CHAPTER 10

Joint modelling of mean and dispersion

10.1 Introduction

In the models so far discussed the variance of a response has been assumed to take the form

$$\mathrm{var}(Y_i) = \phi V(\mu_i),$$

in which $V(\mu)$ is a known variance function. The choice of variance function determines the interpretation of ϕ. For example if $V(\mu) = 1$, ϕ is the response variance: if $V(\mu) = \mu^2$, ϕ is the squared coefficient of variation (noise-to-signal ratio) of the response. Similarly for other variance functions. In the simplest of generalized linear models the dispersion parameter ϕ is a constant, usually unknown, but in circumstances where Y_i is the average of m_i elementary observations it may be appropriate to assume that ϕ_i is proportional to known 'weights' $w_i = 1/m_i$. For one example of this, see the discussion of the insurance-claims data in section 8.4.1. More generally it may be the case that ϕ_i varies in a systematic way with other measured covariates in addition to the weights. In this chapter, therefore, we explore the consequences of constructing and fitting formal models for the dependence of both μ_i and ϕ_i on several covariates, following a suggestion made by Pregibon (1984).

To a large extent the impetus for studying this extended class of models derives from the recent surge of interest in industrial quality-improvement experiments in which both the mean response and the signal-to-noise ratio are of substantive interest. For economy of effort, fractional factorial and related experimental designs are often used for this purpose. The aim very often is to select that combination of factor levels that keeps the mean at a pre-determined 'ideal' value, while at the same time keeping the variability in the product at a minimum. It is thus necessary to study not just how the mean response is affected by factors under study, but also how the variance, or other suitable measure of variability such as the noise-to-signal ratio, is affected by these factors.

For grouped data or ordinal responses, model (5.4) is designed to achieve a similar purpose.

10.2 Model specification

The joint model is specified here in terms of the dependence on covariates of the first two moments. For the mean we have the usual specification

$$\begin{aligned} E(Y_i) &= \mu_i; \quad \eta_i = g(\mu_i) = \sum_j x_{ij}\beta_j, \\ \operatorname{var}(Y_i) &= \phi_i V(\mu_i), \end{aligned} \tag{10.1}$$

in which the observations are assumed independent. The dispersion parameter is no longer assumed constant but instead is assumed to vary in the following systematic way.

$$\begin{aligned} E(d_i) &= \phi_i; \quad \zeta_i = h(\phi_i) = \sum_j u_{ij}\gamma_j \\ \operatorname{var}(d_i) &= \tau V_D(\phi_i). \end{aligned} \tag{10.2}$$

In this specification $d_i \equiv d_i(Y_i; \mu_i)$ is a suitable statistic chosen as a measure of dispersion; $h(\cdot)$ is the dispersion link function; ζ is the dispersion linear predictor, and $V_D(\phi)$ is the dispersion variance function. The dispersion covariates $\mathbf{u}_i$ are commonly, but not necessarily, a subset of the regression covariates $\mathbf{x}_i$.

Two possible choices for the dispersion statistic d_i are

1. the generalized Pearson contribution

$$d_i = r_{\mathrm{P}}^2 = (Y_i - \mu_i)^2 / V(\mu_i)$$

2. the contribution to the deviance of unit i: $d_i = r_{\mathrm{D}}^2$.

For Normal-theory models, but not otherwise, the two forms are equivalent. Note that, when evaluated at the true μ, $E(r_{\mathrm{P}}^2) = \phi$ exactly whereas $E(r_{\mathrm{D}}^2) \simeq \phi$ only approximately.

To fit the extended model we must choose suitable dispersion variance and link functions. If Y is Normal d_i has the $\phi_i\chi_1^2$ distribution, so that a gamma model with $V_D(\phi) = 2\phi^2$ would be chosen. The most natural link functions include the identity, corresponding to additive variance components, and the log, corresponding to multiplicative effects of the covariates. If Y is non-Normal, adjustments to the dispersion model may be necessary to account for the bias in r_{D}^2 or for the excess variability of r_{P}^2. Some possibilities are discussed in section 10.5.

The two models (10.1) and (10.2) are interlinked; that for the mean requires an estimate of $1/\phi_i$ to be used as prior weight, while the dispersion model requires an estimate of μ_i in order to form the dispersion response variable d_i. The form of the interlinking suggests an obvious algorithm for fitting these models, whereby we alternate between fitting the model for the means for given weights $1/\hat{\phi}_i$, and fitting the model for the dispersion using the response variable $d_i = d_i(Y_i, \hat{\mu}_i)$.

10.3 Interaction between mean and dispersion effects

If the data contain replicate observations for each combination of covariate values for the mean, then an estimate of the variance can be formed for each distinct point in the covariate space of the model for $E(Y)$. Suppose now that we fit p parameters in the model for the mean response. With a total of n' distinct x-values this leaves $n' - p$ contrasts having zero mean that contain information about the dispersion. The information from these $n' - p$ contrasts can then be combined with the replicate estimates to improve the model for the dispersion. The practical difficulty is that use of supposedly null contrasts presupposes that the model for the mean be substantially correct. For suppose, for example, that a continuous covariate contributes a term βx to the linear predictor for the mean, but β is small so that the effect is judged insignificant. Nonetheless its omission from the model for the mean may produce relatively large values of $(y - \hat{\mu})^2$ at the two ends of the x-scale and small values at the centre. This characteristic pattern will appear as a quadratic effect of x in the dispersion model. Likewise, omission of an interaction between two factors in the linear predictor for the mean will result in the inflation of supposedly null contrasts used to model the dispersion.

The correct choice of variance function for the mean is also important if distortion of the dispersion model is to be avoided. Thus, in designing an experiment for modelling both mean and dispersion, it is advisable to have estimates of dispersion based on pure replicates. Information from null contrasts can then be combined with the information from replicate contrasts if they prove compatible.

10.4 Extended quasi-likelihood as a criterion

The extended quasi-likelihood, Q^+, developed in section 9.6, provides a possible criterion to be maximized for the estimation of μ and ϕ and for measuring the goodness of fit. We write

$$-2Q^+ = \sum_1^n \frac{d_i}{\phi_i} + \sum_1^n \log(2\pi\phi_i V(y_i)), \tag{10.3}$$

where d_i are the deviance components in the model for the means, i.e.

$$d_i = 2\int_{\mu_i}^{y_i} \frac{y_i - t}{V(t)}\, dt.$$

Suppose now that the two parts of the model are parameterized as $\boldsymbol{\mu} = \boldsymbol{\mu}(\boldsymbol{\beta})$ and $\boldsymbol{\phi} = \boldsymbol{\phi}(\boldsymbol{\gamma})$. Then, from equation (10.3) we see that the estimating equations for $\boldsymbol{\beta}$ are the Wedderburn quasi-likelihood equations

$$\sum_{i=1}^n \frac{y_i - \mu_i}{\phi_i V(\mu_i)} \frac{\partial \mu_i}{\partial \beta_j} = 0, \tag{10.4}$$

except that $1/\phi_i$ must now be included as a weight, the dispersion being non-constant.

The estimating equations for $\boldsymbol{\gamma}$ are given by

$$\sum_{i=1}^n \frac{d_i - \phi_i}{\phi_i^2} \frac{\partial \phi_i}{\partial \gamma_j} = 0. \tag{10.5}$$

These are the Wedderburn quasi-likelihood equations for $V(\mu) = \mu^2$ with the deviance component as response variable. Thus, so far as estimation is concerned, the use of Q^+ as an optimizing criterion is equivalent to assuming that the deviance component has a variance function of the form $V_D(\phi) = \phi^2$, regardless of the variance function for Y. This can only be approximately correct, so we now consider some adjustments to the estimating equations for the dispersion to allow for non-Normal Y and other factors.

10.5 Adjustments of the estimating equations

10.5.1 *Adjustment for kurtosis*

The estimating equations for the dispersion parameters obtained from Q^+ are the same as those that would be obtained by assuming that d_i has the $\phi_i\chi_1^2$ distribution, i.e. that we have gamma errors with a scale factor of 2 (gamma index $= \frac{1}{2}$). In fact, however, the variance of d_i often exceeds the nominal value of $2\phi_i^2$, and appropriate allowance should be made for this excess variability. The correct variance of $(Y-\mu)^2$ is

$$\mathrm{var}\{(Y-\mu)^2\} = \kappa_4 + 2\kappa_2^2 = 2\kappa_2^2(1 + \rho_4/2),$$

where

$$\rho_4 = \kappa_4/\kappa_2^2$$

is the standardized fourth cumulant. Thus the variance of $r_P^2 = (Y-\mu)^2/V(\mu)$ is $2\phi^2(1+\rho_4/2)$. To use this result we need to know the value of ρ_4. However for over-dispersed Poisson and binomial distributions the adjustment can be made provided that the fourth cumulant of Y has a particular pattern in relation to the second cumulant. If condition (9.21) holds up to fourth order then

$$\kappa_2 = \phi V, \qquad \kappa_3 = \phi^2\frac{\partial V}{\partial \theta}, \quad \text{and} \quad \kappa_4 = \phi^3\frac{\partial^2 V}{\partial \theta^2}.$$

Consequently ρ_3 and ρ_4 are expressible in terms of ϕ and the derivatives of the variance function as follows:

$$\rho_3 = \phi^{1/2}V'(\mu)/\{V(\mu)\}^{1/2} \quad \text{and} \quad \rho_4 = \phi V''(\mu) + \rho_3^2,$$

where primes denote differentiation with respect to μ.

Under similar assumptions the approximate mean and variance of the deviance contribution, r_D^2, are

$$E(r_D^2) \simeq \phi(1+b)$$
$$\mathrm{var}(r_D^2) \simeq 2\phi^2(1+b)^2,$$

where $b = b(\phi, \mu) = (5\rho_3^2 - 3\rho_4)/12$ is usually a small adjustment. In general the standardized cumulants $\rho_3^2 = \kappa_3^2/\kappa_2^3$ and ρ_4 depend on both μ and ϕ.

Expressions for these adjustments are shown in Table 10.1.

Table 10.1. *Dispersion adjustments for some standard distributions*

Distribution	$1+\rho_4/2$	$b(\phi,\mu)$
Normal	1	0
Poisson†	$1+\phi/(2\mu)$	$\phi/(6\mu)$
Binomial‡	$1+\dfrac{\phi}{2m}\left(\dfrac{1-6\pi(1-\pi)}{\pi(1-\pi)}\right)$	$\dfrac{\phi}{6m}\left(\dfrac{1-\pi(1-\pi)}{\pi(1-\pi)}\right)$
Gamma	$1+3\phi$	$\phi/6$
Inverse Gaussian	$1+15\phi\mu/2$	0

†with over-dispersion (Section 6.2.3)

‡with over-dispersion (Section 4.5)

The estimating equations for the dispersion parameters may be adjusted by incorporating $(1+\rho_4/2)^{-1}$ or an estimate thereof as a prior weight. The use of such an adjustment has been proposed by Prentice (1988) in the context of over-dispersed binary data.

10.5.2 *Adjustment for degrees of freedom*

The dispersion estimating equations derived from Q^+ make no allowance for the fact that p parameters have been fitted to the means. The effect of fitting is to decrease the average size of the dispersion response variables d_i. A simple adjustment is to multiply the second term in Q^+ by ν/n, where $\nu = n - p$ is the residual degrees of freedom for the deviance. Thus, for the purpose of fitting the dispersion response model, we use the modified Q_M^+

defined by

$$-2Q_M^+ = \sum_i \frac{d_i}{\phi_i} + \frac{\nu}{n}\sum_i \log(2\pi\phi_i V(y_i)). \qquad (10.6)$$

For a model in which the dispersion is constant, ($\phi_i = \phi$), this adjustment gives

$$\hat{\phi} = D/\nu$$

by analogy with the unbiased estimator of variance for Normal-theory linear models. More generally, this adjustment yields approximate restricted maximum likelihood estimates, which are widely preferred to unadjusted maximum-likelihood estimates for the estimation of variance components and covariance functions. See section 7.2. If, as is common, the dispersion link is the logarithm, the modification changes only the intercept, which is often of little interest. However, with a beta-binomial model, for which the dispersion factor is

$$\phi_i = 1 + \theta(m_i - 1),$$

Q_M^+ and the unmodified Q^+ give different estimates of θ.

10.5.3 *Summary of estimating equations for the dispersion model*

The preceding discussion indicates that there is a variety of minor variations among the possible estimating equations for fitting the dispersion model. There are at least $2^3 = 8$ variations based on the following:

1. choice between $d = r_P^2$ and $d = r_D^2$;
2. choice between prior weight 1 and $(1 + \rho_4/2)^{-1}$;
3. adjustment for degrees of freedom or not.

On balance it appears desirable to make the adjustment for degrees of freedom. Adjustment for kurtosis also seems to be desirable provided that a reasonably accurate estimate of ρ_4 is available. The choice between r_P^2 and r_D^2 is less clear-cut.

Note that Q^+ and Q_M^+ provide an optimizing criterion using extended quasi-likelihood for only two of these forms. For the remainder we must rely on the theory of optimum estimating equations. Further work is required to give guidance for selection among these alternatives. Yet another form is based on work by Godambe and Thompson (1988), which we describe next.

10.6 Joint optimum estimating equations

Beginning with the pair of elementary estimating functions

$$g_{1i} = Y_i - \mu_i,$$
$$g_{2i} = (Y_i - \mu_i)^2 - \phi_i V(\mu_i),$$

both of which have zero mean, optimum estimating equations for both the regression parameters and the dispersion parameters may be derived using the method described in section 9.4. In order to derive these equations we require the covariance matrix of $\boldsymbol{g}$ together with the expected derivative matrix of $\boldsymbol{g}$ with respect to the parameters. In carrying out this differentiation it is convenient initially to take the parameters (μ_i, ϕ_i) to be unrestricted.

The covariance matrix of (g_{1i}, g_{2i}) is

$$\mathbf{V}_i = \begin{pmatrix} \kappa_2 & \kappa_3 \\ \kappa_3 & \kappa_4 + 2\kappa_2^2 \end{pmatrix},$$

in which $\kappa_2 \equiv \phi_i V(\mu_i)$ and the subscript i has been omitted from the cumulants. The inverse covariance matrix is

$$\mathbf{V}_i^{-1} = \frac{1}{\Delta}\begin{pmatrix} \kappa_4 + 2\kappa_2^2 & -\kappa_3 \\ -\kappa_3 & \kappa_2 \end{pmatrix},$$

where $\Delta = \det(\mathbf{V}_i) = \kappa_2^3(2 + \rho_4 - \rho_3^2)$.

The negative expected derivative matrix of (g_{1i}, g_{2i}) with respect to (μ_i, ϕ_i) is

$$\mathbf{D}_i = \begin{pmatrix} 1 & 0 \\ \phi V' & V \end{pmatrix}$$

with rows indexed by the components of $\boldsymbol{g}$. Thus

$$\mathbf{D}_i^T \mathbf{V}_i^{-1} = \frac{1}{\Delta}\begin{pmatrix} \kappa_4 + 2\kappa_2^2 - \kappa_3\phi V' & \kappa_2\phi V' - \kappa_3 \\ -\kappa_3 V & \kappa_2 V \end{pmatrix}.$$

Provided that the regression and dispersion models have no parameters in common, the estimating equations thus obtained for $(\boldsymbol{\beta}, \boldsymbol{\gamma})$ are

$$\begin{aligned} \sum_i \left\{ \frac{\kappa_4 + 2\kappa_2^2 - \kappa_3\phi V'}{\Delta} g_{1i} + \frac{\kappa_2\phi V' - \kappa_3}{\Delta} g_{2i} \right\} \frac{\partial \mu_i}{\partial \beta_j} &= 0, \\ \sum_i (g_{2i} - \kappa_3 g_{1i}/\kappa_2) \frac{\kappa_2 V}{\Delta} \frac{\partial \phi_i}{\partial \gamma_r} &= 0. \end{aligned} \tag{10.7}$$

The subscript i has been omitted in all coefficients. These are not the same as the extended quasi-likelihood equations (10.4), (10.5), even after (10.5) is adjusted for kurtosis.

Note that if $\kappa_3 = \kappa_2 \phi V'$, a property of exponential-family distributions, we have

$$\mathbf{D}^T\mathbf{V}^{-1} = \begin{pmatrix} \kappa_2^{-1} & 0 \\ -\kappa_3 V/\Delta & \kappa_2 V/\Delta \end{pmatrix}.$$

It follows that if the regression and dispersion models have no parameters in common then the estimating equations for the regression parameters are

$$\sum_i \frac{y_i - \mu_i}{\phi_i V(\mu_i)} \frac{\partial \mu_i}{\partial \beta_j} = 0, \qquad j = 1, \ldots, p,$$

which is identical to the quasi-likelihood equation (10.4).

10.7 Example: the production of leaf-springs for trucks

These data, taken from Pignatiello and Ramberg (1985), relate to the production of leaf springs for trucks. A heat treatment is to be designed such that the free height y of a spring in an unloaded condition is as close as possible in mean value to eight inches, while having as small a variability as possible. To this end a one half fraction of a 2^5 experiment, with each treatment combination replicated three times, was performed using the factors

B: furnace temperature,
C: heating time,
D: transfer time,
E: hold-down time,
O: quench oil temperature.

The data are given in Table 10.2.

Of the five factors O is somewhat different from the others in that it is apparently less easily controlled. We shall nevertheless follow Nair and Pregibon (1988) by treating it in the same way as the others.

Table 10.2. *Data for a replicated 2^{5-1} factorial experiment to investigate the free height of leaf springs*†

	Factor level							
Run	*B*	*C*	*D*	*E*	*O*	*Free height*		
1	−	−	−	−	−	7.78	7.78	7.81
2	+	−	−	+	−	8.15	8.18	7.88
3	−	+	−	+	−	7.50	7.56	7.50
4	+	+	−	−	−	7.59	7.56	7.75
5	−	−	+	+	−	7.94	8.00	7.88
6	+	−	+	−	−	7.69	8.09	8.06
7	−	+	+	−	−	7.56	7.62	7.44
8	+	+	+	+	−	7.56	7.81	7.69
9	−	−	−	−	+	7.50	7.25	7.12
10	+	−	−	+	+	7.88	7.88	7.44
11	−	+	−	+	+	7.50	7.56	7.50
12	+	+	−	−	+	7.63	7.75	7.56
13	−	−	+	+	+	7.32	7.44	7.44
14	+	−	+	−	+	7.56	7.69	7.62
15	−	+	+	−	+	7.18	7.18	7.25
16	+	+	+	+	+	7.81	7.50	7.59

†Source: Pignatiello and Ramberg (1985).

We require models for both the mean and dispersion effects with a view to finding the factor combination that minimizes the dispersion while keeping the mean close to the target value of eight inches. We begin by modelling the mean assuming homogeneity of the dispersion. The range of the response variable is small in relation to the mean response, so we are unlikely to find evidence to cast doubt on the assumptions of Normality or constancy of variance. From Fig. 10.1, where the run variances are plotted against the run means, there is little evidence that the variance changes with the mean response.

A main-effects model for the mean response shows that D has a negligible effect, and a model with all two-factor interactions shows the effect of E to be independent of the rest. Finally we arrive at a linear predictor of the form

$$M = (B + C).O + E.$$

Note that the defining contrast for this design is

$$I = BCDE,$$

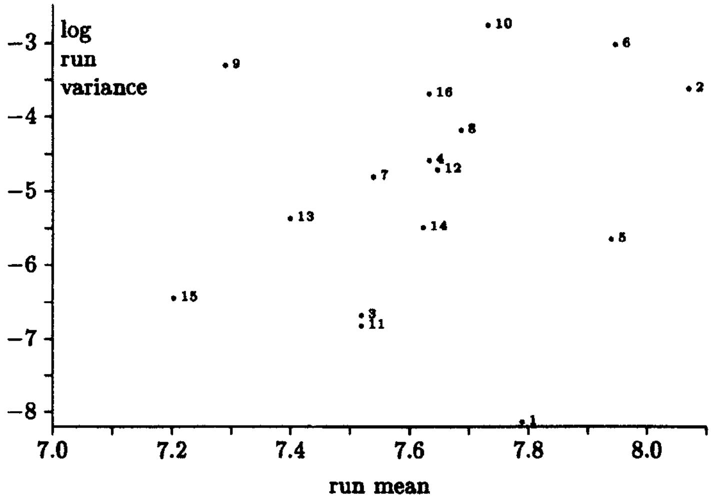

Fig. 10.1 ***Run variances (log scale) plotted against run means for leaf-spring data.***

so that the aliased pairs of two-factor interactions are

$$BD \equiv CE, \quad CD \equiv BE, \quad DE \equiv BC.$$

Fortunately none of these interactions is significant in this analysis.

A shortened analysis of variance table takes the form

Term	*s.s.*	*d.f.*	*m.s.*
$B+C+E+O$	1.898	4	0.4745
$B.O+C.O$	0.414	2	0.2071
Total M	2.312	6	0.3853
Rest of treatments	0.124	9	0.0138
Replicates	0.530	32	0.0166

The table of means for model M shows that C has virtually no effect at the higher level of O, but has a negative effect at the lower level. Increased furnace temperature (B) has a positive effect at all levels of O, but is twice as great at the higher level of O than at the lower level. Hold-down time (E) has a positive effect regardless of the other factors. The combination giving a fitted value closest to the target of eight inches is

$$(B, C, D, E, O) = (+, -, \pm, -, -), \text{ followed by } (+, -, \pm, +, -).$$

The fitted means are 7.953 and 8.057 respectively.

The analysis of dispersion is less clear cut. The original analysis by Pignatiello and Ramberg uses a linear model for the logarithms of the within-replicate sample variances. They selected the largest of the 15 factorial contrasts, which were B, DO, $BCO \equiv DEO$, $CD \equiv BE$ and $CDO \equiv BEO$. However, this selection procedure ignores the marginality conditions discussed in section 3.5, and carries a strong risk of selecting the accidentally large contrasts. A further objection to this analysis is that, with high probability, at least one of the sample variances will be exceptionally small. This may occur because of rounding. In the first run, for example, the sample variance is 0.0003, the next smallest being 0.0012. On the log scale, the sample variance for run 1 is exceptionally large and negative, which explains why such high-order interactions were found in the dispersion model.

If the original data were Normally distributed, the sample variances would be distributed as $s^2 \sim \phi\chi_2^2/2$ with $\phi = \sigma^2$. More generally, for samples of size $r = 3$, we have

$$E(s^2) = \kappa_2 = \phi, \quad \operatorname{var}(s^2) = \frac{\kappa_4}{r} + \frac{2\kappa_2^2}{r-1} = \phi^2(1 + \rho_4/3).$$

Thus we treat the replicate variances as the response, using gamma errors and log link. If ρ_4 is taken as zero the distribution of s^2 is effectively exponential, so the scale factor should then be taken to be unity. The resulting fits show that only B and C have any appreciable dispersion effects. The deviances for selected dispersion models are shown together with the extended quasi-likelihood criterion in Table 10.3.

Table 10.3 *Deviances for selected log-linear dispersion models fitted to the leaf-spring data*

Dispersion model	*Gamma deviance*	*d.f.*	*Quasi-likelihood* $-2Q_M^+$	*d.f.*
1	26.57	15	−40.4	31
B	20.58	14	−46.4	30
C	22.99	14	−44.0	30
$B+C$	16.08	13	−50.9	29
$B+C+D+E+O$	15.00	10	−52.0	26
$B.C$	15.89	12	−51.1	28

For the fits using EQL we take the prior dispersion parameter for the dispersion analysis as 2. The parameter estimates and standard errors for the log-linear dispersion model $B + C$ are

$$\begin{aligned} \hat{b} &= 1.370 \pm 0.500, \\ \hat{c} &= -1.092 \pm 0.500, \end{aligned} \tag{10.8}$$

showing that the variance at the higher furnace temperature (B) exceeds the variance at the lower temperature by an estimated factor of $\exp(1.370) \simeq 3.9$. The effect of increased heating time (C) is to decrease the variance by the factor $\exp(-1.092) \simeq 0.34$. Pearson's X^2 for the dispersion analysis is 28.59 on 29 degrees of freedom, showing that the prior dispersion of 2 is satisfactory.

Table 10.4 *Deviances for selected log-linear dispersion models*

Dispersion model	$-2Q_M^+$	*d.f.*
1	−53.3	40
B	−56.1	39
C	−54.1	39
$B+C$	−56.4	38
$B+C+D+E+O$	−58.0	35
$B.C$	−57.3	37

So far we have not used the information in the null contrasts for the means to augment the replicate variance estimates. We therefore repeat the above exercise using model $M = (B+C).O+E$ for the means. The results as shown in Table 10.4 are in conflict with those in Table 10.3. In particular C now has a negligible effect, and the effect of B is much reduced. The joint effect of B and C, which was highly significant in the analysis based on replicates (deviance reduction $= 10.5$), is now insignificant with a deviance reduction of only 3.1 on two degrees of freedom.

One possible explanation is that the null contrasts and the replicate contrasts are measuring variability of two different types or from different sources. To examine this possibility in more detail we now present an analysis of the null contrasts alone, ignoring the replicate contrasts. To accomplish this we analyse the run means, fitting model M for the means, and trying five dispersion models for the between-run variances. The replicate variance is now ignored,

Table 10.5 *Deviances for selected log-linear dispersion models using the between-runs contrasts*

Dispersion model	$-2Q_M^+$	*d.f.*
1	−22.9	8
B	−25.6	7
C	−26.0	7
$B+C$	−38.4	6
$B.C$	−39.0	5

and the dispersion analysis uses information in the null contrasts alone.

The results of a sequence of fits are as shown in Table 10.5. There are insufficient data available for fitting the dispersion model $B+C+D+E+O$. These results show a large effect for C, and a smaller effect for B. The parameter estimates for the dispersion model $B+C$ are

$$\hat{b} = -1.810 \pm 0.943,$$
$$\hat{c} = 4.794 \pm 0.943,$$

and so are of opposite sign to those derived from the replicate-contrasts analysis (10.8). This reversal of sign explains the apparently small effects of B and C obtained from the combined analysis using all contrasts.

10.8 Bibliographic notes

The idea of using a linked pair of generalized linear models for the simultaneous modelling of mean and dispersion effects was first put forward by Pregibon (1984). For linear models with Normal errors the idea is much older, a simple case being that of heterogeneous variances defined by a grouping factor; see Aitkin (1987) for a general treatment or Cook and Weisberg (1983), who discuss score tests. Smyth (1985) compares different algorithms for the estimation of mean and dispersion effects.

10.9 Further results and exercises 10

10.1 Using expression (15.10) or (C.4) from Appendix C justify the claim following (10.8) that, for exponential observations, the expected value of the mean deviance is approximately 7/6.

10.2 Explain why it is necessary in (10.7) to impose the condition that the regression and dispersion models should have no parameters in common, although they may have covariates and factors in common. Discuss briefly whether this is a reasonable condition in practice.

10.3 Derive the results listed in Table 10.1 using the assumptions of section 10.5.1 for over-dispersed Poisson and binomial distributions.

10.4 Show that the expected Fisher information matrix derived from Q^+ for the parameters (β, γ) is block-diagonal.