Models and Model Building Methods for Case Data

William S. Cleveland, Lorraine Denby, and Chuanhai Liu

Statistics Research Department
Bell Labs
Lucent Technologies
600 Mountain Ave.
Murray Hill, NJ 07974

May 11, 1998

Case data are pervasive in the sciences, engineering, business, and elsewhere. The data consist of a collection of cases and for each case there are measurements of one or more variables. There is typically a case effect that needs to be taken into account in a valid model for the data. Often, the cases are a sample from a population of potential cases, so it is common to model the case effect by a random location effect. But what tends to be overlooked in case data is that there is often a case scale effect as well, which substantially increases the complexity of analyzing the data. This paper describes an approach to modeling case effects by random location and scale effects. The approach includes a class of models for case data as well as tools for *model building*: methods for *identifying* specifications of parts of the model that are left open, and methods for *checking* other specifications that are fundamental to the model. Our methods of identification of the case distributions have an empirical Bayes flavor, and amount to a series of deconvolutions based on certain functions of the data. From the model building stage we move to a Bayesian model for the data because, as a practical matter, the case location and scale distributions fit readily into a hierarchical Bayesian framework.

1 Introduction

1.1 Case Location and Scale Effects

Many sets of data consist of cases, each of which has associated measurements of a response variable and explanatory variables. The goal is to describe the dependence of the response on the explanatory variables, but the cases have an effect as well, which complicates the study of dependence. Cases arise in many circumstances, for example, subjects in clinical trials, experimental units in designed experiments, and respondents in sample surveys.

It typically makes most sense to model case effects as random effects. This means that there are one or more collections of parameters, and for each collection there is one parameter for each case that is assumed to be generated by a probability distribution. It is common in models for case data to account for cases by random location effects, for example, the models of Bradlow and Zaslavsky (1997), Johnson (1997), and Rossi, McCulloch and Allenby (1995). Case scales can vary as well, and often do. But random case scale effects are seldom invoked in models for case data (Brady 1985). Model identification is complex because the scale effect is typically entangled with the location effect and the error variables; and sorting out the separate pieces in building a model for the data can be quite complicated. Once a model is identified, model complexity makes estimation difficult.

One type of case data that is pervasive in the social and business sciences is rater data (e.g., (Torgerson 1958, Krantz, Luce, Suppes and Tversky 1971, Brady 1989, Longford 1995, Bradlow and Zaslavsky 1997, Clark, Cleveland, Denby and Liu 1998)). Respondents in a survey rate attributes of a number of items on a subjective scale, say 1 to 10. The raters are the cases and the ratings are the response. In a business survey, the raters might be customers, the items might be companies, and the attributes might be company attributes such as product quality and on-time delivery. Typically, a scale of 1 to 10 can only be nebulously defined, so raters use the scale in different ways. This creates case effects. Raters position themselves at different points of a rating-scale, and they use up different amounts of the rating-scale, creating both location and scale effects. Often, rater effects are ignored in rater surveys altogether, which can invalidate analyses, biasing the estimation of other effects (Brady 1985).

1.2 Models and Model Building Methods for Case Data

This paper introduces a class of models and model-building methods for case data. The basic idea of our models is that they begin with a class of current models for case data, which provide random location effects, and add to them aspects that account for random scale effects. In addition, we provide model building tools. The tools are used to identify specifications of parts of the model that are left open, and to check other specifications that are fundamental to the model. In practice, the distinction between identifying and checking is often somewhat blurred, so we will use the term "model building", which covers both.

Our model class and our model building methods are meant to convey an approach to case data. Case data are common and come in many different forms in many different disciplines. In any specific application, it might be necessary to transform or otherwise restructure the data so that it is amenable to using the class, or it might make sense to change some of the details of the class. Still, most of what we present here is likely to be relevant.

An important part of our model building tools are methods of data exploration: methods for fitting mathematical structures (partial models) to the data and methods of visualization. The exploration methods allow us to reason about the likelihood of the data given specifications (Cleveland and Liu 1998). We combine our judgment of likelihood based on the data exploration with our judgment of likelihood based on information outside of the data to form a combined likelihood that is the basis for decisions about specifications. One visualization that will be used extensively in the paper is trellis display, a framework for the visualization of multivariable data (Becker, Cleveland and Shyu 1996).

In this paper, we will motivate and explain our ideas with examples, and then address general issues. Section 2 cites other work related to ours. In Section 3, we discuss the very simple situation in which there are no explanatory variables. In practice, for case data, there are typically explanatory variables, and the important goal is to describe the dependence of the response on the explanatory variables, but it is instructive to think through the simple case. Then in Sections 4 to 8, we discuss our models and model building for an actual set of data: rater data from a customer survey. Section 4 describes the data. Section 5 describes some guidelines for identifying the dependence of the response on the explanatory variables. Section 6 describes methods for identifying the distributions of the case location and scale effects and the error distributions. Section 7 describes methods for checking several dependencies that could occur in the data that are specified not to be present in the model class. Section 8 describes the full Bayesian model that emanates from the model building processes. In Section 9 we comment on how our approach should be thought of in general by giving a general model formulation and summarizing the model building tasks for this general model.

2 Related Work: Models and Model Building

The ideas in this paper are associated with at least three areas of work on case data: empirical Bayesian, Bayesian, and rater data.

2.1 Empirical Bayes and Deconvolution

The exploratory fitting methods used in our model building result in functions of the data that convolve the case location distribution, the case scale distribution, and sampling distributions. So our methods have an empirical Bayes flavor (e.g., (Maritz and Lwin 1989)), and amount to deconvolution procedures (e.g., (Laird 1978, Mendelsohn and Rice 1982, Carroll and Hall 1988, Cordy and Thomas 1997)). The deconvolution for our model building is complex because we must disentangle three distributions — case location, case scale, and sampling — that are related in an additive and multiplicative way. Furthermore, we must do this in the presence of explanatory variables, which can affect these distributions.

2.2 Bayesian Models, Sensitivity Analysis, and Posterior Predictive Distributions

Case data have been modeled by Bayesian hierarchical models with random case location effects as well as a dependence on explanatory variables, for example, in (Johnson 1997, Bradlow and Zaslavsky 1997, Rossi et al. 1995). The Bayesian models that result from the models and model building methods of this paper add case scale effects and provide model identification and checking methods. Adding the random case scale effects is a simple step from the model specification point of view, but adds substantial complexity to model building and to the computation of posterior distributions.

Our model building tools are complements to existing methods. Once a Bayesian model is fully specified, it is important to check it through sensitivity analysis, posterior predictive checking, and model expansion (Rubin 1984, Kass, Tierney and Kadane 1989, Gelman, Carlin, Stern and Rubin 1995, Gelman, Meng and Stern 1996). Johnson (1997) provides a particularly interesting example of using these techniques for model checking for case data on university course grades; the students are the cases, the grades are the response, and the two explanatory variables are the classes and the course instructors. There is a random case location effect, which becomes the student achievement index. The error variance depends on the course instructor, so there is a dependence of the error variance on an explanatory variable. And the grades for each class are determined by cutoffs, with probabilities in each range determined by a normal distribution, so the cutoffs amount to a transformation estimated from the data and allowed to vary with class. The case scales are assumed to be constant.

2.3 Rater Data

For rater data, described earlier in Section 1, Longford (1995) (p. 52) points out that location adjustment often is either a full adjustment, which means that a fixed-effect rater location estimate is subtracted, or is not done at all. Fixed-effect modeling of rater data might be reasonable if each rater has a very large number of ratings, but in general, to use sampling-theory language, it uses up too many degrees of freedom. Our modeling approach captures rater effects, both location and scale, through random effects, which provides a more parsimonious description, using, typically, fewer degrees of freedom.

The distortion that rater location and scale effects can induce is often not appreciated. Brady (1989) states: "If the manifest responses of individuals differ from their underlying true responses by monotonic transformations which vary from person to person, then the covariances of the manifest responses may bear little relationship to the covariances of the true responses so that standard analytical tools such as factor analysis may yield incorrect results." Building a statistical model for rater data provides a mechanism that automatically takes rater effects into account and renders estimated relationships accurate, provided, of course, the model accurately reflects the structure of the data.

3 A Simple Example of Case Effects: No Explanatory Variables

3.1 Case Location Effects: The Standard Deconvolution Problem

Suppose we have measurements of a response r_{ac} for $c = 1$ to n cases and $a = 1$ to m repeat measurements for each case (and no missing data). Let us suppose that

$$r_{ac} = \theta + \mu_c + \epsilon_{ac} \tag{1}$$

where θ is a fixed unknown parameter, the ϵ_{ac} are independent random variables with mean zero and variance σ_ϵ^2, the μ_c are independent random variables with distribution F_μ, and the collections μ_c and and ϵ_{ac} are mutually independent. $\theta = E(r_{ac})$ is a fixed location parameter for the r_{ac}. The μ_c are a random case location effect:

$$E(r_{ac}|\mu_c) = \theta + \mu_c.$$

The variance of the μ_c is nonzero when different cases establish different locations on the measurement scale. The ϵ_{ac} are random errors. Equation (1) is the the standard deconvolution model (Laird 1978, Mendelsohn and Rice 1982, Carroll and Hall 1988, Maritz and Lwin 1989, Cordy and Thomas 1997).

3.2 Random Scale Effects

One important assumption in the above deconvolution model is that the variances of the ϵ_{ac} are constant. But for case data, the scale of r_{ac} can also change with c as well as the location. Different cases can establish different amounts of variability on the scale. We can accommodate a case scale effect simply by adding a scale parameter that depends on c:

$$r_{ac} = \theta + \mu_c + \alpha_c \epsilon_{ac} \tag{2}$$

where the α_c are independent random variables with variance one and distribution F_α, and are independent of the μ_c and ϵ_{ac}. It is arbitrary how we assign an overall level of variance to the two variables α_c and ϵ_{ac}, so we have taken the variance of α_c to be one.

3.3 More Assumptions

We will add to the enlarged model of Equation (2) an additional assumption: the ϵ_{ac} have a normal distribution with variance σ_ϵ^2. While we focus here on the normal, it seems likely that our general approach would work for other distributions, but details would need to be worked out.

In addition, we need some soft assumptions about sample sizes. To use the data to identify the case distributions we do need more than just very small samples. As we shall see, we need nm is be sufficiently large to get a reasonable overall variance estimate. Also, the smaller the value of n, the less we can rely on the data for specifying the case distributions and the more we must rely on information from outside the data for specification. This point will be taken up again in Section 9.

3.4 Model Building: General Tasks

In the model in Equation (2) there are certain specific assumptions about the structure of the data, for example, the mutual independence of the collections μ_c, α_c and ϵ_{ac}, and the normality of ϵ_{ac}. However, certain specifications remain open — the form of the two distributions, F_μ and F_α. There are two model building tasks. One is to check the specifications that have been made. The other is to identify characteristics of the F_μ and F_α that then become additions to the model; for example, we might specify parametric families for each. We can rely on information from outside the data to guide this specification, but we would like to explore the data as well to help build the model.

3.5 Model Building: σ^2 and The Case Regression

The basis of our model building will be a *case regression*. First, we can estimate σ^2 by

$$\hat{\sigma}^2 = \frac{\sum_{a,c}(r_{ac} - \bar{r}_c)^2}{n(m-1)}$$

where

$$\bar{r}_c = \frac{\sum_a r_{ac}}{m}.$$

Next, because of our soft assumptions about sample size, $n(m-1)$ is large, so we will suppose, for model building purposes, that the estimate is the true value, $\hat{\sigma}^2 = \sigma^2$, and divide Equation (2) by σ, which results in

$$y_{ac} = \mu_c x_c + \alpha_c \zeta_{ac} \quad (3)$$

where

$$\begin{aligned} y_{ac} &= r_{ac}/\sigma \\ x_c &= 1/\sigma \\ \zeta_{ac} &= \epsilon_{ac}/\sigma. \end{aligned}$$

We treat Equation (3) for each fixed c as a regression equation with normal errors, and estimate μ_c and α_c by least squares. We have

$$\hat{\mu}_c = \frac{\sum_a y_{ac}}{m}.$$

The residuals are

$$\hat{\zeta}_{ac} = y_{ac} - \hat{\mu}_c x_c.$$

s_c^2, the usual estimate of residual variability, estimates α_c^2:

$$s_c^2 = \alpha_c^2 \frac{\sum_a \hat{\zeta}_{ac}^2}{m-1}.$$

so conditional on α_c, $\mathrm{E}s_c^2 = \alpha_c^2$. The studentized residuals are

$$\hat{\phi}_{ac} = \frac{\hat{\psi}_{ac}}{(m-1)s_c}.$$

3.6 Model Building: Checking the Normality of the Errors and Independence, and Identifying the Case Distributions

We will now briefly describe our model building methods for the simple case. More detail will be given in the example of the coming sections.

If the ϵ_{ac} are normal, then the ϕ_{ac} have a known distribution related to the beta-distribution. So we can check the normality of the ϵ_{ac} through quantile plots of the ϕ_{ac} to see if they are well approximated by the quantiles of the known distribution.

Suppose it is decided that the distribution of the ϵ_{ac} is well approximated by the normal. Then s_c^2 is a multiplicative convolution of the distribution of the unknown distribution of the α_c^2 and a constant times a $\chi^2\{m-1\}$. If we posit a distribution for the α_c^2, we can check it by using a quantile plot to study how well the distribution of the s_c^2 is approximated by the multiplicative convolution distribution.

Once we have identified a distribution for the α_c^2, we can use it, along with the assumption of normality of the ϵ_{ac}, to identify a distribution for the μ_c. Now

$$\hat{\mu}_c = \mu_c + \alpha_c \bar{\zeta}_c,$$

where

$$\bar{\zeta}_c = \frac{\sum_a \zeta_{ac}}{m},$$

has a normal distribution with mean zero and variance one. If we posit a distribution for μ_c we can check it by a quantile plot of the $\hat{\mu}_c$ against a distribution that is the convolution of the posited distribution plus the product of the normal and the identified distribution for α_c.

Our simple case model specifies that the variables μ_c, α_c^2, and ϵ_{ac} are mutually independent. There are a variety of checks we can carry out to see if dependencies occur. For example, we can check for a dependence of α_c^2 on μ_c by a scatterplot of cube root s_c^2 against $\hat{\mu}_c$. We will see more checks in the next sections, where the presence of explanatory variables increases the potential for dependencies.

4 An Example: Customer Survey Data

Next, we will discuss our case model and model building methods in the context of data from a customer survey. This example is typical of case data in that explanatory variables are present, and the goal of the analysis is to discover the dependence of the response on the explanatory variables. Clark et al. (1998) provide a detailed discussion of the background of the data, the final model for the data, and the subject matter conclusions based on the model (but not the model building). In this section we will briefly describe the data and the initial case model for them; then, in the next three sections, we will detail the model building, checking and identification.

4.1 Data

The data are from a quarterly survey of 19 supplier companies in the information technology industry. The data cover 9 quarters. Customers are polled and rate the suppliers on 9 attributes on a scale of 1 to 10. The attributes are

prod-qual product quality

over-qual overall quality of supplier processes

delivery delivery of the product

cost cost of the product to the customer and the process of working with the customer to establish the cost

features technological excellence of product

pre-sup providing product information before delivery to support customer processes

response responsiveness of the supplier to customer needs

value value of the product relative to the cost of the product

service service provided by the sales team

There are 3505 pollings, which are the cases. If there were no missing data, we would have $3505 \times 9 = 31545$ ratings. But because of missing data, there are 25309 ratings. The survey design results in no respondent rating all nine attributes; only purchasing agents rate `delivery` and only product and process designers rate `pre-sup`. The remaining 7 attributes are intended to be rated by all respondents, but some respondents choose not to rate all attributes about which they are asked. Most respondents are polled just once; that is, they rate just one supplier for one quarter. 109 of the pollings are repeats, but since the fraction is so small, we will treat the repeat pollings as if they are different people. Each person rates attributes for one supplier in one quarter. The rating is the response variable. The attribute, supplier, and quarter variables are explanatory variables.

In the next sections we will study certain functions of the data whose likelihood given our general model for the data depend on the number of attributes, $n(c)$, rated by each respondent. The values of these functions will be grouped according to $n(c)$. Table 1 shows counts of the number of attributes rated in the 3505 pollings.

Exploratory plots of the ratings reveal a skewness to the left and so we will work with transformed ratings. We will use a square root transformation, $s\sqrt{(11 - \text{rating})} + t$ where s and t are chosen so that a rating of 1 becomes 1 on the transformed scale and a rating of 10 becomes 10.

4.2 A Case Model for the Customer Survey Data

For the customer survey data, we begin with the following case model:

$$r_{ac} = \theta_{asq} + \mu_c + \alpha_c \epsilon_{ac}, \qquad (4)$$

where r_{ac} is the rating (on the transformed scale) by person c on attribute a, μ_c is the person (case) location effect, and α_c is the person (case) scale effect; θ_{asq} expresses the dependence of the expectation of r_{ac} on the attribute, supplier, and quarter; that is, it is a supplier-attribute effect in one quarter. (Note that s and q depend on c, so a more explicit but cumbersome notation for θ_{asq} would be $\theta_{as(c)q(c)}$.) ϵ_{ac} is the error term. Information from sources outside of the data suggest that the variance of the errors likely depends on the attribute a. Hence, we will suppose $E\epsilon_{ac}^2 = \sigma_a^2$.

5 Model Building: Some Comments on Modeling θ_{asq}

The θ_{asq} in our model for the customer survey data, Equation (4), capture a dependence of the rating on the explanatory variables that is not associated with the case effects. The specification of the dependence in this way — separate, unrelated coefficients for each combination of a, s, and q — is quite general. Our knowledge of customer perceptions makes it clear that we can model θ_{asq} further, that is, specify further structure that relates the different values of the θ_{asq} to one another. But we would like to explore the data to determine the adequacy of likely specifications. It would be inappropriate here to go into great detail about the identification of θ_{asq} for our customer survey data, since much of it depends of the specifics of the application, but we will discuss a few points that apply more generally.

5.1 Iteration

The model of Equation (4) provides a guide for developing appropriate ways to explore the data for the purpose of identifying θ_{asq}. But at this point, we have not established the validity of the model, so we are relying on unverified specifications to check other specifications. We will later check the initial model specifications, relying then on our specification for θ_{asq}. Should we find the initial specifications to be untenable, we would need to return to the beginning, alter the basic model, then study the specification of θ_{asq} again. This iterative process of checking and respecifying is an important and necessary part of model building (Box 1980).

5.2 Guidance from the Model: 1

The model of Equation (4) shows that we can use a location estimate based on the values of r_{ac} across c for fixed a, s, and q to provide information about the θ_{asq}. The person, or case effects, do not interfere with this process because the expected value of the averaged case effects across people is zero. (Of course, the number of people will affect the precision of the estimate.) Figure 1 is one illustration of such estimates. It is the first page of a 5-page trellis display of sample means across c for each value of a, k, and q.

Number of Attributes Rated	1	2	3	4	5	6	7	8
Number of Pollings	5	9	26	64	150	350	1106	1795

Table 1: Counts of Pollings According to Number of Attributes Rated

Each panel displays the means for a fixed value of a and s against q together with a loess curve fitted to the all of the ratings (Cleveland and Devlin 1988, Cleveland 1993). The loess smoothing parameter is 1, and local linear fitting is used.

We combine information from displays such as Figure 1 with our knowledge from sources other than the data, to model the θ_{asq}. We expect companies to have quite different ratings and that differences will depend on the attribute; in other words, we expect a company-by-attribute interaction. We expect company ratings to change through time, but we expect change through time to be largely smooth, that is, almost exclusively made up of low-frequency behavior. Exploration of the data by displays such as Figure 1 strongly support these expectations. Ultimately, in the analysis of these data, the θ_{asq} for fixed a and s were modeled by taking them to be an integrated moving process largely made up of low-frequency behavior (Clark et al. 1998). The process is described in Section 8.

5.3 Guidance from the Model: 2

There is also a substantial amount of exogenous information about relationships among the attributes, that is, relationships among the θ_{asq} across a for fixed values of s and q. But exploration of the r_{ac} to help identify these relationships is problematic; we cannot use functions of the data that require an inner product of two vectors of values of r_{ac} across c, one vector for one value of a and another vector for another value of a. In such a case, the expectation involves $E(\mu_c^2)$, which biases the estimation of attribute relationships. In other words, the case effects interfere. This problem, as we have discussed in Section 2, is often ignored in analyses of case data, particularly rater data. The solution is to aggregate first across sufficiently large numbers of values of c, averaging across values of μ_c, and then study relationships. For example, we can use the quarterly means of Figure 1.

5.4 Removing the Effect of θ_{asq}

To identify the person and error distributions we will remove the effect of the θ_{asq}. We will do this simply by subtracting from the ratings the loess curves used in Figure 1, ignoring the effect of the adjustment on the person and error distributions, since the number of degrees of freedom of the fitting, 355, is small compared with the number of observations, 25,309. The adjusted ratings, with a slight abuse of notation, will be denoted by r_{ac}. The model of Equation (4) becomes

$$r_{ac} = \mu_c + \alpha_c \epsilon_{ac}. \qquad (5)$$

6 Model Building: Identifying the Case Effect and Error Distributions

6.1 Case Effects and Error Effects

The subjective rating scale used by customers is anchored by giving respondents the guidelines that 1 is poor and 10 is excellent. People position themselves at different places on the rating scale and use up different amounts of it. In other words, we can expect case location and scale effects. These are modeled by the μ_c and α_c.

Even apart from these repeatable case effects, we expect further variation in a person's rating of the attribute of a company. First, there is variation induced by reporting integers. We have been supposing that a customer has a real-valued rating that is rounded to an integer, which induces variation of up to ± 0.5. But, in addition, it is sensible to expect another form of error; a person only sees in a very fuzzy way the performance of a company on a particular attribute. Were we able to poll the person a short time later and somehow remove the memory of the first rating, we would expect a different unrounded rating. This variation is modeled by the ϵ_{ac}.

6.2 Case Regressions

Suppose for the moment that the σ_a^2 are known. Dividing Equation (5) by σ_a yields

$$r_{ac}/\sigma_a = \mu_c/\sigma_a + \alpha_c \zeta_{ac} \qquad (6)$$

or

$$y_{ac} = \mu_c x_a + \alpha_c \zeta_{ac} \qquad (7)$$

where the ζ_{ac} are independent with mean 0 and variance 1.

Let us fix c. In other words, we are conditioning on μ_c and α_c. The index a runs over the set of attributes $A(c)$ answered by person c. Now let us estimate μ_c by least-squares from Equation (7):

$$\hat{\mu}_c = \frac{\sum_{A(c)} \sigma_a^{-2} r_{ac}}{\sum_{A(c)} \sigma_a^{-2}}.$$

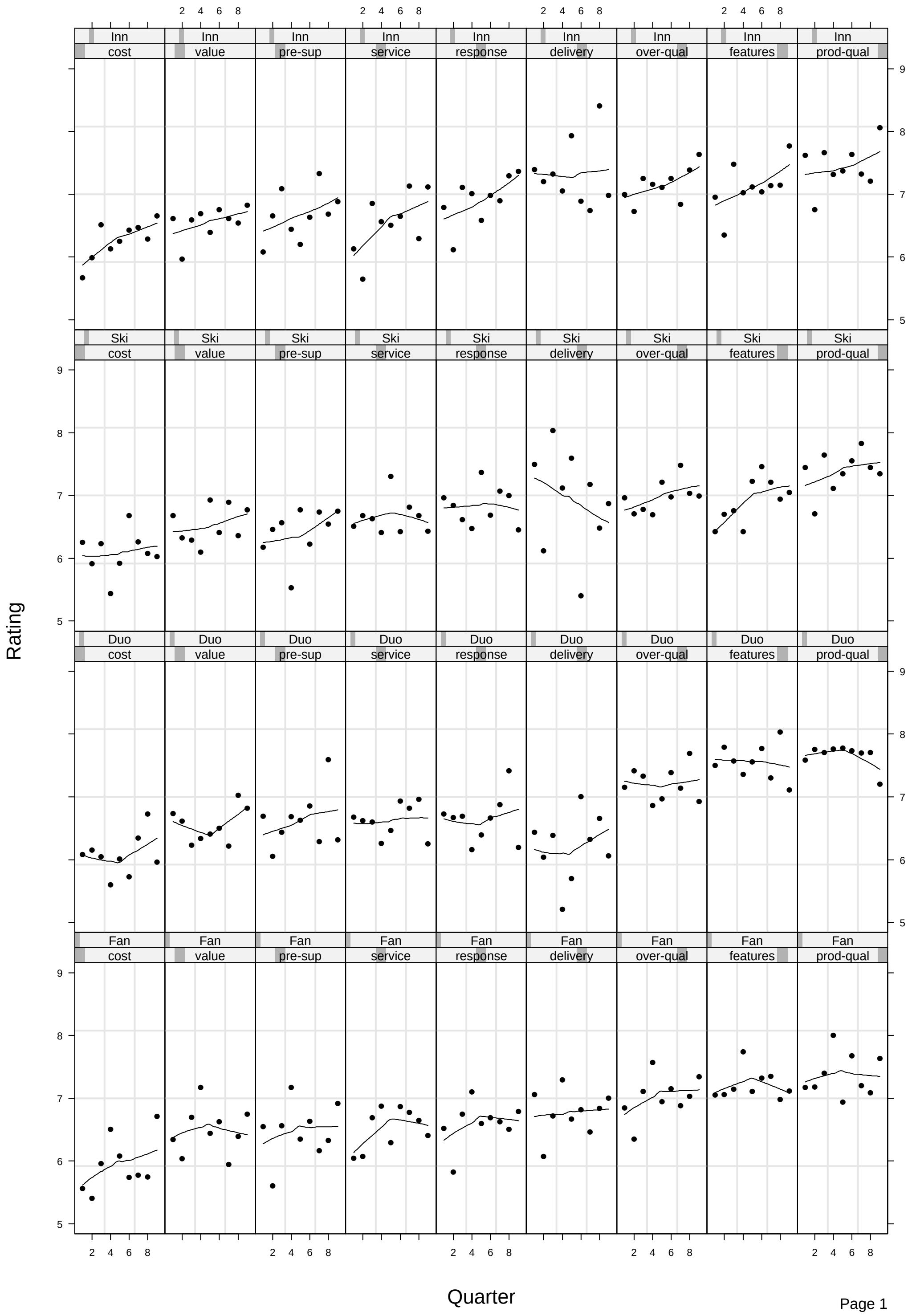

Figure 1: Time Plots of Quarterly Mean Rating Given Attribute and Company and Loess Fits

The residuals from this regression are

$$y_{ac} - \hat{\mu}_c x_a = \frac{r_{ac} - \hat{\mu}_c}{\sigma_a} = \alpha_c \hat{\zeta}_{ac}$$

where

$$\hat{\zeta}_{ac} = \zeta_{ac} - \sigma_a^{-1} \frac{\sum_{A(c)} \sigma_j^{-1} \zeta_{jc}}{\sum_{A(c)} \sigma_j^{-2}}.$$

The expected squared residuals are

$$E(\alpha_c \hat{\zeta}_{ac})^2 = \alpha_c^2 (1 - \frac{\sigma_a^{-2}}{\sum_{A(c)} \sigma_j^{-2}}).$$

The standardized residuals are defined to be

$$\hat{\psi}_{ac} = \frac{\alpha_c \hat{\zeta}_{ac}}{\sqrt{1 - \frac{\sigma_a^{-2}}{\sum_{A(c)} \sigma_j^{-2}}}}$$

Let s_c^2 be the usual estimate of residual variability, the residual sum of squares divided by the degrees of freedom:

$$s_c^2 = \alpha_c^2 \frac{\sum_{A(c)} \hat{\zeta}_{ac}^2}{n(c) - 1},$$

where $n(c)$ is the number of attributes rated by person c, that is, the size of the set $A(c)$. Then

$$E s_c^2 = \alpha_c^2.$$

The studentized residuals are defined to be

$$\hat{\phi}_{ac} = \frac{\hat{\psi_{ac}}}{s_c} = \frac{\sqrt{n(c) - 1}}{\sqrt{1 - \frac{\sigma_a^{-2}}{\sum_{A(c)} \sigma_j^{-2}}}} \frac{\hat{\zeta}_{ac}}{\sqrt{\sum_{A(c)} \hat{\zeta}_{ac}^2}}$$

6.3 Estimating and Adjusting for the Error Variances

For each a, we will sum the squared residuals across those people $P(a)$ who answer a and take expectations (no longer conditioning on μ_c and α_c):

$$E \sum_{P(a)} ((r_{ac} - \hat{\mu}_c)/\sigma_a)^2 = \sum_{P(a)} (1 - \frac{\sigma_a^{-2}}{\sum_{A(c)} \sigma_j^{-2}}) \quad (8)$$

We now have a complicated set of 9 equations in the 9 values of σ_a. Given the $\hat{\mu}_c$ from a previous set of σ_a we can solve these equations for a new set, compute new $\hat{\mu}_c$, and solve again, iterating this process until convergence. The solution yields the estimates of the σ_a in Table 2. Since these estimates of the attribute standard deviations are formed from a large number of observations, we will treat the estimates in the ensuing discussion as known, and denote them by σ_a.

6.4 Distribution of the Errors

We now turn our attention to the distribution of the errors, ζ_{ac}. Because we took care at the outset of the modeling to symmetrize the distributions of the ratings for each combination of company and attribute we can at least hope that the error distributions are symmetric, perhaps even normal.

Suppose

$$\zeta_{ac} \sim N\{0, 1\}.$$

Then, conditional on μ_c and α_c, for the standardized residuals,

$$\hat{\psi}_{ac} \sim \alpha_c N\{0, 1\},$$

and for the studentized residuals,

$$\hat{\phi}_{ac} \sim \sqrt{n(c) - 1} \beta_{\pm}^2 \{1/2, (n(c) - 2)/2\}$$

where $\beta_{\pm}^2\{r, s\}$ is a symmetric distribution on $[-1, 1]$ whose square is distributed $\beta\{r, s\}$. We will call this the symmetrized squared-beta distribution.

We will check the normality of the ζ_{ac} by checking to see if the $\hat{\phi}_{ac}$ have the prescribed distribution. Since this distribution depends on $n(c)$, we will group the $\hat{\phi}_{ac}$ according to the value of $n(c)$. Since there is only a small number of values of p for which $n(c)$ is 1 or 2, we will eliminate c for which $n(c) < 3$.

Figure 2 is a trellis display given $n(c)$. On each panel, $\hat{\phi}_{ac}$ is plotted against quantiles of the prescribed distribution. The solid line on each panel has intercept 0 and slope 1. The horizontal lines are drawn at the 0.01, 0.1, 0.9, and 0.99 quantiles to give a clear picture of where the majority of the data lie. As we go from left to right and then bottom to top through the panels, $n(c)$ increases from 3 to 8. The quantiles of the data follow those of the $(n(c) - 1)\beta_{\pm}^2\{1/2, (n(c) - 2)/2\}$ quite closely in each case. On this basis, we will specify the distribution of the ζ_{ac} to be normal.

6.5 Distribution of the Person Variances

The case variances, α_c^2, are individually estimated by s_c^2 in the sense that conditional on α_c, $E s_c^2 = \alpha_c^2$. But across people, the distribution of s_c^2 has two components: the distribution of the α_c^2, which we seek to model, and the distribution of

$$\frac{\sum_{A(c)} \hat{\zeta}_{ac}^2}{n(c) - 1},$$

which we can now take to be $(n(c) - 1)^{-1} \chi^2\{n(c) - 1\}$ because we are prepared to entertain the $N\{0, 1\}$ distribution for ζ_{ac}. Thus the s_c^2 distribution is a multiplicative convolution of the distribution of α_c^2 with a known distribution. We need to carry out a deconvolution to form a picture of the α_c^2 distribution. We will do this indirectly using quantile plots. We will try a specific parametric distribution

over-qual	response	prod-qual	pre-sup	value	features	delivery	cost	service
0.79	1.17	1.19	1.42	1.46	1.61	1.88	1.90	1.93

Table 2: Estimates of Attribute Standard Deviations

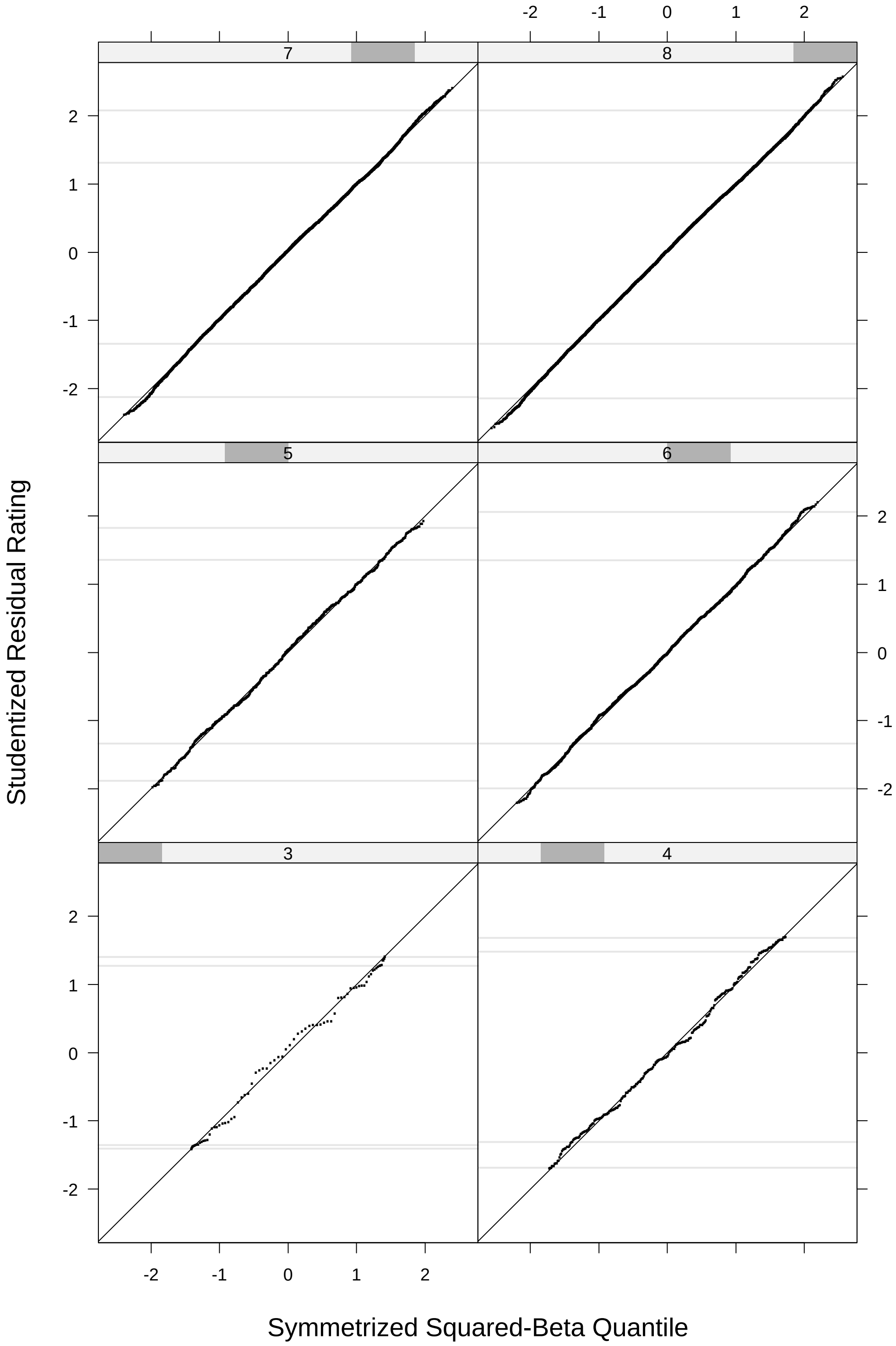

Figure 2: Quantile Plots of Studentized Residuals Given $n(p)$

for α_c^2, convolve with the known distribution, and make a quantile plot of the s_c^2 against the convolution quantiles. If the convolution quantiles do not match those of s_c^2, the plot will likely provide information on changing the parametric distribution to provide a better match.

In our discussion thus far we have assumed, based on our exogenous information, that the α_c^2 do vary by case and are not equal to a constant. It makes sense to first see if the data support our assumption. In other words, the first distribution for α_c^2 to check is the simple one, a constant, which would be 1 since $\mathrm{E}\alpha_c^2 = 1$. This means that

$$s_c^2 \sim \frac{\chi^2\{n(c)-1\}}{n(c)-1}.$$

Figure 3 is a trellis display of quantile plots given $n(c)$. On each panel, the cube roots of the s_c^2 quantiles are plotted against the cube roots of the quantiles of the prescribed distribution. The solid line is drawn through the lower and upper quartiles. The horizontal lines are drawn at the 0.01, 0.1, 0.9, and 0.99 quantiles. The s_c^2 do not follow the prescribed distribution; among other things, their dispersion is far greater due, clearly, to a convolving with a nondegenerate α_c^2 distribution.

We will make another quantile plot to help us guess a form for the distribution of α_c^2. For the standardized residuals,

$$\hat{\psi}_{ac} \sim \alpha_c N\{0,1\}.$$

Figure 4 is a normal quantile plot of the standardized residuals for respondents with $n(c) > 2$. The solid line is drawn through the lower and upper quartiles, and the horizontal lines are drawn at the 0.01, 0.1, 0.9, and 0.99 quantiles. The display shows the $\hat{\psi}_{ac}$ have a distribution with longer tails than the normal.

The tails of the distribution of the $\hat{\psi}_{ac}$ would be longer than the normal if the distribution of α_c^2 were a constant times the inverse of a $\chi^2\{\nu\}$. Then $\hat{\psi}_{ac}$ would have a $t\{\nu\}$ distribution times a constant. Since the mean of an inverse $\chi^2\{\nu\}$ is $\nu/(\nu-2)$, the constant is $(\nu-2)/\nu$. This implies

$$s_c^2 \sim \frac{\nu-2}{\nu} F\{n(c)-1,\nu\}.$$

We made quantile plots with with varying values of ν. The best was $\nu = 6$. Figure 5 is a trellis display of quantile plots given $n(c)$. On each panel, cube root s_c^2 quantiles are graphed against cube root $2/3F\{n(c)-1,6\}$ quantiles. The solid line is drawn through the quartiles. The horizontal lines are at the 0.01, 0.1, 0.9, and 0.99 quantiles. The agreement is poor; the upper tails of the two distributions do not agree. In other words, the plot allows us to judge the likelihood of the data — the sample case variances — given an inverse chi-square distribution times a constant for the case variance distribution; the likelihood is small.

We searched for another distribution of the α_c^2 that would produce an acceptable convolution quantile plot. We found that a $\chi^2\{2\}/2$, which is a unit exponential distribution, $exp\{1\}$, gave a reasonable approximation. This means that α_c has a Weibull distribution, $W\{1,2\}$. Figure 6 shows quantile plots given $n(c)$. On each panel, cube root s_c^2 is plotted against the cube roots of quantiles of a distribution that is the product of an $exp\{1\}$ variable and a $\chi^2\{n(c)-1\}/(n(c)-1)$ variable. The agreement is quite good.

6.6 Distribution of the Person Means

We now have in place two distributional assumptions:

$$\begin{aligned} \zeta_{ac} &\sim N\{0,1\} & (9)\\ \alpha_c &\sim W\{1,2\}. & (10) \end{aligned}$$

From Section 6.2

$$\hat{\mu}_c = \mu_c + \alpha_c \hat{\omega}_c$$

where

$$\hat{\omega}_c = \frac{\sum_{A(c)} \sigma_a^{-2} \zeta_{ac}}{\sum_{A(c)} \sigma_a^{-2}}.$$

Because of our distributional assumption for ζ_{ac},

$$\hat{\omega}_c \sim \frac{N\{0,1\}}{\sqrt{\sum_{A(c)} \sigma_a^{-2}}}.$$

Thus, $\hat{\mu}_c$, is the convolution of μ_c with random variables that have a known distribution, so we will search for a good approximating distribution for μ_c using the same graphical deconvolution approach as that for studying the distribution of α_c^2.

Recall that 3505 is the number of sample case means. Now, taking unconditional expectations,

$$\begin{aligned} E\frac{1}{3505}\sum_c \hat{\mu}_c^2 &= E\mu_c^2 + E\alpha_c^2 E\hat{\omega}_c^2 \\ &= E\ \mu_c^2 + \frac{1}{3505}\sum_c \frac{1}{\sum_{A(c)} \sigma_a^{-2}}. \end{aligned}$$

Thus we can estimate the variance of the μ_c distribution by subtracting

$$\frac{1}{3505}\sum_c \frac{1}{\sum_{A(c)} \sigma_a^{-2}} = 0.19$$

from

$$\frac{1}{3505}\sum_c \hat{\mu}_c^2 = 1.69,$$

which is 1.50. Thus most of the variability in $\hat{\mu}_c$ is due to the variability in μ_c, which makes it clear, as expected, that we have a nondegenerate distribution for the μ_c. The value of $\sqrt{1.50} = 1.22$ is important to us because we will be able to specify the scale of the μ_c distribution in our deconvolution search.

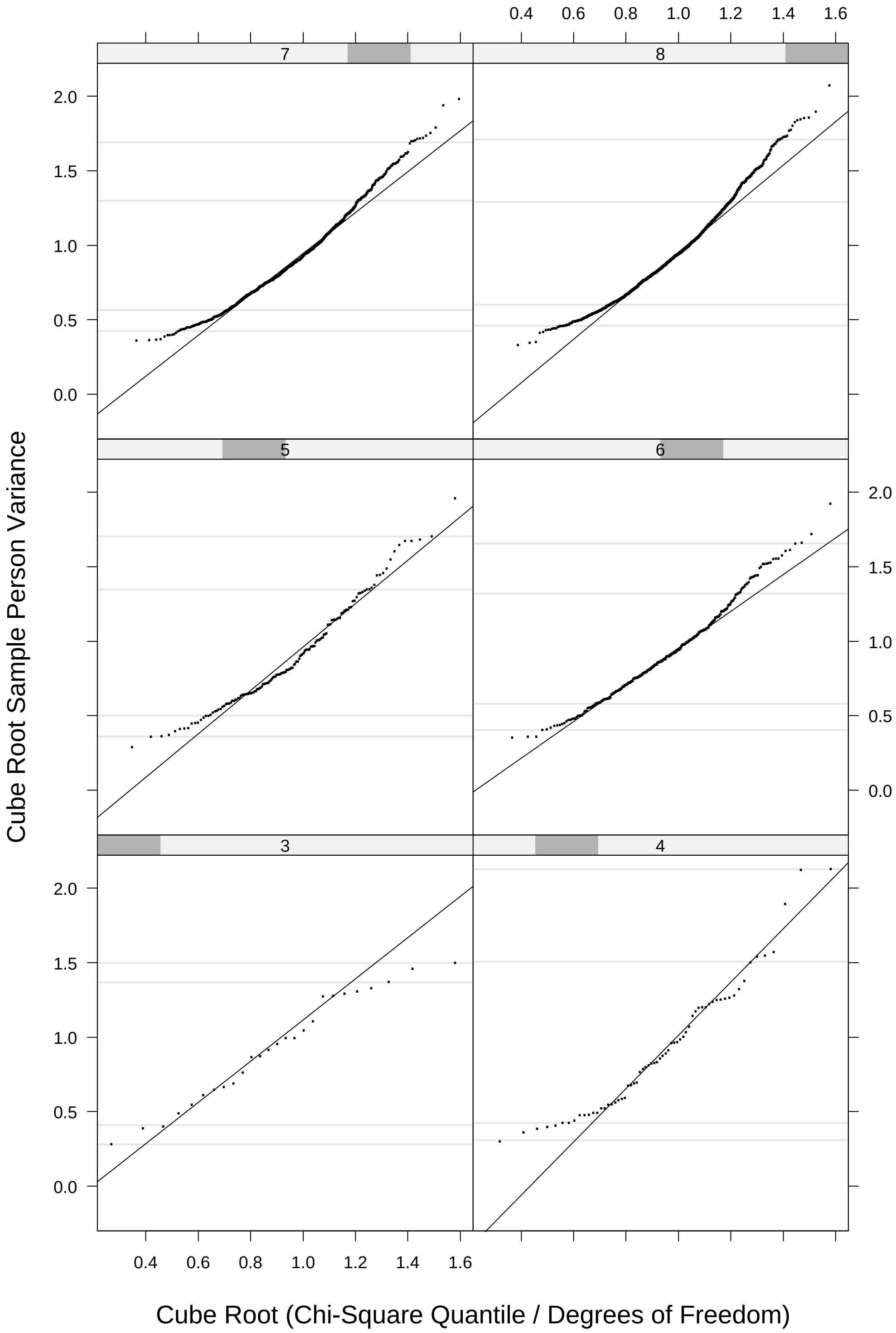

Figure 3: Quantile Plots of Cube Root Sample Person Variances Given $n(p)$

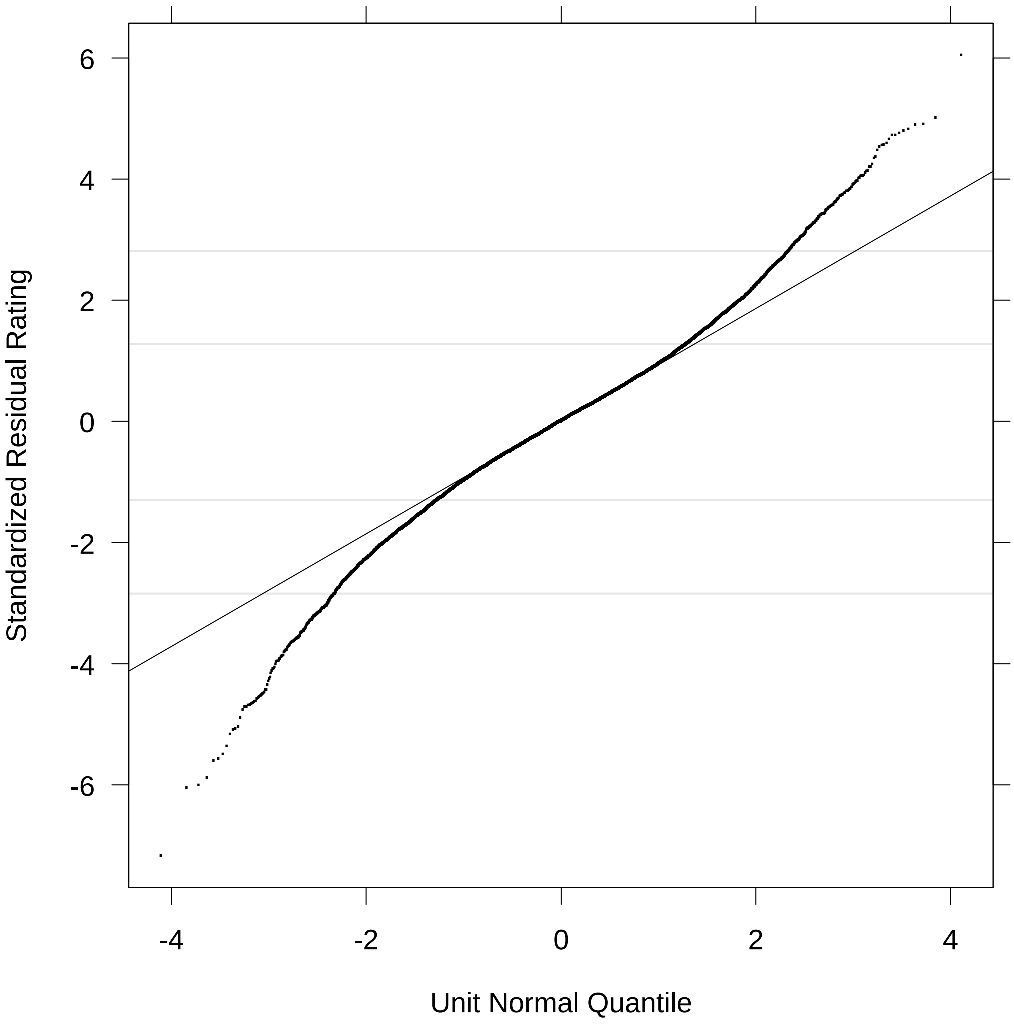

Figure 4: Normal Quantile Plot of Standardized Residuals

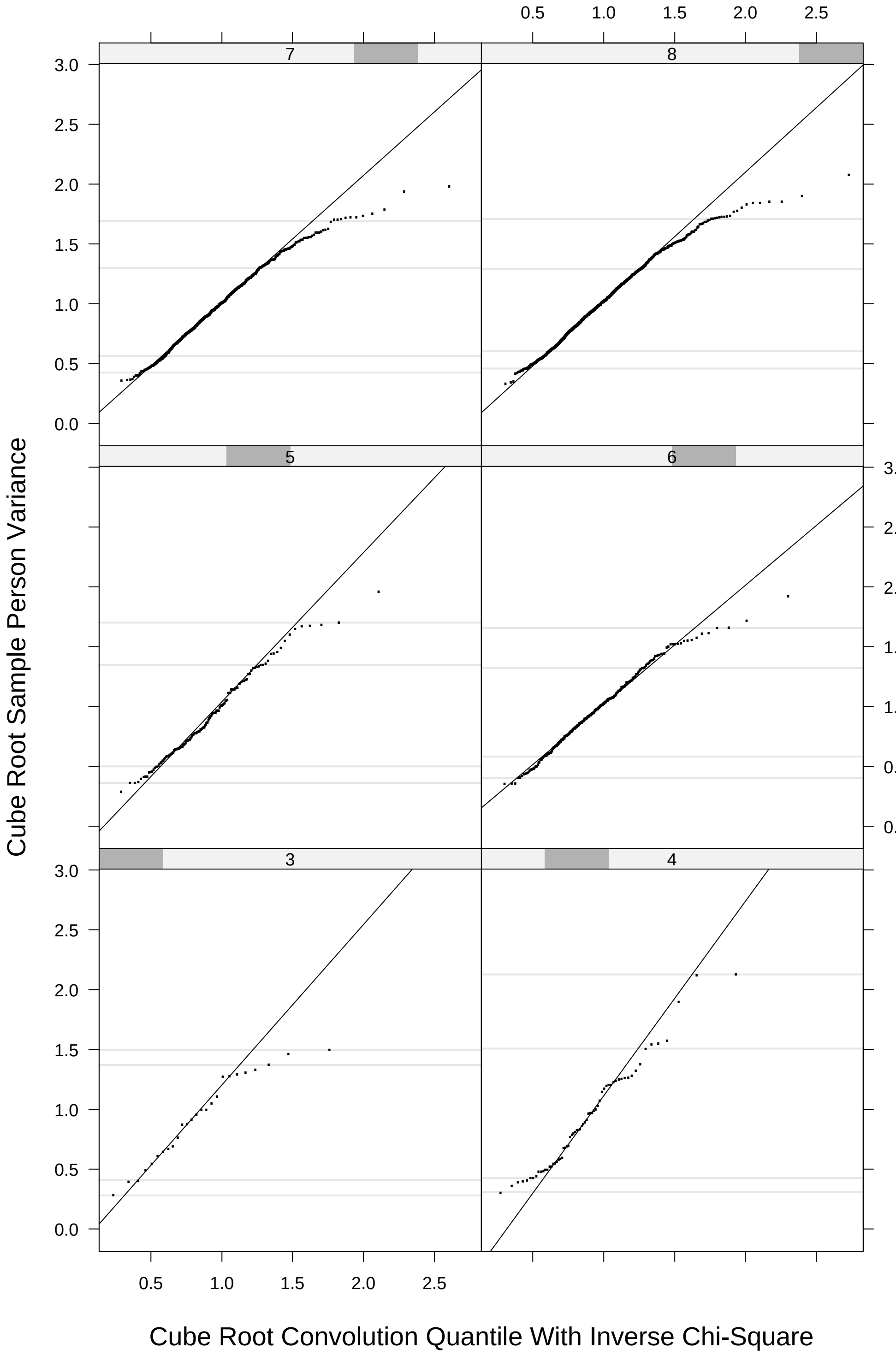

Figure 5: Quantile Plots of Cube Root Sample Person Variances Given $n(p)$ vs. Convolution Quantiles with Inverse Chi-Square

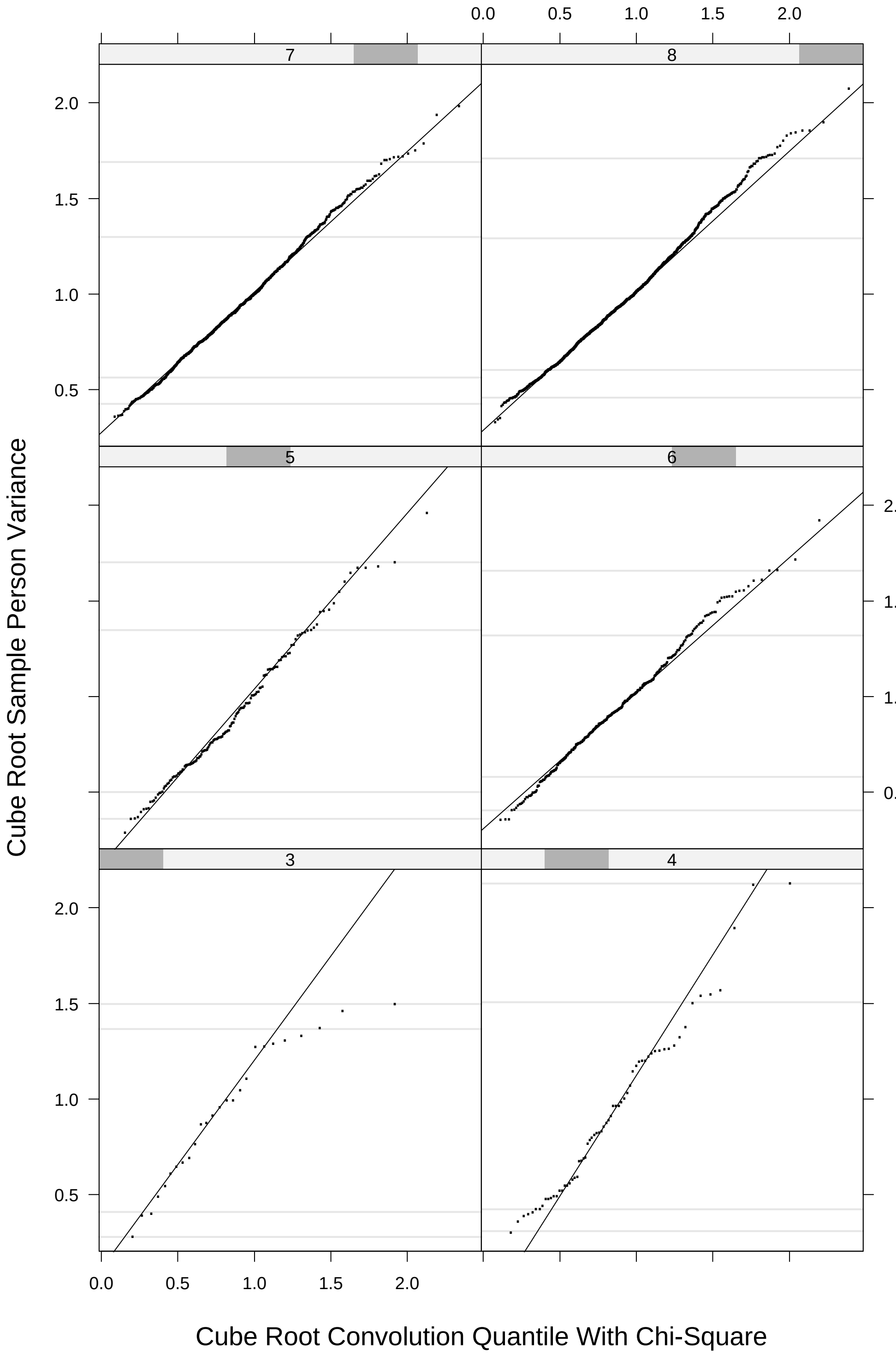

Figure 6: Quantile Plots of Cube Root Sample Person Variances Given $n(p)$ vs. Convolution Quantiles with Chi-Square

In carrying out our deconvolution by quantile plots, we must contend with the variation in the values of

$$v(c) = \frac{1}{\sum_{A(c)} \sigma_a^{-2}}.$$

This is the variance of the random variable that is added to μ_c to yield s_c^2. We think of the σ_a as known but the mechanism of missing values produces an empirical distribution of the $v(c)$. The minimum $v(c)$ is 0.167, the 0.95 quantile is 0.254 and the maximum is 1.924. To reduce variability in our plots, and restrict the variation of the $v(c)$, we will delete the 198 c for which $v(c) > 0.25$.

We will first try $N(0, 1.5)$ for the distribution of μ_c. Figure 7 is a quantile plot. The quantiles of the data, $\hat{\mu}_c$, are graphed against the corresponding quantiles of the convolution distribution,

$$N(0, 1.5) + W\{1, 2\} \frac{N\{0, 1\}}{\sqrt{\sum_{A(c)} \sigma_a^{-2}}}.$$

The agreement on the plot is not particularly good. At the extreme of the upper tail, at the 0.99 quantile and beyond, the data quantiles increase by only a small amount but the calculated quantiles increase at a greater rate. In this region, distribution of the sample case means is reduced because of the boundary of 10 on the rating scale. But this departure is not particularly damaging because it involves a small fraction of the data. (The horizontal lines help us to see that the phenomenon is much less dramatic than the fraction of visible ink given to it by the plot.) Were the data otherwise normal we would ignore this extreme behavior. The real problem is that the remaining 99% of the distribution is not well approximated by the normal. In fact, for this portion of the distribution, the data show leptokurtosis.

We tried the $t\{\nu\}$ family for μ_c and found a good match for 6 degrees of freedom. Since the variance of a $t\{6\}$ with unit scale is $3/2$, a scale parameter of 1.00 matches the estimated variance, 1.50, of the μ_c. Figure 8 is the quantile plot. We will endure the shortened tail of the data at the extreme upper tail.

7 Model Building: Ruling Out Dependencies

7.1 Possible Dependencies in the Customer Survey Data

The customer survey model of Equation (4) has three collections of variables — μ_c, α_c, and ϵ_{ac} — that are assumed to be mutually independent. Furthermore, their distributions are assumed not to depend on the explanatory variables except as explicitly specified in the model. (The one specification in the model is that the variance of ϵ_{ac} depends on a). In this section we describe diagnostic methods for checking those independence specifications that seem, based on our knowledge of customer perceptions and of the measurement process and scale.

7.2 Dependence of Person Variance on Person Mean

One likely possibility is a dependence of the variability of the error variables on the rating level; for example, at the ends of the scale, the variability may be reduced. Were it not for the transformation carried out at the beginning of the analysis, the danger would be far greater. The transformation had the purpose of symmetrizing the company-attribute rating distributions. The transformation pulls out the top of the rating scale relative to the bottom. One might expect that this would have the effect of equalizing variability as well.

Figure 9 is trellis display given $n(p) = 4$ to 9. Each panel is an s-l plot (Cleveland 1993): cube root s_c^2 is plotted against $\hat{\mu}_c$ with loess smooths fitted by local linear fitting and a smoothness parameter of 1. No dependence appears, which suggests that the person variances do not depend on the person means.

7.3 Dependence of Error Variance on Rating Level

Another plausible dependence is correlation among the ϵ_{ac} for different a for a fixed c (which also fixes s and q) — in other words, correlation induced at the respondent level. The μ_c go a long way to eliminating this correlation but might not be enough.

Figure 10 is a trellis display given $n(p) = 4$ to 9. Each panel shows an s-l plot: square root absolute studentized residuals is plotted against the fitted ratings,

$$\hat{\theta}_{asq} + \hat{\mu}_c,$$

where the $\hat{\theta}_{asq}$ are the values of the loess fits described in Section 5, and the curve on each panel is a loess smooth with local linear fitting and a smoothness parameter of 1. No dependence appears which suggests the error variance does not depend on the rating level. This is a great relief. Our initial transformation has done its job.

7.4 Correlation in the Residuals

Finally, there is a danger that for the μ_c and α_c, the distribution of one depends on the other; for example, respondents who center themselves in one region of the rating scale may use up a different amount of the scale that those in another region.

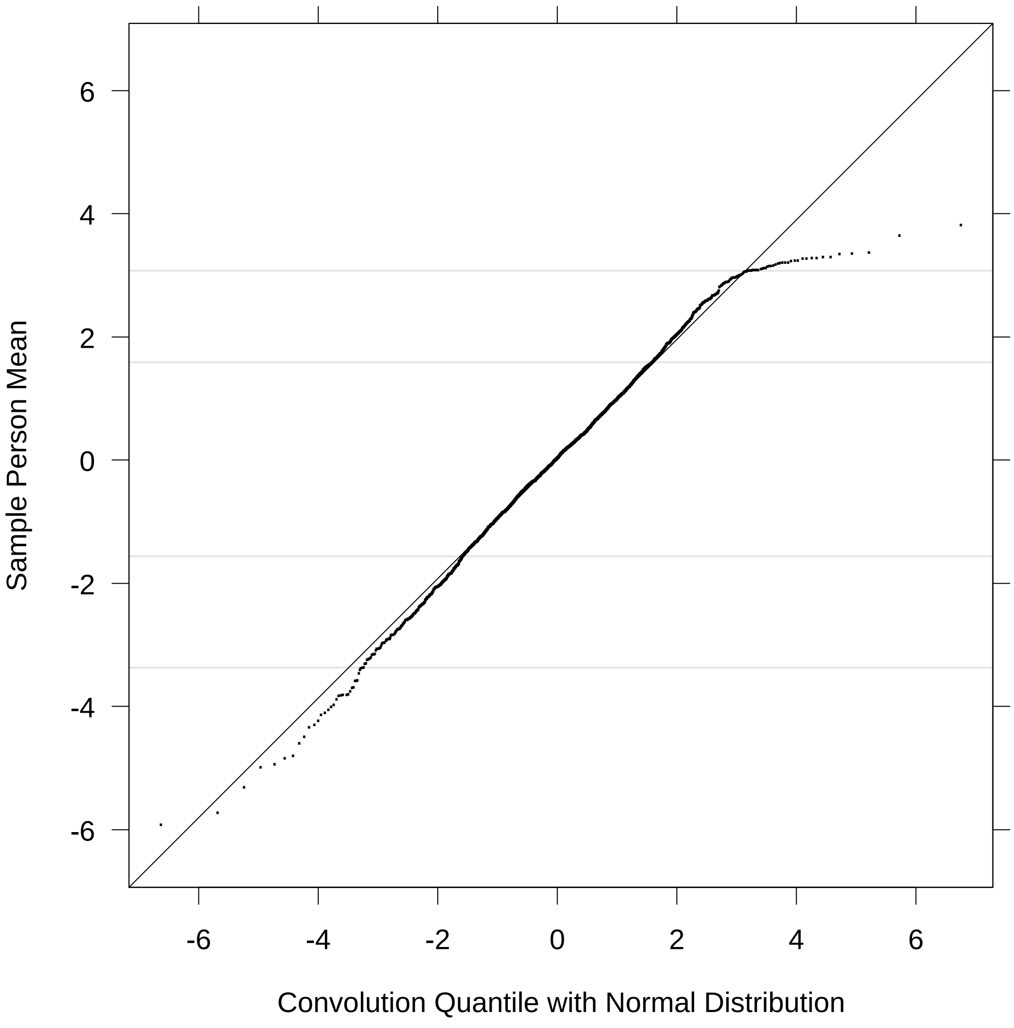

Figure 7: Quantile Plot of Sample Person Means

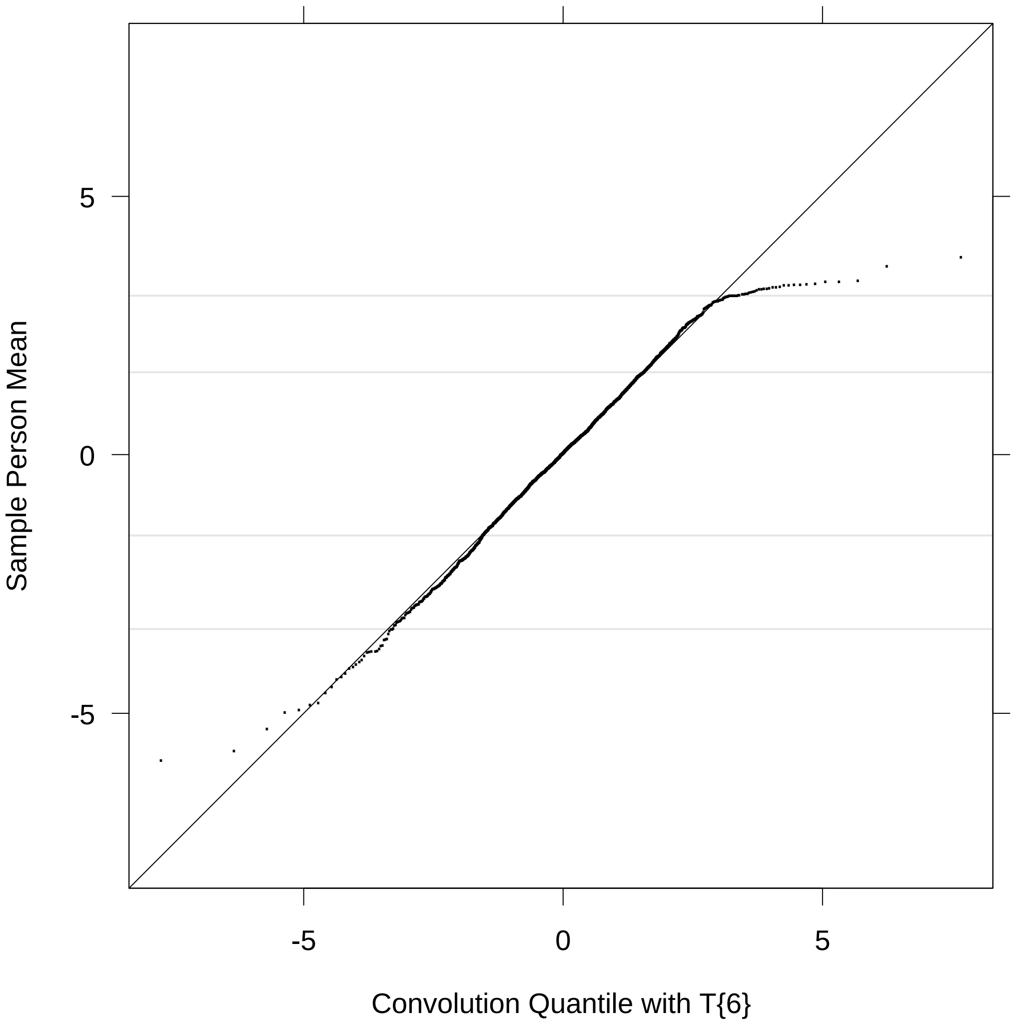

Figure 8: Quantile Plot of Sample Person Means

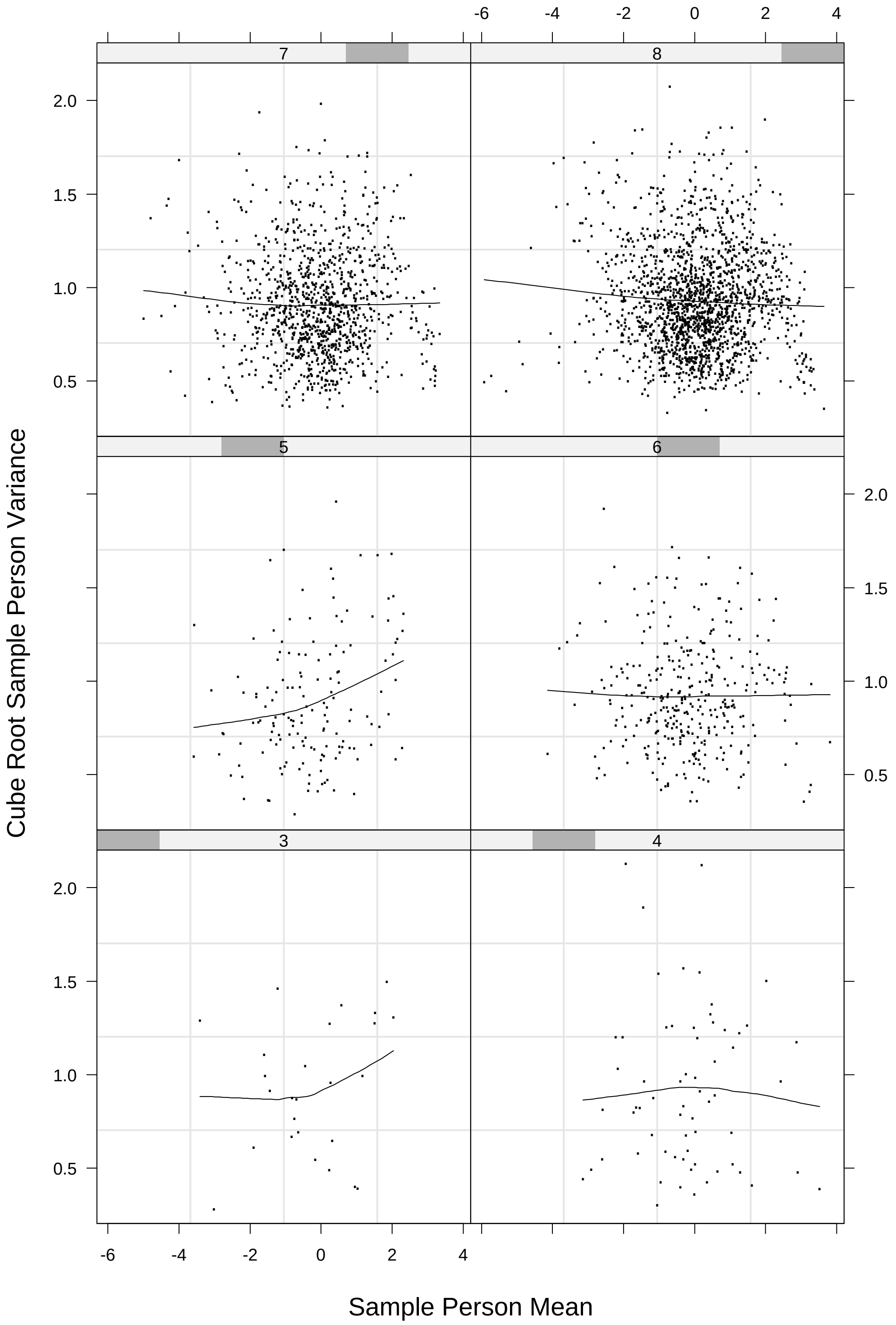

Figure 9: Cube Root Sample Person Variance vs. Sample Person Mean Given $n(p)$

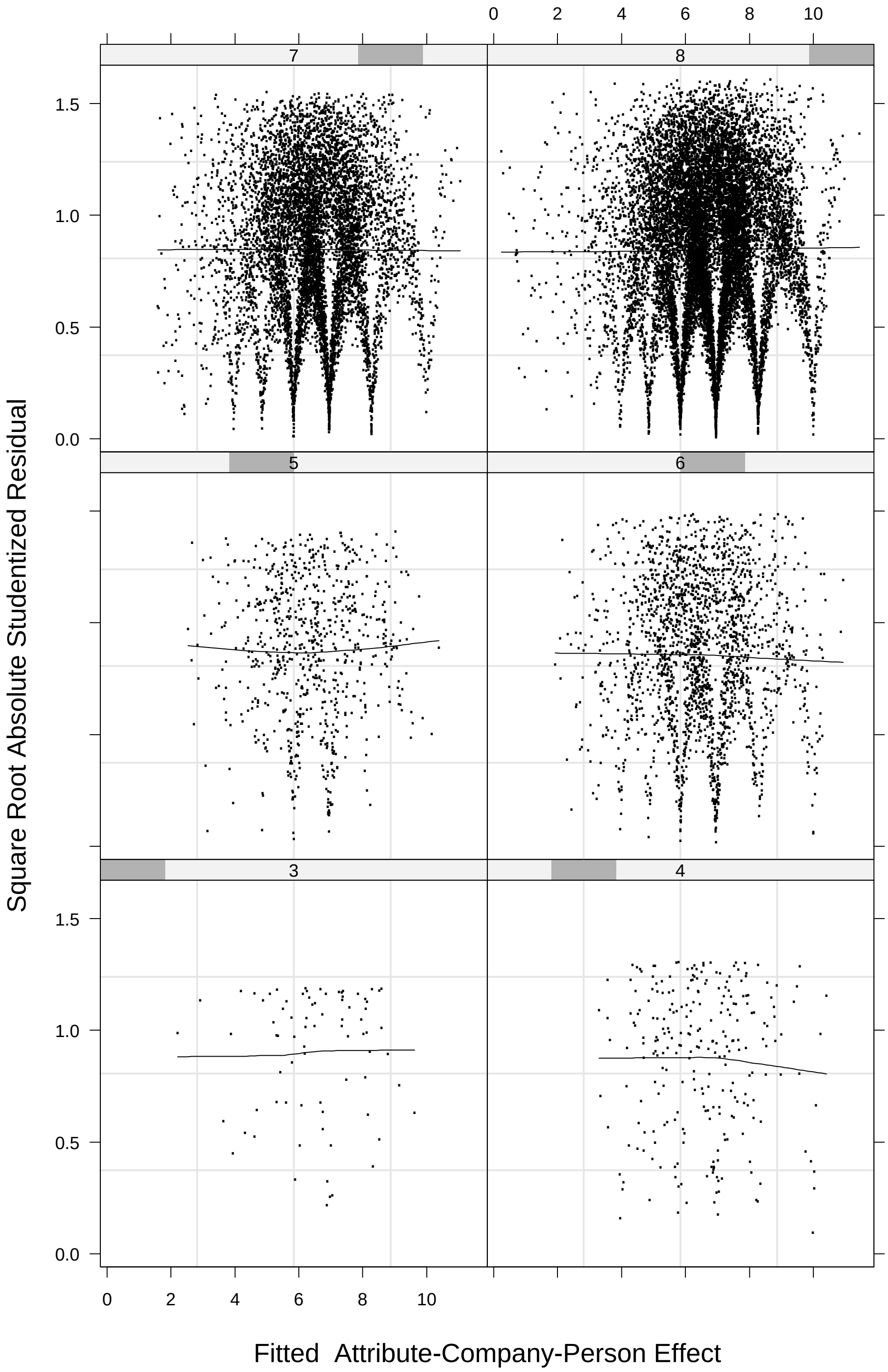

Figure 10: Square Root Absolute Studentized Residual vs. Fitted Rating Given $n(p)$

We estimated correlations of the studentized residuals for all pairs of attributes across respondents (and therefore companies and quarters) by a simple cross product:

$$n_{ab}^{-1} \sum_c \hat{\phi}_{ac} \hat{\phi}_{bc}$$

where n_{ab}^{-1} is the number of terms in the sum. We then subtracted expected correlations under an assumption of no correlation. The difference measure provides a check for correlation in the errors.

The differences are displayed in Figure 11, a trellis display that shows the correlation measure given each attribute; each panel shows the measure between the panel attribute and all other attributes for which there is a measure (some pairs never appear because of the survey plan). We found a small level of correlation that is in accord with our exogenous knowledge. Figure 11 portrays a simple clustering of the attributes; the breaks between the panels divide the attributes into four clusters. Also, the x's show differences for attribute pairs in the same cluster and the •'s show differences for attribute pairs in different clusters. We formed the clusters by the following method. We began with all attributes in separate clusters and then at each stage merged the two clusters that had the largest single value of the difference measure for all pairs of attributes, one in one cluster and one in the other. We continued until there were no singletons. The clustering was successful in that for each cluster, the differences within the cluster are greater than the differences for one attribute in the cluster and one out; the right vertical line on the panels is drawn at the minimum within-cluster measure.

The clustering is quite sensible given our exogenous information. The first cluster at the top of Figure 11 concerns the physical product, its quality and its technology. The second concerns the company-customer interaction. The third concerns the price of the product and whether it is worth what the customer paid for it.

While we have detected believable correlation, the magnitude is exceedingly small. If we can rely on the magnitudes of our estimates, we are simply showing that we have a lot of data; the amount of correlation is no threat to an assumption of independence and it would simply increase complexity at little benefit to include it in the model. (In saying this we must remember that a correlation of 0.2 results in an variance reduction of 4%.)

But the question is whether we can rely on the magnitude of the correlation in the studentized residuals. Let us couch the question in likelihood terms. The data displayed in Figure 11, the actual minus expected cross products, have high likelihood given no correlation or a small amount. But is it possible that the likelihood of the data given a large amount of correlation is also high? Can the processing that results in the studentized residuals mask correlation? We investigated this for several correlation structures. If all correlations among attributes were equal and large, then our exploratory method would in fact reduce them substantially and thus conceal them. In other words, the likelihood of the data in Figure 11 is large given large equal correlations. But we rule this case out based on our exogenous knowledge. We also investigated the case of equal high correlation within each of several clusters of attributes and no correlation between clusters; this has some plausibility given our exogenous information. More specifically we looked at a clustering that corresponds to the observed one in Figure 11. In such a case the magnitude of the correlation is not masked, so the likelihood of the data given such block correlation with high correlations is low.

8 From Model Building to a Bayesian Model

8.1 The General Case

Once our model identification is completed, moving to a full Bayesian model from a specified member of our class of models requires additional specification of distributions for unknown parameters that arise in the specified model. Such a Bayesian model is in part hierarchical with the case distributions at the first level of the hierarchy and the distributions of the unknown parameters at the second level.

8.2 The Rater Data

We will use the rater data to illustrate moving to a Bayesian model. The specified general model for the rater data resulting from the model building of earlier sections is

$$r_{ac} = \theta_{asq} + \mu_c + \alpha_c \epsilon_{ac}. \tag{11}$$

where

$$E(\mu_c) = E(\epsilon_{ac}) = E(\rho_c) = 0,$$

and

$$E(\alpha_c^2) = 1. \tag{12}$$

The distributions of the case effects and errors are

$$\begin{aligned} \mu_c &\sim \sigma(\mu) t\{6\} \\ \alpha_c^2 &\sim exp\{1\} \\ \epsilon_{ac} &\sim \sigma_a(\epsilon) N\{0,1\}. \end{aligned}$$

All of these random variables are independent of one another and of the θ_{asq}.

For fixed a and s we suppose that for $q = 1$ to 9, the time series θ_{asq} is an ISD(1,1) process (Cleveland and Liu, 1998c):

$$\theta_{asq} = \kappa_{asq} + \iota_{asq} \tag{13}$$

$$\kappa_{asq} - \kappa_{as(q-1)} = \lambda_{asq} + \lambda_{as(q-1)}, \tag{14}$$

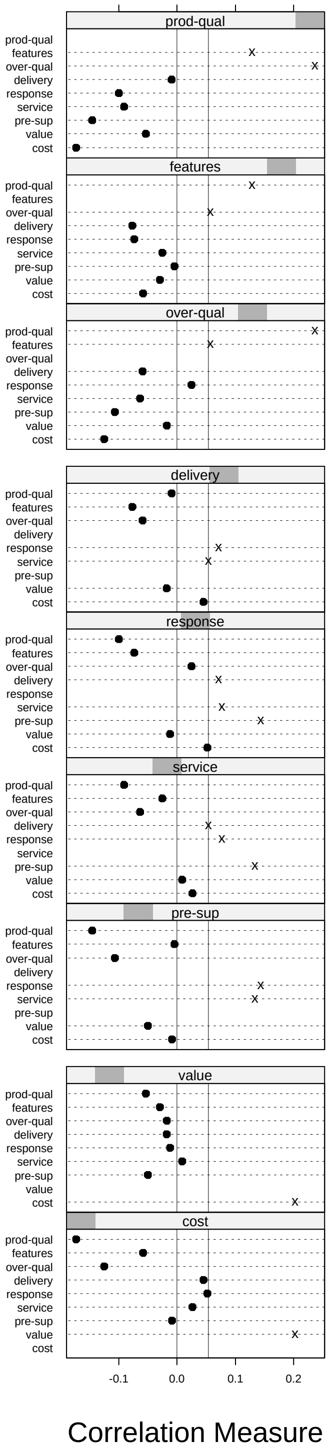

Figure 11: Correlation Measures of Studentized Residuals for Pairs of Attributes Given Attribute

where the λ_{asq} and ι_{asq} are independent of one another, and each is a gaussian white noise series with variances $\sigma(\kappa)$ and $\sigma(\iota)$. This specification is in keeping our exogenous information that company effects through time are largely smooth. While there is substantial exogenous information about the θ_{asq} across a for fixed c and q only mild assumptions were made, a hierarchical distribution on the 19×9 values of κ_{as1} for all combinations of a and s. We take them to be independent with

$$\kappa_{as1} \sim N\{\tau_a, 1\} \tag{15}$$

where the 9 τ_a to have improper uniform distributions.

The above model specifications have given rise to several variances: $\sigma_a^2(\epsilon)$, $\sigma^2(\mu)$, $\sigma^2(\rho)$, $\sigma^2(\kappa)$, and, $\sigma^2(\iota)$. We wish to place little restriction on the these parameters based on our exogenous information; we specify distributions for them by taking their inverses to be unit exponentials. Further discussion of this model, including computational methods, is given in (Clark et al. 1998).

9 The General Approach

We have conveyed models and model building methods in the context of two examples. The basic ideas and approach, however, are more generally applicable, although in any specific application the data analyst needs to tailor details to the specifics of the explanatory variables. In this section we will present the general model, and then discuss the two examples and three other data sets as instances of the general model.

9.1 The General Model

Let x_{ac} be a vector of values of v explanatory variables corresponding to r_{ac}. The general model for the data is

$$r_{ac} = g(x_{ac}; \beta) + \mu_c(x_{ac}) + \alpha_c \epsilon_{ac}(x_a c). \tag{16}$$

$g(x_{ac}; \beta) = E(r_{ac})$ models a dependence of the expected value of the response on the explanatory variables where the β are unknown parameters. The β, μ_c, α_c, and ϵ_{ac} are mutually independent. The $\mu_c(x_{ac})$, whose means are zero, have distributions $F_\mu(x_{ac})$ that can depend on the explanatory variables. The $\epsilon_{ac}(x_{ac})$ are normally distributed with zero means, but the variances $\sigma^2_{ac}(x_{ac})$ can depend on the explanatory variables. α_c has a distribution F_α where $\mathrm{E}(\alpha_c^2) = 1$.

The general model and the model building methods are meant to convey an approach rather than one fixed equation about relationships. We can alter aspects of the model and methods in many ways to accommodate the data and stay within our general framework. For example, while we focus here on normal distributions for the ϵ_{ac}, it seems likely that our general approach would work for other distributions, but details would need to be worked out.

By model identification, we mean filling in specifications not stipulated explicitly by the general model such as the form of g_{ac}, $F_\mu(x_a c)$, and F_α. By model checking, we mean any additional checking of specifications once they have been made, either specifications made at the beginning such as the additivity of the general model, or specifications made during the model building based on data exploration. The model building, identification and checking, consists of the following:

- identify $g(x_{ac}; \beta)$;
- identify $F_\mu(x_{ac})$;
- identify F_α;
- identify $\sigma^2(x_{ac})$;
- check the assumption of normality of $\epsilon_{ac}(x_{ac})$;
- check the assumption of additivity of the terms $g(x_{ac}; \beta)$, $\mu_c(x_{ac})$, and $\alpha_c \epsilon_{ac}(x_{ac})$;
- check the assumption of the mutual independence of the four collections of variables: β, $\mu_c(x_{ac})$, α_c, and ϵ_{ac}.

9.2 Examples

For the simple example discussed in Section 3, there are no explanatory variables, $g(\beta)$ is an unknown constant, and each of the three collections of variables — μ_c, α_c, and ϵ_{ac} — are identically distributed.

For the customer survey data studied in Sections 5 to 7, x_{ac} takes on three values: the attribute a, the supplier s, and the quarter q. $g(x_{ac}; \beta)$, to start, is simply θ_{asq}, but information from sources outside of the data together with exploration of the data leads to a time series specification of the behavior of θ_{asq} for fixed a and s across the quarters q. The distributions of μ_c and α_c do not depend on x_{ac}. But the variance of $\epsilon_{ac}(x_{ac})$ depends on a.

Bradlow and Zaslavsky (1997) builds a model for customer survey data where 20 attributes are judged on a scale of 1 to 10 for one company. So as with the data in Sections 5 to 7, attribute is an explanatory variable. The attributes fall into six categories. The case location effect is taken to be

$$\mu_c(x_{ac}) = \sigma_\nu \nu_c + \sigma_\eta \eta_{u(a)c}$$

where $u(a)$ is the group for attribute a, and $\sigma_\nu \nu_c$ and $\sigma_\eta \eta_{u(a)c}$ are independent normal variables with mean zero and variance one. The result of this location effect is to account for a case correlation in the responses where the correlation for two attributes within a category is different from the correlation of two attributes in different categories.

In this last example, the case location effect has a distribution that depends on an explanatory variable. This is a special case of a more general linear regression location effect with case regression coefficients; the case location effect is

$$\mu_c(x_{ac}) = \beta_{ac} x_{ac}$$

The model building methods of Section 5 extend readily to this example with the case regressions estimating the regression coefficients.

References

Becker, R. A., Cleveland, W. S. and Shyu, M. J. (1996). The Design and Control of Trellis Display, *Journal of Computational and Statistical Graphics* **5**: 123–155. Black and white postscript — *http://cm.bell-labs.com/stat/doc/trellis.jcgs.bw.ps.* Color postscript — *http://cm.bell-labs.com/stat/doc/trellis.jcgs.col.ps.*

Box, G. E. P. (1980). Sampling and Bayes Inference in Scientific Modelling and Robustness, *Jour. Roy. Stat. Soc., A* **143**: 383–430.

Bradlow, E. T. and Zaslavsky, A. M. (1997). A Hierarchical Latent Variable Model for Ordinal Data from a Customer Satisfaction Survey with "No Answer" Responses, *Technical report*, Department of Marketing, Wharton School, University of Pennsylvania.

Brady, H. (1985). The Perils of Survey Research: Interpersonally Incomparable Responses, *Political Methodology* **11**: 269–292.

Brady, H. (1989). Factor and Ideal Point Analysis for Interpersonally Incomparable Data, *Political Methodology* **11**: 269–292.

Carroll, R. and Hall, P. (1988). Optimal Rates of Convergence for Deconvolving a Density, *jasa* **83**: 1184–1186.

Clark, L., Cleveland, W. S., Denby, L. and Liu, C. (1998). Modeling Customer Survey Data, *in* C. Gatsonis, R. Kass, A. Carriquiry, A. Gelman, I. Verdinelli and M. West (eds), *Case Studies in Bayesian Statistics IV*, Springer, New York.

Cleveland, W. S. (1993). *Visualizing Data*, Hobart Press, Summit, New Jersey, U.S.A., books@hobart.com.

Cleveland, W. S. and Devlin, S. J. (1988). Locally-Weighted Regression: An Approach to Regression Analysis by Local Fitting, *Journal of the American Statistical Association* **83**: 596–610.

Cleveland, W. S. and Liu, C. (1998). An Approach to Model Building Based on a Bayesian Theory of Data Exploration, **in preparation**.

Cordy, C. B. and Thomas, D. R. (1997). Deconvolution of a Distribution, *Political Methodology* **11**: 269–292.

Gelman, A., Carlin, J. B., Stern, H. S. and Rubin, D. B. (1995). *Bayesian Data Analysis*, Chapman and Hall, New York.

Gelman, A., Meng, X.-L. and Stern, H. (1996). Posterior Predictive Assessment of Model Fitness Via Realized Discrepancies, *Statistica Sinica* **6**: 733–807.

Johnson, V. E. (1997). An Alternative to Traditional GPA for Evaluating Student Performance (with discussion), *Statistical Science* **12**: 251–278.

Kass, R. E., Tierney, L. and Kadane, J. B. (1989). Approximate Methods for Assessing Influence and Sensitivity in Bayesian Analysis, *Biometrika* **67**: 663–674.

Krantz, D., Luce, D., Suppes, P. and Tversky, A. (1971). *Foundations of Measurement*, Academic Press, New York.

Laird, N. (1978). Nonparametric Maximum Likelihood Estimation of a Mixing Distribution, *jasa* **73**: 805–811.

Longford, N. T. (1995). *Models for Uncertainty in Educational Testing*, Springer, New York.

Maritz, J. S. and Lwin, T. (1989). *Empirical Bayes Methods*, Chapman and Hall, New York.

Mendelsohn, J. and Rice, J. (1982). Deconvolution of Microfluorometric Histograms with B Splines, *jasa* **77**: 748–753.

Rossi, P. E., McCulloch, R. E. and Allenby, G. M. (1995). Hierarchical Modelling of Consumer Heterogeneity: An Application to Target Marketing, *in* C. Gatsonis (ed.), *Case Studies in Bayesian Statistics III*, Springer, New York, pp. 323–349.

Rubin, D. B. (1984). Bayesianly Justifiable and Relevant Frequency Calculations for The Applied Statistician, *The Annals of Statistics* **12**: 1151–1172.

Torgerson, W. S. (1958). *Theory and Methods of Scaling*, Wiley, New York.