigation of such speed issues see Smith and Kohn (1996a).

3. Univariate nonparametric regression

3.1 Regression splines

Suppose

$$y_i = f(x_i) + e_i \quad i = 1, \dots, n \quad (6)$$

where y_i is the i th observation, f is a smooth function which is to be estimated and, in this section, e_i is an independent $N(0, \sigma^2)$ error sequence. We propose to approximate $f(x)$ by the cubic regression spline

$$b_0 + b_1x + b_2x^2 + b_3x^3 + \sum_{k=1}^m \beta_k (x - \tilde{x}_k)_+^3 \quad (7)$$

where $\tilde{x}_1, \dots, \tilde{x}_m$ are the m 'knots' placed along the domain of the independent variable x , such

that $\min(x_i) < \tilde{x}_1 < \dots < \tilde{x}_m < \max(x_i)$, while $(z)_+ = \max(0, z)$. Let $r = m + 4$, $\beta = (b_0, b_1, b_2, b_3, \beta_1, \dots, \beta_m)'$, let $\mathbf{x} = (x_1, \dots, x_n)'$, let $\mathbf{1}$ be a $n \times 1$ vector of ones, and let the $n \times r$ matrix $X = (\mathbf{1}, \mathbf{x}, \mathbf{x}^2, \mathbf{x}^3, (\mathbf{x} - 1\tilde{x}_1)_+^3, \dots, (\mathbf{x} - 1\tilde{x}_m)_+^3)$. Then the nonparametric regression problem at (6) can be written as the linear model at (1) where $f(x)$ has been replaced by the cubic regression spline (7).

The most important question associated with fitting such regression splines is the choice of both the number and location of the knots $\tilde{x}_1, \dots, \tilde{x}_m$. If the knots are badly located, details of the curve can be missed, while if too many knots are included the fitted spline based on these knots will have high local variance. One way to solve the problem is to introduce a large number of potential knots (a ‘saturated’ regression spline) from which a significant subset can be selected, e.g. Friedman and Silverman (1989).

The problem then becomes one of subset selection, where each knot corresponds to a column of a design matrix from which a significant subset is to be determined. By applying the Bayesian approach to subset selection outlined in the previous section, uncertainty about the significant knots can be represented by the parameter γ . It is by considering the distribution of this parameter that a parametric cubic regression spline becomes a nonparametric estimator of the function f .

3.2. Estimation

We estimate the function f by approximating the posterior mean $E(f|y)$ by

$$\hat{f}(z) = \mathbf{z}' \sum_{k=1}^K E(\beta|\gamma^{[k]}, y)$$

Here, $\mathbf{z} = (1, z, z^2, z^3, (z - \tilde{x}_1)_+^3, \dots, (z - \tilde{x}_m)_+^3)$ is the vector of regression spline basis terms evaluated at an arbitrary point z . This is then multiplied by the mixture estimate of $E(\beta|y)$, which uses the Gibbs iterates $\gamma^{[1]}, \dots, \gamma^{[K]}$ to integrate out γ in a Monte Carlo manner. Therefore, the function f is a cubic regression spline, but where the distribution of potential knot positions has been smoothed over.

Although the number of knots introduced, m , will typically be large so that r will be large, the number of significant variables, q_γ , required to obtain a good approximation with a regression spline will usually be quite small (see Smith and Kohn, 1996a). In general the potential knot sites we use follow the observed density of the independent variable which helps ensure that a knot is at, or near, the positions required to capture the curvature in the regression function.

3.3. Benchmarking

To benchmark this procedure to construct the Bayesian estimator we conducted a small univariate simulation, comparing it to a global bandwidth kernel smoother (Ruppert, Sheather and Wand, 1995) and a locally adaptable kernel smoother (LocFit by Clive Loader). Two nonparametric regression models were used, one for each of the two functions $f_1(x) = \cos(4\pi x)$ and $f_2(x) = \phi(x, 0.15, 0.05)/4 + \phi(x, 0.6, 0.2)/4$ where $\phi(x, a, b)$ is a normal density with mean a and standard deviation b , evaluated at x . The standard deviation of the model errors was $\sigma = 1/2$ and $n = 100$. One hundred data sets were generated from each model, with $x \sim \text{Uniform}(0, 1)$. Both kernel smoothers, plus the Bayesian estimator proposed in this paper, were fit to each data set. The integrated squared error criterion $ISE = \frac{1}{800} \sum_{i=1}^{800} (f(z_i) - \hat{f}(z_i))^2$, where $z_i = i/800$, was calculated as a quantitative measure of the quality of each fit. Boxplots of the ISE values are given in panels (a) and (b) of Figure 1 and reveal the competitive nature of the approach compared against the two kernel smoothers. To visually indicate the nature of the fits, panels (c) and (d) plot the median (as ranked by ISE) fit for the Bayesian estimator, along with the two kernel fits to the corresponding data. Panel (d) illustrates the lack of ‘local adaptability’ in the global bandwidth smoother. This is because f_2 requires a different degree of smoothing on the left of the domain of the independent variable, compared to the right hand side. Both LocFit and the Bayesian estimator are locally adaptive, producing a smooth fit across all the domain.

It should be noted here that LocFit requires an initial estimate of σ , something we set to the true value of $1/2$. The fits were created using the Bayesian regression program, (br), available from the World Wide Web Statlib repository at <http://lib.stat.cmu.edu/S/br>

4. A general nonparametric methodology

A crucial strength of the nonparametric estimator outlined in the previous section is that it is easily extended to a wide variety of complex regression models. This is achieved by modeling the additional processes involved and constructing an appropriate Markov chain Monte Carlo scheme to estimate the model. The following sections briefly outline various estimators that can result.

4.1. Robust nonparametric regression

Sometimes the errors in a regression model do not appear to follow the normal process suggested in Section 2, with outlying values apparent in the data. One popular Bayesian approach to robustifying the method in

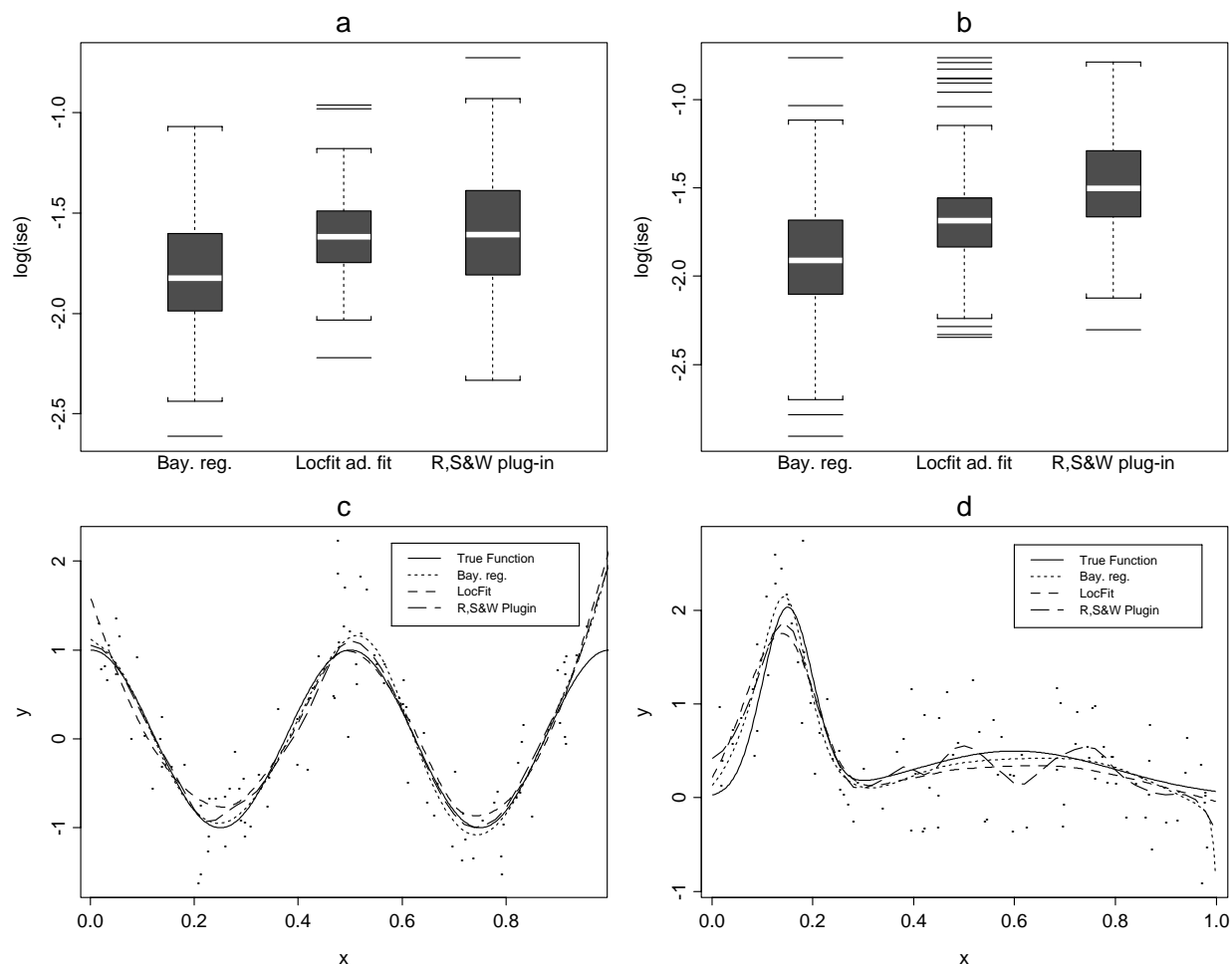


Figure 1. Panels (a) and (c) correspond to f_1 , while (b) and (d) are for f_2 . The boxplots are of the $\log(\text{ISE})$ for (from left to right) the Bayesian estimator, LocFit and the global bandwidth based kernel estimator. The scatter plots are of the data corresponding to the median (ISE) fit for the Bayesian estimator. This fit, the true function and kernel estimates from the same data are also plotted.

Section 2 to such outliers would be to model the errors with a mixture of two normals (Box and Tiao, 1968). Binary indicators $\omega = (\omega_1, \dots, \omega_n)'$ can be introduced to represent whether, or not, each observation was an outlier. If so, then the error can be defined as resulting from a distribution with a large variance. For example, where $e_i \sim N(0, \omega_i \sigma^2)$, with $\omega_i = 100$ if e_i is an outlier, and $\omega_i = 1$ if it is not. This approach appears very successful in tackling a large range of outlier problems, challenging the small sample performance of traditional robust regressors in a linear model (see Smith, Sheather and Kohn, 1996).

The key to accommodating this approach within the hierarchical model found in Section 2, is to develop an analogous conditional prior for β_γ . Smith and Kohn (1996a) suggest using

$$\beta_\gamma | \gamma, \sigma^2, \omega, y \sim N(0, c\sigma^2(X_\gamma' \Omega_\omega^{-1} X_\gamma)^{-1})$$

where $\Omega_\omega = \text{diag}\{\omega\}$. This utilizes the variance of the

generalized least squares estimator of β_γ , blown up by the scale factor c .

Using the other prior assumptions found in Smith and Kohn (1996a), a Markov chain Monte Carlo scheme can be constructed. Smith, Sheather and Kohn (1996) illustrate how this sampler can be implemented in an efficient manner and a Monte Carlo sample $\gamma^{[1]}, \omega^{[1]}, \dots, \gamma^{[K]}, \omega^{[K]}$ from $\gamma, \omega | y$ obtained. If a saturated regression spline, of the type outlined in Section 3.1, is coupled with this robust subset selection procedure then a robust nonparametric estimate of f at (6) is given by

$$\hat{f}(z) = \mathbf{z}' \sum_{k=1}^K E(\beta | \gamma^{[k]}, \omega^{[k]}, y)$$

Notice that f is obtained by smoothing over the distribution of the outlier process ω and uncertainty of knot significance γ .

Smith and Kohn (1996a) show the favorable empirical performance of this estimator when compared to a robust local quadratic nonparametric estimator, while Smith, Sheather and Kohn (1996) provide a similar comparison against quantile smoothing splines.

4.2. Nonparametric regression with autocorrelated errors

The methodology can also be extended to the case where the errors in the regression are potentially autocorrelated of order R . That is, where $e_i = \sum_{j=1}^R \theta_j e_{i-j} + \tilde{e}_i$, $\tilde{e}_i \sim N(0, \sigma^2)$ and $\theta_1, \dots, \theta_R$ are the autoregressive parameters. Assuming stationarity, the autocorrelation can be reparameterized in terms of its partial autocorrelations $\psi = (\psi_1, \dots, \psi_R)'$ and the variance of $e = (e_1, \dots, e_n)'$ written in the form of a banded matrix, so that $\text{var}(e) = \sigma^2 \Omega_\psi$ (see Barnett, Kohn and Sheather, 1996). Here, Ω_ψ is a function of only the data and ψ , allowing for a generalization of the conditional prior for β_γ in Section 2, so that

$$\beta_\gamma | \gamma, \sigma^2, \psi, y \sim N(0, c\sigma^2(X'_\gamma \Omega_\psi^{-1} X_\gamma)^{-1})$$

Thus, along with minor additional prior assumptions found in Smith, Wong and Kohn (1996), the hierarchical model in Section 2 can be adjusted to incorporate an autocorrelated error structure. A Markov chain Monte Carlo scheme can be implemented and a sample $\gamma^{[1]}, \psi^{[1]}, \dots, \gamma^{[K]}, \psi^{[K]}$ obtained from the joint distribution $\gamma, \psi | y$ (Smith, Wong and Kohn, 1996).

Again, if the linear model is a saturated regression spline, then a nonparametric estimate for f at (6) that smooths over the distribution of the autocorrelation process ψ and uncertainty of knot significance γ , is given by

$$\hat{f}(z) = \mathbf{z}' \sum_{k=1}^K E(\beta | \gamma^{[k]}, \psi^{[k]}, y)$$

Smith, Wong and Kohn (1996) show how this procedure can be extended to incorporate estimation of the autocorrelation order R . They also demonstrate that when $x_i = i$ in equation (6) (that is, the independent variable is the discrete index itself), then the procedure is competitive against alternative nonparametric estimators, such as that forwarded by Hart (1994).

More importantly, it is noted that this method provides a nonparametric estimator for the hitherto unaddressed case where x_i is a general variable, with no restriction with regard to the index i . It is also noted that, as outlined in the following section, the methodology provides the first comprehensive solution to the additive nonparametric autocorrelated error problem.

4.3. Additive models

Consider a multivariate nonparametric regression with an additive structure, so that

$$y_i = \alpha + f_1(x_{1i}) + f_2(x_{2i}) + \dots + f_p(x_{pi}) + e_i$$

Here, x_{1i}, \dots, x_{pi} is the i th observation of p independent variables and f_1, \dots, f_p are signals to be estimated simultaneously in a nonparametric fashion. If each component f_i is modeled using a saturated regression spline basis of the type discussed in Section 3, then the columns of the resulting design matrices can be collected together to form a single design matrix, but with a single global intercept. Application of the subset selection to a linear model with such a design matrix results in the simultaneous estimation of the components f_1, \dots, f_p . This is because the distribution of γ now represents the joint distribution of knot uncertainty for all components.

Such estimation can be accomplished under any of the three different error distributions discussed. Smith and Kohn (1996a) and Smith, Sheather and Kohn (1996) apply this additive estimator to real data under the assumption of normal errors (Section 3) and a mixture of two normals (the robust estimator discussed in Section 4.1); Smith, Wong and Kohn (1996) do so for additive models with autocorrelated data (as discussed in Section 4.2).

4.4. Further extensions

Further extensions of the methodology are possible, such as including estimation of a data transformation (Smith and Kohn, 1996a), or application to different function bases, such as B-splines. Extension to multivariate function bases, so that nonparametric multivariate surfaces can be estimated without imposing additivity, is also possible (Smith and Kohn, 1996b)

Acknowledgments

Robert Kohn's work was partially supported by an Australian Research Grant. We thank Matt Wand and Clive Loader for making their programs available to us.

References

- Barnett, G., Kohn R., and Sheather S. (in press), "Bayesian estimation of an autoregressive model using Markov chain Monte Carlo," *Journal of Econometrics*.
- Box, G.E.P., and Tiao, G.C. (1968), "A Bayesian approach to some outlier problems," *Biometrika*, 55, 119–129.
- Friedman, J.H., and Silverman, B.W. (1989), "Flexible parsimonious smoothing and additive modeling," *Technometrics*, 31, 3–39.
- Gelfand, A.E., and Smith, A.F.M. (1990), "Sampling-based approaches to calculating marginal densities,"

Journal of the American Statistical Association, 85, 398–409.

George, E.I., and McCulloch, R.E. (1993), “Variable selection via Gibbs sampling,” *Journal of the American Statistical Association*, 88, 881–889.

Hart, J.D. (1994), “Automated kernel smoothing of dependent data using time series cross-validation,” *Journal of the Royal Statistical Society, Series B*, 56, 529–542.

Loader, C, *LocFit*, available at <http://cm.bell-labs.com/stat/project/locfit/adap.html>

Mitchell, T.J., and Beauchamp, J.J. (1988), “Bayesian variable selection in linear regression,” *Journal of the American Statistical Association*, 83, 1023–1036.

Ruppert, D., Sheather S., and Wand M. (1995), “An effective bandwidth selector for local least squares regression,” *Journal of the American Statistical Association*, 90, 1257–1270.

Smith, M., and Kohn, R. (in press), “Nonparametric regression using Bayesian variable selection,” *Journal of Econometrics*.

Smith, M., and Kohn, R. (1996b), “Nonparametric bivariate regression,” Manuscript.

Smith, M., Sheather S., and Kohn R. (in press), “Finite sample performance of robust Bayesian regression,” *Computational Statistics*.

Smith, M., Wong C., and Kohn R. (1996), “Additive nonparametric regression with autocorrelated errors,” under revision for *Journal of the Royal Statistical Society, Series B*.

Mike Smith
Department of Econometrics
Monash University
mikes@smith.ecom.monash.edu.au

Robert Kohn
Australian Graduate School
of Management
R.Kohn@unsw.edu.au



TOPICS IN INFORMATION VISUALIZATION

Plot Interpretation and Information Webs: A Time-Series Example From the Bureau of Labor Statistics

By Daniel B. Carr, Richard Valliant, and Daniel Rope

1. Plot Interpretation, Metadata, and Software

Plot interpretation is an interesting topic because the ability to interpret plots is a measure of scientific literacy and because plots themselves can be informative. Tufte (1983) observes that the popular media in the U.S. considers the scatterplot beyond the common reader. Noting that some elementary school curricula include Cartesian plots, and even box and whisker plots, there is hope that the statistical literacy movement will eventually force a well-justified reassessment.

There is one form of scatterplot that routinely appears in the newspapers, the times series plot. This implies that when one variable is time something magical happens and people, especially those with economic interests, can understand. Section 2 in this paper suggests that when the other variable is measured in floating units, such as dollars, understanding the times series at more than a superficial level requires background information.

Interpretation of times series is problematic when articles suppress required metadata. Popular articles typically advocate a position and strive to be entertaining. Guidance from Entertainment 101 reads: “How to entertain with statistics: Omit 95%.” Too often the role of statistics in an article is to suggest credibility and little more. Actual content may not be required. For the omissions one can chose from the seven basic building blocks of background information (metadata): who, what, where, when, why, how, and how well (W^5H^2). Who and why are important to omit whenever the answers raise issues of bias. How and how well are extremely important to omit in an advocacy context because they so often raise the undesirable issues of representativeness, complexity and uncertainty. In fairness, the motivation behind omissions may be brevity more than it is advocacy, but the negative effect can still be the same.

Publications at the federal statistical agencies often provide a striking contrast to popular media articles. At the federal agencies appropriate interpretation of plots is important and publications can devote significant space to metadata descriptions. However, even in lengthy government publications, the metadata description is often provided in text that is separated from summary plots and tables. This puts agencies at risk of having their graphics pulled out of context. It is easy to copy a plot and reproduce it without all the surrounding documentation. The article including the graphic will likely cite the source agency as part of the credibility gambit. Few are likely to pursue the information trail to the metadata to see if the graphic really supports the article.

Attaching metadata to every plot is problematic. In the static document context, Carr (1994a) suggests attaching icons that provide immediate warnings or serve as a cue for finding additional description. The emerging world of interactive documents provides additional alternatives that allow brevity and information access. To provide access to metadata and better graphics at the Bureau of Labor Statistics (BLS), Dan Rope and Dan Carr are developing a Java™-based Graphics Production Library (GPL). The resulting software provides metadata access through mousable metadata icons.

Historically BLS developed a table producing language (TPL). If there is an analogous graph producing language for the federal agencies, it is spreadsheet graphics. Spreadsheet graphics have remained to a large extent 1970's business graphics. Since the 1970's the statistical graphics community has provided much guidance about graphics and developed replacements for inferior graphics. Unfortunately this has had limited effect. For example the elegant dot plots described by Cleveland (1985, 1993a) remain hard to find in either the public media or government publications.

While education is important, the biggest barrier to routine use of preferable graphics is the limited availability of convenient software. We hope that the Java™-based GPL will provide easy access to sound graphics and that use of inferior graphics, such as 3-D bar plots, will diminish. A prototype applet derived from the GPL is scheduled for completion in August. What appears here is suggestive but preliminary. The carefully designed object-oriented library facilitates revision and the software will evolve as BLS statisticians, economists, cognitive scientists and icon designers provide input.

In what follows, Section 2 discusses the web of information behind a BLS time series and provides some interesting templates for time series graphics. Section 3 describes current web access to BLS series. Section 4

returns to the topic of metadata and provides an example of an applet with metadata access through icons and user interactivity.

2. Toward Understanding a BLS Time Series

Time series at BLS can be quite complicated. Computational techniques include adjusting series at benchmark periods, weighted averaging, seasonal adjustments, intervention analysis, scaling, and lagged ratioing (e.g. percent change over 12 months). Behind all this lies sampling methodology that determines the benchmarks, the weights, and other adjustments. Much of this is embedded in the time series shown in Figure 1. This figure, adapted from Employment Cost Indexes and Levels, 1975-1995, shows two series. The series are quarterly values depicted as occurring at mid-March, mid-June, mid-September and mid-December. The omission of values prior to 1980 provides resolution in subsequent monthly plots. The first series in the plot is marked "Current" and gives the 12 month change in wages and salaries for people in private industry as expressed in current dollars. The units are percent change and this calculation involves lagged ratioing.

Tufte (1983) warns that a series expressed in dollars can deliver that wrong message unless it is re-expressed in terms of constant dollars. The second series in Figure 1 is reexpressed in constant dollars. The message is quite different. Although the percentage changes in current dollar salaries are always positive, ranging from about 3% to 9%, the changes in constant dollars are often negative because of high inflation. The current dollar change of 9% in the first quarter of 1980 is actually more than a 4%

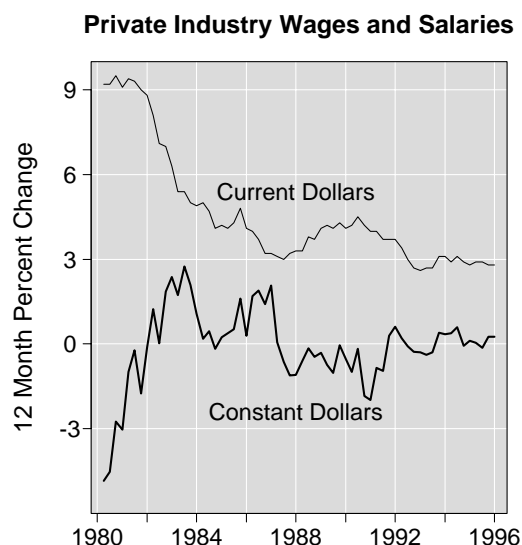


Figure 1: Simple Time Series?

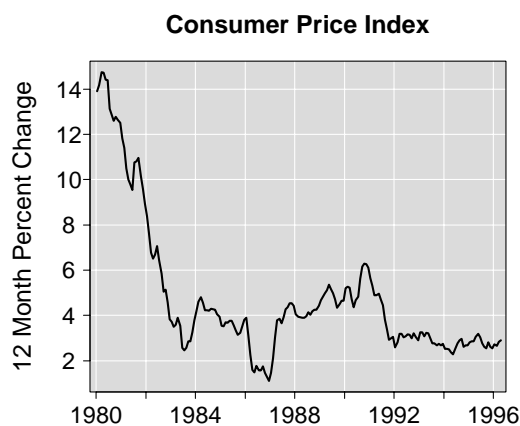


Figure 2: Key Series For Conversion To Constant Dollars

decrease in real dollars. The 1980 to mid-1983 period of deceleration in current dollar pay raises translates into a period of accelerating constant dollar pay changes.

Figure 1 suggests that the constant dollar salaries are pretty flat in the recent years. Of course this may not hold for all subsets of people, such as chief executive officers (CEOs). The decomposition of the wages and salaries series is of interest but the focus here is on the adjustment to constant dollars.

The adjustment to constant dollars involves scaling the salary series (before lagged ratioing) by the Consumer Price Index (CPI). Figure 2 shows the 12 month change in the CPI with values depicted at the middle of each month. Note the variability of the CPI as compared to wages and salaries. Since the CPI is variable and provides the basis for scaling to constant dollars, interested readers will want to know how the CPI is derived.

2.1 The CPI Subseries and Graphical Design

The CPI is the weighted average of seven major subseries. Figure 3 shows the CPI, the weights and the seven subseries. (These series are not seasonally adjusted—a procedure described in Section 2.2.) The bar plots to the left of the series show the weights. For example the housing cost has the largest weight of around 42%. In terms of graphic design, Figure 3 is a derivative of the row-labeled plots discussed in Carr (1994a). The design places the summary (the CPI series) first and places the subseries with the largest weights closest to this series. The figure communicates this graphically via the bars in the bar plots.

Figure 3 shows the time series as filled polygons. The fill is relative to the value 100. This calls attention to that value and the time period when the series were close to 100. As suggested by the plot, the benchmark-

ing for the series occurred during the time period 1982 to 1984 where indexes were set to 100. Increases in the index reflect the relative increase in price for benchmark goods. The grid lines in the plot help in judging values and making comparisons (see Cleveland 1993b and Carr 1994b). Adding the polygon fill risks hiding the useful grid lines below the series. Figure 3 gets around this by using a translucent fill. That is, the grid lines show through the polygon fill.

Perhaps the most controversial design feature is the inclusion of text in the bar plots. Common media practice is to include both graphics and text in the attempt to communicate to both right brain and left brain, respectively. Unfortunately text is a distraction to right-brained pattern perception. The design here works towards a compromise by placing the text in the top right of the graph. Sorting of bars in decreasing order typically leaves this space free. An earlier version of this figure drew strong attention to the text by plotting white letters on a black background. Figure 3 attempts to reduce the text contrast with the rest of the graph. The box surrounding the text still draws undesirable attention, but omitting the box risks having the text misinterpreted as a special grid line label. The purest approach, of course, is to remove the text from the plot and put it with the series label.

The graph design provides grid labels for the CPI bar plot, but avoids repeating the labels for the identically scaled bar plots in the seven subseries. The reader may be reluctant to assume the scales are identical at first glance, but the similarity of plots and written values for bar heights provide strong evidence of identical scales. The inference that the highlighted bar represents the weight for the series should be straightforward. The grid-line labels are on the right for the time series. This facilitates reading the most recent values that are likely to be of greatest interest.

The series in Figure 3 show several interesting patterns. The increase in medical costs is dramatic and the subject of considerable media attention. However, the weight is small compared to that of housing. Controlling the cost of housing would have more effect on the CPI.

2.2 Seasonality

The wiggle in the Apparel and Upkeep Series draws attention. There are four cycles between the two year grid lines indicating two cycles per year. The first dip is in January when after-holiday sales are held. Prices then rise to a peak, typically in April, as winter fashions are replaced by ones for spring and summer. The second valley is in July when summer fashions are being phased out. The second peak then occurs in October or November as stores stock up on winter clothing.

The CPI and Seven Component Series

Indices = 100 in Reference Period 1982-1984

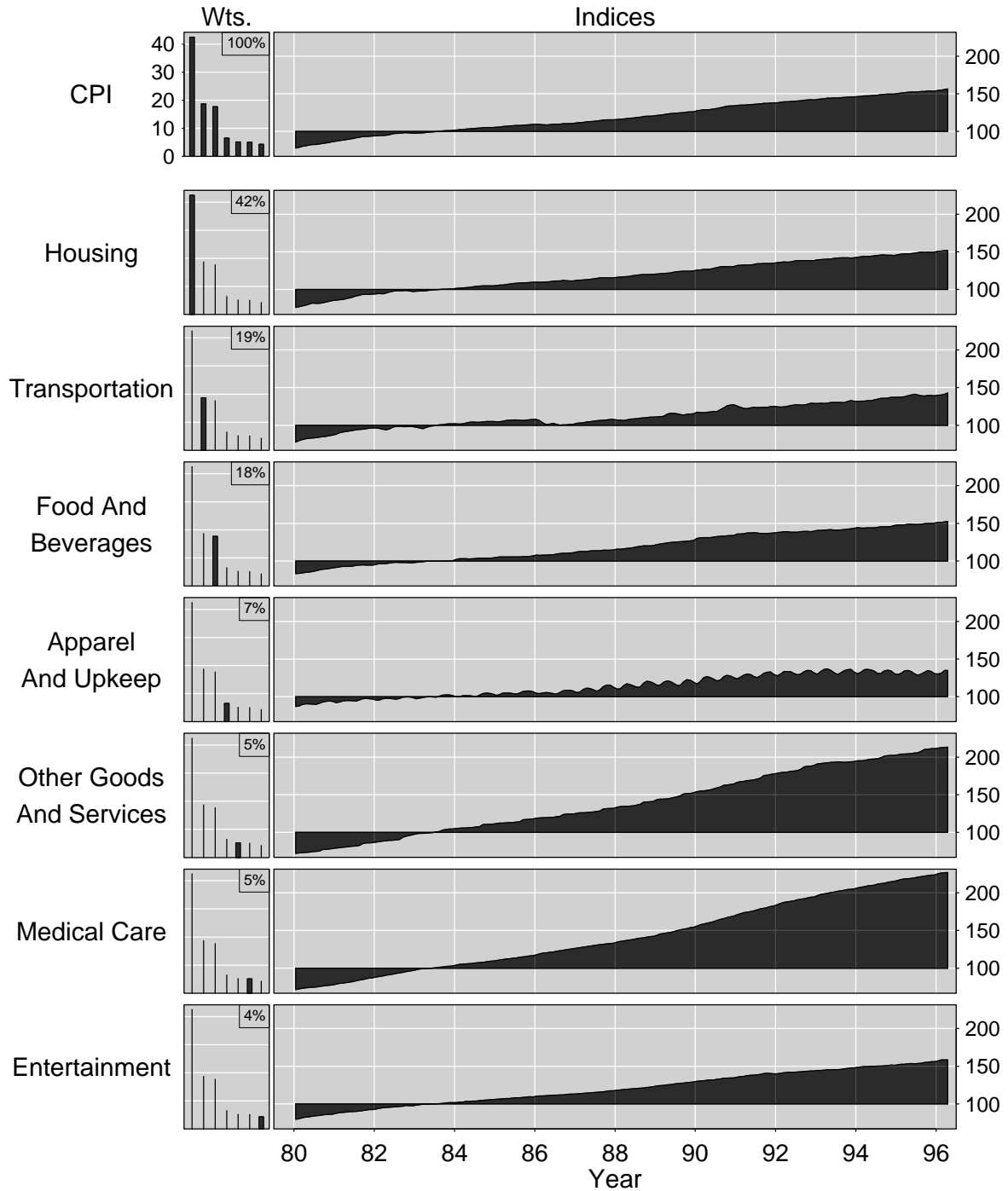


Figure 3: Multiple Time Series
With Relative Weights

Many BLS series have regular hops or skips that can be predicted with some accuracy. The unemployment rate, for example, rises every year in early summer as more students try (and fail) to find jobs. Prices drop in late December and early January for clothing as stores run after-holiday sales. Gasoline prices typically jump up in May or June when people begin to do more vacation driving. Price changes that normally occur about the same time and in about the same magnitude every year do not usually signal any general change in economic trends. For that reason, data series that have these seasonal fluctuations removed may be better for some analyses. Seasonal adjustments are additive or multiplicative factors applied to time series to reduce or eliminate the effects of regularly occurring short-term blips. Whether or not series have been seasonally adjusted is important metadata for the appropriate interpretation of BLS series.

2.3 Intervention Analysis and Seasonal Adjustment

The Transportation series has a distinct nonseasonal hump in the latter half of 1990. This occurred shortly after Iraq invaded Kuwait. The rise in prices was mainly associated with uncertainties in the supply of motor fuel and oil. From July through November, 1990, the Transportation index rose 7.2% but then decreased 3.9% from December, 1990, through April, 1991, as supply stabilized.

Other nonseasonal economic phenomena can also happen that cause a shift in level in a time series that is more long-lasting. Sales tax increases or oil embargoes are examples. Before doing seasonal adjustment, these level shifts are deleted from a series in a process called intervention analysis. Seasonal adjustment is then done on the series after which the level shift is injected back into the series.

2.4 Further Series Decomposition

Each of the seven series are decomposed into more detailed series that are also available through the BLS Homepage as indicated in Section 3. Food and Beverages, for example, is broken down into these subseries:

Food

Food at home

Cereals and bakery products

Meats, poultry, fish, and eggs

Dairy products

Fruits and vegetables

Other food at home

Sugar and sweets

Fats and oils

Nonalcoholic beverages

Other prepared food

Food away from home

Alcoholic beverages

2.5 Prices for the CPI

The inputs to the decomposed subseries are individual items for which prices can be determined. BLS collects prices for the CPI through a set of sample surveys that are probably some of the more complex surveys conducted by the Federal government. BLS collects prices in 85 urban areas across the country for about 7,300 housing units per month and 22,500 retail establishments-department stores, supermarkets, hospitals, gasoline stations, and other outlets that sell goods and services. The items that are priced are selected in several stages including probability sampling of geographic areas, establishments, and specific items within establishments. Prices on most of the sampled items are then collected by personal visit or telephone either monthly or bimonthly for up to five years, at which time new samples are rotated-in for each geographic sample area.

The fact that the CPI is a sample survey means that the time series BLS publishes are sample estimates and, consequently, subject to variability. How to graphically portray the variance or standard errors of time series is a problem we have not yet considered, but one that definitely needs attention.

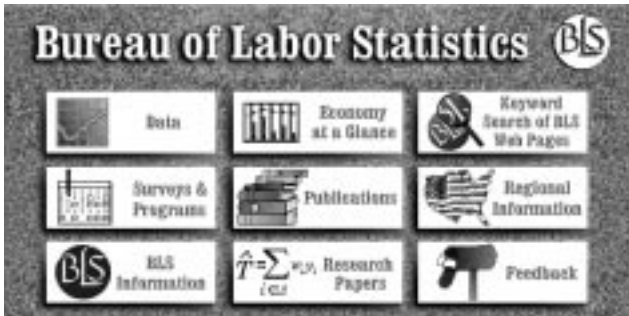
2.6 Relative Importance of Subseries

BLS publications often list the “relative importance” of each subseries. To compute the current relative importance, the procedure is to multiply each subseries benchmark weight by the current subseries index and normalize so the weights sum to one. One criticism of the CPI is that the benchmark weights in Figure 3 are not updated often enough to reflect changes in consumers’ buying habits. As the relative price for gasoline falls over time, for example, consumers may increase their purchases and, as a result, gasoline’s share in consumer budgets may exceed its benchmark weight in the CPI. The benchmark weights are updated every 10 years and will next be revised in 1998 as part of a major, regularly scheduled revision of the CPI.

3. Web Access to BLS Time Series

The Bureau of Labor Statistics (BLS) is the principal fact-finding agency for the Federal Government in the area of labor economics and statistics. It has a dual role as the statistical arm of the Department of Labor and

as an independent national statistical agency that collects, processes, analyzes, and disseminates economic and statistical data to the American public, Congress, other Federal agencies, State and local governments, business, and labor. In addition to the wage, salary, and consumer price data already mentioned, BLS data covers the areas of employment and unemployment, producer prices and prices of internationally traded good, compensation and benefits for productivity and technology, and employment projections.



BLS Homepage.

Much of the published data is available on the BLS Homepage (<http://stats.bls.gov>) along with links to general information on the Bureau, lists of paper publications, research papers, and other links. Users can retrieve data through their web browsers or gopher clients. For example, the unemployment rate for every month since January 1948, the monthly Consumer Price Index from 1913 to the present, and quarterly export and import price indexes beginning in 1983 are all retrievable time series published by BLS.

4. Future Graphics at BLS

BLS is not content with the current publications of tables and network distribution of data. BLS also seeks to provide user access to good graphical presentation of its time series and to provide more convenient access to metadata. The work in progress includes the development of a Graphics Production Library.

4.1 The Graphics Production Library

The GPL stems from a need for convenient tools for producing quality statistical graphics. Feedback from seminars shows that there is interest in applying new ideas for statistical graphics, but as suggested earlier, the reality is that people will use what is most convenient to create statistical graphs. The problem of producing quality graphics worsens when these graphics are included in World Wide Web pages. The rendered bitmaps are often barely legible on low resolution computer screens. To solve both of these problems at once we are developing platform-independent, extensible Java™ class libraries

for producing graphics.

The library itself is composed of a set of reusable components (objects) that encapsulate drawing graphics. These objects can be “glued” together to create applets or applications, and the objects’ underlying methods are used to change high level behavior and appearance. Applets are miniature software applications that run within Web browsers, and applications are standalone software products.

Beyond the ability to draw directly on the user’s computer screen, one of the most significant advantages of this approach is that we can introduce a whole new level of user interaction with statistical graphics. The user will be able to directly manipulate the graphic to more quickly answer his or her questions about the data. In a sense, the graphic will become the interface, since the interface is the graphic. For instance, as illustrated in Figure 4, moving the mouse over a data point yields its value in the extreme lower left corner of the display. Other ideas include:

- The ability to drag & drop actual time series from different panels onto one panel to facilitate comparison.
- The ability to drag panels around with the mouse to reorder them.
- Providing windows that pop-up with more information about a data value (such as metadata or hyperlinks to further information) when the user clicks on a particular data point.
- The ability to navigate through the data by dragging or stretching the dark gray box located in the “Pan & Zoom” control.
- The ability to highlight points of interest. For instance, clicking the “A” in the “Highlight Month” control in Figure 4 yields a small tic on every April in all the time series.

Other advantages of using Java™ as the development language include database connectivity and object embedding. As part of the Java™ Enterprise API, the JDBC™ (Java Database Connectivity) will allow for applets with seamless connectivity to existing proprietary database formats including Sybase, Oracle, and ODBC compliant databases. The proposed Java™ “Beans” API will allow applications to be embedded within other software products. So, for example, graphics applications built using the library could be embedded directly within popular word processing packages. Our plan is to have an initial prototype of an applet ready sometime during August for feedback. For more information, see http://www.mnsinc.com/dan_ropo/gpl

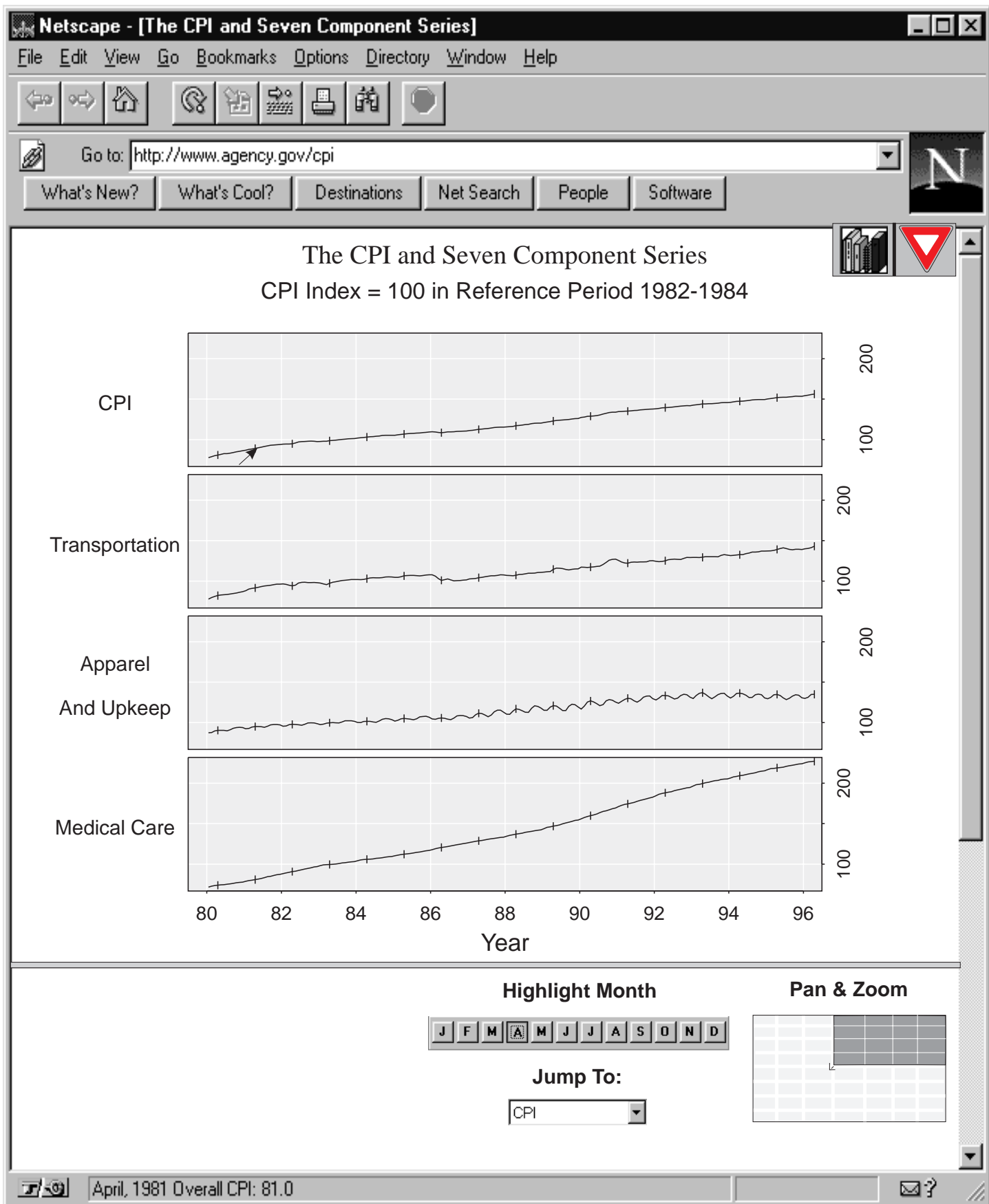


Figure 4. An example of a possible applet built using GPL

4.2 Metadata Icons and Access

Figure 4 includes just two metadata icons. The first is a set of books that provide access to a technical guide to BLS procedures, the BLS Handbook of Methods (1992). The second is a yield sign that provides access to interpretation warnings. Another icon might call attention to features in the graphics. Experience suggests that people sometimes miss basic features in the graphics. Spotting the basic features is a function of graphical experience, background in the subject matter, motivation, and time.

BLS has many types of metadata to represent. Sarndal, Swensson and Wretman (1992) discuss quality declaration for survey data. They suggest providing quality descriptors for a number of areas including:

- Coverage
- Sampling Error
- Response Rates
- Comparability over Time

- Benchmarking and Revisions
- Comparability with Other Data Sources
- Total Variance and/or its Components by source
- Non-response Bias

- Response Bias
- Edit and Imputation Effects
- Seasonal Adjustment

We do not elaborate the descriptors here. Suffice it to say, that providing access to such descriptors is important and the BLS research is addressing the topic.

5. Access and Comments

Data, Splus functions and script files producing the first three figures are available via anonymous ftp to galaxy.gmu.edu. Change directory to `submissions/newsletter/bls`. The page layout function, text functions, and panel access function called `talk()` may be of interest to Splus users. Examples from other newsletter articles are also stored under the newsletter directory. As always, I (Dan Carr) welcome constructive comments plus new graphics challenges.

Acknowledgments

The current research was supported by an ASA/NSF/BLS fellowship. It builds on methodology developed under EPA cooperative research agreement CR820820-01-0 with George Mason University. Thanks go to Sallie Keller-McNulty who suggested improvements to Figure 3.

References

- Carr, D.B. (1994a), "Converting Plots To Tables," Technical Report No. 101, Center for Computational Statistics, George Mason University, Fairfax Va 22030.
- Carr, D.B. (1994b), "Using Gray in Plots," Statistical Computing & Graphics Newsletter, Vol. 5, No. 2, pp. 11-14.
- Cleveland, W.S. (1985), *The Elements of Graphing Data*, Monterey CA: Wadsworth Advanced Books and Software.
- Cleveland, W.S. (1993a), *Visualizing Data*, Summit NJ: Hobart Press.
- Cleveland, W.S. (1993b) "A Model for Studying Display Methods of Statistical Graphics," *Journal of Computational and Graphical Statistics*, Vol 2, No. 4, pp. 323-343.
- Sarndal, C.E., Swensson, B., and Wretman, J., (1992), *Model Assisted Survey Sampling*, New York, N.Y.: Springer-Verlag.
- Tufte, E.R. (1983), *The Visual Display of Quantitative Information*, Cheshire CT.: Graphics Press.
- U.S. Department of Labor, (1992), *BLS Handbook of Methods*, Washington, D.C.: Government Printing Office.
- U.S. Department of Labor, (1995), *Employment Cost Indexes and Levels, 1975-1995*, Washington, D.C.: Government Printing Office.

Daniel B. Carr
George Mason University
dcarr@voxel.galaxy.gmu.edu

Richard Valliant
Bureau of Labor Statistics
vallianr@ore.psb.bls.gov

Daniel Rope
Bureau of Labor Statistics
rope_d@bls.gov



Why Perl?

By Phil Spector

In an ideal world, statisticians could focus on statistical problems, without having to worry about getting programs to communicate with each other, extracting information from overly verbose output or rearranging data files to correspond to the needs of a particular program. Unfortunately, these kinds of problems do occur, and they are often crucial to the success of a project. Furthermore, with the increasing spread of the World Wide Web, many people are finding a need to write programs processing interactive web pages or using interprocess communication, as well as providing easier-to-use front ends for other programs. The perl programming language, developed by Larry Wall in the late 1980s, is emerging as the tool of choice for these tasks and many others. In this article I'm going to try to explain why perl is so popular – and so daunting to learn. I'll also provide some online resources which you may find helpful in your study of perl. In the next article in the series, I'll focus on some more advanced applications of perl.

Why People Love/Hate perl

A quick glance at the perl manual page or at some perl programs makes it clear that perl has borrowed quite a bit from many other languages. Its overall command structure is strongly reminiscent of the C programming language, while at the same time it contains many features generally only found in shell programming languages. But perl is neither an imitation of some other language or a slave to the latest programming trends.

Accordingly, it has some features which may not be familiar to even an experienced programmer. For example, perl functions can be called without arguments and parentheses are optional. When a function is called without arguments, the default argument is a variable known as "\$_", which is generally the most recently read line from an input source. Although it takes some time to get used to these sorts of rules, it makes it very simple to perform many tasks. For example, to print each line of a file which is longer than 80 characters, the following one line perl program could be used:

```
perl -ne 'print if length > 80' filename
```

where filename is the name of the file in question. The -n flag tells perl to read each line of the file in turn, putting the current line's contents in \$_, and the -e flag tells perl to read its program from the command line, not from a named file. Finally, the snippet of code tells perl

to print those lines from file filename whose length (actually length(\$_)) is greater than 80 .

Another unusual operator is the input operator, which, in its simplest form is a pair of angle brackets (<>). When this operator is "called" it reads a line of input into \$_. Thus, a perl program like the following can be quite unnerving to an unsuspecting programmer:

```
while(<>){
    chop;
    @T = split;
    print if $T[2] eq "one";
}
```

The program keeps reading until it reaches an end-of-file, removes the newline from each line (chop), breaks the line up into words (stored in the array @T), and prints the line if the third word was "one". (Perl arrays start at index 0, not 1). An additional feature of the input operator is that it is very smart about where it gets its input. If you invoke the program with no command line arguments, the operator reads from standard input, but if you list filenames on the command line, it will open each of the files in turn. Getting familiar with these features is well worth the effort, because you'll then have a single extensible programming tool which can meet all your scripting, text processing and network communication needs.

Prefix	Meaning	Example
none	built-in function	@T = split;
&	subroutine	\$result = &mysub(\$x);
\$	scalar	\$color = "red";
@	array	join(" ",@values);
%	associative array	@names = keys(%info);

Table 1. Prefixes of perl identifiers.

Every identifier (except for built in functions and commands) must be preceded by a symbol to identify its type to the perl interpreter. These different prefixes are displayed in Table 1. Each type of object has its own namespace. That means that there can be an array, a subroutine, a scalar and an associative array all named x; perl can tell them apart by the way that they are used. One subtlety illustrated by the simple example program above is that when you refer to whole arrays (like @T in the example above), you precede them with an @ sign, while the individual elements of the array (like \$T[2] in the example) are preceded by a \$ sign. Note that the scalar object \$T is a totally different thing, and has not been created or accessed by the above program.

Using perl under UNIX

There are a variety of ways to invoke perl in the UNIX environment. You can invoke perl directly from the command line, followed by the name of a perl program, but by far the most common way of running a perl program under UNIX is to put a line like the following as the first line of your perl program:

```
#!/usr/local/bin/perl
```

This line, which must be the first line of your file and which must begin in the first column, signals to the shell that the commands which follow should be interpreted by a program other than the shell. (The technique works for any language where, like perl, the pound sign (#) serves as a comment delimiter.) Then, by making the perl script executable by using the UNIX `chmod` command, you can execute your perl script by simply typing its name. In addition, as shown in previous examples, you can execute a short perl script through the command line using the `-e` flag.

Perl contains an excellent debugger, which is invoked by the `-d` flag, making it easy to pinpoint problems in your perl programs. One additional feature of the debugger should not be overlooked. If you invoke the debugger with a script consisting of just a 0, the debugger acts as an interactive perl "shell," allowing you to try different perl commands and see what the results are. Simply type

```
perl -de 0
```

at the UNIX prompt to get started.

A Quick Example

To give you a feel for why perl is so useful, I'll relate a problem which I saw a few days ago. A data tape in an unusual format had been downloaded to our system, and, although it was claimed to have a record length of 233, a quick check showed that somewhere along the line, the file had been broken up into lines of length 80, along with their associated newlines. Thus some fields were arbitrarily broken up between lines, making it very difficult to read the file correctly with even the most flexible program. But by using perl, it was easy to restore the file to its proper format. Here's the program I used:

```
#!/usr/local/bin/perl
```

```
while(<>){
    chop;
    $line .= $_;
    $n += length;
```

```
    if($n > 233){
        $out = substr($line,0,233);
        $n = length($line) - 233;
        $line = substr($line,233,$n);
        print "$out\n";
    }
}
print $line if length($line) > 0;
```

I simply concatenated the input lines (after removing the trailing newlines with `chop`), and then wrote the file out in chunks of 233 characters. A more general solution would accept the new record length as a command line argument, as explained in a previous article in this series on Unix commands.

Regular Expressions

Much of perl's text processing power comes from its novel use of regular expressions. Central to the use of regular expressions is the special operator `=~`. This operator takes its right hand argument and applies it to its left hand argument. In addition, statements using this operator return a 1 if a match takes place and a 0 otherwise. The right hand argument can either be a regular expression enclosed in slashes (`//`), in which case the left hand argument is unchanged, or it can be of the form `s/old/new/`, in which case the first occurrence of a string matched by `old` is changed to `new`. (A `g` after the final slash changes all occurrences, not just the first; similarly, an `i` performs a case insensitive search and can also be used with the `//` operator. As might be expected, if either of these arguments are presented without the `=~` the searching or changing takes place with respect to `$_`, the usual default input. Once again, this can lead to some deceptively powerful one line programs. To change each occurrence of `one` to `two` in a file we could write:

```
perl -pe 's/one/two' filename
```

The `-p` flag is like the `-n` flag but it prints each line of the input file before reading the next.

To count the occurrences of the string "perl" in a file you could write

```
perl -ne '$i++ while s/perl//; \
print "$i\n" if eof' filename
```

Since the `s///` operator returns a value of one when it makes a change, `$i` gets incremented each time the string is found.

Online Resources

The start for most perl explorations should be the CPAN (Comprehensive Perl Archive Network), modeled after

the invaluable CTAN archive for the T_EX programming language. To find the location of the site nearest you, check <http://www.perl.com/perl/CPAN/CPAN.html>. Three tutorial introductions which are worth looking at are:

1. Patrick Ryan's brief and excellent "Introduction to Perl" which can be found at <http://www.eecs.nwu.edu/perl/SAG-perl.ps>
2. Doug Bates' paper from the 1992 Interface meetings: ftp://wingra.stat.wisc.edu/pub/doc/Data_in_perl.ps.gz
3. The slides from my 1994 Interface tutorial on perl: ftp://www.stat.berkeley/pub/users/spector/perl_intro.ps.Z

Phil Spector
Department of Statistics
UC at Berkeley
spector@stat.berkeley.edu



CALL FOR PAPERS

2nd IASC World Conference on Computational Statistics and Data Analysis

...on the Eve of the 21st Century

February 19-22, 1997

Pasadena, California

You are invited to participate in the Second World Conference of the International Association for Statistical Computing (IASC), hosted by the University of Southern California and the University of California, Los Angeles. The IASC, a section of the International Statistical Institute, promotes the theory, methods and practice of statistical computing, and fosters interest in the knowledge of effective statistical computing through international contacts among statisticians, computing professionals, organizations, governments and the general public throughout the world.

Cooperating Organizations

The Interface Foundation of North America, the Computing and Graphics Sections of the American Statistical

Association, and the Southern California Chapter of the American Statistical Association.

Keynote and Plenary Speakers

Professor Bradley Efron, Max H. Stein Professor of Humanities and Sciences at Stanford University, will give the keynote address on "The Statistical Century." Professor Malcolm Pike, Flora L. Thornton Chair of Preventive Medicine at the University of Southern California, will give a plenary luncheon address on "Computational and Statistical Risk Factors in Epidemiological Research."

Invited Paper Sessions

Over twenty invited paper sessions comprising over 70 speakers have been organized by internationally distinguished statisticians and researchers. A preliminary list of invited speakers includes: John Adams, Niall Adams, Tim Arnold, Tim Barnet, Doug Bates, Bo Bergman, Abraham Bovas, Adrian Bowman, David Chickering, Jan DeLeeuw, Holger Dette, Hideyuki Douke, Usama Fayyad, V. Fedorov, W. James Gauderman, Leo Goodman, Timothy Haas, Chikio Hayashi, Joan Hilton, Karim Hirji, Paul Holland, M.S. Juhn, Genshiro Kitagawa, William Li, Shili Lin, Robin Lock, David Madigan, Charles Marshal, Thomas Mazzuchi, Christopher Meek, Cyrus Mehta, Max Mendel, William Navidi, Bill Owen, C.J. Portier, Daryl Pregibon, Charles Quesenberry, Adrian Raftery, Lewis Sheiner, Roberta Siciliano, E. Simoudis, Glenn Stone, Michael Turmon, Hiroyuki Uesaka, Antony Unwin, Paul Velleman, Jon Wakefield, Stephen Walker, Ellen Wijsman, and Carl Wunsch.

Call for Contributed Papers

Interested persons may contribute papers encompassing areas such as: algorithms for complex statistical problems; artificial intelligence and symbolic data analysis; certification and quality assessment of software; computational methods in design of experiments; computational statistical methods for model selection; computer graphics; computer networks, future directions for statistical applications; computers in statistics education; extracting information from extensive research databases; multivariate statistical modeling and data analysis; neural networks and statistical inference; statistical and computational methods in genetics; statistical and computing issues in pharmacokinetics; the internationalization of statistics; or other related topics.

All contributed papers will be published in the *Conference Proceedings*. In addition, the official journal of the IASC, *Computational Statistics and Data Analysis* (CSDA), will publish a special issue devoted to key conference papers.

For more information on registration and submission deadlines, access the IASC World Wide Web Sites <http://www.stat.unipg.it/iasc.html> and <http://icarus2.hsc.usc.edu/iasc>.



Fresh Ideas in Statistical Graphics: Special Contributed Paper Session at 1997 the JSM

By Dianne Cook

If you are working on an interesting statistical graphics method or application please send us an abstract describing the work. Or if you know someone else doing some interesting graphics work please encourage them to send an abstract.

We will select the 5 abstracts describing the freshest ideas and organize these to be presented in a special contributed paper session at the 1997 Joint Meetings, August 10-14, Anaheim, California. The equipment available will be state-of-the-art, so think freely about high-tech presentations or low-tech, as the work dictates.

Email submission to: Lorraine Denby
ld@bell-labs.com
Cc to: Dianne Cook
dicook@iastate.edu
Deadline for submission: November 15, 1996

Include: Title, Author name(s) and address (including email), Abstract (Up to 500 words)

Abstracts not selected for the special session we hope will be submitted as regular contributed papers. Please note that the deadline for regular contributed paper submissions is February 1, 1997, and the submission form will be available October 1996 at <http://www.amstat.org>.

Dianne Cook
Iowa State University
dicook@iastate.edu



Data Exposition

By Colin Goodall

You are invited to participate in the 1997 *Data Exposition*, sponsored by the Statistical Graphics Section, and to be held at the 1997 Joint Statistical Meetings in Anaheim.

What is the Data Exposition?

The Data Exposition is sponsored every other year by the Statistical Graphics Section. Participants present their findings in a Special Poster Session at the Annual ASA Meetings. All papers presented in the session must address the provided dataset and present an analysis that makes use of graphical methods.

The dataset will be available November 1, 1996

In past years, we've offered a variety of intriguing datasets, including ones on colleges and universities, Alaskan crabs, breakfast cereals, and baseball. The 1997 data is from health care, and is used to build report cards for hospitals. One way to access the data will be over the internet. Link to the ASA homepage (<http://www.amstat.org/>), then link through <sections>, then <Section on Statistical Graphics>. A preliminary view of the data may be found at <http://www.quadramed.com> in the data laboratory.

Hospital report cards data

A hospital is a complex institution, serving a diverse patient population in many different ways. The choice that a prospective patient makes to seek admission to a particular hospital often is based solely on overall reputation, advertisements run by the hospitals, and a physician's recommendation. A physician may be better informed, but both patient and physician are affected by the choices that payer organizations, notably HMOs, make in choosing to contract with particular hospitals.

In response to the perceived need for more objective data, US News and World Report has published a ranking of hospitals in 16 speciality areas (America's Best Hospitals, John Wiley 1996) with help from the National Opinion Research Center. Individual states, such as Missouri and New Jersey, and organizations, such as the Cleveland Clinic, have published report cards for consumers.

The hospital report cards data are the raw material for constructing report cards. For each of 1088 US hospitals in 10 states, data are given on around 30 patient study populations, including obstetrics, pediatrics, respiratory, cardiovascular, etc. Variables comprise num-

ber of cases, outcome rate (where outcome is C-section, complications, or mortality, depending on the study population), average length of stay, average cost, and average charges. These summaries may also be available by primary payer, and by age and sex. The data were prepared from state data sources for part of St. Anthony's Press/Quadramed's publication "The Guide to Benchmarking Hospital Value."

The objective of the data analysis is to prepare a series of tabular and graphical presentations, and possibly interactive multimedia or web-based presentations, of the data. These will comprise a prototype report card, that will help prospective patients select a hospital. The design of the report card should recognize the differing needs of consumers, and further that fine distinctions (rankings) among hospitals may not be possible or reliable.

A report card for public release would include much explanatory text written by medical specialists, some reference to variables not included in the hospital report card data, and a detailed review of the sources of the data and the assumptions underlying the generation of the data set. While welcome for the Data Exposition, such material is not essential or anticipated. The objective of the Data Exposition is to investigate practical data analysis and presentation aspects of report card preparation.

To participate

To participate in the Data Expo, submit a contributed paper abstract by February 1, 1997 to the ASA. Use the same form as for all other submissions to the annual meetings. The abstract form will contain a special box you should check to indicate the paper is a Data Expo submission. Please note that a Data Expo submission does NOT count against the limit of one contributed paper per speaker. Thus, you can submit both a contributed paper and a Data Expo paper. However, you must submit the registration fee along with your abstract as for all contributed papers. Please note that your presentation will be in a Poster Session. To help encourage computer-based presentations, the organizers (Goodall, Cook and Morton listed below) welcome correspondence on special equipment needs.

For more information

Connect to the Statistical Graphics Section website via the ASA homepage, or contact

Colin Goodall
1997 Data Exposition Co-Organizer
Health Process Management
email: colin@hdsys.com

Dianne H. Cook
1997 Statistical Graphics Program Chair
Iowa State University
email: dicook@iastate.edu

Sally C. Morton,
1997 Statistical Graphics Section Chair
RAND Corporation
email: Sally_Morton@rand.org.

Colin R. Goodall
Health Process Management
(formerly at QuadraMed)
colin@hdsys.com



CONFERENCE NOTICES

Notes from the Interface

Statistical Computing and Graphics Down Under

Interface Conference, Sydney, Australia

July 8-10, 1996

For attendees from North America or Europe, it was a long plane ride, but the Sydney International Statistical Congress was well worth the trip. Colleagues from all over the world met, discussed recent statistical work, and visited one of the world's most beautiful cities. Nick Fisher, head of the organizing committee, even arranged for great weather throughout the conference.

The 28th Symposium on the Interface between Computer Science and Statistics was one part of SISC '96. Other organizations participating in the Congress included the Statistical Society of Australia and IMS. This made for a large group of statisticians (approximately 900) and 3 parallel invited session tracks along with many contributed sessions – something for everyone.

The Third International S Workshop, one of SISC satellite meetings, was held the following week at Macquarie University and organized by Glenn Stone. This was substantially smaller (about 40 people) but filled with good talks and discussion. Bill Venables won the coveted S mug for best contribution, a polynomial and rational arithmetic library that is available at <ftp://arcola.stats.adelaide.edu.au:pub/pub/Polynom/>. Should you get the chance, don't hesitate to take a trip to Australia. There is nowhere else you can meet such friendly people and also cuddle a koala.

INSIDE

A WORD FROM OUR CHAIRS

Statistical Computing	1
Statistical Graphics	1

EDITORIAL 2

BAYESIAN FUNCTION ESTIMATION

Bayesian Strategies for Wavelet Analysis	1
Trees à la Bayes	9
Nonparametric Regression: A General Methodology	14

TOPICS IN INFORMATION VISUALIZATION

Plot Interpretation and Information Webs: A Time-Series Example From the Bureau of Labor Statistics	19
---	----

UNIX COMPUTING

Why Perl?	27
---------------------	----

CALL FOR PAPERS

2nd IASC World Conference on Computational Statistics and Data Analysis	29
Fresh Ideas in Statistical Graphics: Special Contributed Paper Session at 1997 the JSM	30
Data Exposition	30

CONFERENCE NOTICES

Interface '96	31
-------------------------	----

Statistical

COMPUTING & GRAPHICS

The *Statistical Computing & Statistical Graphics Newsletter* is a publication of the Statistical Computing and Statistical Graphics Sections of the ASA. All communications regarding this publication should be addressed to:

Mark Hansen
Editor, Statistical Computing Section
Statistics Research
Bell Laboratories
Murray Hill, NJ 07030
(908) 582-3869 • FAX: 582-3340
cocteau@bell-labs.com

Mario Peruggia
Editor, Statistical Graphics Section
Department of Statistics
The Ohio State University
Columbus, OH 43210-1247
(614) 292-0963 • FAX: 292-2096
peruggia@stat.ohio-state.edu

All communications regarding membership in the ASA and the Statistical Computing or Statistical Graphics Sections, including change of address, should be sent to:

American Statistical Association
1429 Duke Street
Alexandria, VA 22314-3402 USA
(703) 684-1221 • FAX (703) 684-2036
asainfo@amstat.org



American Statistical Association
1429 Duke Street
Alexandria, VA 22314-3402
USA

Nonprofit Organization U. S. POSTAGE PAID Permit No. 50 Summit, NJ 07901
