Model Selection Strategies for Experiments with Dispersion Effects

Transformations vs. Generalized Linear Models

Martina Erdbrügge
University of Dortmund
Germany

Abstract

Recently, instead of transforming responses for the analysis of designed experiments, i. e. Taguchi-type-experiments like product or combined arrays, generalized linear models as suggested by Nelder and Lee (1991) have been used to jointly model the mean and dispersion of the response. For fitting these models, variance functions and link functions for both the mean and dispersion submodels need to be specified. In this paper, a graphical method based on the quasi-deviance of the joint model is presented which gives visual help in discriminating between transformation models and generalized linear models. Furthermore, the choice of appropriate variance and link functions or the transformation parameter is supported, respectively.

1 Introduction

In many technical applications, off-line process control in terms of design of experiments is used for process optimization. Robust parameter design has been coined by Taguchi (1986), who suggests product array designs, where an outer array (variation of noise factors) is carried out for every combination of design factor levels in an inner array. The aim is to determine factor level combinations which lead to production on target while minimizing response variation. Taguchi (1986) therefore distinguishes among "dispersion factors", i. e. control factors that influence the variance, "location factors", i. e. factors that affect the mean, and those which neither have an effect on the mean nor the variance of the response. Factors influencing only the mean are called “adjustment factors”. These notations also transmit to (fractional) factorial designs with replications instead of noise array runs. Even for unreplicated experiments, methods for the identification of dispersion effects have been proposed by Box and Meyer (1986); Bergman and Hynen (1997) and others.

Transformations - either data driven or based on user knowledge - are commonly used to achieve model simplicity and meet model assumptions. In terms of dispersion and adjustment factors, we seek a metric in which only a few factors are influencing the variance and a larger number of factors can be used to adjust the mean on target after minimizing the variation (assumption of separation).

Recently, also the generalized linear models theory (GLM) has been applied to experiments with dispersion factors, in particular by Nelder and Lee (1991), (1998) and by Engel and Huele (1996). This approach is based on the experience that it might be impossible to find a single transformation which leads to additivity and normality and also removes dependencies between the mean and variance of the response at the same time. Therefore this approach covers location and dispersion factors that are additive in different scales, resulting in two generalized linear models for the mean and dispersion. Therefore these models are also called “double generalized linear models”, compare Smyth and Verbyla (1999). Models using transformations can be viewed as an approximate special case of double generalized linear models.

In Section 2 transformation models are presented. Section 3 covers the theory of double generalized linear models and the estimation algorithm. In Section 4, graphical tools for the model choice are suggested and applied to some simulated and real examples. The paper closes with an outlook and discussion in Section 5.

2 Transformation Models

Transformation models have been examined in detail by Box and Cox (1964), who consider power transformations and present a method for the estimation of the transformation parameter based on the error sums of squares. Assuming an underlying data transformation belonging to the family of Box-Cox-transformations

$$T_\lambda(y_i) = \begin{cases} \dfrac{y_i^\lambda - 1}{\lambda} & : \quad \lambda \neq 0 \\ \ln(y_i) & : \quad \lambda = 0 \end{cases} , \tag{1}$$

where the response y_i, $i = 1, \dots, n$, is assumed to be positive, they use the simple linear model

$$T_\lambda(y_i) = \alpha_0 + \sum_{p \in L} \alpha_p \, x_{ip} + e_i \ , \quad e_i \ \text{ i.i.d. } N(0, \sigma^2) \, ,$$

with $L := \{p \ : \ \text{factor } p \text{ affects the mean of } T_\lambda(\mathbf{y})\}$ the set of location factors.

Especially for data from experiments with factors that might influence the variance of the response, the assumption of homogeneous variances after transformation is too restrictive. Therefore, this model has been extended by many authors to allow for dispersion effects:

$$T_\lambda(y_i) = \alpha_0 + \sum_{p \in L} \alpha_p \, x_{ip} + \left(\prod_{q \in D} \exp\{\gamma_q/2\}^{x_{iq}} \right) e_i \ , \quad e_i \ \text{ i.i.d. } N\big(0, \ \exp\{\gamma_0\}\big) \, ,$$

with $D := \{q \ : \ \text{factor } q \text{ affects the variation of } T_\lambda(\mathbf{y})\}$ the set of dispersion factors, compare e. g. Nair and Pregibon (1988). For the expectation and variance of the transformed response this model implies

$$\begin{aligned} E\left(T_\lambda(y_i)\right) &= \alpha_0 + \sum_{p \in L} \alpha_p \, x_{ip} \, , \\ Var\left(T_\lambda(y_i)\right) &= \exp\{\gamma_0\} \prod_{q \in D} \exp\{\gamma_q/2\}^{2 \cdot x_{iq}} = \exp\left\{ \gamma_0 + \sum_{q \in D} \gamma_q x_{iq} \right\} . \end{aligned}$$

In most applications, there is no straightforward or subject-given choice of the transformation parameter, but this parameter needs to be estimated from the data. By estimating this parameter appropriately, we hope to find a metric in which dispersion factors (possibly few or none) and additional additive adjustment factors may be identified (assumption of separation).

3 Generalized Linear Models

The use of generalized linear models instead of transformation models has been suggested by Nelder and Lee (1991). Not only for counts and proportions data, but also for continuous data, it may not be possible to achieve additivity, normality and to remove dependencies between mean and variance of a response variable at the same time by a single transformation. Often different transformations are needed to achieve these objectives. Considering gamma random variables y_i, $i = 1, \ldots, n$, approximate normality can be achieved by transforming the observations according to $y_i^{1/3}$. On the other hand, transforming the data according to $\log(y_i)$ yields constant variances. In addition, the transformation needed to achieve additivity of design factor effects depends on the true underlying structure and not on the distribution assumption. In the following, generalized linear models are reviewed. These allow other than normal distributions and can handle different kinds of relations between mean and variance of the response as well as mean and dispersion depending on covariates by using extensions of the log-likelihood function.

Double generalized linear models needed for experiments with dispersion factors are composed of two generalized linear models. The model for the mean is based on the observed values while the model for the dispersion is based on residuals of the mean model or on replicates. Nelder and Lee (1991) suggest the following mean and dispersion model.

Mean Model:

$$\begin{aligned} E(y_i) &= \mu_i, \qquad g(\mu_i) = \eta_i = \alpha_0 + \sum_{p \in L} \alpha_p x_{ip} \\ Var(y_i) &= \phi_i V(\mu_i), \end{aligned} \tag{2}$$

where $L := \{p : \text{factor } p \text{ affects the linear predictor } \eta\}$ is the set of location factors and $g(\cdot)$ a monotonic function, the so called link function for the mean. This function specifies the scale for the mean model. The variance function $V(\mu_i)$ is used to model dependencies between mean and variance.

For the dispersion parameter ϕ_i a second generalized linear model is set up to model dependencies on (presumably few of) the design factors. It is not possible to observe ϕ_i directly, therefore a dispersion component d_i will be used complying with the condition $E(d_i) = \phi_i$.

Dispersion Model:

$$
\begin{aligned}
E(d_i) &= \phi_i, \qquad h(\phi_i) = \xi_i = \gamma_0 + \sum_{q \in D} \gamma_q \, x_{iq} \\
Var(d_i) &= \tau_i \, V_D(\phi_i) \,,
\end{aligned} \tag{3}
$$

where $D := \{q \,:\, \text{factor } q \text{ affects the linear predictor } \xi\}$ is the set of dispersion factors (compare also Lee und Nelder (1998)). Common choices for the link $h(\cdot)$ and variance function $V_D(\cdot)$ as well as the dispersion parameter τ_i are given by $\tau_i \equiv 2$, $V_D(\phi_i) = \phi_i^2$ and $h(\phi_i) = \ln(\phi_i)$ which ensures that $\hat{\phi}_i$ will always be positive, resulting in a positive variance for y_i. This corresponds to fitting a gamma generalized linear model with log link. We will stick to this approach.

3.1 Parameter Estimation in (classical) GLMs

To summarize inference and parameter estimation in (classical) GLMs, we use the mean model and assume that the link and variance function are known and the dispersion parameter $\phi_i \equiv \phi$ is constant for all observations, i. e. the model $\eta_i = g(\mu_i) = \alpha_0 + \sum \alpha_p \, x_{ip}$, $Var(y_i) = \phi \, V(\mu_i)$.

For the analysis of GLMs, the response is assumed to be a random variable following an exponential family distribution. Inference is based on the maximum likelihood principle, necessary regularity conditions are assumed to be met.

The log likelihood can be written as

$$
\ell(\boldsymbol{\theta}, \mathbf{y}) = \sum_{i=1}^{n} \left[\frac{y_i \theta_i - b(\theta_i)}{\phi} + c(y_i, \phi) \right] .
$$

From the properties of exponential families it can be shown that $\mu_i = \frac{\partial}{\partial \theta_i} \, b(\theta_i)$ and $V(\mu_i) = \frac{\partial}{\partial \theta_i} \, \mu_i = \frac{\partial^2}{(\partial \theta_i)^2} \, b(\theta_i)$, compare McCullagh and Nelder (1989), p. 28 f.

The maximum likelihood estimates for $\boldsymbol{\alpha} = (\alpha_0, [\alpha_p]_{p \in L})'$ can then be obtained by solving the score-equations

$$
s(\alpha_p) \;=\; \sum_{i=1}^{n} \frac{\partial \ell(\theta_i, y_i)}{\partial \alpha_p} = \sum_{i=1}^{n} \frac{\partial \ell(\theta_i, y_i)}{\partial \theta_i} \, \frac{\partial \theta_i}{\partial \mu_i} \, \frac{\partial \mu_i}{\partial \eta_i} \, \frac{\partial \eta_i}{\partial \alpha_p} = \sum_{i=1}^{n} \frac{y_i - \mu_i}{\phi} \, \frac{1}{V(\mu_i)} \, \frac{\partial \mu_i}{\partial \eta_i} \, x_{ip} = 0
$$

for instance by using the Fisher scoring algorithm (see McCullagh and Nelder (1989), p. 40 ff). The resulting vector of coefficient estimates $\hat{\boldsymbol{\alpha}}$ is consistent and asymptotically normal with

$$\hat{\boldsymbol{\alpha}} \;\overset{a}{\sim}\; N\left(\boldsymbol{\alpha}, E\left(-\frac{\partial s(\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}'}\right)\right) = N\big(\boldsymbol{\alpha}, I(\boldsymbol{\alpha})^{-1}\big)\,,$$

where the approximate variance-covariance matrix is the Fisher information matrix.

Measures for the goodness of fit and residuals for generalized linear models are among others the (scaled) deviance Dev and deviance residuals r_D or the χ^2_P-Pearson-statistic and corresponding Pearson-residuals r_P given by

$$\begin{aligned}
Dev(\mathbf{y}, \hat{\boldsymbol{\mu}})/\phi &= \sum_{i=1}^{n} r_D^2(y_i, \hat{\mu}_i) \\
\text{with} \quad r_D(y_i, \hat{\mu}_i) &= \operatorname{sign}(y_i - \hat{\mu}_i)\sqrt{-2\int_{y_i}^{\hat{\mu}_i} \frac{y_i - t}{V(t)}\, dt}\,, \\
\chi^2_P(\mathbf{y}, \hat{\boldsymbol{\mu}}) &= \sum_{i=1}^{n} r_P^2(y_i, \hat{\mu}_i) \\
\text{with} \quad r_P(y_i, \hat{\mu}_i) &= \frac{y_i - \hat{\mu}_i}{\sqrt{V(\hat{\mu}_i)}}\,.
\end{aligned}$$

Both the deviance and the Pearson-statistic are asymptotically χ^2-distributed, compare Firth (1991). The deviance is commonly used for testing nested hypothesis. For the test of H_A vs. H_B of dimension $A < B$, the difference of deviances

$$Dev(\hat{\boldsymbol{\mu}}_B, \hat{\boldsymbol{\mu}}_A) = Dev(\mathbf{y}, \hat{\boldsymbol{\mu}}_A) - Dev(\mathbf{y}, \hat{\boldsymbol{\mu}}_B) \overset{a}{\sim} \chi^2_{B-A}$$

can be used as a test statistic.

3.2 Parameter Estimation in double GLMs using Extended Quasi Likelihood or Pseudo Likelihood

In GLMs, distributions belonging to a specific exponential family involve specific dispersion parameters ϕ and variance functions $V(\mu)$, i. e. for the normal distribution $\phi = \sigma^2$, $V(\mu) = 1$. The quasi (log) likelihood approach allows for a separation of the mean and variance.

A quasi likelihood can be defined for any desired mean and variance connection

$$\begin{aligned}
E(\mathbf{y}|\mathrm{X}_L) &= \boldsymbol{\mu} = g^{-1}(\mathrm{X}_L\boldsymbol{\alpha}) \\
Var(\mathbf{y}|\boldsymbol{\mu}, \phi) &= \phi V(\boldsymbol{\mu})\,,
\end{aligned}$$

where $V(\mu)$ and $g(\mu)$ are assumed to be known and ϕ allows for over- or underdispersion. The quasi likelihood Q is defined so that the first derivative of Q with respect to μ is the same as for a log-likelihood function:

$$\frac{\partial Q(y_i, \mu_i)}{\partial \mu_i} = \frac{y_i - \mu_i}{\phi V(\mu_i)} .$$

This results in

$$Q(\mathbf{y}, \boldsymbol{\mu}) \;=\; \sum_{i=1}^{n} \left[\int_{y_i}^{\mu_i} \frac{y_i - t}{\phi V(t)}\, dt \,+\, f(y_i) \right] .$$

Most of log likelihood properties transmit to QLs. (Quasi) score equations and (quasi) deviances can be specified analogous to the ones obtained from log-likelihoods. For the (quasi) deviance we get

$$QDev(\mathbf{y}, \boldsymbol{\mu}) = -2\{Q(\mathbf{y}, \boldsymbol{\mu}) - Q(\mathbf{y}, \mathbf{y})\} = -2 \sum_{i=1}^{n} \int_{y_i}^{\mu_i} \frac{y_i - t}{V(t)}\, dt = \sum_{i=1}^{n} r_D^2(y_i, \mu_i) .$$

The idea of quasi likelihood can further be extended for the case of dispersion parameters depending on covariates or for unknown variance functions for the mean. Therefore Nelder and Pregibon (1987) define an extended quasi likelihood (EQL) function Q^+ by

$$\begin{aligned} Q^+(y_i, \mu_i, \phi_i) \;&=\; Q(y_i, \mu_i) + k(y_i, \phi_i) \\ &=\; -\frac{1}{2}\frac{d_i}{\phi_i} - \frac{1}{2} \ln \left(2\pi\phi_i V(y_i)\right) . \end{aligned}$$

It can be derived via saddlepoint-approximation to the log-density and complies the following two conditions:

$$\begin{aligned} E\left(\frac{\partial Q^+}{\partial \mu_i}\right) \;&=\; E\left(\frac{y_i - \mu_i}{V(\mu_i)} \cdot \frac{1}{\phi_i}\right) = 0 , \\ E\left(\frac{\partial Q^+}{\partial \phi_i}\right) \;&=\; E\left(\frac{d_i}{2\phi_i^2} - \frac{1}{2\phi_i}\right) = 0 . \end{aligned}$$

Therefore the EQL can be viewed as a log likelihood function for the dispersion parameter if $\boldsymbol{\mu}$ is known and as a quasi likelihood function for the mean parameter if $\boldsymbol{\phi}$ is known. The EQL for all observations is given by

$$Q^+(\mathbf{y}, \boldsymbol{\mu}, \boldsymbol{\phi}) = -\frac{1}{2} \sum_{i=1}^{n} \ln \left(2\pi\phi_i V(y_i)\right) - \frac{1}{2} \sum_{i=1}^{n} \frac{d_i}{\phi_i} .$$

The dispersion model will be specified according to equation (3) and for the variance function of the mean model we assume a power function $V_\beta(\boldsymbol{\mu}) = \boldsymbol{\mu}^{2\beta}$ with β unknown.

The fitting algorithm maximizes Q^+ resp. minimizes the (extended quasi) deviance $-2\,Q^+$ by alternating between fitting the mean and dispersion model assuming the actual estimates $\hat{\boldsymbol{\mu}}$ or $\hat{\boldsymbol{\phi}}$ as fixed, respectively. This alternating procedure is motivated by the orthogonality $E(\partial^2 Q^+/\partial \boldsymbol{\alpha}\, \partial \boldsymbol{\gamma}) = 0$. For fitting the dispersion model, $d_i = r_D^2(y_i, \mu_i)$ is used as a response variable, although $E(r_D^2(y_i, \mu_i)) \simeq \phi_i$ only approximately.

As an alternative, the Pearson-residuals $d_i = r_P^2(y_i, \mu_i)$ satisfying $E(r_P^2(y_i, \mu_i)) = \phi_i$ may be used instead of the deviance residuals in the dispersion submodel. These correspond to deviance residuals in the case of constant variance function, for instance $V^*(\mu_i) = 1$. Firth (1991) suggests that in this case the dispersion parameter should also contain the variation of the variance due to the dependence on the mean. This idea leads to the normal density because of the variance function, but with dispersion parameter and therefore variance depending on the mean. As a result, we get the so-called Pseudo Likelihood (PL), which is given by

$$P(\mathbf{y}, \mu) = -\frac{1}{2}\sum_{i=1}^{n} \ln\left(2\pi\phi_i V(\mu_i)\right) - \frac{1}{2}\sum_{i=1}^{n} r_P^2(y_i, \mu_i)/\phi_i\,.$$

compare also Engel and Huele (1996) and McCullagh and Nelder (1989), Ch. 10. The PL can therefore be regarded as being the likelihood of a normal distribution with variance function depending on the mean. Lee and Nelder (1998), for instance, point out that the EQL provides better dispersion estimates than the PL, but there is no general agreement on this issue in the literature.

3.3 Transformation Models as an approximate special case

In the case of $g(\mu_i) = T_\lambda(\mu_i)$ and $V(\mu_i) = \mu_i^{2\,(1-\lambda)}$ with $T_\lambda(\cdot)$ as defined in (1), modeling mean and variance by using a GLM yields approximately the same results as analyzing a transformed response instead of y_i and assuming $T_\lambda(y_i) \sim$ i.i.d. $N(\eta_i, \phi_i)$ since the mean and variance for both models are first-order equivalent, compare Nelder and Pregibon (1987). For $T_\lambda(y_i)$, mean and variance can be approximated by

$$\begin{aligned} E\left(T_\lambda(y_i)\right) &\approx T_\lambda(E(y_i)) = g(\mu_i) = \eta_i \\ Var\left(T_\lambda(y_i)\right) &\approx \left[\left.\frac{\partial T_\lambda(y_i)}{\partial y_i}\right|_{y_i=E(y_i)}\right]^2 Var\,(y_i) = \left[\left. y_i^{\lambda-1}\right|_{y_i=E(y_i)}\right]^2 \phi_i V(\mu_i) = \phi_i\,. \end{aligned}$$

This connection between transformation models and GLMs is used by many authors for suggesting methods to choose variance and link functions in generalized linear models, compare i. e. Nelder and Lee (1991).

On the contrary, we apply methods for fitting mean and dispersion models in GLMs to transformation models for estimating the transformation parameter. In particular in unreplicated experiments, there are no methods yet available that allow the consistent estimation of λ when dispersion effects are present.

In the following, we deal with the approximate special case of a transformation model as well as with GLMs that allow for independent link and variance functions for the mean model.

4 Determination of transformation parameters

For the simple transformation model, the transformation parameter λ may be determined by using the SSE-Plot proposed by Box and Cox (1964) or the β-Method mentioned by Logothetis (1990). For the extended transformation model, the so called λ-Plot has been introduced by Box (1988) and a variant of the β-Method for this situation has been suggested by Kunert and Lehmkuhl (1998).

For the more general case of generalized linear models, our main concern will be the specification of the link and variance function for the mean model. In order to simplify notation, these parameters will also be called transformation parameters motivated by the connection between transformation models and GLMs. Nelder and Lee (1991) suggest the use of a mean-variance-plot respectively the β-Method to specify the variance function for the mean. For an appropriate link function, usually a power transformation is used and identified by the method proposed by Box and Cox (1964) or the λ-Plot by Box (1988). Nelder and Pregibon (1987) use a contour plot of the extended quasi likelihood to check the chosen transformation parameters in retrospect.

We prefer to allow for possible dispersion parameters in advance and therefore propose a graphical method for the estimation of transformation parameters. It is based on the extended quasi likelihood respectively the pseudo likelihood which can be used for experiments with and without replication and approximately reduces to the SSE-Plot by

Box and Cox for the transformation model and equal variances. We use EQL-plots (Nelder and Pregibon, 1987) for different dispersion models, but these are considered in advance rather than in retrospect since the estimation of transformation parameters may heavily depend on the choice of the dispersion model, compare Carroll and Ruppert (1988). It is therefore in general not advisable to use the model with equal variances or fixed $\phi_i \equiv \phi$ at the stage of estimating the transformation parameters and then to fit the mean and variance model afterwards. If important dispersion factors are not considered while choosing the link and variance function, in particular the parameter of the variance function may be heavily biased.

The plot of SSE and the contour plot of the extended quasi deviance both do not allow the consideration of dispersion effects, therefore generalizations seem reasonable. For the following exploration, we will assume that we know which of the factors affect the mean and that there are active dispersion effects (commonly one). The assumption of a known set of location effects is quite restrictive, but necessary to avoid the consideration of aliasing problems between location and dispersion effects. In practical applications, however, the set of location factors has to be estimated.

4.1 Extension of SSE-Plot

The extended transformation model as introduced before, assumes normality after transformation. Therefore the model of the transformed response corresponds to a GLM with variance function $V(\mu) = 1$ and dispersion parameter $\phi_i = \exp\{\gamma_0 + \sum \gamma_q x_{iq}\} = \sigma_i^2$. We can fit this model by using an iteratively weighted least squares procedure. Here we alternate between a weighted least squares fit for the mean and a gamma generalized linear model for the dispersion. The weights for the mean model are given by $1/\sigma_i^2$ (compare Aitken, 1987; Engel and Huele, 1996). For normal models and constant variance function, pseudo and extended quasi deviance are identical and given by

$$\begin{aligned} \Rightarrow \quad D(\lambda, \mathbf{y}, \hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\phi}}) &= \sum_{i=1}^{n} \ln\left(2\pi\hat{\phi}_i\right) + \sum_{i=1}^{n} \frac{\left(T_\lambda(y_i) - \hat{\mu}_i\right)^2}{\hat{\phi}_i} \\ &= \sum_{i=1}^{n} \ln\left(2\pi\hat{\phi}_i\right) + \sum_{i=1}^{n} \frac{r^2(y_i, \hat{\mu}_i)}{\hat{\phi}_i}, \end{aligned}$$

where $r^2(y_i,\hat{\mu}_i) = r_D^2(y_i,\hat{\mu}_i) = r_P^2(y_i,\hat{\mu}_i)$ for this model. We can see that for equal variances this coincides with the quantity of interest for the original SSE-plot suggested by Box and Cox (1964) by

$$D(\lambda, \mathbf{y}, \hat{\boldsymbol{\mu}}, \widehat{\sigma^2}) = n \ln\left(2\pi\widehat{\sigma^2}\right) + n \qquad \text{for } \sigma_i^2 \equiv \sigma^2 .$$

Alternatively, the mean and variance of the untransformed response may be modeled directly in a GLM. In this case, the parameters of the link function λ and the variance function β are connected according to $\lambda = 1-\beta$. In general, PL and EQL differ (unless $V(\mu)=1$). The PL is given by

$$\begin{aligned} -2\,P(\lambda, \mathbf{y}, \hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\phi}}) &= \sum_{i=1}^{n} \ln\left(2\pi\hat{\phi}_i\hat{\mu}_i^{2(1-\lambda)}\right) + \sum_{i=1}^{n} \frac{\left(y_i - T_\lambda^{-1}\left(\mathbf{x}_i^\mathsf{T}\hat{\boldsymbol{\alpha}}\right)\right)^2}{\hat{\phi}_i\,\hat{\mu}_i^{2(1-\lambda)}} \\ &= \sum_{i=1}^{n} \ln\left(2\pi\hat{\phi}_i\hat{\mu}_i^{2(1-\lambda)}\right) + \sum_{i=1}^{n} \frac{r_P^2(y_i,\hat{\mu}_i)}{\hat{\phi}_i} . \end{aligned}$$

The EQL is given by

$$-2\,Q^+(\lambda, \mathbf{y}, \hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\phi}}) = \sum_{i=1}^{n} \ln\left(2\pi\hat{\phi}_i\hat{y}_i^{2(1-\lambda)}\right) + \sum_{i=1}^{n} \frac{d_i}{\hat{\phi}_i} ,$$

where d_i can be either $d_i = r_D^2(y_i,\hat{\mu}_i)$ or $d_i = r_P^2(y_i,\hat{\mu}_i)$. Minimizing the Pseudo-Deviance or the Extended-Quasi-Deviance yields in most cases very similar results to minimizing the Deviance for the transformation model as can be seen from the Figures accompanying the following examples.

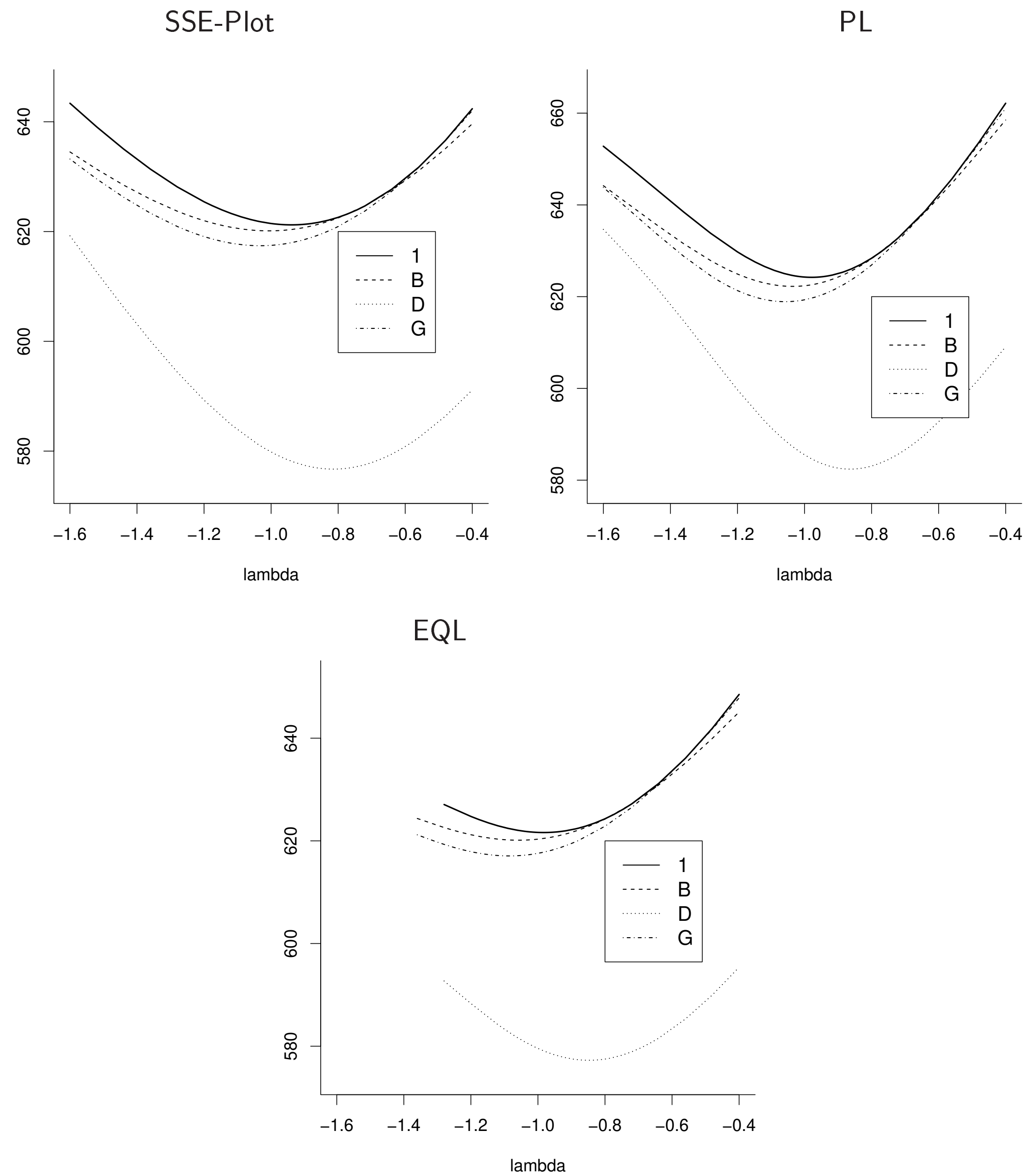
When using double GLMs, we suggest the consideration of different (simple) dispersion models in advance in the stage of estimating the (unconnected) parameters β and λ of the variance and link function, respectively. This can be achieved by applying the extended quasi deviance plot as used by Nelder and Pregibon (1987) prior to the analysis. A comparison of these estimates for different dispersion model assumptions allows the choice of an appropriate model.

In the following, some examples will be given to visualize the effect of neglected dispersion factors on the transformation parameter estimation and how such dispersion effects can be identified from the suggested plots.

Examples from simulations

Our first data set (**Example 1**) has been simulated according to the transformation model with transformation parameter $\lambda = -1$ and has been reported by Box (1988). The set of location factors is given by $L = \{\mathtt{B},\ \mathtt{D},\ \mathtt{G}\}$, factor D also affects the dispersion, i. e. $D = \{\mathtt{D}\}$.

Figure 1: Extended SSE-Plot and PL and EQL-Plots for Example 1



In Figure 1, the extended SSE-Plots as well as the respective plots of the EQL (with $d_i = r_D^2$) and PL when assuming $\lambda = 1 - \beta$ are given. Obviously, a notedly reduction of the deviance will be achieved when assuming a dispersion effect of factor D. All three figures yield very similar results in terms of transformation parameter estimation.

Our second example (**Example 2**) has been simulated according to an underlying double GLM with the sets of location and dispersion effects given by $L = \{\texttt{A, B, C}\}$ and $D = \{\texttt{A}\}$. The mean and variance structure for this example is determined by the two parameters $\lambda = 0$ and $\beta = 0$.

Figure 2: Extended SSE-Plot and PL and EQL-Plots for Example 2

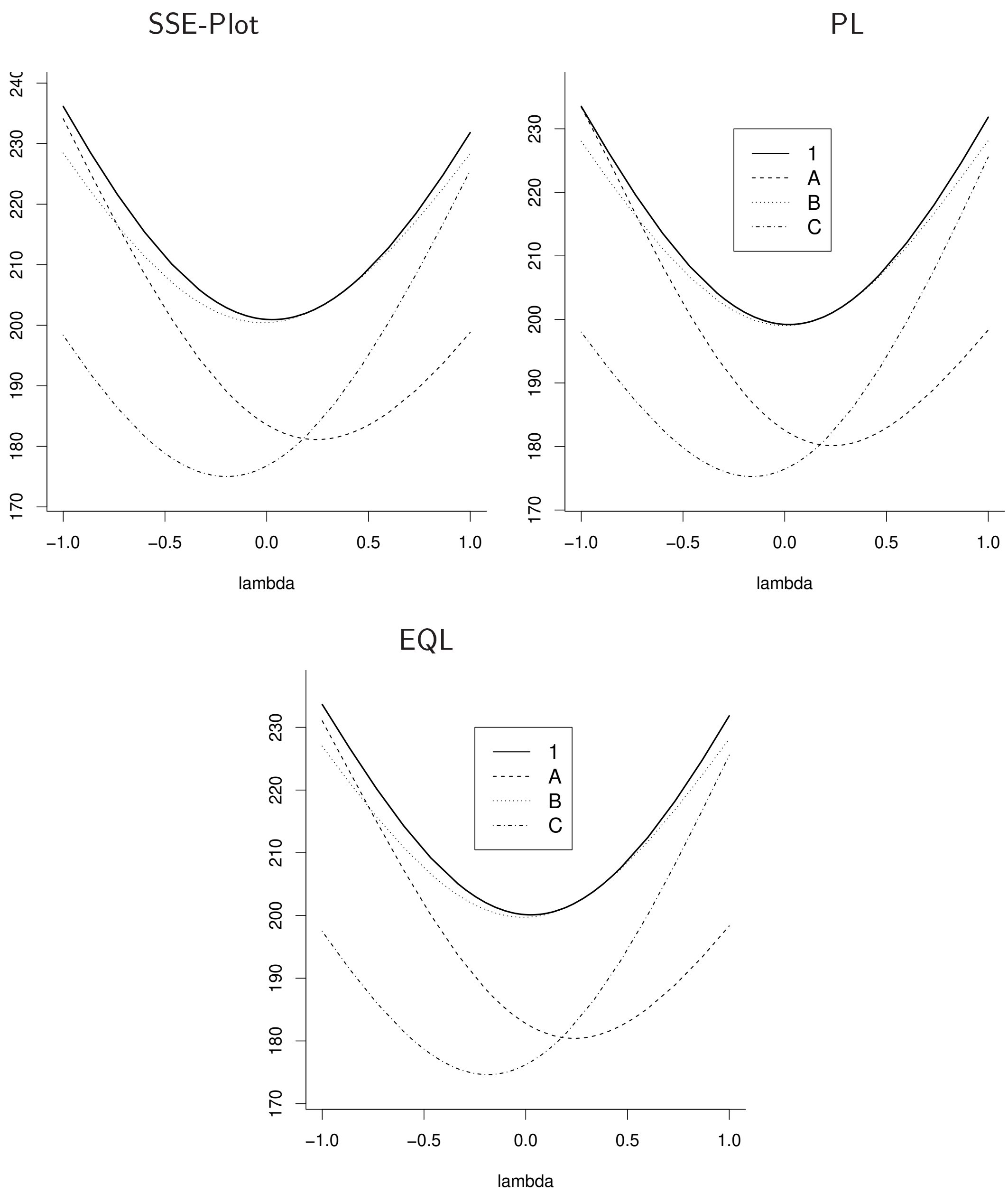
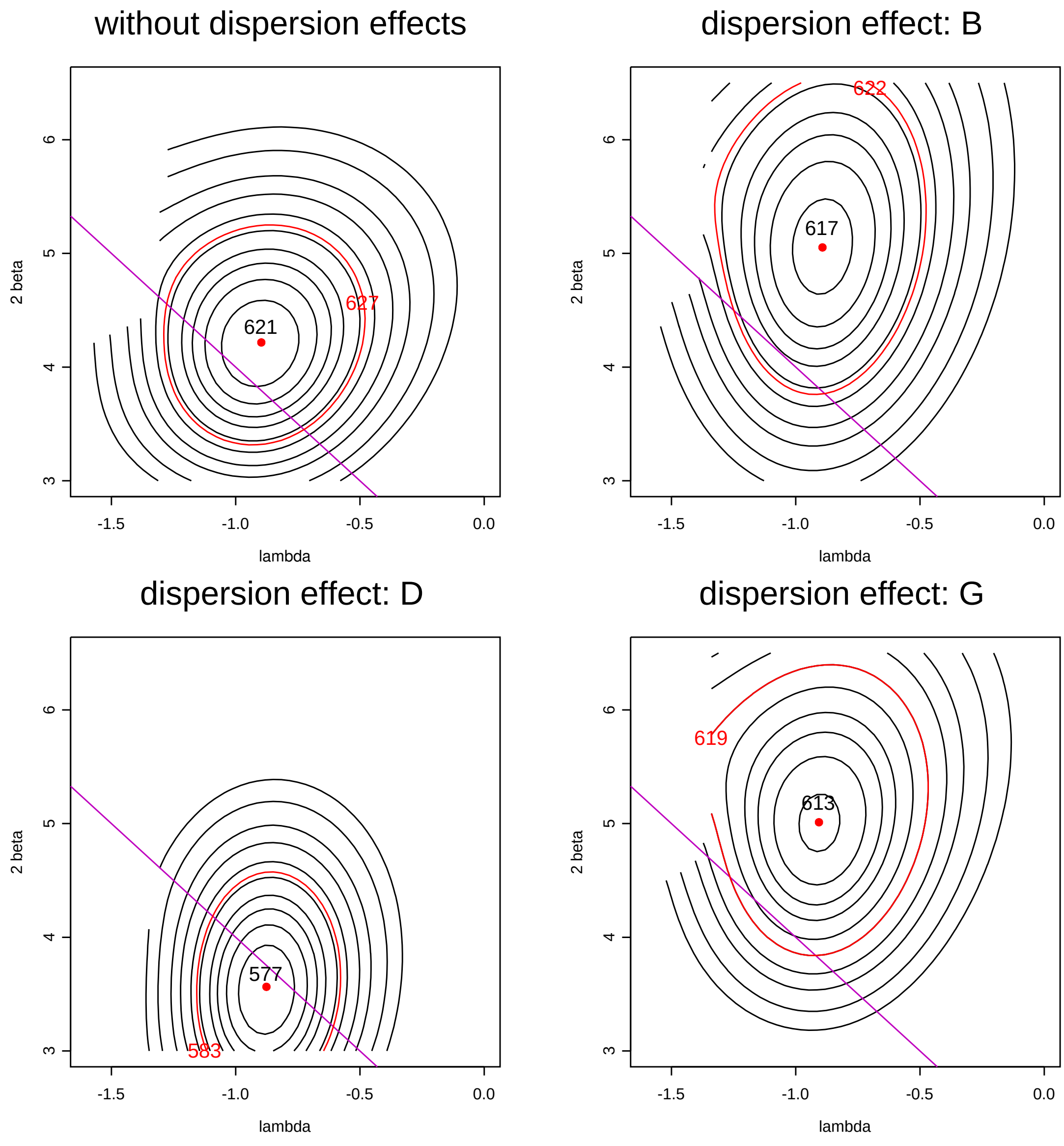


Figure 2 contains the extended SSE-Plots and the PL and EQL-Plot for this data set (assuming $\lambda = 1 - \beta$). Again, all three applied methods lead similar results, but for this example both `A` and `C` might be considered as being dispersion effects.

Since we also want to consider double GLMs in our analysis, extended quasi deviance

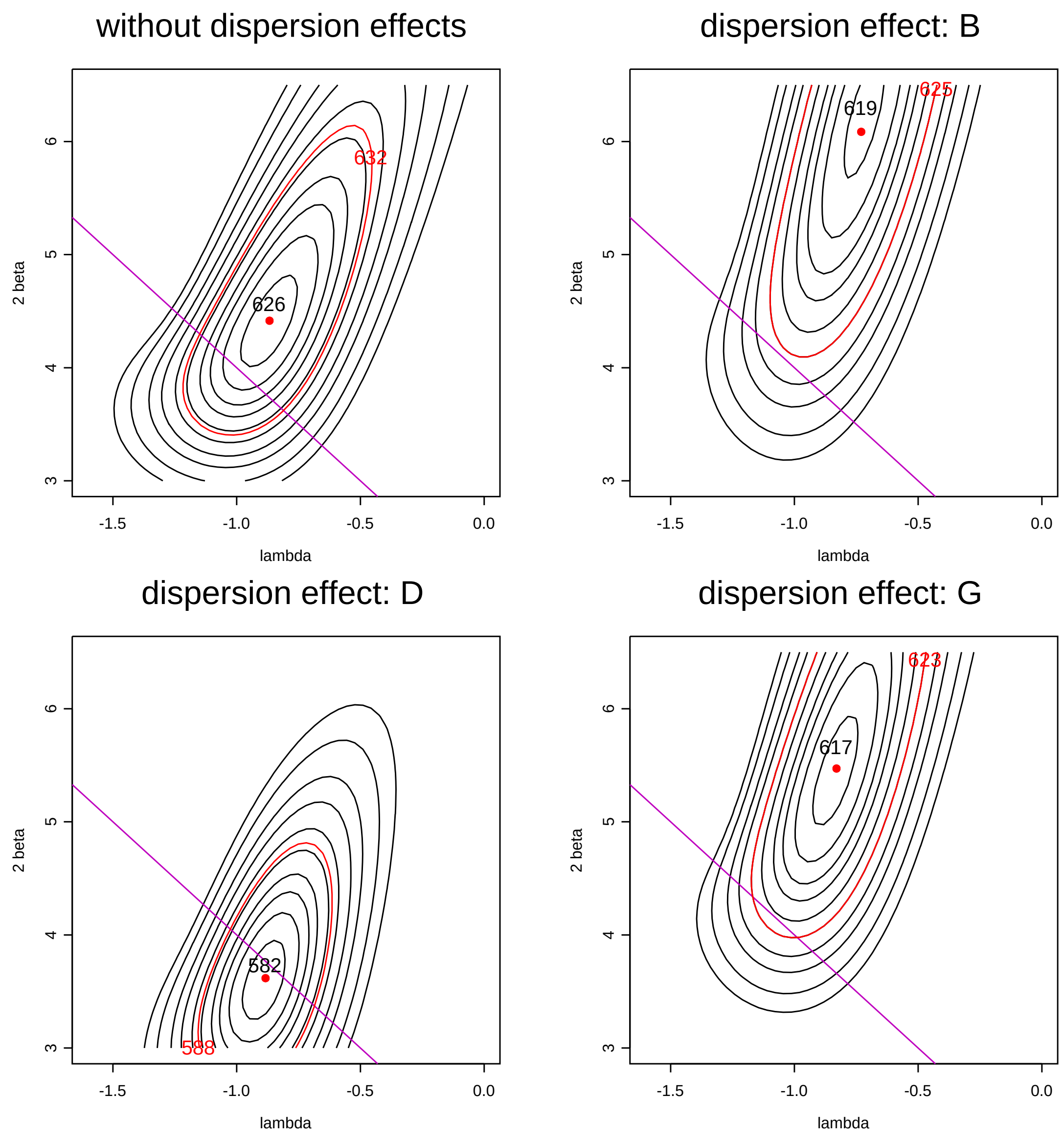
$(-2\,Q^+)$ and pseudo deviance $(-2\,P)$ plots for independent values of λ and β are also created. Contour plots for Example 1 are given in Figures 3 and 4. The line intersecting the plots corresponds to $\lambda = 1 - \beta$, i. e. transformation models, the brighter contour line reflects a 95 % confidence ellipse for the pair of transformation parameters.

Figure 3: Plot of $-2\,Q^+$ for different dispersion model assumptions, Example 1



From both figures, the estimates of λ and β minimizing the Deviances are nearest to the true values, if the factor `D` is assumed as a dispersion effect. The line corresponding to transformation models intersects the confidence ellipses for all considered dispersion models, but in particular when assuming `B` or `G` as dispersion effects, the estimates differ

Figure 4: Plot of $-2\,P$ for different dispersion model assumptions, Example 1



from the true parameters.

The shape of the contour lines differ for the two deviances based on the EQL or the PL. When applying the EQL (Figure 3), the parameters λ and β seem nearly independent, while for the PL (Figure 4), a positive correlation between these parameters is observed.

For our second example, the data are simulated according to a PL from a GLM with $\lambda = \beta = 0$. There is almost no difference between the plots using the EQL and the PL, compare Figures 5 and 6.

From both figures we can see that the true underlying parameter combination is not covered by the confidence ellipses, if a wrong dispersion model is assumed. In particular if equal variances are assumed, i. e. no dispersion effects, a transformation model with $\lambda = 0$ will be suggested. Only the consideration of `A` as a dispersion factor yields estimates near by the true transformation parameters.

Figure 5: Plot of $-2\,Q^{+}$ for different dispersion model assumptions, Example 2

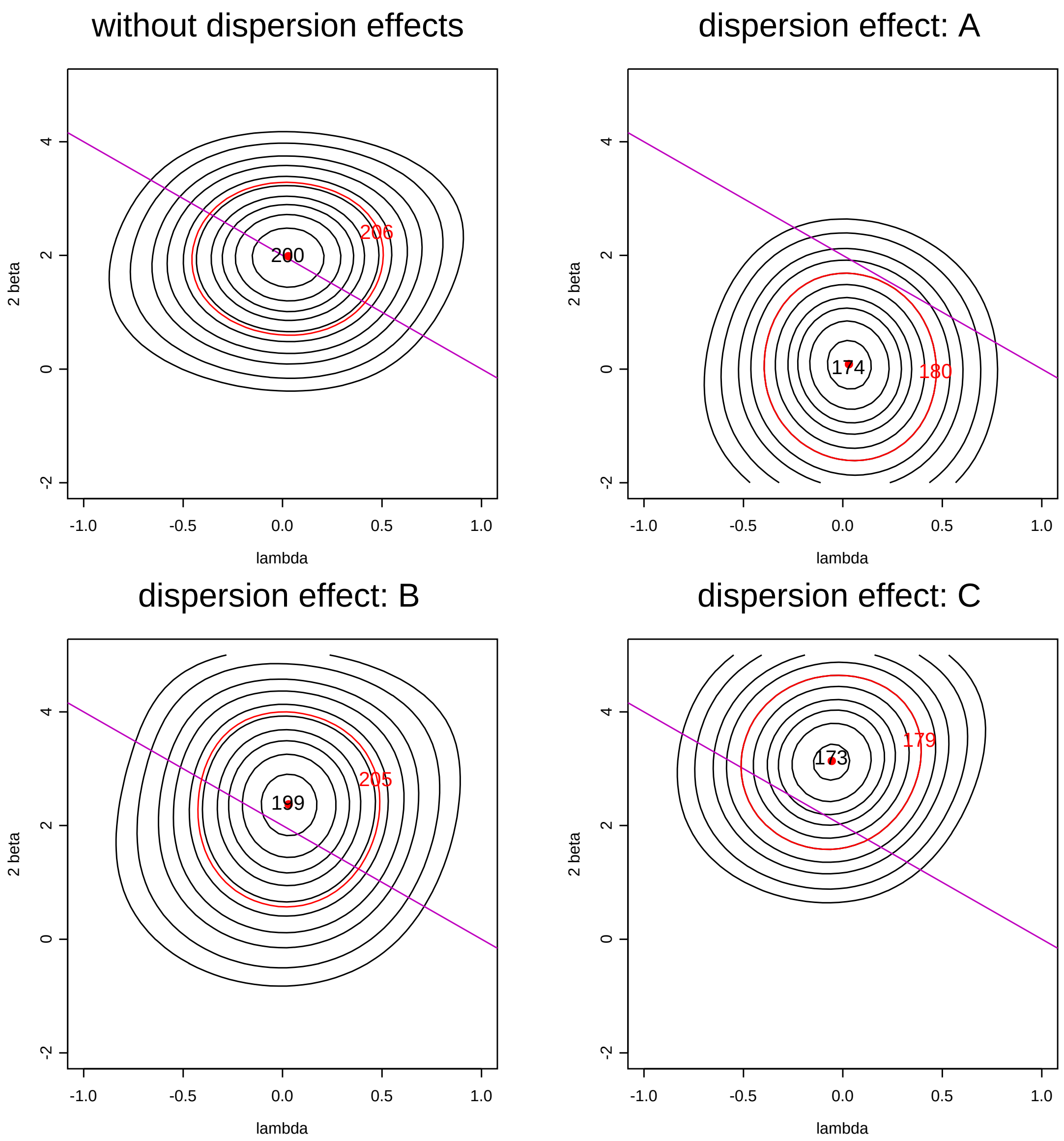
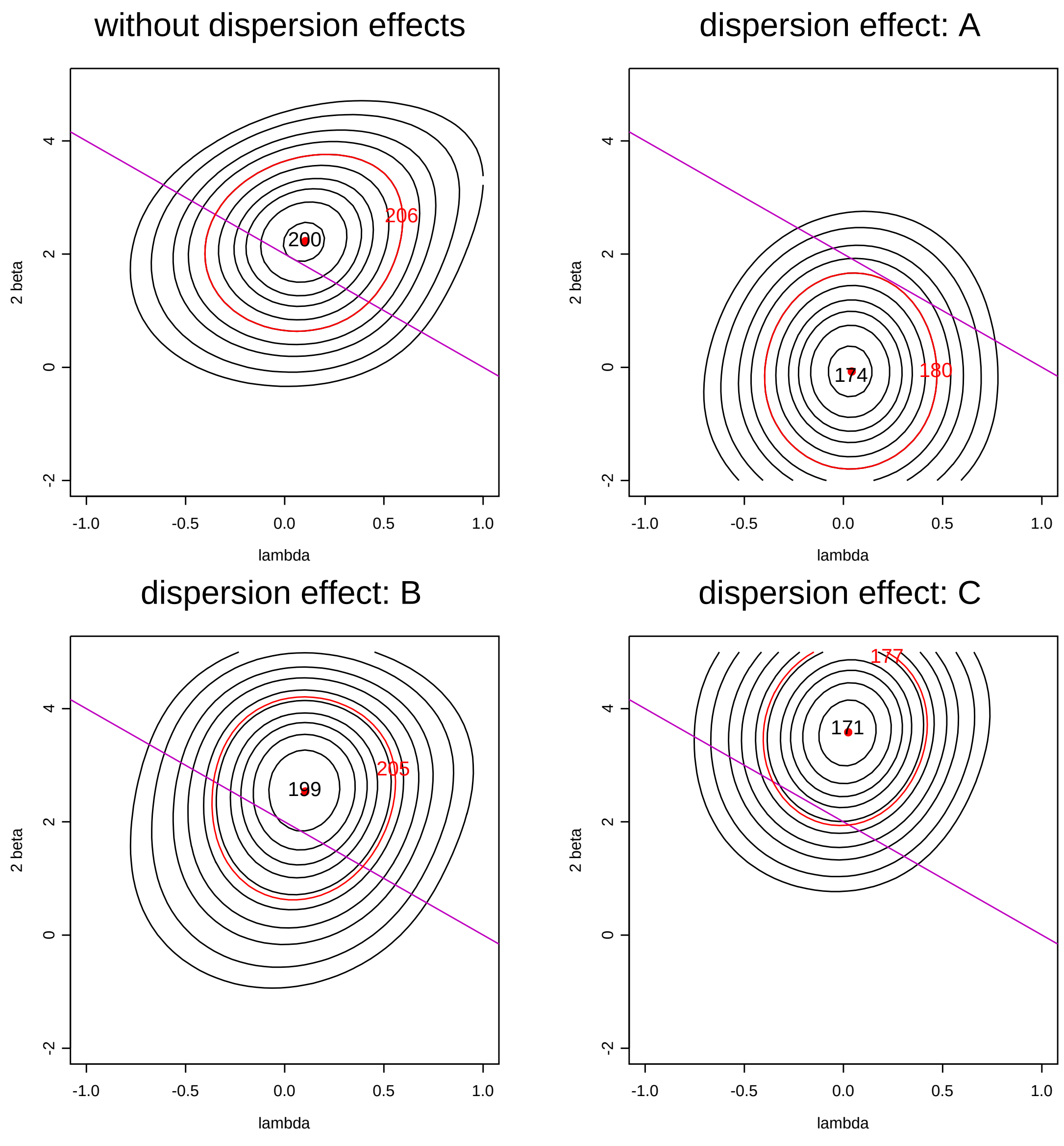


Figure 6: Plot of $-2\,P$ for different dispersion model assumptions, Example 2



The last example considered, **Example 3**, is the textile data set reported by Box and Cox (1964). We assume the following set of location factors: $L = \{x_1,\ x_2,\ x_3,\ x_1\,x_2,\ x_1\,x_3,\ x_2\,x_3\}$, namely the set of all main effects and two factor interactions of the three factors.

Figure 7: Extended SSE-Plot and PL and EQL-Plots for the textile data (Box, Cox, 1964)

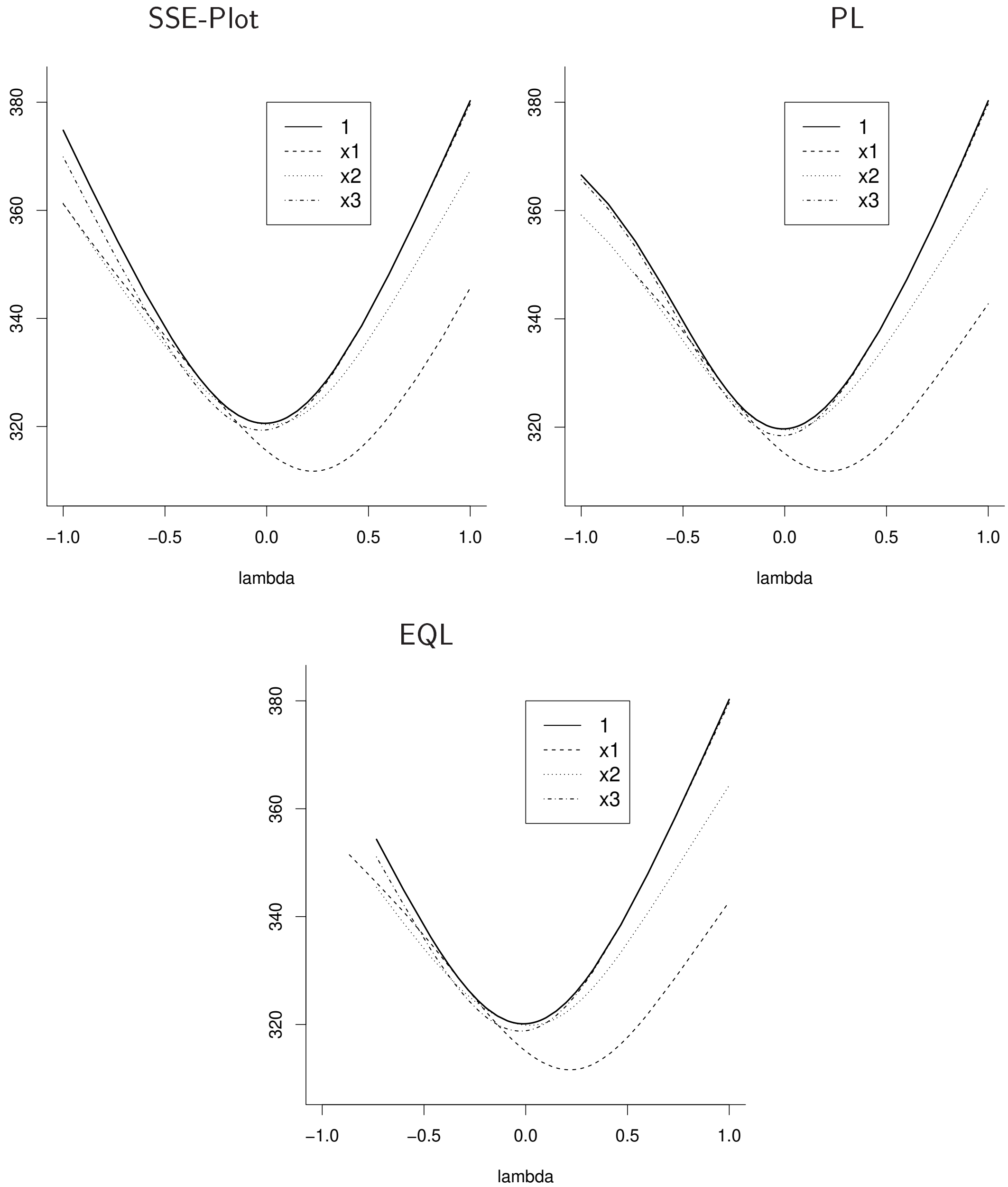
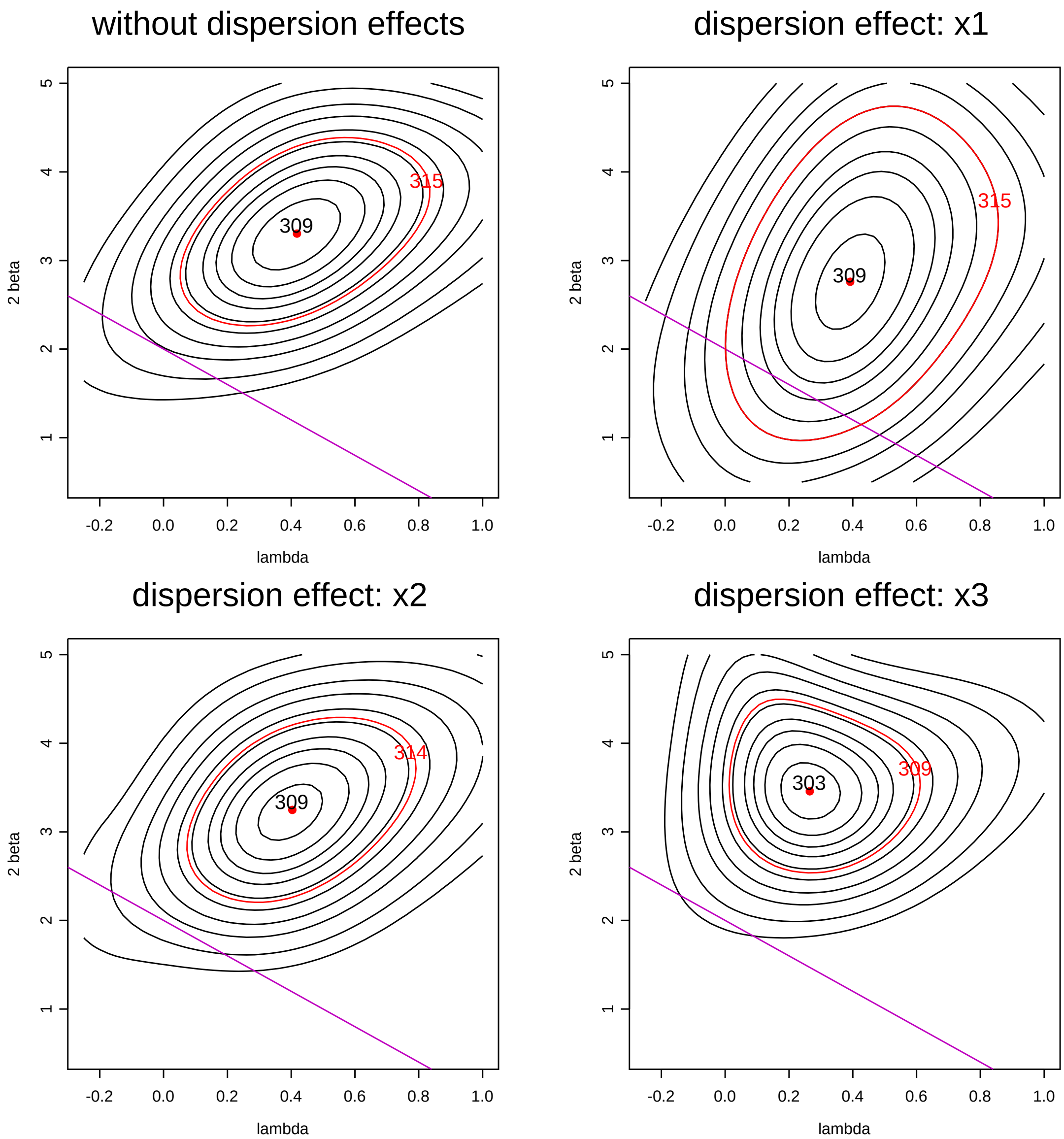


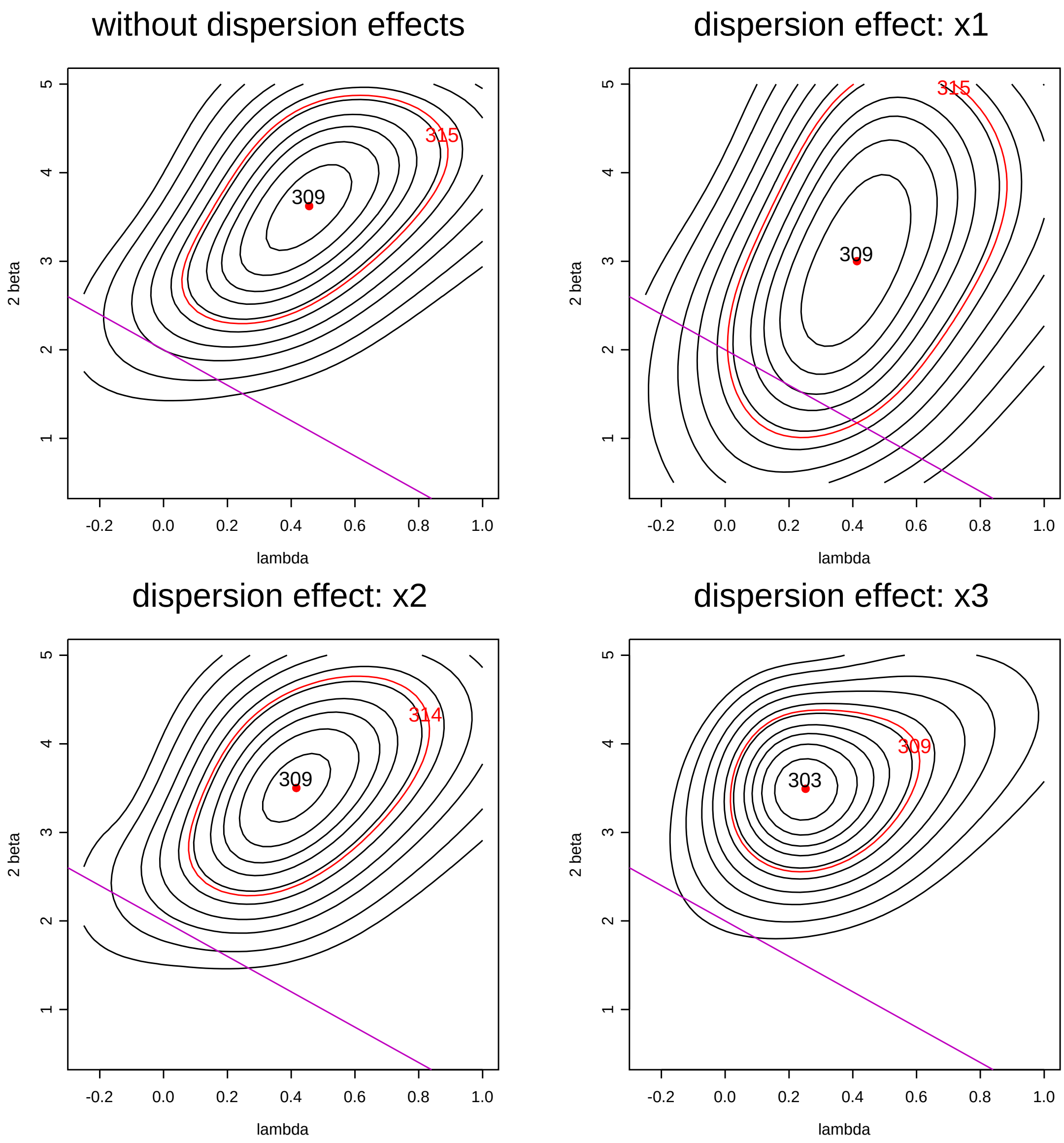
Figure 7 displays the comparison of deviances for different dispersion models and different model assumptions, i. e. Transformation Models and GLM for the special case $\lambda = 1 - \beta$. Only the main effects are considered as dispersion factors. The model with dispersion factor x_1 clearly yields the smallest deviance. Based only on these plots, we would conclude a dispersion effect of factor x_1 and an underlying transformation parameter of $\lambda = 1/4$.

Figure 8: Plot of $-2\,Q^+$ for different dispersion model assumptions, textile data (Box, Cox)



However, if we compare the plots of Figure 7 with the extended quasi deviances and pseudo deviances for the generalized linear model, i. e. independent transformation parameters λ and β as displayed in Figures 8 and 9, we can see that this clear structure for the transformation model is due to the different shape of contours for the model with dispersion factor x_1 compared to the remaining three considered models. This model is the only one supporting a transformation model to some extend, for all other

Figure 9: Plot of $-2\,P$ for different dispersion model assumptions, textile data (Box, Cox)



dispersion models, the line representing these transformation models doesn't intersect the confidence region. All figures suggest transformation parameters about $2\beta \approx 3.3$ and $\lambda \approx 0.4$. Assuming these parameter values, factor x_3 will be the only one with a significant dispersion effect (10 %-level).

From all three examples we get the impression that the consideration of dispersion effects in the stage of model selection is very important since the estimation of transformation parameters can be biased due to wrong assumptions. This may then result in wrong location and dispersion factor models.

5 Outlook and Discussion

Motivated by the three examples resulting from simulated as well as real data sets, a closer examination of the extended quasi likelihood function or the pseudo likelihood function and its behaviour for different dispersion model assumptions seems to be of interest. Standard likelihood theory suggests that under certain regularity conditions, consistent estimates for the transformation parameters are achieved when assuming the true underlying dispersion structure. The objective of further investigations will be quantify the bias of transformation parameter estimates if wrong dispersion models are assumed.

So far, only few simulations have been carried out. These suggest that for generalized linear models the estimation of λ is hardly affected, but the estimation of β is biased. The bias direction seems to be related to the product of signs of location and dispersion effect of dispersion factors.

This is not very surprising, since in the case of active dispersion effects, the estimation of $V(\mu)$ and $g(\mu)$ also depends on ϕ_i. For fixed value λ, it can in fact be shown that estimating β via β-method minimizes the extended quasi deviance. However, since λ and β are not orthogonal, i. e. $E\left(\frac{\partial^2 Q^+}{\partial\lambda\partial\beta}\right) \neq 0$ and $E\left(\frac{\partial^2 P}{\partial\lambda\partial\beta}\right) \neq 0$, iterative estimation of both parameters does not necessarily end up at the global minimum.

Acknowledgement

Financial support of the "Deutsche Forschungsgemeinschaft" is gratefully acknowledged. The author is also thankful to G. Smyth for providing the S-plus function `dglm()` which has been used for the extended quasi deviance and pseudo deviance plots in this work.

References

Bergman, B. and Hynén, A. (1997). Dispersion Effects from Unreplicated Designs in the 2^{k-p} Series. *Technometrics 39*, 191–198.

Box, G. E. P. (1988). Signal-to-noise Ratios, Performance Criteria, and Transformations. *Technometrics 30*, 1–17.

Box, G. E. P. and Cox, D. R. (1964). An Analysis of Transformations. *Journal of the Royal Statistical Society 26*, 211–252.

Box, G. E. P. and Meyer, R. D. (1986). Dispersion Effects from Fractional Design. *Technometrics 28*, 19–27.

Carroll, R. J. and Ruppert, D. (1988). Discussion of "Signal-to-noise Ratios, Performance Criteria, and Transformations". *Technometrics 30*, 30–31.

Engel, J. and Huele, A. F. (1996). A Generalized Linear Modeling Approach to Robust Design. *Technometrics 38*, 365–373.

Firth, D. (1991). Generalized linear models. In *Hinkley, Reid, Snell (Eds.): Statistical Theory and Modelling, Chapter 11*, pp. 55–82.

Kunert, J. and Lehmkuhl, F. (1998). *MODA 5 - Advances in Model-Oriented Data Analysis and Experimental Design; The Generalized β-Method in Taguchi Experiments.*, pp. 223–230. Atkinson, Pronzato, Wynn (Eds.), Physica-Verlag.

Lee, Y. and Nelder, J. A. (1998). Generalized Linear Models for the Analysis of Quality-Improvement Experiments. *The Canadian Journal of Statistics 26*, 95–105.

Logothetis, N. (1990). Box–Cox Transformations and the Taguchi Method. *Applied Statistics 39*, 31–48.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models.* Chapman & Hall.

Nair, V. N. and Pregibon, D. (1988). Discussion of "Signal-to-noise Ratios, Performance Criteria, and Transformations". *Technometrics 30*, 24–29.

Nelder, J. A. and Lee, Y. (1991). Generalized Linear Models for the Analysis of Taguchi-type experiments. *Applied Stochastic Models and Data Analysis 7*, 107–120.

Nelder, J. A. and Pregibon, D. (1987). An extended quasi-likelihood function. *Biometrika 74*, 221–232.

Smyth, G. K. and Verbyla, A. P. (1999). Adjusted Likelihood methods for Modelling Dispersion in Generalized Linear Models. *Environmetrics 10*, 695 – 709.

Taguchi, G. (1986). *Introduction to Quality Engineering: Designing Quality Into Products and Processes.* White Plains, NY: Kraus International Publications.