A GENERALIZATION OF THE GAUSS-NEWTON METHOD THAT SOLVES EXTENDED LEAST SQUARES PROBLEMS[1]

BRADLEY M. BELL[2] AND ALAN SCHMITZKY[3]

Abstract. Modeling the mean of a random variable as a function of unknown parameters leads to a nonlinear least-squares objective function. The Gauss-Newton method reduces nonlinear least-squares problems to a sequence of linear least-squares problems and requires only first-order information about the model functions. In a more general heteroscedastic setting, there are also unknown parameters in a model for the variance. This leads to an objective function that is no longer a sum of squares. We present an extension of the Gauss-Newton method that minimizes this objective function by reducing the problem to a sequence of linear least-squares problems and requires only first-order information. This represents a new result because other methods that reduce this problem to a sequence of linear least-squares problems do not necessarily converge.

Key words. Iteratively reweighted least squares, extended least squares, asymptotic statistics

AMS subject classifications. 65K05, 62F12

1. Introduction

1.1. Background. Heteroscedastic models often have a variance component that changes in a systematic way. One approach to such models is to maximize a quasi-likelihood of a transformation of the data [12]. Another approach is to model the variance as a function of the mean of the data; for example, see [9]. It is often important to fit the mean and the variance simultaneously [7]. We present a new algorithm that does this together with a proof of its convergence. We also present the asymptotic statistics for the corresponding estimator. A formal statement of our model is contained in the assumption below.

Assumption 1.1. We are given an open set $X \subseteq \mathbf{R}^n$, a compact convex set $C \subset X$, an open set $T \subseteq \mathbf{R}$, and an open set $U \subseteq \mathbf{R}^m$. For $j = 1, \ldots, N$ we are also given independent samples y_j and the twice continuously differentiable functions $S_j : X \to T$ and $V_j : T \times U \to \mathbf{R}$, such that for unknown values $\overline{x} \in C$ and $\overline{u} \in U$, $S_j(\overline{x})$ is the mean of y_j and $V_j(S_j(\overline{x}), \overline{u})$ is the variance of y_j. Define $L_N : X \times U \to R$ by

$$L_N(x, u) = \frac{1}{2} \sum_{j=1}^{N} \log \left[V_j(S_j(x), u) \right] + \frac{(y_j - S_j(x))^2}{V_j(S_j(x), u)} \ .$$

We assume there is a twice continuously differentiable function $\widehat{u}_N : X \to U$ such that $V_j(S_j(x), \widehat{u}_N(x)) > 0$ and $\widehat{u}_N(x)$ is a minimizer of $L_N(x, u)$ with respect to u.

Remark 1.1. If the sequence of random variables $\{y_j\}$ is normally distributed, $L_N(x, u)$ is the negative log-likelihood minus $(N / 2) \log(2\pi)$. The function $L_N(x, u)$ is very similar to the objective in [5, Eq. 3.1]. In this paper we present an algorithm that estimates $\overline{x}$ and $\overline{u}$ by minimizing $L_N(x, u)$. This estimation procedure is consistent and asymptotically normal.

Our algorithm uses the derivatives of V_j and S_j to reduce the problem of minimizing $L_N(x, u)$ to a sequence of linear least-squares subproblems. The iteratively reweighted least-squares method presented in [6, p. 30] or [11, p. 399] estimates $\overline{x}$ and $\overline{u}$ by a similar reduction but there is no proof of its convergence (see Counter Example 2.3). The nonlinear extension of the Newton-Raphson method presented in [10] also reduces the problem to a sequence of linear least-squares problems (provided the approximating function is convex).

In addition, our algorithm assumes that given a value of x, it is easy to calculate the minimizer with respect to the variance parameters, i.e., $\widehat{u}_N(x)$. Minimizing the variance parameters is a form of Bender's decomposition and has been considered for a similar problem in [13]. We prove convergence of our algorithm by showing it

is a special case of the algorithm presented in [2]. This proof should extend to the method presented in [10] provided that a line search is included and the approximating function is convex.

1.4 Notation.

y_j	The j-th data value
$S_j(x)$	Model for the mean value of y_j
X	Domain for the function S_j
C	The constraint set for the parameters in the model for the mean
$V_j(s, u)$	Model for the variance of y_j
$T \times U$	Domain for the function V_j
$\overline{x}$	True but unknown value for the mean parameters
$\overline{u}$	True but unknown value for the variance parameters
$L_N(x, u)$	Negative loglikelihood corresponding to N data points
$\widehat{u}_N(x)$	Value that minimizes $L_N(x, u)$ with respect to u
$G(x)$	The minimum of $L_N(x, u)$ with respect to u, i.e., $L_N(x, \widehat{u}_N(x))$
$A(x, h)$	An approximation for $G(x + h)$ near $h = 0$
e_j	The residual in the j-th data value, i.e., $y_j - S_j(\overline{x})$
σ_j	The standard deviation of e_j, i.e., the square root of $V_j(S_j(\overline{x}), \overline{u})$
$\mathbf{R}^p$	The real vector space with p components
z^T	The transpose z
$E[z]$	The expected value of z
$\|f(z)\|_Z$	The maximum of $|f(z)|$ with respect to $z \in Z$
$f^{(k)}(z)$	The k-th derivative of $f(z)$
$\partial_h A(x, h)$	partial derivative of $A(x, h)$ with respect to h

2. Examples

2.0. A Biological example. In the example in Section 2.1 of Davidian and Giltian the concentration of indomethacin in plasma was measured at eleven time points following an injection of the drug. The data for the 5th subject is included below and comes from Table 2.1 of Davidian and Glitian [**??**].

t_j (hours)	y_j (concentration)	t_j (hours)	y_j (concentration)
0.25	2.05	3.00	0.13
0.50	1.04	4.00	0.11
0.75	0.81	5.00	0.08
1.00	0.39	6.00	0.10
1.25	0.30	8.00	0.06
2.00	0.23		

The model for the mean of the data

$$S_j(x) = \exp(x_1)\exp(-\exp(x_2)t_j) + \exp(x_3)\exp(-\exp(x_4)t_j)$$

Ordinary least squares (OLS) models the variance as $V_j(S_j(x), u) = u$ where u is a scalar. The corresponding estimates and standard errors of the estimates (SE) are

	estimate	SE
x_1	+1.27	.082
x_2	+1.04	.147
x_3	−1.23	.491
x_4	−1.51	.642

The standard errors are the square root of the diagonal element of the covariance matrix which can be approximated using the formula for D in Theorem 4.1 with $\bar{x}$ and $\bar{u}$ replaced by their estimated values. These values agree with those in Table 2.2 of [**??**]. As mentioned in [**??**], it appears that the variance model should be $V_j(S_j(x), u) = u^2 S_j(x)^2$. The corresponding estimates and standard errors for the extended least squares method (ELS) are

	estimate	SE
x_1	+1.19	.234
x_2	+0.94	.159
x_3	−1.44	.210
x_4	−1.76	.229

The iteratively reweighted least squares estimates (called GLS in Table 2.2 of [**??**]) are

	estimate	SE
x_2	+1.21	.24
x_3	+0.95	.16
x_3	−1.45	.21
x_4	−1.74	.24

Note that the result for ELS and GLS are very similar. Example 2.2 and Remark 3.1 continue the comparison between these two methods.

2.1. Multiple data sets. Suppose that the odd indices correspond to one data set and the even indices to another data set, and $V_j(s, u)$ is equal to u_1 if j is odd and u_2 if j is even. In this case, when N is even,

$$
\begin{aligned}
L_N(x, u) &= \frac{N}{4} \log(u_1) + \frac{1}{2u_1} \sum_{j=1}^{N/2} (y_{2j-1} - S_{2j-1}(x))^2 \\
&\quad + \frac{N}{4} \log(u_2) + \frac{1}{2u_2} \sum_{j=1}^{N/2} (y_{2j} - S_{2j}(x))^2 .
\end{aligned}
$$

This is the objective function for the two-data-set case of the multiple-data-set problem considered in [2]. Note that given x, we can determine the corresponding value of $\widehat{u}_N(x)$ that minimizes $L_N(x, u)$. If $\widehat{u}_{1,N}(x)$ and $\widehat{u}_{2,N}(x)$ denote the first and second components of $\widehat{u}_N(x)$,

$$
\widehat{u}_{1,N}(x) = \frac{2}{N} \sum_{j=1}^{N/2} (y_{2j-1} - S_{2j-1}(x))^2
$$

$$
\widehat{u}_{2,N}(x) = \frac{2}{N} \sum_{j=1}^{N/2} (y_{2j} - S_{2j}(x))^2 .
$$

The assumption that $V_j(S_j(x), \widehat{u}_N(x)) > 0$ is equivalent to neither $\widehat{u}_{1,N}(x)$ nor $\widehat{u}_{2,N}(x)$ being zero; i.e., no value of x fits all the data in one of the measurement sets perfectly.

2.2. Model-dependent variances. Suppose that $V_j(s, u)$ is equal to $u\,s$. In this case,

$$
L_N(x, u) = \frac{1}{2} \sum_{j=1}^{N} \log(uS_j(x)) + \frac{(y_j - S_j(x))^2}{uS_j(x)} .
$$

Note that given x, we can determine the corresponding value $\widehat{u}_N(x)$, which minimizes $L_N(x, u)$:

$$
\widehat{u}_N(x) = \frac{1}{N} \sum_{j=1}^{N} \frac{(y_j - S_j(x))^2}{S_j(x)} .
$$

The condition that $V_j(S_j(x), \widehat{u}_N(x)) > 0$ is true if neither $\widehat{u}_N(x)$ nor $S_j(x)$ is zero; i.e., no value of x fits all the data perfectly and $S_j(x) > 0$ for $j = 1, \ldots, N$ and for all $x \in X$.

Counter Example 2.3. The iteratively reweighted least-squares method for estimating $\overline{x}$ and $\overline{u}$ is described in [6, p. 30] and [11, p. 399]. In this section we present a simple example where this method does not converge. Convergence of our method is

guaranteed by Theorem 3.2. In addition, the actual iterates for our method are listed at the end of this section. For this example $N = 2$, and for $j = 1, 2$ we have

j	y_j	$S_j(x)$	$V_j(s, u)$
1	2	x	s
2	0	x	$2 - s$.

The iteratively reweighted least-squares method for solving this problem estimates x by fixing the variance, then modifies the variance to correspond to the new value of x, and iterates until convergence. This method does not converge for this example. If x^k is the current iterate, the next iterate minimizes

$$\frac{(y_1 - S_1(x))^2}{V_1(S_1(x^k), u)} + \frac{(y_2 - S_2(x))^2}{V_2(S_2(x^k), u)} = \frac{(2 - x)^2}{x^k} + \frac{(0 - x)^2}{(2 - x^k)} .$$

Taking the derivative of the right-hand side with respect to x and setting it equal to zero, we obtain the formula for the next iterate x^{k+1}:

$$0 = 2\frac{(2 - x^{k+1})}{x^k}(-1) + 2\frac{x^{k+1}}{(2 - x^k)}$$

$$\frac{2}{x^k} = x^{k+1}\left[\frac{1}{x^k} + \frac{1}{(2 - x^k)}\right]$$

$$x^{k+1} = \frac{2}{x^k}\frac{x^k(2 - x^k)}{x^k + 2 - x^k} = 2 - x^k .$$

If x^0 is any value in the interval $(0, 2)$ other than 1, this method oscillates and never gets any closer to the solution. Thus both x^0 and $2 - x^0$ are limit points of the sequence $\{x^k\}$ generated by the iteratively reweighted least-squares algorithm even though neither satisfies the first-order condition for a minimum.

If we use the constraint set $C = [0.1, 1.9]$ and the parameters $\gamma = 0.5$, $\mu = 0.01$ in Step 0 with the approximation $A(x, h)$ from Theorem 3.2, the following iterates are generated by the algorithm in Theorem 3.1.

k	x^k	k	x^k
0	0.500	5	1.082
1	1.200	6	0.938
2	0.833	7	1.047
3	1.135	8	0.964
4	0.894	9	1.027

3. Algorithm. In this section, we review the algorithm presented in [2], §4. For a set $C \subset \mathbf{R}^n$ we use the notation $N_C(x)$ for the normal cone [8, p. 15], which is defined by

$$N_C(x) = \{z : z^{\mathrm{T}}(x' - x) \leq 0 \text{ for all } x' \in C\} .$$

The following theorem is direct consequence of [2, Theorem 3] which is more general because it allows for $G(x)$ to be minus infinity. The condition $0 \in \nabla G(\widehat{x}) + N_C(\widehat{x})$ is the first-order necessary condition for a minimum of $G(x)$ subject to $x \in N_C(\widehat{x})$ [4, p. 52]. In the special case where $\widehat{x}$ is an interior point of C, this means that the derivative of G is zero.

THEOREM 3.1. *Suppose that $G : C \to \mathbf{R}$, and G is twice continuously differentiable on an open set X containing the nonempty compact convex set $C \subset \mathbf{R}^n$. Further suppose that, $A : C \times \mathbf{R}^n \to \mathbf{R}$ is such that $A(x, h)$ is convex with respect to h, the partial of A with respect to h is continuous on $X \times \mathbf{R}^n$, and for each $x \in X$*

$$G^{(1)}(x) = \partial_h A(x, 0) \ .$$

If $\widehat{x}$ is a limit point of the sequence $\{x^k\}$ generated by the algorithm below, then $0 \in \nabla G(\widehat{x}) + N_C(\widehat{x})$.

> *Step 0:* Choose $x^0 \in \mathbf{R}^n$, $\gamma \in (0, 1)$, and $\mu \in (0, 1)$. Set $k = 0$.
> *Step 1:* Set h^k equal to a minimizer of $A(x^k, h)$
> with respect to h and subject to $x^k + h \in C$.
> *Step 2:* $\lambda^k = 1$.
> While $G(x^k + \lambda^k h^k) - G(x^k) > \mu \lambda^k \left[A(x^k, h^k) - A(x^k, 0) \right]$
> do $\lambda^k = \gamma \lambda^k$.
> *Step 3:* $x^{k+1} = x^k + \lambda^k h^k$.
> *Step 4:* Set $k = k + 1$ and go to Step 1.

THEOREM 3.2. *Suppose that Assumption 1.1 holds, define $G(x) = L_N(x, \widehat{u}_N(x))$ and*

$$A(x, h) = \frac{1}{2} \sum_{j=1}^{N} \frac{1}{V_j(S_j(x), \widehat{u}_N(x))}$$

$$\left\{ y_j - S_j(x) - S_j^{(1)}(x) h - 0.5 \left[1 - \frac{(y_j - S_j(x))^2}{V_j(S_j(x), \widehat{u}_N(x))} \right] \partial_s V_j(S_j(x), \widehat{u}_N(x)) \right\}^2 \ .$$

If we apply the algorithm in Theorem 3.1 for this $G(x)$ and $A(x, h)$, any limit point $\widehat{x}$ of the sequence generated by the algorithm satisfies $0 \in \nabla G(\widehat{x}) + N_C(\widehat{x})$.

Remark: The calculation of h^k in Step 1 of the algorithm requires the minimization of $A(x, h)$ with respect to h which, for this definition of $A(x, h)$, is a linear least square problem. Thus Theorem 3.2 enables us to minimizes the function $L_N(x, u)$ (defined in Assumption 1.1) by solving a sequence of linear least squares problems.

8

Proof. We need to demonstrate the assumptions of Theorem 3.1. The function $G(x)$ is twice continuously differentiable because L_N and $\widehat{u}_N$ are. The function $A(x,h)$ is convex in h because it is the sum of the terms, each of which is the square of a linear function of h. The partial of $A(x,h)$ with respect to h is given by

$$
\partial_h A(x,h) = -\sum_{j=1}^{N} \frac{S_j^{(1)}(x)}{V_j(S_j(x), \widehat{u}_N(x))}
$$
$$
\left\{ y_j - S_j(x) - S_j^{(1)}(x)h - 0.5\left[1 - \frac{(y_j - S_j(x))^2}{V_j(S_j(x), \widehat{u}_N(x))}\right] \partial_s V_j(S_j(x), \widehat{u}_N(x)) \right\} .
$$

The continuity of this partial follows directly from assumptions of this theorem. We complete the proof by demonstrating that

$$
G^{(1)}(x) = \partial_h A(x, 0) .
$$

We begin by evaluating the partial of $L_N(x,u)$ with respect to x, which is equal to

$$
\frac{1}{2}\sum_{j=1}^{N}\left[\frac{1}{V_j(S_j(x), u)} - \frac{(y_j - S_j(x))^2}{V_j(S_j(x), u)^2}\right]\partial_s V_j(S_j(x), u)S_j^{(1)}(x) - \frac{2(y_j - S_j(x))}{V_j(S_j(x), u)}S_j^{(1)}(x) .
$$

Grouping common terms, we obtain

$$
\partial_x L_N(x, u) = \sum_{j=1}^{N}\left\{0.5\left[1 - \frac{(y_j - S_j(x))^2}{V_j(S_j(x), u)}\right]\partial_s V_j(S_j(x), u) - y_j + S_j(x)\right\}\frac{S_j^{(1)}(x)}{V_j(S_j(x), u)} .
$$

Using the formula for the partial of $A(x,h)$ above, we obtain

$$
\begin{aligned}
G^{(1)}(x) &= \partial_x L_N(x, \widehat{u}_N(x)) + \partial_u L_N(x, \widehat{u}_N(x))\widehat{u}_N^{(1)}(x) \\
&= \partial_x L_N(x, \widehat{u}_N(x)) = \partial_h A(x, 0) ,
\end{aligned}
$$

where we have used the fact that the partial of $L_N(x,u)$ with respect to u is zero when $u = \widehat{u}_N(x)$ because $\widehat{u}_N(x)$ minimizes $L_N(x,u)$ with respect to $u \in U$ and U is open. This completes the proof. $\square$

Remark 3.1. There are two aspects to the algorithm in Theorem 3.2 that are new for this problem. One is the line search. The other is the inclusion of the term

$$
5\left[1 - \frac{(y_j - S_j(x))^2}{V_j(S_j(x), \widehat{u}_N(x))}\right]\partial_s V_j(S_j(x), \widehat{u}_N(x))
$$

in the objective function for the approximate subproblem. It is interesting to note that the expected value of

$$
\left[1 - \frac{(y_j - S_j(\overline{x}))^2}{V_j(S_j(\overline{x}), \overline{u})}\right]
$$

9

is zero. This is probably why the iteratively reweighted least-squares method in [6, p. 30] often converges.

4. Statistics. In the theorems below, the notation $\|S_j(x)\|_X$ denotes the maximum of $|S_j(x)|$ with respect to $x \in X$. This notation is extended to functions other than $S_j(x)$ and to sets other than X. The notation $E[L_N(x,u)]$ denotes the expected value of $L_N(x,u)$ with respect to the probability space that the sequence $\{y_j\}$ is sampled from. This notation is extended to functions other than $L_N(x,u)$. The notation e_j denotes $y_j - S_j(\overline{x})$ and σ_j denotes the standard deviation of e_j; i.e., $\sigma_j^2 = V_j(S_j(\overline{x}), \overline{u})$. The following theorem is a direct consequence of [2, Theorem 3].

THEOREM 4.1. *Suppose Assumption 1.1 holds with $\overline{x}$ in the interior of C and in addition we have*

1. *The sequence $\{y_j\}$ are independent normally distributed random vectors defined a the complete probability space and there is a constant M such that*

$$E\left[|y_j|^6\right] \leq M$$

2. *The functions $\{S_j(x)\}$, $\{V_j(s,x)\}$ are three times continuously differentiable on $X \times U$ such that $E[y_j] = S_j(\overline{x})$ and $Var[y_j] = V_j(S_j(\overline{x}), \overline{u})$. In addition there is a constant M such that for $i = 0, 1, 2$*

$$\left\|S_j^{(i)}(x)\right\|_X \leq M$$
$$\left\|V_j^{(i)}(S_j(x), u)\right\|_{X \times U} \leq M$$
$$\left\|V_j(S_j(x), u)^{-1}\right\|_{X \times U} \leq M$$

3. *There is are functions $L(x,u)$ and $H(x,u)$ defined on $X \times U$ such that $H(\overline{x}, \overline{u})$ is positive definite and*

$$\left\|L(x,u) - \frac{1}{N}E\left[L_N(x,u)\right]\right\|_{X \times U} \to 0 \text{ as } N \to \infty$$

$$\left\|H(x,u) - \frac{1}{N}E\left[L_N^{(2)}(x,u)\right]\right\|_{X \times U} \to 0 \text{ as } N \to \infty$$

4. *There is a positive definite matrix D such that*

$$\frac{1}{N}E\left[\left(L_N^{(1)}(\overline{x}, \overline{u})\right)^T \left(L_N^{(1)}(\overline{x}, \overline{u})\right)\right] \to D \text{ as } N \to \infty$$

It follows that $\sqrt{N}\begin{pmatrix} x_N - \overline{x} \\ u_N - \overline{u} \end{pmatrix}$ *converges in distribution to a normal random vector with mean zero and covariance* D^{-1}*where*

$$H_N(\overline{x}, \overline{u}) \rightarrow D \text{ as } N \rightarrow \infty$$

$$H_N(x, u) = \sum_{j=1}^{N} \frac{1}{2\sigma_j^4} \begin{pmatrix} D_{xx}(x, u) & D_{xu}(x, u) \\ D_{xu}^T(x, u) & D_{uu}(x, u) \end{pmatrix}$$

$$D_{xx}(x, u) = 2\left[\sigma_j^2 + \partial_s V_j(S_j(x), u)^2\right] S_j^{(1)}(x)^T S_j^{(1)}(x)$$

$$D_{xu}(x, u) = \partial_s V_j(S_j(x), u) S_j^{(1)}(x)^T \partial_u V_j(S_j(x), u)$$

$$D_{uu}(x, u) = \partial_u V_j(S_j(x), u)^T \partial_u V_j(S_j(x), u)$$

Proof. Note that the data vector y_j is scalar valued and there is compact set that contains $C \times \widehat{u}(C)$ such that $(\overline{x}, \overline{u})$ is in its interior. If one makes the following identifications

$$\theta = (x, u)$$

$$\theta_N = (x_N, u_N)$$

$$F_j(\theta) = S_j(x)$$

$$W_j(\theta) = V_j(S_j(x), u)$$

then using [2, Theorem 3 and 4] we can conclude that $\sqrt{N}(\theta_N - \overline{\theta})$ converges in distribution to a normal random vector with mean zero and covariance $D^{-1}$where

$$\sum_{j=1}^{N} \frac{F_j^{(1)}(\overline{\theta})^T F_j^{(1)}(\overline{\theta})}{W_j(\overline{\theta})} + \frac{1}{2} \frac{W_j^{(1)}(\overline{\theta})^T W_j^{(1)}(\overline{\theta})}{W_j(\overline{\theta})^2} \rightarrow D \text{ as } N \rightarrow \infty$$

Subsisting the definitions $W_j(\theta) = V_j(S_j(x), u)$ and $\sigma_j = \sqrt{V_j(S_j(\overline{x}), \overline{u})}$ together with the relations

$$F_j^{(1)}(\theta) = \left(S_j^{(1)}(x), 0\right)$$

$$W_j^{(1)}(\theta) = \left(\partial_s V_j(S_j(x), u) S_j^{(1)}(x), \partial_u V_j(S_j(x), u)\right)$$

we obtain the conclusion of this theorem.□

5. Conclusion. We have presented a new algorithm that estimates both variance and model parameters when scalar measurement values are independent. This raises the question of how one should estimate the parameters when the measurements are vector valued and there is covariance between the components of each measurement. Our method reduces the original problem to a sequence of linear least-squares problems. Another question is how one should choose the norm for the least-squares

problems so that convergence of the algorithm is accelerated. The extended nonlinear Newton-Rapson methods use a Hessian that is related to the likelihood function for their quadratic term.

References

[1]

[2] B. M. BELL, J. V. BURKE, AND A. SCHUMITZKY, *A relative weighting method for estimating parameters and variances in multiple data sets*, Comput. Statist. Data Anal., 22 (1996), pp. 119–135.

[3] B. M. BELL AND A. SCHUMITZKY, *Asymptotic statistics for a parameteric estimator that simultaneously fits mean and variance parameters*, unpublished manuscript (1997).

[4] F. H.CLARKE, *Optimization and Nonsmooth Analysis*, Princeton University Press, Princeton, NJ, 1970.

[5] M. DAVIDSON AND R. J. CARROLL, *Variance function estimation*, J. Amer. Statist. Ass., 82 (1987), pp. 1079–1091.

[6] M. DAVIDIAN AND D. M. GILTINAN, *Nonlinear Models for Repeated Measurement Data*, Chapman and Hall, New York, 1995.

[7] M. DAVIDIAN AND P.D. HAALAND, *Regression and calibration with nonconstant error variance*, Chemometircs and Intelligent Laboratory Systems, 9 (1990), pp. 231–248.

[8] R. T. ROCKAFELLAR, *Convex Analysis*, Princeton University Press, Princeton, NJ, 1970.

[9] W. A. SADLER AND M. H. SMITH, *A computer program for estimating imprecision characteristics of immunoassays*, Comput. Biomed. Res., 23 (1990), pp. 105–114.

[10] W. D. Stirling, *Fitting generalized linear models and their nonlinear extensions with least squares calculations*, Comput. Statist. Data Anal., 6 (1988), pp. 7–19.

[11] E. F. Vonesh and V. M. Chinchilli, *Linear and Nonlinear Models for the Analysis of Repeated Measurements*, Marcel Dekker, New York, 1997.

[12] R. A. J. Wixley, *The use of asymmetric power transformation in data-based quasi-likelihood analysis*, Comput. Statist. Data Anal., 7 (1988), pp. 1–21.

[13] R. Wolke and H. Schwetlick, *Iteratively reweighted least squares: Algorithms, convergence analysis, and numerical comparisons*, SIAM J. Sci. Statist. Comput., 9 (1988), pp. 907–921.