# VARIANCE FUNCTION ESTIMATION IN HETEROSCEDASTIC REGRESSION MODELS

by

Marie Davidian

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Statistics

Chapel Hill

1986

Approved by

Raymond J. Carroll
Advisor

David Ruppert
Reader

Gordon Simons
Reader

MARIE DAVIDIAN. Variance Function Estimation in Heteroscedastic Regression Models. (Under the direction of RAYMOND J. CARROLL.)

Heteroscedastic regression models are used to analyze data in a variety of fields, including economics, engineering and the biological and physical sciences. Often, the heteroscedasticity is modeled as a function of the regression and other structural parameters. Standard asymptotic theory implies that under reasonable conditions, how one estimates the variance function, in particular the structural parameters, has no effect on the first order properties of the estimates of the regression parameters; however, it has been noted in practice that how one estimates the variance function does matter. Furthermore, in some settings, estimation of the variance function is of independent interest or plays an important role in the properties of estimates of other quantities besides the regression parameters.

We develop a general theory for variance function estimation in regression which includes most methods in common use. In particular, we focus on estimation of the structural parameters. In our development, we note that most variance function estimation procedures can be looked upon as regressions with "responses" being transformations of absolute residuals from a preliminary fit or sample standard deviations from replicates at a design point. Our theory allows us to conclude that, typically, using residuals is more efficient than using sample standard deviations, but not uniformly so. For variance function estimates based on transformations of absolute residuals, efficiency is a monotone function of the efficiency of the preliminary fit when the errors are symmetric, so that one should

iterate so that the residuals are based on generalized least squares. Robustness issues are of even more importance for variance function estimation than for estimation of a regression function.

To illustrate the implications of these results and to show how variance function estimates play a role in the estimation of other important quantities, we focus on the analysis of assay data, which are often fit by a nonlinear heteroscedastic regression model. An additional component of assay analysis is the estimation of auxillary constructs such as the minimum detectable concentration, for which many definitions exist. We consider one such definition and show how the properties of the standard estimate for minimum detectable concentration are dependent to first order on how one estimates the structural variance parameters. Simulation results and an example support the asymptotic theory.

## ACKNOWLEDGEMENTS

While I could probably go on for several pages to thank individually each of the kind souls without whom this dissertation would have never been completed, I would like especially to acknowledge those people whose help and understanding were of the highest order of magnitude.

My most heartfelt appreciation goes to my advisor, Ray Carroll. His vast expertise, availability for help and guidance, boundless enthusiasm (sometimes to excess) and, most importantly, his patience were instrumental in my completion of my degree.

I cannot do justice with mere words to the depth of my gratitude to my roommate, compatriot, partner in misery (and sometimes crime) and friend, Stena Kettl. This makes two degrees we've been through together, not to mention all the associated pain, disgust and self-doubt; as fun as it's been, though, I don't think I'm up for it again. See, Stena, I didn't even mention the Virginian (or Papagayo's, or Molly Maguire's, or our living room floor, or ...).

The understanding and support of my friends has been far above and beyond the call of duty; their phone calls, visits and patient tolerance of my sometimes hysterical ravings have been a lifeline for me throughout my stay here. True "best" friends are rare, and I consider myself blessed that I have two -- Joyce Maxwell and Ellen Bierschenk. I cannot even begin to express how important their friendship and love have been these past four years.

If it had not been for the support and love of Jed Frees, I think I'd be aiming for a tenure-track position at a different kind of state institution than the kind that awards degrees. His encouragement and understanding were nothing short of legendary.

Along the way, I had the opportunity to meet people whose friendship I will continue to cherish long after I've left Chapel Hill. The mere existence of June Maxwell, Courtney Stark, Doug Vass and Jane Wille made the day-to-day travail a bit more bearable, maybe even enjoyable. The importance of their patient endurance of my ongoing tirade on the concept of doom cannot be emphasized enough.

The appreciation and gratitude I have for the love, encouragement and tolerance of my family during the past four years is simply inexpressible. To my mother and brother, thank you for believing in me, and, more importantly, putting up with me. To my wonderful father, who did not get the opportunity to actually see me complete my degree, even though you didn't say it often, thank you for your love and encouragement; knowing you were up there rooting for me left me no option but to finish my degree. I dedicate this dissertation to you.

This research was supported by the Air Force Office of Scientific Research AFOSR-F-49620-85-C-0144.

## TABLE OF CONTENTS

# CHAPTER I

# INTRODUCTION

## 1.0 Introduction and overview

Heteroscedastic regression models are accepted as appropriate in a wide variety of fields, including radioimmunoassay (Rodbard and Frazier (1975), Finney (1976), Raab (1981)), econometrics (Hildreth and Houck (1968), Amemiya (1977)), pharmokinetic modeling (Bates, Wolf and Watts (1985)), enzyme kinetics (Haaland, et al. (1986)) and chemical kinetics (Pritchard, Downie and Bacon (1977)) among others. In such settings, the mean response is modeled as a possibly nonlinear function of known explanatory variables and unknown regression parameters. The heteroscedasticity may be regarded as of unknown form or may be modeled as a function of the explanatory variables, known constants exogenous to the model and the regression parameters. This function may be completely known, specified up to additional unknown parameters or completely unknown. The usual goal is to obtain estimates of the regression parameters in order to investigate the character of the mean response.

Due to efficiency considerations, estimation methods for the regression parameters incorporate a method for estimation of the variances; the variance estimates also allow for better understanding of the variability in the data. As will be discussed shortly, it has become increasingly apparent that, despite the implications of standard asymptotic theory, the better one's estimates of the variances, the better one's estimates of the regression parameters will be. Furthermore, in some applications, estimation of the variances is of independent interest or plays an important theoretical role in estimation of quantities other than the regression parameters. Thus, while much effort has been focused on the study of properties of estimators for the regression parameters, it is of some interest to investigate the properties of estimators for variance as well.

We shall be interested, then, in investigating and comparing the properties of different methods for estimating variances in heteroscedastic regression models. In particular, we focus on settings for which variance is modeled as a known function of explanatory and exogenous variables, the regression parameters and additional unknown structural parameters; our analysis reduces to determining the properties of various estimators for these additional variance function parameters. As will become clear shortly, many such estimators, while motivated by different considerations, have similar formulations; we exploit this fact to pursue a simple and concise unified theory for variance function estimators from which specific estimators arise as special cases. This theory makes general observations and comparisons particularly straightforward.

## 1.1 Model and basic assumptions

Consider a general possibly nonlinear heteroscedastic regression model for observable data Y given by

$$(1.1) \qquad E\, Y_i = \mu_i = f(x_i,\beta); \qquad \text{var } Y_i = \sigma_i^2; \qquad i = 1,\ldots,N.$$

Here, $\{x_i\ (k \times 1)\}$ are the design vectors, $\beta$ $(p \times 1)$ is the regression parameter, f is the possibly nonlinear mean response function, and the $\{\sigma_i\}$ express the heteroscedasticity. N is the total sample size.

The model (1.1) is quite general. When replicate observations are available at each design point, one might choose to regard the form of the $\{\sigma_i\}$ as entirely unknown and use the replicate observations to estimate the $\{\sigma_i\}$; see Jacquez, Mather and Crawford (1968), Jacquez and Norusis (1973), Fuller and Rao (1978) and the discussion of Section 1.2. Often, however, a model to describe the heterogeneity may be suggested by the data, application or convention. Indeed, if no replication is available the above approach is not feasible and some additional assumption must be made. One common situation in practice is that in which $\sigma_i^2$ increases as a function of explanatory or exogenous variables or the mean response. As a result, much recent interest has focused on modeling the variance as a function of these quantities; see Box and Hill (1974), Rodbard and Frazier (1975), Finney (1976), Raab (1981) and Carroll and Ruppert (1982a) among many others. A general parametric model for the variance which subsumes all of the above can be written as

$$(1.2) \qquad \sigma_i^2 = \sigma^2 g^2(z_i,\beta,\theta),$$

where $\sigma$ is an unknown scale parameter, and the variance function g expresses the heteroscedasticity, where $\{z_i\ (\ell \times 1)\}$ are known vectors, possibly containing the $\{x_i\}$, and $\theta$ $(r \times 1)$ is an unknown parameter. In practice as well as for theoretical investigations, g is taken to satisfy appropriate smoothness conditions.

In a model such as (1.2), estimation of the variances essentially reduces to estimation of $\theta$, since $\beta$ will be estimated routinely and the final estimates of $\beta$ and $\theta$ may be used to obtain a final estimate of $\sigma$. Thus, our investigation of properties of variance estimators for (1.2) focuses on properties of estimators for $\theta$.

Another approach to modeling variances as a function of the above-mentioned quantities is to assume that

$$\sigma_i^2 = h(z_i) \text{ or } h(\mu_i),$$

where h is an unknown but "smooth" function. The smoothness assumption separates this approach from that in which the $\{\sigma_i\}$ are completely unknown and invites estimation schemes from the realm of nonparametric function estimation; see Carroll (1982) and Matloff, Rose and Tai (1985). We do not consider this model here as we restrict our investigation to properties for parametric models.

## 1.2 Estimation in heteroscedastic regression models - background

Before we consider methods for variance estimation in (1.1) when (1.2) holds, we briefly review the literature for estimation in heteroscedastic regression, as the well known results we describe provide more concrete justification for our endeavor. We discuss particular methods for estimation of $\theta$ in the next section. Much of the literature focuses on the linear model $\mu_i = x_i'\beta$ with $k = p$; however, with sufficient smoothness conditions on f we may consider most of the results for the linear model applicable for (1.1).

In what follows, when we consider the possibility of replication at each design point, we write in obvious fashion $\{Y_{ij}\}$, $j = 1,\ldots,m_i$, to denote the $m_i \geq 2$ observations at $x_i$ and let $N = \Sigma_{i=1}^{M} m_i$ be the total sample size, where M is the number of design points. In the discussion below, as $N \to \infty$, consider the $\{m_i\}$ as fixed.

When (1.1) holds and the $\{\sigma_i\}$ are known, Gauss-Markov shows that, under regularity conditions as $N \to \infty$,

$$N^{1/2}(\hat{\beta}_{LS} - \beta) \xrightarrow{\mathcal{L}} \mathcal{N}(0,\Sigma_{LS}),$$

$$N^{1/2}(\hat{\beta}_{WLS} - \beta) \xrightarrow{\mathcal{L}} \mathcal{N}(0,\Sigma_{WLS}),$$

$$\Sigma_{LS} \geq \Sigma_{WLS} \text{ in the sense of nonnegative definiteness,}$$

where $\hat{\beta}_{LS}$ and $\hat{\beta}_{WLS}$ are the ordinary and weighted least squares estimators, respectively, and $\mathcal{N}(a,B)$ is the appropriate multivariate normal distribution with mean a and covariance matrix B. When the $\{\sigma_i\}$ are unknown, a natural attempt to capitalize on this result is to replace the $\{\sigma_i\}$ by estimates $\{\hat{\sigma}_i\}$ and perform weighted least squares to obtain the generalized least squares estimator $\hat{\beta}_{GLS}$. In this popular approach, the only difference between various estimation schemes for $\beta$ is the choice of the $\{\hat{\sigma}_i\}$.

Early suggestions for choosing the $\{\hat{\sigma}_i\}$ require replication. Jacquez, Mather and Crawford (1968) and Jacquez and Norusis (1973) study $\hat{\beta}_{GLS}$ empirically when

$$(1.3) \qquad \hat{\sigma}_i^2 = s_i^2 = (m_i - 1)^{-1} \Sigma_{j=1}^{m_i} (Y_{ij} - \overline{Y}_{i.})^2,$$

where $\overline{Y}_{i.} = m_i^{-1} \Sigma_{j=1}^{m_i} Y_{ij}$. They find that using these weights can be disastrous in the estimation of $\beta$, particularly if the $\{m_i\}$ are small. Fuller and Rao (1978) exploit the form of the mean response by choosing

$$(1.4) \qquad \hat{\sigma}_i^2 = m_i^{-1} \Sigma_{i=1}^{m_i} \{Y_{ij} - f(x_i,\hat{\beta}_{LS})\}^2.$$

They show that when the data are normally distributed, $\hat{\beta}_{GLS}$ computed under (1.4) satisfies

$$N^{1/2}(\hat{\beta}_{GLS} - \beta) \xrightarrow{\mathscr{L}} \mathscr{N}(0,\Sigma_{FR}), \quad \Sigma_{FR} \geq \Sigma_{WLS}.$$

As we discuss shortly, it is possible to construct estimators for the $\{\sigma_i\}$ such that $\hat{\beta}_{GLS}$ is asymptotically equivalent to $\hat{\beta}_{WLS}$, so in an asymptotic sense it is possible to improve on (1.3) and (1.4).

The difficulty with these naive forms of nonparametric estimation of the $\{\sigma_i\}$ is that the estimated variances can be widely different even when the $\{\sigma_i\}$ are close, especially if the $\{m_i\}$ are small. The above results thus suggest that poor, "nonsmooth" estimation of the variances can lead to poor estimation of $\beta$.

Because of results such as the above as well as considerations such as those described in Section 1.1, more recent work has focused on making assumptions about the form of the $\{\sigma_i\}$. Most often, the approach is parametric as in (1.2). Examples are plentiful; a very abbreviated list consists of the references in Sections 1.0 and 1.1 as well as Jobson and Fuller (1980) and Nelder and Pregibon (1986) among a host of others. Other nonparametric approaches not to be considered here are discussed, for example, by Carroll (1982).

When (1.2) holds, clearly one only need estimate $g(z_i,\beta,\theta)$ to obtain $\hat{\beta}_{GLS}$. The most common method for estimating $\beta$ is that in which one obtains estimates of $g(z_i,\beta,\theta)$ by using an estimate of $\theta$ and a preliminary estimate of $\beta$ and then performs weighted least squares; see, for example, Carroll and Ruppert (1982a) and Box and Hill (1974). We will henceforth use the term generalized least squares to refer exclusively to an estimator of this type. The following result is proved precisely by Jobson and Fuller (1980) and Carroll and Ruppert (1982a).

<u>Theorem A</u>. Suppose that (1.2) holds and that $\hat{\theta}$ and $\hat{\beta}_*$ are estimators for $\theta$ and $\beta$ such that $N^{1/2}(\hat{\theta} - \theta) = \mathcal{O}_p(1)$ and $N^{1/2}(\hat{\beta}_* - \beta) = \mathcal{O}_p(1)$. Then, under regularity conditions as $N \to \infty$, the generalized least squares estimator computed with $\hat{\sigma}_i = g(z_i,\hat{\beta}_*,\hat{\theta})$ satisfies

$$N^{1/2}(\hat{\beta}_{GLS} - \beta) \xrightarrow{\mathcal{L}} \mathcal{N}(0,\Sigma_{WLS}). \quad \square$$

This result shows that while we can improve on (1.3) and (1.4) under the assumptions that (1.2) holds and $\hat{\beta}_*$ and $\hat{\theta}$ are "good" estimates, standard asymptotic theory is not very helpful for deciding which estimates $\hat{\beta}_*$ and $\hat{\theta}$ to use.

Despite standard asymptotics as in Theorem A, there is evidence for finite samples to suggest that the choices of $\hat{\theta}$ and $\hat{\beta}_*$ do matter.

It has become increasingly apparent that the better one's estimate of $\theta$, the better one's final estimate of $\beta$ will be. Williams (1975) states that "both analytic and empirical studies...indicate that... ordering of efficiency (of estimates of $\beta$)...in small samples is in accordance with the ordering by efficiency (of estimates of $\theta$)." Rothenberg (1984) shows via second order calculations that in case g does not depend on $\beta$, when the data are normally distributed the covariance matrix of the generalized least squares estimator of $\beta$ is an increasing function of the covariance matrix of the estimator of $\theta$; see also Carroll and Ruppert (1985). Carroll and Ruppert (1982a,b) and Carroll, Ruppert and Wu (1986) cite evidence to suggest that the standard asymptotic theory for $\beta$ can be highly optimisitic. Carroll, Ruppert and Wu (1986) consider iteration of the generalized least squares method by letting $\hat{\beta}_* = \hat{\beta}_{GLS}$ after initial estimation by generalized least squares and repeating the process $\mathcal{C} - 1$ more times, the aim being to reduce the effect of an initial inefficient preliminary estimator for $\beta$ such as $\hat{\beta}_{LS}$. Their results suggest that the choice of $\hat{\theta}$ and $\hat{\beta}_*$ will play a role in determining the optimal number of cycles $\mathcal{C}$. A Monte-Carlo study of Goldfeld and Quandt (1972, pages 96-120) shows that it is possible to construct a badly inefficient generalized least squares estimator as well as a quite efficient one.

Second order asymptotics provide only part of the justification for studying the properties of variance function estimators. Estimation of $\theta$ is of independent interest and has important consequences in many settings. In some applications, estimation of $\beta$ is not the only problem of interest. In chemical and biological assay problems, for example, issues of prediction and calibration arise. In such problems, the estimator of $\theta$ plays a central role; in radioimmunoassay the statistical properties of prediction intervals and constructs such as the minimum detectable concentration are highly dependent on how one estimates $\theta$. We discuss this application and exhibit this result precisely in Chapter 5.

In engineering applications an important goal is to estimate the error made in predicting a new observation; this can be obtained directly from the variance function estimate. In off-line quality control, the emphasis is not only on the mean response but also on its variability; Box and Meyer (1986) state that "one distinctive feature of Japanese quality control improvement techniques is the use of statistical experimental design to study the effect of a number of factors on variance as well as the mean," see also Taguchi and Wu (1980). Effective estimation of the variance function could play a major role in this application.

The above discussion suggests that there are numerous practical situations in which choice of the method for estimating the $\{\sigma_i\}$ will be crucial. In the case of our model (1.2), this choice is defined by how we choose to estimate the variance function g, and in particular, $\theta$. Our brief review indicates that the parameter $\theta$ can play an important part in a statistical analysis far beyond that of a nuisance parameter.

Variance function estimation may be thought of as a type of regression problem in which we try to understand variance as a function of known or estimable quantities, where $\theta$ is a "regression parameter." Many of the methods for estimation of $\theta$ that have been proposed in the literature are (possibly weighted) regression methods based on functions of either absolute residuals from the current regression fit or, in the case of replication at each design point, sample standard deviations (1.3). Still other methods are joint estimation methods based on assumptions about the underlying distributions in which $(\sigma,\beta,\theta)$ are in principle estimated simultaneously. In Section 1.3 we describe specific approaches to estimation of $\theta$.

## 1.3 Estimation of $\theta$

We now discuss the form and motivation for several estimators of $\theta$ in (1.1) and (1.2). We confine our attention to methods which are simple or are in common use; in particular, we do not discuss the robust methods of Carroll and Ruppert (1982a) or Giltinan, Carroll and Ruppert (1986). In what follows, let $\hat{\beta}_*$ be a preliminary estimator for $\beta$. This could be unweighted least squares or the current estimate in an iterative reweighted least squares calculation. Denote the residuals by

$$r_i = Y_i - f(x_i,\hat{\beta}_*),$$

and define the errors $\{\epsilon_i\}$ by

$$\epsilon_i = \{Y_i - f(x_i,\beta)\} / \{\sigma g(z_i,\beta,\theta)\}.$$

### 1.3.1 Regression methods

Pseudo-likelihood. Given $\hat{\beta}_*$, the pseudo-likelihood estimator for $\theta$ maximizes the normal log-likelihood $\ell(\hat{\beta}_*,\theta,\sigma)$, where

$$\ell(\beta,\theta,\sigma) = -N \log \sigma - \Sigma_{i=1}^{N} \log g(z_i,\beta,\theta) - (2\sigma^2)^{-1}\Sigma_{i=1}^{N} \{Y_i - f(x_i,\beta)\}^2/g^2(z_i,\beta,\theta), \tag{1.5}$$

see Carroll and Ruppert (1982a). The terminology is borrowed from Gong and Samaniego (1981). While this method does not appear in this form to be based on a regression using absolute residuals, examination of the estimating equations for $\theta$ and $\sigma$ based on (1.5) show that they have the form of equations for weighted regression. Generalizations of pseudo-likelihood have been studied by Carroll and Ruppert (1982a) and Giltinan, Carroll and Ruppert (1986).

Least squares on squared residuals. Besides pseudo-likelihood, other methods using squared residuals have been proposed. The motivation for these methods is that the squared residuals have approximate expectation $\sigma^2 g^2(z_i,\beta,\theta)$, see Jobson and Fuller (1980) and Amemiya (1977). This suggests a nonlinear regression problem in which the "responses" are $\{r_i^2\}$ and the "regression function" is $\sigma^2 g^2(z_i,\hat{\beta}_*,\theta)$. The estimator $\hat{\theta}_{SR}$ minimizes in $\sigma$ and $\theta$

$$\Sigma_{i=1}^{N} \{r_i^2 - \sigma^2 g^2(z_i,\hat{\beta}_*,\theta)\}^2.$$

For normal data, the squared residuals have approximate variance $\sigma^4 g^4(z_i,\beta,\theta)$; in the spirit of generalized least squares, this suggests the weighted estimator which minimizes in $\theta$ and $\sigma$

$$(1.6) \qquad \Sigma_{i=1}^{N} \{r_i^2 - \sigma^2 g^2(z_i,\hat{\beta}_*,\theta)\}^2 \,/\, g^4(z_i,\hat{\beta}_*,\hat{\theta}_*),$$

where $\hat{\theta}_*$ is a preliminary estimator for $\theta$, $\hat{\theta}_{SR}$, for example. Full iteration, when it converges, would be equivalent to pseudo-likelihood.

Accounting for the effect of leverage. One objection to methods such as pseudo-likelihood and least squares based on squared residuals is that no compensation is made for the loss of degrees of freedom associated with preliminary estimation of $\beta$. For example, the effect of applying pseudo-likelihood directly seems to be a bias depending on p/N. For settings such as fractional factorials where p is large relative to N this bias could be substantial.

Bayesian ideas have been used to account for loss of degrees of freedom; see Harville (1977) and Patterson and Thompson (1974). When g does not depend on $\beta$, the restricted maximum likelihood approach of the latter authors suggests in our setting that one estimate $\theta$ from the mode of the marginal posterior density for $\theta$ assuming normal data and a prior for the parameters proportional to $\sigma^{-1}$. When g depends on $\beta$, one may extend the Bayesian arguments and use a linear approximation as in Box and Hill (1974) and Beal and Sheiner (1986) to define a restricted maximum likelihood estimator.

Let Q be the $N \times p$ matrix with ith row $f_\beta(x_i,\beta)^t \,/\, g(z_i,\beta,\theta)$, where $f_\beta(x_i,\beta) = \partial/\partial\beta \;\{f(x_i,\beta)\}$, and let $H = Q(Q^tQ)^{-1}Q^t$ be the "hat" matrix with diagonal element $h_{ii} = h_{ii}(\beta,\theta)$; the values $\{h_{ii}\}$ are the leverage values. See Cook and Weisberg (1982) for a discussion of leverage. It turns out that the restricted maximum likelihood estimator is equivalent to an estimator obtained by modifying pseudo-likelihood to account for the effect of leverage; this characterization, while not unexpected, is new. We exhibit the derivation of this estimator and its equivalence to a modification of pseudo-likelihood in Section 1.5.

The least squares approach using squared residuals can also be modified to show the effect of leverage. Jobson and Fuller (1980), for example, essentially note that for nearly normally distributed data we have the approximations

$$E\ r_i^2 \approx \sigma^2(1 - h_{ii})g^2(z_i,\beta,\theta),$$

$$\text{var}\ r_i^2 \approx \sigma^4(1 - h_{ii})^2 g^4(z_i,\beta,\theta).$$

To exploit these approximations modify (1.6) to minimize in $\theta$ and $\sigma$

$$(1.7) \qquad \Sigma_{i=1}^N \{r_i^2 - \sigma^2(1-\hat{h}_{ii})g^2(z_i,\hat{\beta}_*,\theta)\}^2/\{(1-\hat{h}_{ii})^2 g^4(z_i,\hat{\beta}_*,\theta)\},$$

where $\hat{h}_{ii} = h_{ii}(\hat{\beta}_*,\hat{\theta}_*)$ and $\hat{\theta}_*$ is a preliminary estimator for $\theta$. An asymptotically equivalent variation of this estimator in which one sets the derivatives of (1.7) with respect to $\theta$ and $\sigma$ equal to zero and then replaces $\hat{\theta}_*$ by $\theta$ can easily be seen to be equivalent to pseudo-likelihood in which one replaces standardized residuals by studentized residuals. While this estimator also takes into account the effect of leverage, it is different from restricted maximum likelihood.

Least squares on absolute residuals. Squared residuals are skewed and long-tailed, which has lead many authors to propose using absolute residuals to estimate $\theta$; see Glejser (1969) and Theil (1971). Assume that

$$E\ |Y_i - f(x_i,\beta)| = \eta g(z_i,\beta,\theta),$$

which is satisfied if the errors $\{\epsilon_i\}$ are independent and identically distributed. Mimicking the least squares approach based on squared residuals, one obtains the estimator $\hat{\theta}_{AR}$ by minimizing in $\eta$ and $\theta$

$$\Sigma_{i=1}^{N}\ \{|r_i| - \eta g(z_i,\hat{\beta}_*,\theta)\}^2.$$

In analogy to (1.6), the weighted version is obtained by minimizing

$$\Sigma_{i=1}^{N}\ \{|r_i| - \eta g(z_i,\hat{\beta}_*,\theta)\}^2 \ /\ g^2(z_i,\hat{\beta}_*,\hat{\theta}_*),$$

where $\hat{\theta}_*$ is a preliminary estimator for $\theta$, probably $\hat{\theta}_{AR}$. As for least squares estimation based on squared residuals, one could presumably modify this approach to account for the effect of leverage.

Logarithm method. The suggestion of Harvey (1976) is to exploit the fact that the logarithm of the absolute residuals has approximate expectation $\log\ \{\sigma g(z_i,\beta,\theta)\}$. Estimate $\theta$ by ordinary least squares regression of $\log\ |r_i|$ on $\log\ \{\sigma g(z_i,\hat{\beta}_*,\theta)\}$, since if the errors are independent and identically distributed, the regression should be approximately homoscedastic. If one of the residuals is near zero the regression could be adversely affected by a large "outlier," hence in practice one might wish to delete a few of the smallest absolute residuals, perhaps trimming the smallest few percent, for example. The logarithm method invites further interpretation for $\theta$ when, for example, g can be written

$$(1.8) \qquad g(z_i,\beta,\theta) = g(\mu_i,z_i,\theta);$$

if we have the "power of the mean" model $g(\mu_i,z_i,\theta) = \mu_i^{\theta}$, the natural interpretation of $\theta$ is as the slope parameter.

### 1.3.2 Other methods

Maximum likelihood. In a parametric model such as (1.1) and (1.2), joint maximum likelihood estimation is possible, where we use the term maximum likelihood to mean normal theory maximum likelihood. When the variance function does not depend on $\beta$, it can be easily shown that the maximum likelihood estimator of $\theta$ is asymptotically equivalent to weighted least squares methods based on squared residuals; in the important situation in which the variance function depends on $\beta$, as in (1.8), this is not the case. These results will be exhibited in Chapter 3. Jobson and Fuller (1980) have in fact shown that under normality the maximum likelihood estimator of $\beta$ has asymptotic covariance matrix at least as small as that of any generalized least squares estimator of Theorem A. However, it has been observed by Carroll and Ruppert (1982b) and McCullagh (1983) that while maximum likelihood estimators enjoy asymptotic optimality when the model and distributional assumptions are exactly correct, the maximum likelihood estimator of $\beta$ can suffer severe problems under departures from these assumptions. This suggests that joint maximum likelihood estimation should not be applied blindly, although we consider the properties of the maximum likelihood estimator of $\theta$ for comparative purposes because of optimality when the assumptions are correct.

Extended quasi-likelihood. When $\theta$ is known and the variance function has form (1.8), quasi-likelihood estimation of $\beta$ is a form of generalized least squares which is iterated such that $\hat{\beta}_* = \beta$; see Wedderburn (1974), McCullagh (1983) and McCullagh and Nelder (1983). The extended quasi-likelihood method of Nelder and Pregibon (1986) is a joint estimation scheme which attempts to extend the notion of quasi-likelihood to include estimation of $\theta$ assuming (1.8). The method is based on the assumption that the data arise from a class of distributions depending on $\theta$ and involves estimation of $\theta$ by minimizing in $\beta$, $\theta$ and $\sigma$ the "extended quasi-likelihood"

$$(1.9) \qquad Q^+ = -(1/2)\,\Sigma_{i=1}^{N} \left[\, \log\{2\pi\sigma^2 g^2(Y_i, z_i, \theta)\} + D_\theta(Y_i, \mu_i, z_i)/\sigma^2 \,\right],$$

where

$$D_\theta(y, \mu, z) = -2 \int_y^\mu \frac{y - u}{g^2(u, z, \theta)}\, du.$$

The major motivation for this method is that it includes an extended parametric family which "nearly" includes the gamma and Poisson distributions. For example, when $g(\mu_i, z_i, \theta) = \mu_i^\theta$ and $\theta = 1/2$, $\sigma = 1$,

$Q^+$ differs from the Poisson log-likelihood by replacing $Y_i!$ by its Stirling approximation; for $\theta = 1$, $Q^+$ differs from the gamma log-likelihood by a factor depending on $\sigma$.

For a related formulation, see Efron (1985).

Methods requiring replication. Methods requiring $m_i \geq 2$ replicate observations at each $x_i$ have been proposed in the assay literature; for simplicity, we consider only the case of equal replication $m_i \equiv m \; \forall i$. These methods do not depend on the postulated form of the regression function; one reason that this might be advantageous is that in many assays along with observed pairs $\{Y_{ij}, x_i\}$ there will also be pairs in which only $Y_{ij}$ is observed.

A popular and widely used method in radioimmunoassay is that of Rodbard and Frazier (1975). If we assume (1.8), the method is identical to the logarithm method previously discussed except that one replaces $|r_i|$ by the sample standard deviation $s_i$ and $f(x_i, \hat{\beta}_*)$ in the "regression function" by the sample mean $\overline{Y}_{i.}$.

Under the assumption of independence and (1.8), the modified maximum likelihood method of Raab (1981) estimates $\theta$ by joint maximization of the "modified" normal likelihood

$$(1.10) \quad \Pi_{i=1}^{M} \{2\pi\sigma^2 g^2(\mu_i, z_i, \theta)\}^{(m-1)/2} \exp[-\Sigma_{j=1}^{m} (Y_{ij}-\mu_i)^2 / \{2\sigma^2 g^2(\mu_i, z_i, \theta)\}]$$

in the $(M + r + 1)$ parameters $\sigma^2$, $\theta$, $\mu_1, \ldots, \mu_M$; the modification serves to make the estimator of $\sigma^2$ unbiased.

## 1.4 Summary

The discussion of Sections 1.2 and 1.3 suggests the need for a unified investigation of variance functions in a model such as (1.2), in particular, estimation of the structural parameter $\theta$. Previous work in the literature tends to be scattered in that it appears in the efforts of researchers from a variety of fields of application and it treats special cases of (1.2) as different models with their own estimation methods. Our intent is to study parametric variance function in a unified way so that we may make general observations about different methods as a class and comparisons among the methods of Section 1.3 under various conditions. The major insight which allows for a unified study is that the absolute residuals or the sample standard deviations in the case of replication are the basic building blocks for analysis.

In Chapter 2 we develop an asymptotic theory for a general class of estimators for $\theta$ whose construction encompasses the regression methods of Section 1.3.1. In Chapter 3 we investigate the properties of the estimators of Section 1.3.2 which do not fit into this general class. Chapter 4 contains comparisons of the theoretical properties of the estimators of Section 1.3 based on the results of Chapters 2 and 3 and a general discussion of the implications of our work. In Chapter 5 we discuss the notion of minimum detectable concentration from the field of radioimmunoassay and show precisely how the estimator of $\theta$ plays a central role in the estimation of this important quantity. Throughout, our presentation is brief and in many cases heuristic in order that general insights and results are not overshadowed by laborious technical details.

## 1.5 Characterization of restricted maximum likelihood

We now specify the form of the restricted maximum likelihood estimator for $\theta$ in (1.1) and (1.2) and show its equivalence to a modification of pseudo-likelihood.

Let $\hat{\beta}_*$ be a preliminary estimator for $\beta$ and define

$$\hat{\sigma}_G^2(\theta) = (N-p)^{-1}\Sigma_{i=1}^N r_i^2 / g^2(z_i,\beta,\theta),$$

$$S_G(\theta) = N^{-1}\Sigma_{i=1}^N f_\beta(x_i,\hat{\beta}_*)f_\beta(x_i,\hat{\beta}_*)^t / g^2(z_i,\hat{\beta}_*,\theta).$$

Assume first that g does not depend on $\beta$. Thus, writing in this case $g_i(\theta)$ to denote g at $z_i$, the likelihood is proportional to

$$p(\beta,\theta,\sigma)$$

$$=\{\Pi_{i=1}^N g_i^2(\theta)\}^{-1/2}\sigma^{-N}\exp[-(2\sigma^2)^{-1}\Sigma_{i=1}^N\{Y_i-f(x_i,\beta)\}/g_i^2(\theta)].$$

Let the prior distribution $\pi(\beta,\theta,\sigma)$ for the parameters be proportional to $\sigma^{-1}$. Then, by Bayes' theorem, the joint posterior density $p(\beta,\theta,\sigma|\underline{Y})$ is proportional to

$$(1.11)\qquad p(\beta,\theta,\sigma)\ \pi(\beta,\theta,\sigma).$$

The marginal posterior for $\theta$ may be computed by integration of $p(\beta,\theta,\sigma|\underline{Y})$. For a nonlinear regression model the integral is hard to compute in closed form, so, following Box and Hill (1974) and Beal and Sheiner (1986), note that if $\hat{\beta}_*$ is a generalized least squares estimator evaluated at the true $\theta$, we have the linear approximation

$$f(x_i,\beta) \approx f(x_i,\hat{\beta}_*) + f_\beta(x_i,\hat{\beta}_*)^t(\beta - \hat{\beta}_*).$$

Replacing $f(x_i,\beta)$ by its linear expansion in (1.11), one can compute the marginal posterior for $\theta$ exactly as proportional to

$$(1.12) \qquad p(\theta) = \frac{\{\Pi_{i=1}^N g_i^2(\theta)\}^{1/2}}{\hat{\sigma}_G^{(N-p)}(\theta)\ \{\text{Det } S_G(\theta)\}^{1/2}},$$

where Det A = determinant of A. If the variances depend on $\beta$, we extend the Bayesian arguments by replacing $g_i(\theta)$ by $g(z_i,\hat{\beta}_*,\theta)$; see Box and Hill (1974) and Beal and Sheiner (1986) for related discussion.

Let $\tilde{H}$ be the hat matrix H evaluated at $\hat{\beta}_*$ and let $\tilde{h}_{ii} = h_{ii}(\hat{\beta}_*,\theta)$. From (1.5), pseudo-likelihood solves in $(\theta,\sigma)$

$$(1.13) \qquad \Sigma_{i=1}^N\ [r_i^2/\{\sigma^2 g^2(z_i,\hat{\beta}_*,\theta)\}] \begin{bmatrix} 1 \\ \nu_\theta(i,\hat{\beta}_*,\theta) \end{bmatrix} = \Sigma_{i=1}^N \begin{bmatrix} 1 \\ \nu_\theta(i,\hat{\beta}_*,\theta) \end{bmatrix},$$

where $\nu_\theta(i,\beta,\theta) = \partial/\partial\theta\ \{\log\ g(z_i,\beta,\theta)\}$. From the discussion in

Section 1.3 and the fact that $\tilde{H}$ is idempotent, the left-hand side of (1.13) has approximate expectation

$$(1.14) \qquad \begin{bmatrix} N - p \\ \Sigma_{i=1}^{N} \nu_{\theta}(i,\hat{\beta}_*,\theta)(1 - \tilde{h}_{ii}) \end{bmatrix}.$$

To modify pseudo-likelihood to account for loss of degrees of freedom, the suggestion is to equate the left hand side of (1.13) to its approximate expectation; i.e., solve (1.13) with the right-hand side replaced by (1.14). We now show that this is equivalent to restricted maximum likelihood. From (1.13) and (1.14), the modified pseudo-likelihood estimator solves

$$(1.15) \qquad \begin{aligned} &-\Sigma_{i=1}^{N} \nu_{\theta}(i,\hat{\beta}_*,\theta)\ (1 - \tilde{h}_{ii}) \\ &\quad + \{\Sigma_{i=1}^{N} r_i^2\ \nu_{\theta}(i,\hat{\beta}_*,\theta)\ /\ g_i^2(z_i,\hat{\beta}_*,\theta)\}\ /\ \hat{\sigma}_G^2(\theta) = 0. \end{aligned}$$

Taking logarithms in (1.12) and setting equal to zero the derivative with respect to $\theta$ yields (1.15) with the first term on the left-hand side replaced by

$$\Sigma_{i=1}^{N} \nu_{\theta}(i,\hat{\beta}_*,\theta) - (1/2)\ \partial/\partial\theta\ [\log\{\mathrm{Det}\ S_G(\theta)\}];$$

thus, to show equivalence of the two estimators we wish to show that

$$(1.16) \qquad (1/2)\ \partial/\partial\theta\ [\log\{\mathrm{Det}\ S_G(\theta)\}] = -\Sigma_{i=1}^{N} \nu_{\theta}(i,\hat{\beta}_*,\theta)\tilde{h}_{ii}.$$

From Nel (1980),

$$(1.17) \qquad \partial/\partial\theta\ [\log\{\mathrm{Det}\ S_G(\theta)\}] = \mathrm{trace}[S_G(\theta)^{-1}\ \partial/\partial\theta\ \{S_G(\theta)\}].$$

Letting $V = \mathrm{diag}[[\ \nu_\theta(i,\hat{\beta}_*,\theta)]]$ gives

$$\partial/\partial\theta\ \{S_G(\theta)\} = -2N^{-1}\tilde{Q}^t V\tilde{Q},$$

where $\tilde{Q}$ is Q of Section 1.3 evaluated at $\hat{\beta}_*$. Plugging this into (1.17) and using the fact that $NS_G(\theta) = \tilde{Q}^t\tilde{Q}$ and the definition of $\tilde{H}$ yields (1.16).

## CHAPTER II

## AN ASYMPTOTIC THEORY OF VARIANCE FUNCTION ESTIMATION

### 2.0 Introduction

In this chapter we construct an asymptotic theory for a general class of regression-type estimators for $\theta$ based on residuals from some regression fit. This formulation includes all the estimators of Section 1.3.1 as well as maximum likelihood. We comment briefly on the properties of the estimator of Rodbard and Frazier and Raab's functional maximum likelihood estimator in the course of our discussion as well. We also investigate the effect of replacing in the methods of Section 1.3.1 absolute residuals by sample standard deviations and, when (1.8) holds, of replacing predicted values $f(x_i,\hat{\beta}_*)$ by sample means $\overline{Y}_{i.}$ in the "response function" part of the regression.

The technical assumptions necessary for investigating a nonlinear model such as (1.1) and (1.2) are detailed; see, for example, Jennrich (1969) and Carroll and Ruppert (1982a). Since our major interest lies in obtaining general insights, we do not explicitly state such technical assumptions so that important results are not obscured by a complicated exposition. The nature of the necessary assumptions is apparent in the proofs of major results in Section 2.6. Assume throughout that the $\{\epsilon_i\}$ are independent random variables.

## 2.1 Methods based on functions of absolute residuals

Write $d_i(\beta) = |Y_i - f(x_i,\beta)|$. Let $H_1$ be a smooth function and define $H_{2,i}$ by

$$H_{2,i} = H_{2,i}(\eta,\theta,\beta) = E\ [\ H_1\{d_i(\beta)\}\ ],$$

where $\eta$ is a scale parameter which is usually a function of $\sigma$ only. If $\hat{\eta}_*$, $\hat{\theta}_*$ and $\hat{\beta}_*$ are any estimators for $\eta$, $\theta$, and $\beta$, define $\hat{\eta}$ and $\hat{\theta}$ to be the solutions of

$$(2.1) \qquad N^{-1/2}\ \Sigma_{i=1}^{N}\ H_{4,i}(\eta,\theta,\hat{\beta}_*)\ \left\{\frac{H_1\{d_i(\hat{\beta}_*)\} - H_{2,i}(\eta,\theta,\hat{\beta}_*)}{H_{3,i}(\hat{\eta}_*,\hat{\theta}_*,\hat{\beta}_*)}\right\} = 0,$$

where $H_{3,i}(\eta,\theta,\beta)$ is a smooth function and $H_{4,i}$ is usually the partial derivative of $H_{2,i}$ with respect to $(\eta,\theta)$.

The class of estimators solving (2.1) includes directly or includes an asymptotically equivalent version of the estimators of Section 1.3.1; we exhibit this in Section 2.5. Note that for methods which account for the effect of leverage, $H_{2,i}$, $H_{3,i}$ and $H_{4,i}$ will depend on the $\{h_{ii}\}$. It may be shown by that in this case we need the additional assumption that if $h = \max_{1\le i\le N} h_{ii}$, then $N^{1/2}h$ converges to zero. Such an assumption is typical in studies of robust regression estimators.

<u>Theorem 2.1</u>. Let $\hat{\eta}_*$, $\hat{\theta}_*$ and $\hat{\beta}_*$ be $N^{1/2}$ consistent for estimating $\eta$, $\theta$ and $\beta$. Let $\dot{H}_1$ be the derivative of $H_1$ and define

$$C_i = H_{4,i} \left[ \frac{H_1\{d_i(\beta)\} - H_{2,i}}{H_{3,i}} \right];$$

$$B_{1,N} = N^{-1}\Sigma_{i=1}^N H_{4,i}H_{4,i}^t/H_{3,i};$$

$$B_{2,N} = -N^{-1}\Sigma_{i=1}^N (H_{4,i}/H_{3,i})\ \partial/\partial\beta\ \{H_{2,i}(\eta,\theta,\beta)\};$$

$$B_{3,N} = -N^{-1}\Sigma_{i=1}^N (H_{4,i}/H_{3,i})f_\beta(x_i,\beta)\ E\ [\dot{H}_1\{d_i(\beta)\}\mathrm{sign}(\epsilon_i)].$$

Then, under regularity conditions as $N \to \infty$,

$$B_{1,N}\, N^{1/2} \begin{bmatrix} \hat{\eta} - \eta \\ \hat{\theta} - \theta \end{bmatrix}$$

(2.2)

$$= N^{-1/2}\Sigma_{i=1}^N C_i + (B_{2,N} + B_{3,N})\, N^{1/2}(\hat{\beta}_* - \beta) + o_p(1). \quad \square$$

We may immediately make some general observations about the estimator $\hat{\theta}$ solving (2.1). Note that if the variance function does not depend on $\beta$, then $H_{2,i}$ does not depend on $\beta$ and hence $B_{2,N} \equiv 0$. For the estimators of Section 2.1, $\dot{H}_1$ is an odd function. Thus, if the errors $\{\epsilon_i\}$ are symmetrically distributed, $E[\ \dot{H}_i\{d_i(\beta)\}\mathrm{sign}(\epsilon_i)\ ] = 0$ and hence $B_{3,N} \equiv 0$. The following result is then immediate.

Corollary 2.1(a). Suppose that the variance function does not depend on $\beta$ and the errors are symmetrically distributed. Then the asymptotic distributions of the regression estimators of Section 1.3.1 do not depend on the method used to obtain $\hat{\beta}_*$. If both of these conditions do not hold simultaneously, then the asymptotic distributions will depend in general on the method of estimating $\beta$. □

The implication is that in the situation for which the variance function does not depend on $\beta$ and the data are approximately symmetrically distributed, for large sample sizes the preliminary estimator for $\beta$ will play little role in determining the properties of $\hat{\theta}$. Note also from (2.2) that for weighted methods, the effect of the preliminary estimator of $\theta$ is asymptotically negligible regardless of the underlying distributions.

The preliminary estimator $\hat{\beta}_*$ will be in general the unweighted least squares estimator, a generalized least squares estimator or some robust estimator. See, for example, Huber (1981) and Giltinan, Carroll and Ruppert (1986) for examples of robust estimators for $\beta$. It is easily shown that for some vectors $\{v_{N,i}\}$, these estimators admit an asymptotic expansion of the form

$$(2.3) \qquad N^{1/2}(\hat{\beta}_* - \beta) = N^{-1/2}\Sigma_{i=1}^{N} \Psi(v_{N,i},\epsilon_i) + o_p(1).$$

Here $\Psi$ is odd in the argument $\epsilon$. In case the variance function depends on $\beta$, $B_{2,N} \neq 0$ in general; however, if the errors are symmetrically distributed and $\hat{\beta}_*$ has expansion of form (2.3), then the two terms on the right-hand side of (2.2) are asymptotically independent. The following is then immediate.

Corollary 2.1(b). Suppose that the errors are symmetrically distributed and that $\hat{\beta}_*$ has an asymptotic expansion of the form (2.3). Then for the estimators of Section 1.3.1, the asymptotic covariance matrix of $\hat{\theta}$ is a monotone nondecreasing function of the asymptotic covariance matrix of $\hat{\beta}_*$. □

By the Gauss-Markov theorem and the results of Jobson and Fuller (1980) and Carroll and Ruppert (1982a), the implication of Corollary 2.1(b) is that in general using unweighted least squares estimates of $\beta$ will result in inefficient estimates of $\theta$. This phenomenon is exhibited in small samples in the Monte Carlo study of Chapter 5. The result suggests that if one has available as $\hat{\beta}_*$ the unweighted least squares estimate, one ought to iterate the process of estimating $\theta$ -- use the current value $\hat{\beta}_*$ to estimate $\theta$ from (2.1), use these $\hat{\beta}_*$ and $\hat{\theta}$ to obtain an updated $\hat{\beta}_*$ by generalized least squares and repeat the process $\mathscr{C} - 1$ more times. It is clear that the asymptotic distribution of $\hat{\theta}$ will be the same for $\mathscr{C} \geq 2$ with larger asymptotic covariance for $\mathscr{C} = 1$, so in principle, asymptotically at least, one ought to iterate this process at least twice.

## 2.2 Methods based on sample standard deviations

Assume at each of M design points we have $m \geq 2$ replicate observations so that $N = Mm$ represents the total number of observations. We may compute then the sample standard deviations $\{s_i\}$, where $s_i^2 = (m-1)^{-1}\Sigma_{j=1}^{m} (Y_{ij} - \overline{Y}_{i\cdot})^2$. Sample standard deviations themselves have been proposed as estimators of the variance with disasterous results as described in Chapter 1. When replication exists, however, practitioners feel comfortable with the notion that the $\{s_i\}$ may be used as a basis for estimating variances; thus, one might reasonably seek to estimate $\theta$ by replacing $d_i(\hat{\beta}_*)$ by $s_i$ in the general equation (2.1).

The following result is almost immediate from the proof of Theorem 2.1 in Section 2.6. Here we let $N \to \infty$ such that m remains fixed.

<u>Theorem 2.2</u>. If $d_i(\hat{\beta}_*)$ is replaced by $s_i$ in (2.1), then under the conditions of Theorem 2.1 the resulting estimator for $\theta$ satisfies (2.2) with $B_{3,N} \equiv 0$ and the redefinitions

$$C_i = (H_{4,i}/H_{3,i})\{H_1(s_i) - H_{2,i}\};$$

(2.4)

$$H_{2,i} = E\,\{H_1(s_i)\} = H_{2,i}(\eta,\theta,\beta). \quad \square$$

If the errors are symmetrically distributed, then from (2.2) and the result of Theorem 2.2 we have that whether one is better off using absolute residuals or sample standard deviations in the methods of Section 1.3.1 depends only on the differences between the expected values and variances of $H_1\{d_i(\beta)\}$ and $H_1(s_i)$. In Chapter 4 we exhibit such comparisons explicitly and show that absolute residuals can be preferred to sample standard deviations in situations of practical importance.

## 2.3 Methods not depending on the regression function

We assume throughout this discussion that the variance function has form (1.8) and that $m \geq 2$ replicates are available at each $x_i$. From Section 1.3.1 we see that the "regression function" part of the estimating equations depends on $f(x_i,\hat{\beta}_*)$, so that in the general equation (2.1) $H_{2,i}$, $H_{3,i}$ and $H_{4,i}$ all depend on $f(x_i,\hat{\beta}_*)$. In some settings, one may not postulate a form for the $\mu_i$ for estimating $\theta$; the method of Rodbard and Frazier (1975), for example, uses $s_i$ in place of $d_i(\hat{\beta}_*)$ as in Section 2.2 and replaces $f(x_i,\hat{\beta}_*)$ by the sample mean $\overline{Y}_{i.}$. We now consider the effect of replacing predicted values by sample means for the general class (2.1).

The presence of the sample means in the variance function in (2.1) requires more complicated and restrictive assumptions than the usual large sample asymptotics applied heretofore. The method of Rodbard and Frazier and more generally the general method (2.1) with sample means are functional nonlinear errors in variables problems as studied by Wolter and Fuller (1982) and Stefanski and Carroll (1985). In our setting, the standard asymptotics for these problems correspond to letting $\sigma$ go to zero at rate $N^{-1/2}$. In Section 2.4 we discuss the practical implications of assuming $\sigma$ to be small; for now, we state the following result.

<u>Theorem 2.3.</u>. Suppose that we replace $f(x_i,\hat{\beta}_*)$ by $\overline{Y}_{i.}$ in $H_{2,i}$, $H_{3,i}$ and $H_{4,i}$ in Theorems 2.1 and 2.2 and adopt the assumptions of those theorems. Further, suppose that as $N \rightarrow \infty$, $\sigma \rightarrow 0$ simultaneously and

(i) $N^{1/2}\sigma \to \lambda,\ 0 \le \lambda < \infty$;

(ii) $N^{1/2}\Sigma_{i=1}^{N} C_i$ has a nontrivial asymptotic normal limit distribution;

(iii) The $\{\epsilon_i\}$ are independent and identically distributed random variables. Either the $\{\epsilon_i\}$ are symmetric or $\lambda = 0$.

(iv) $\{|\overline{Y}_{i.} - \mu_i| / \sigma\}^2$ has uniformly bounded k moments for some $k > 2$.

Then the results of Theorems 2.1 and 2.2 hold with $B_{2,N} = B_{3,N} \equiv 0$. □

This result shows that under certain restrictive assumptions, one may replace predicted values by sample means under replication; however, it is important to realize that the assumption of small $\sigma$ is not generally valid and hence the use of sample means may be disadvantageous in situations where these asymptotics do not apply.

The estimator of Raab (1981) discussed in Section 1.3.2 is also a functional nonlinear errors in variables estimator, complicated by a parameter space with size which is $\mathcal{O}(N)$. Sadler and Smith (1985) have observed that the Raab estimator is often indistinguishable from the same estimator with $\mu_i$ replaced by $\overline{Y}_{i.}$ in (1.10); such an estimator is contained in the general class (2.1). Note that (iv) is trivially satisfied in this case. In Chapter 3 we show that under the asymptotics of Theorem 2.3 and additional regularity conditions that the two estimators are asymptotically equivalent. We may thus consider the result of Theorem 2.3 relevant to this estimator as well.

## 2.4 Small $\sigma$ asymptotics

In Section 2.3 we were forced by technical considerations to pursue an asymptotic theory in which $\sigma$ is small. It turns out that in some situations of practical importance these asymptotics are relevant. In particular, in assay data values for $\sigma$ are often observed which are quite small relative to the means and for which assumption (i) of Theorem 2.3 is reasonable. Further, such asymptotics are used in the study of data transformations in regression. It is thus worthwhile to consider the effect of small $\sigma$ on the results of Sections 2.1 and 2.2 and to comment on some other implications of letting $\sigma \rightarrow 0$.

The following observations can be made from the discussion in Section 2.5. In the situation of Theorem 2.1, if the errors are symmetrically distributed, then for the estimators of Section 1.3.1, if $\sigma \rightarrow 0$ as $N \rightarrow \infty$, there is no effect for estimating the regression parameter $\beta$. In the situation of Theorem 2.2, the errors need not even be symmetrically distributed. The major insight provided by these results is that in certain practical situations in which $\sigma$ is small, the choice of $\hat{\beta}_*$ may not be too important even if the variance function depends on $\beta$.

Small $\sigma$ asymptotics may be used also to provide insight into the behavior of other estimators for $\theta$ which do not fit into the general framework of (2.1). We use small $\sigma$ asymptotics in Chapter 3 to evaluate the properties of the extended quasi-likelihood estimator and to gain insight into the behavior of the maximum likelihood estimator. We show that the extended quasi-likelihood estimator is asymptotically equivalent to weighted regression estimators based on squared residuals and the maximum likelihood estimator when $\sigma \to 0$.

## 2.5 Examples and further results.

The asymptotic theory constructed in this chapter has allowed us to state some general characteristics of regression-type estimators of $\theta$. In this section we use the theory to exhibit the specific forms for the various estimators of Section 1.3.1. Throughout, define

$$\nu(i,\beta,\theta) = \log g(z_i,\beta,\theta),$$

and let $\nu_\theta(i,\beta,\theta)$ and $\nu_\beta(i,\beta,\theta)$ be the column vectors of partial derivatives of $\nu$ with respect to $\theta$ and $\beta$. Further, let

$$\tau(i,\beta,\theta) = \begin{bmatrix} 1 \\ \nu_\theta(i,\beta,\theta) \end{bmatrix},$$

$$D_{1,N} = N^{-1}\Sigma_{i=1}^{N} \tau(i,\beta,\theta)\, \tau(i,\beta,\theta)^t,$$

$$D_{2,N} = -N^{-1}\Sigma_{i=1}^{N} \tau(i,\beta,\theta)\, \nu_\beta(i,\beta,\theta)^t \text{ and}$$

$$\varsigma(\beta,\theta) = N^{-1}\Sigma_{i=1}^{N} \{\nu_\theta(i,\beta,\theta) - \bar{\nu}_\theta(\beta,\theta)\}\,\{\nu_\theta(i,\beta,\theta) - \bar{\nu}_\theta(\beta,\theta)\}^t,$$

where $\bar{\nu}_\theta(\beta,\theta)$ is the mean of the $\nu_\theta(i,\beta,\theta)$. For simplicity, assume that the errors $\{\varepsilon_i\}$ are independent and identically distributed with kurtosis $\kappa$; $\kappa = 0$ for normality.

2.5.1 <u>Pseudo-likelihood, restricted maximum likelihood and weighted squared residuals.</u>

If when accounting for the effect of leverage we let $h \to 0$ such that $N^{1/2}h \to 0$, then these methods are asymptotically equivalent. Writing $\eta = \log \sigma$, we have $H_1(x) = x^2$,

$$H_{2,i} = \exp(2\eta)\ g^2(z_i,\beta,\theta), \quad H_{3,i} = H_{2,i}^2$$

and $E\ [\ \dot{H}_1\{d_i(\beta)\}\ \mathrm{sign}(\epsilon_i)\ ] = 2\ E\ [\ Y_i - f(x_i,\beta)] = 0$ so that $B_{3,N} \equiv 0$ regardless of the form of the underlying distributions. Algebra yields $B_{1,N} = 4D_{1,N}$, $B_{2,N} = 4D_{2,N}$ and $C_i = 2\ (\epsilon_i^2 - 1)\ \tau(i,\beta,\theta)$. If g does not depend on $\beta$, or if as $\sigma \to 0$

$$(2.5) \qquad N^{1/2}(\hat{\beta}_* - \beta)/\sigma = \sigma_p(1),$$

which is satisfied by unweighted least squares and generalized least squares estimators and the maximum likelihood estimator, then $\hat{\theta}$ is asymptotically normally distributed with mean $\theta$ and covariance matrix

$$(2.6) \qquad (2 + \kappa)\ \{4N\ \varsigma(\beta,\theta)\}^{-1}.$$

Thus, the regression estimators considered here are asymptotically equivalent to the maximum likelihood estimator when $\sigma$ is small or g does not depend on $\beta$; we exhibit an asymptotic expansion for the maximum likelihood estimator in Chapter 3.

As mentioned in Section 2.4, under the small $\sigma$ asymptotics of Theorem 2.3, the extended quasi-likelihood estimator of $\theta$ is asymptotically equivalent to the estimators here with asymptotic covariance matrix (2.6). Thus, under these conditions, pseudo-likelihood, weighted squared residuals, restricted maximum likelihood, maximum likelihood and, if $\sigma \to 0$, extended quasi-likelihood, are all asymptotically equivalent. In addition, all of these estimators have influence functions which are linear in the squared errors, indicating substantial nonrobustness.

We may also observe that efficiency considerations dictate that these methods are preferable to unweighted regression on squared residuals. To see this note that we may write (2.6) as

$$(2.7) \qquad \frac{(2+\kappa)}{4}\,(V^t W^{-1} V)^{-1},$$

where W is the $N \times N$ diagonal matrix with elements $H_{3,i}$ and V is the $N \times p$ matrix with $i^{th}$ row $H^t_{4,i}$. For the unweighted estimator based on squared residuals, calculations similar to those above show that the asymptotic covariance matrix when either $\sigma \to 0$ or g does not depend on $\beta$ is given by

$$(2.8) \qquad \frac{(2+\kappa)}{4}\,(V^t V)^{-1}(V^t W V)(V^t V)^{-1}.$$

The comparison between (2.7) and (2.8) is simply that of the Gauss-Markov theorem, so that (2.7) is no larger than (2.8) in the sense of nonnegative definiteness.

2.5.2 Logarithms of absolute residuals

We do not consider deletion of the few smallest absolute residuals. Here $H_1(x) = \log x$ so that $\dot{H}_1(x) = x^{-1}$. Letting $\eta = \log \sigma$ and assuming independent and identically distributed errors we have $H_{2,i} = \eta + \nu(i,\beta,\theta) + E \log |\epsilon|$, $H_{3,i} \equiv 1$ and $H_{4,i} = \tau(i,\beta,\theta)$, so that $B_{1,N} = D_{1,N}$, $B_{2,N} = D_{2,N}$, $C_i = \{ \log |\epsilon_i| - E \log |\epsilon| \} \tau(i,\beta,\theta)$ and

$$(2.9) \qquad B_{3,N} = -\frac{N^{-1}}{\sigma} \Sigma_{i=1}^{N} g^{-1}(z_i,\beta,\theta)\, \tau(i,\beta,\theta)\, f_\beta(x_i,\beta)^t\, E\left[ \frac{\text{sign}(\epsilon_i)}{|\epsilon_i|} \right].$$

Under the assumption of symmetry of the errors, with g not depending on $\beta$ or $\sigma \to 0$, and assuming (2.5), algebra shows that $\hat{\theta}$ is asymptotically normally distributed with mean $\theta$ and covariance matrix

$$(2.10) \qquad \text{var}\,\{\log |\epsilon|^2\}\ \{4N\, \varsigma(\beta,\theta)\}^{-1}.$$

The influence function for this estimator is linear in the logarithm of the absolute errors, indicating less severe nonrobustness than for the squared residual estimators. Note that in general for the squared residuals estimators of Section 2.5.1, $B_{3,N} \equiv 0$, whereas for the logarithm method $B_{3,N} \neq 0$ in general unless the errors are symmetrically distributed. From (2.9) we see that even if $\sigma \to 0$ and

(2.5) holds, if the errors are not symmetric then there will be an additional effect due to estimating $\beta$ not present for the methods of Section 2.5.1, even if g does not depend on $\beta$.

2.5.3 Absolute residuals

Here, assume that the errors are independent and identically distributed and let $\exp(\eta) = \sigma E|\epsilon|$. Consider the weighted estimator. We have $H_1(x) = x$, $\dot{H}_1(x) = 1$, $H_{2,i} = \exp(\eta)\, g(z_i,\beta,\theta)$ and $H_{3,i} = H^2_{2,i}$ so that $B_{1,N} = D_{1,N}$, $B_{2,N} = D_{2,N}$,

$$C_i = [\ \{\ |\epsilon_i|/E\ |\epsilon|\ \} - 1\ ]\ \tau(i,\beta,\theta) \text{ and}$$

$$B_{3,N} = \frac{N^{-1}}{\sigma E|\epsilon|}\ \Sigma_{i=1}^{N}\ g^{-1}(z_i,\beta,\theta)\ \tau(i,\beta,\theta)\ f_\beta(x_i,\beta)^t E\ [\mathrm{sign}(\epsilon_i)].$$

Thus, if the errors are symmetrically distributed and either g does not depend on $\beta$ or $\sigma \to 0$, $\hat{\theta}$ is asymptotically normally distributed with mean $\theta$ and covariance matrix

$$(2.11) \qquad \{\delta/(1-\delta)\}\ \{N\ \varsigma(\beta,\theta)\}^{-1},$$

where $\delta = \mathrm{var}\ |\epsilon|$.

The influence function for this estimator is linear in the absolute errors, indicating less lack of robustness than methods based on squared residuals. Note that if the errors are not symmetric, even if g does not depend on $\beta$ there will be an additional effect due to estimating $\beta$ not shared by the estimators based on squared residuals.

By an argument similar to that at the end of Section 2.5.1, we may conclude that when the effect of $\hat{\beta}_*$ is negligible one should use a weighted estimator and iterate the method.

## 2.6 Proofs of major results

We now present sketches of the proofs of Theorems 2.1, 2.2 and 2.3. Our exposition is brief and nonrigorous as our goal is to provide general insights. In what follows, we assume that $\hat{\beta}_*$ satisfies 2.5 for all $\sigma$ and that

$$(2.12) \qquad N^{1/2} \begin{bmatrix} \hat{\eta} - \eta \\ \\ \hat{\theta} - \theta \end{bmatrix} = \mathcal{O}_p(1);$$

under sufficient regularity conditions it is possible to prove (2.12). Such a proof would be long, detailed and essentially noninformative; see Carroll and Ruppert (1982a) for a proof of $N^{1/2}$ consistency in a special case.

Sketch of proof of Theorem 2.1: From (2.1), a Taylor series, the fact that $E\ [\ H_1\{d_i(\beta)\}\ ] = H_{2,i}$ and laws of large numbers, we have

$$(2.13) \qquad 0 = N^{-1/2}\Sigma_{i=1}^{N}\ (H_{4,i}/H_{3,i})[H_1\{d_i(\hat{\beta}_*)\} - H_{2,i}(\hat{\eta},\hat{\theta},\hat{\beta}_*)] + o_p(1)$$

By the arguments of Ruppert and Carroll (1980) or Carroll and Ruppert (1982a),

$$(2.14)\quad N^{-1/2}\Sigma_{i=1}^{N}\ (H_{4,i}/H_{3,i})[H_1\{d_i(\hat{\beta}_*)\} - H_1\{d_i(\beta)\}]$$

$$= N^{-1/2}\Sigma_{i=1}^{N}\ (H_{4,i}/H_{3,i})\ \dot{H}_1\{d_i(\beta)\}\{d_i(\hat{\beta}_*) - d_i(\beta)\} + o_p(1)$$

$$= B_{3,N}\ N^{1/2}(\hat{\beta}_* - \beta) + o_p(1).$$

Applying this result to (2.13) along with a Taylor series in $H_{2,i}$ gives

$$0 = N^{-1/2}\Sigma_{i=1}^{N}\ C_i + (B_{2,N} + B_{3,N})\ N^{1/2}(\hat{\beta}_* - \beta)$$

$$- B_{1,N}\ N^{1/2}\begin{bmatrix} \hat{\eta} - \eta \\ \hat{\theta} - \theta \end{bmatrix} + o_p(1),$$

which is (2.2). □

The result of Theorem 2.2 follows by a similar argument; in this case the representation (2.14) is unnecessary.

<u>Sketch of proof of Theorem 2.3</u>: We consider Theorem 2.2; the proof for Theorem 2.1 is similar. Recall here that (1.8) holds. In the following, all derivatives are with respect to the mean $\mu_i$ and the definitions of $C_i$ and $H_{2,i}$ are as in (2.4).

Assumption (iv) implies that $N^{1/2} \max_{1\le i\le N} |\overline{Y}_{i.} - \mu_i|^2 \xrightarrow{p} 0$ so that a Taylor series in $\eta$, $\theta$ and $\overline{Y}_{i.}$ gives

$$(2.15) \qquad B_{1,N}\, N^{1/2}\begin{bmatrix} \hat{\eta} - \eta \\ \hat{\theta} - \theta \end{bmatrix}$$

$$= N^{-1/2}\Sigma_{i=1}^{N}\, C_i \; - \; N^{-1/2}\Sigma_{i=1}^{N}\, (\dot{H}_{2,i}H_{4,i}/H_{3,i})(\overline{Y}_{i\cdot} - \mu_i)$$

$$+ \; N^{-1/2}\Sigma_{i=1}^{N}\, \{(\dot{H}_{4,i}/H_{3,i}) - (\dot{H}_{3,i}/H_{3,i})\}(\overline{Y}_{i\cdot} - \mu_i)$$

$$+ \; o_p(1).$$

Since $\overline{Y}_{i\cdot} - \mu_i = \sigma\, g(\mu_i, z_i, \theta)\, \overline{\epsilon}_{i\cdot} \approx \lambda N^{-1/2} g(\mu_i, z_i, \theta)\, \overline{\epsilon}_{i\cdot}$, where $\overline{\epsilon}_{i\cdot}$ is the mean of the errors at $x_i$, we can write the last two terms on the right-hand side of (2.15) as

$$(2.16) \qquad \lambda N^{-1}\Sigma_{i=1}^{N}\, \overline{\epsilon}_{i\cdot}(q_{i,1} + q_{i,2}C_i)$$

for constants $\{q_{i,j}\}$. Since $\overline{\epsilon}_{i\cdot}$ has mean zero, (2.16) converges in probability to zero if $E(\overline{\epsilon}_{i\cdot}C_i) = 0$, which holds under the assumption of symmetry. If $\lambda = 0$, then (2.16) is trivially equal to zero. Thus, (2.16) converges to zero which from (2.15) completes the proof. Note that if we drop the assumption of symmetry and if $\lambda \neq 0$, (2.16) implies that the asymptotic normal distribution of $N^{1/2}(\hat{\theta} - \theta)$ will have mean $\text{p-lim}_{N\to\infty} \{ \lambda\, B_{1,N}^{-1}\, N^{-1}\Sigma_{i=1}^{N}\, (\overline{\epsilon}_{i\cdot}C_i q_{i,2})\}$. □

# CHAPTER III

# OTHER ESTIMATORS

## 3.0 Introduction

In this chapter we investigate the properties of some of the estimators for $\theta$ described in Section 1.3.2. While the maximum likelihood estimator fits into the framework of Section 2.1, we treat it separately here. The extended quasi-likelihood estimator and the modified maximum likelihood estimator of Raab are not explicitly included in the formulation of Chapter 2; the properties of these estimators are derived in this chapter.

We show in this chapter that under certain conditions the maximum likelihood and extended quasi-likelihood estimators are asymptotically equivalent to the estimators of Section 2.5.1. We also verify that under certain conditions the Raab estimator is asymptotically equivalent to that of Sadler and Smith as described in Section 2.3. Further results and discussion of all the estimators are given in Chapter 4. Assume throughout that the $\{\epsilon_i\}$ are independent.

## 3.1 Maximum likelihood

The maximum likelihood estimator for $\theta$ is obtained by solving (1.5) for $\sigma$ and $\theta$, where $\hat{\beta}_*$ is the maximum likelihood estimator of $\beta$. Thus, we can use the result of Theorem 2.1 to obtain an asymptotic expansion for the maximum likelihood estimator. To do this, we derive an expansion for $N^{1/2}(\hat{\beta}_* - \beta)$ in terms of $N^{1/2}(\hat{\eta} - \eta\ ,\ \hat{\theta} - \theta)^t$ and substitute the result into (2.2) when $B_{1,N}$, $B_{2,N}$, $B_{3,N}$ and $C_i$ are as in Section 2.5.1. Write $\eta = \log \sigma$, let $B_{1,N}$, $B_{2,N}$, $B_{3,N}$ and $C_i$ be defined as in Section 2.5.1, and let the generics $(\hat{\beta},\hat{\theta},\hat{\eta})$ denote the maximum likelihood estimators of $(\beta,\theta,\eta)$. Define

$$\rho(i,\beta,\theta) = f_\beta(x_i,\beta)\ /\ g(z_i,\beta,\theta);$$

$$G_{1,N} = N^{-1}\ \Sigma_{i=1}^{N}\ \rho(i,\beta,\theta)\ \rho(i,\beta,\theta)^t;$$

$$G_{2,N} = N^{-1}\ \Sigma_{i=1}^{N}\ \nu_\beta(i,\beta,\theta)\ \nu_\beta(i,\beta,\theta)^t \text{ and}$$

$$G_N = G_{1,N} + 2\sigma^2\ G_{2,N}.$$

From (1.5), the maximum likelihood estimator of $(\beta,\theta,\eta)$ solves

$$(3.1) \qquad 0 = \begin{bmatrix} M_{1,N} \\ M_{2,N} \end{bmatrix},$$

where

$$M_{1,N} = N^{-1/2} \Sigma_{i=1}^{N} \left[ \frac{r_i \rho(i,\hat\beta,\hat\theta)}{\exp(2\hat\eta)\, g(z_i,\hat\beta,\hat\theta)} + \left\{ \frac{r_i^2}{\exp(2\hat\eta)\, g^2(z_i,\hat\beta,\hat\theta)} - 1\right\} \nu_\beta(i,\hat\beta,\hat\theta) \right]$$

and

$$M_{2,N} = N^{-1/2} \Sigma_{i=1}^{N} \left\{ \frac{r_i^2}{\exp(2\hat\eta)\, g^2(z_i,\hat\beta,\hat\theta)} - 1\right\} \tau(i,\hat\beta,\hat\theta).$$

Under regularity conditions, the maximum likelihood estimator is consistent, thus, by a Taylor series in (3.1) and laws of large numbers, we obtain by tedious calculations that

$$\begin{bmatrix} \frac{2}{\sigma^2} G_N & -B_{2,N}^t \\ -B_{2,N} & B_{1,N} \end{bmatrix} N^{1/2} \begin{bmatrix} \hat\beta - \beta \\ \hat\eta - \eta \\ \hat\theta - \theta \end{bmatrix}$$

(3.2)

$$= \begin{bmatrix} 2N^{-1/2} \Sigma_{i=1}^{N} \{ \epsilon_i \rho(i,\beta,\theta)/\sigma + (\epsilon_i^2 - 1)\nu_\beta(i,\beta,\theta) \} \\ N^{-1/2} \Sigma_{i=1}^{N} C_i \end{bmatrix} + o_p(1).$$

The second equation in (3.2) is simply (2.2) for weighted squared residual estimators, as it should be. The first equation in (3.2) yields, upon rearrangement,

$$N^{1/2}(\hat{\beta} - \beta) = \frac{\sigma^2}{2} G_N^{-1} B_{2,N}^t N^{1/2} \begin{bmatrix} \hat{\eta} - \eta \\ \hat{\theta} - \theta \end{bmatrix}$$

(3.3)

$$+ \sigma^2 G_N^{-1} N^{-1/2} \Sigma_{i=1}^N \{\epsilon_i \rho(i,\beta,\theta)/\sigma + (\epsilon_i^2 - 1)\nu_\beta(i,\beta,\theta)\} + o_p(1).$$

(3.3) affirms our earlier assertion that (2.5) holds for the maximum likelihood estimator of $\beta$. Substituting into (2.2) gives the following result.

<u>Theorem 3.1</u>. Under regularity conditions, as $N \to \infty$, the maximum likelihood estimator admits the expansion

$$( B_{1,N} - \frac{\sigma^2}{2} B_{2,N} G_N^{-1} B_{2,N}^t ) N^{1/2} \begin{bmatrix} \hat{\eta} - \eta \\ \hat{\theta} - \theta \end{bmatrix}$$

(3.4)

$$= 2 N^{-1/2} \Sigma_{i=1}^N (\epsilon_i^2 - 1) \{ \tau(i,\beta,\theta) + \frac{\sigma^2}{2} B_{2,N} G_N^{-1} \nu_\beta(i,\beta,\theta) \}$$

$$+ \sigma N^{-1/2} \Sigma_{i=1}^N \epsilon_i B_{2,N} G_N^{-1} \rho(i,\beta,\theta) + o_p(1). \qquad \square$$

To compare (3.4) with the result for the methods of Section 2.5.1, we state explicitly the expansion for these methods. To do so we need an expression for $N^{1/2}(\hat{\beta}_* - \beta)$, where here again $\hat{\beta}_*$ is a preliminary estimator for $\beta$. The discussion at the end of Section 2.1 suggests that we consider as $\hat{\beta}_*$ a generalized least squares estimator of $\beta$ as in Theorem A of Chapter 1. Then (2.3) becomes

$$(3.5) \qquad N^{1/2}(\hat{\beta}_* - \beta) = \sigma N^{-1/2} \Sigma_{i=1}^N v_{N,i} \epsilon_i + o_p(1),$$

where $v_{N,i} = G_{1,N}^{-1}\ \rho(i,\beta,\theta)$. (3.5) shows how (2.5) holds for the generalized least squares estimator of $\beta$. If (3.5) holds, we obtain the following expansion for weighted squared residuals methods:

$$
\begin{aligned}
&B_{1,N}\ N^{1/2}\begin{bmatrix} \hat{\eta} - \eta \\ \hat{\theta} - \theta \end{bmatrix} \\
(3.6)\qquad &= 2\ N^{-1/2}\ \Sigma_{i=1}^{N}\ (\epsilon_i^2 - 1)\ \tau(i,\beta,\theta) \\
&+ \sigma\ N^{-1/2}\ \Sigma_{i=1}^{N}\ \epsilon_i\ B_{2,N}\ G_{1,N}^{-1}\ \rho(i,\beta,\theta)\ +\ o_p(1).
\end{aligned}
$$

(3.4) and (3.6) show that the maximum likelihood estimator and the estimators of Section 2.5.1 are in general different asymptotically. Thus, if normality obtains and the model is correct, the latter (as well as any other estimators in Chapter 2) are not efficient. In many situations, however, the following result is relevant.

<u>Corollary 3.1(a)</u>. Suppose that either the variance function does not depend on $\beta$, so that $G_N = G_{1,N}$ and $B_{2,N} = 0$, or that $\sigma \to 0$ so that $G_N \to G_{1,N}$. Then the maximum likelihood estimator of $\theta$ and the estimators of Section 2.5.1 are asymptotically equivalent. □

The corollary is a reaffirmation of the implication of Corollary 2.1(a) and what we would expect from the discussion of Sections 2.4 and 2.5.1 when $\sigma \to 0$. Note that no symmetry is required here. The result implies that in many situations of practical importance the computationally simpler weighted squared residual type estimators equal the asymptotic performance of the maximum likelihood estimator and so when the data are approximately normally distributed, the weighted squared residual methods are asymptotically efficient. This result has special relevance to the analysis of assay data to which the small $\sigma$ asymptotics apply.

A further comparison of (3.4) and (3.6) yields the following curious result. From (3.6), we see that the asymptotic covariance matrix of the estimators of Section 2.5.1 when $\hat{\beta}_*$ is the generalized least squares estimator and asymptotic normality obtains increases without bound in $\sigma$. In (3.4), however, note that

$$\lim_{\sigma \to \infty} \sigma^2 G_N^{-1} = (2\, G_{2,N})^{-1}.$$

This shows that the asymptotic covariance of the maximum likelihood estimator of $\theta$ will stay bounded for all $\sigma$. While this observation may have implications favoring maximum likelihood in some situations, the result of Corollary 3.1(a) is of more practical relevance in application. Further, the difficulties associated with maximum likelihood discussed in Section 1.3.2 make preference for this estimator tenuous.

## 3.2 Extended quasi-likelihood

Recall for the extended quasi-likelihood estimator we assume that (1.8) holds; we write $g(\mu_i, z_i, \theta)$ and $\mu_i = f(x_i, \beta)$ to emphasize this fact. Throughout this section, let $B_{1,N}$ and $C_i$ be as in Section 2.5.1, let $\eta = \log \sigma$ and let the generics $(\hat\beta, \hat\theta, \hat\eta)$ represent the joint extended quasi-likelihood estimators for $(\beta, \theta, \eta)$, the true values. Write

$$(3.7) \qquad H_\theta(y, \mu, z) = \int_y^\mu \frac{y - u}{g^2(u, z, \theta)} \, du .$$

From (1.9), the extended quasi-likelihood estimator for $(\beta, \theta, \eta)$ solves in $\hat\beta$, $\hat\theta$, and $\hat\eta$

$$(3.8) \qquad 0 = \begin{bmatrix} Q_{1,N}(\hat\beta, \hat\theta, \hat\eta) \\ Q_{2,N}(\hat\beta, \hat\theta, \hat\eta) \\ Q_{3,N}(\hat\beta, \hat\theta, \hat\eta) \end{bmatrix},$$

where

$$Q_{1,N}(\beta, \theta, \eta) = N^{-1/2} \Sigma_{i=1}^N (Y_i - \mu_i) \, \rho(i, \beta, \theta) \, / \, g(\mu_i, z_i, \theta);$$

$$(3.9) \qquad Q_{2,N}(\beta, \theta, \eta) = N^{-1/2} \Sigma_{i=1}^N \{-2 \, e^{-2\eta} \, H_\theta(Y_i, \mu_i, z_i) - 1\}; \text{ and}$$

$$Q_{3,N}(\beta, \theta, \eta) = N^{-1/2} \Sigma_{i=1}^N \, [ \, e^{-2\eta} \, \partial/\partial\theta \, H_\theta(Y_i, \mu_i, z_i)$$

$$- \{\partial/\partial\theta \log g(Y_i, z_i, \theta)\} \, ].$$

The fact that the extended quasi-likelihood estimator is consistent as $N \to \infty$, $\sigma \to 0$ follows from the result below.

Lemma 3.2. Under regularity conditions,

$$(3.10)\qquad N^{-1}\,\Sigma_{i=1}^{N}\,H_\theta(Y_i,\mu_i,z_i) = -\frac{\sigma^2}{2}\,N^{-1}\Sigma_{i=1}^{N}\,\epsilon_i^2$$

$$+\,\sigma^3\,N^{-1}\Sigma_{i=1}^{N}\,s_{1,i}\,\epsilon_i^3 + \mathcal{O}_p(\sigma^4);$$

$$(3.11)\qquad N^{-1}\,\Sigma_{i=1}^{N}\,\partial/\partial\theta\;H_\theta(Y_i,\mu_i,z_i) = \sigma^2\,N^{-1}\Sigma_{i=1}^{N}\,\epsilon_i^2\nu_\theta(i,\beta,\theta)$$

$$+\,\sigma^3\,N^{-1}\Sigma_{i=1}^{N}\,s_{2,i}\,\epsilon_i^3 + \mathcal{O}_p(\sigma^4);$$

$$(3.12)\qquad N^{-1}\,\Sigma_{i=1}^{N}\,\partial^2/\partial\theta^2\;H_\theta(Y_i,\mu_i,z_i)$$

$$= -\,\sigma^2\,N^{-1}\Sigma_{i=1}^{N}\,\epsilon_i^2\,[\,3\,\nu_\theta(i,\beta,\theta)\,\nu_\theta^t(i,\beta,\theta)$$

$$-\,\{g_{\theta\theta}(\mu_i,z_i,\theta)\,/\,g(\mu_i,z_i,\theta)\}\,]$$

$$+\,\sigma^3\,N^{-1}\Sigma_{i=1}^{N}\,s_{3,i}\,\epsilon_i^3 + \mathcal{O}_p(\sigma^4),$$

for constants $\{s_{j,i}\}$, $j = 1, 2, 3$, where

$$g_{\theta\theta}(\mu_i,z_i,\theta) = \partial^2/\partial\theta\;\partial\theta^t\;\{g(\mu_i,z_i,\theta)\}.$$

Sketch of proof: The proofs of (3.10) - (3.12) follow from Taylor series in $\sigma$ about 0. Write

$$\epsilon = (y - \mu)\,/\,\{\sigma\,g(\mu,z,\theta)\},$$

and note that

$$(3.13) \qquad H_\gamma(y,\mu,z) = -\int_{\mu}^{\mu + \sigma g(\mu,z,\theta)\epsilon} (y - u) / g^2(u,z,\gamma)\, du.$$

Here, $\theta$ represents the true value of the parameter. For (3.11) and (3.12), assuming that g is sufficiently regular so that the interchange of differentiation and integration is legitimate, we obtain

$$(3.14) \qquad \partial/\partial\gamma\ H_\gamma(y,\mu,z)$$

$$= 2 \int_{\mu}^{\mu + \sigma g(\mu,z,\theta)\epsilon} (y - u)\ g_\gamma(u,z,\theta)/g^3(u,z,\theta)\, du$$

$$(3.15) \qquad \partial^2/\partial\gamma^2\ H_\gamma(y,\mu,z)$$

$$= 2 \int_{\mu}^{\mu + \sigma g(\mu,z,\theta)\epsilon} (y - u)[\ g_{\gamma\gamma}(u,z,\gamma)/g^3(u,z,\gamma)$$

$$- 3\ g_\gamma(u,z,\gamma)\ g_\gamma^t(u,z,\gamma)/g^4(u,z,\gamma)\ ]\, du.$$

For illustration, we show the form of the expansion for (3.10); the others are similar. A Taylor series in $\sigma$ about 0 in (3.13) gives, upon simplification and under regularity conditions that

$$H_\gamma(y,\mu,z) = -\sigma\ (y-u)\ g(\mu,z,\theta)\ \epsilon\ /\ g^2(\mu,z,\gamma)$$

$$+ \frac{\sigma^2}{2}\{ \frac{g^2(\mu,z,\gamma) + 2\ g'(\mu,z,\gamma)\ (y-\mu)}{g^4(\mu,z,\gamma)} \}\ g^2(\mu,z,\theta)\ \epsilon^2$$

$$+ \frac{\sigma^3}{3}[ \frac{g^2(\mu,z,\gamma)\ g'(\mu,z,\gamma) - g(\mu,z,\gamma)\ g''(\mu,z,\gamma)}{g^5(\mu,z,\gamma)}$$

$$+ \frac{4\ \{g'(\mu,z.\gamma)\}^2\ (y-u)}{g^5(\mu,z,\gamma)}]\ g^3(\mu,z,\theta)\ \epsilon^3$$

$$+ \mathcal{O}(\sigma^4),$$

where $g'(\mu,z,\theta)$ and $g''(\mu,z,\theta)$ are the first and second derivatives of g with respect to its first argument. Thus, under regularity conditions, for some constant $s_1$,

$$H_\theta(y,\mu,z) = -\frac{1}{2}\epsilon^2\sigma^2 + s_1\ \epsilon^3\sigma^3 + \mathcal{O}(\sigma^4).$$

Applying the above calculations to the left-hand side of (3.10) yields under regularity conditions the result. □

From (3.10) and (3.11) of Lemma 3.2, it is easily seen that with sufficient smoothness conditions on g, the estimating equations (3.8) are unbiased as $N \to \infty$, $\sigma \to 0$, i.e., as $N \to \infty$, $\sigma \to 0$,

(3.16) $$N^{-1/2}\ Q_{j,N}(\beta,\theta,\eta) \xrightarrow{p} 0,\ j = 1,\ 2,\ 3,$$

implying consistency of the solution. The theory of M-estimation (see, for example, Huber (1981) or Serfling (1980)) implies that, under sufficient regularity conditions, if the solution to (3.8) is consistent in general for all fixed $\sigma$, then (3.16) should hold as $N \to \infty$, $\sigma$ fixed. This is clearly not apparent from (3.9). A simple heuristic argument to see that the extended quasi-likelihood estimator of $\theta$ need not be consistent in general can be pursued as follows. For simplicity, assume $\beta$ is known so that we need only consider the second two equations of (3.8). These equations imply that we solve in $\theta$

$$K_N = N^{-1} \Sigma_{i=1}^{N} \partial/\partial\theta \; H_\theta(Y_i,\mu_i,z_i)$$

$$+ \{2\, N^{-1} \Sigma_{i=1}^{N} H_\theta(Y_i,\mu_i,z_i)\}\{N^{-1} \Sigma_{i=1}^{N} \partial/\partial\theta \; \log g(Y_i,z_i,\theta)\}.$$

If the solution $\hat{\theta}$ is consistent, then for all fixed $\sigma$ we should have that as $N \to \infty$

$$K_N \xrightarrow{p} 0. \tag{3.17}$$

Lemma 3.2 implies that as $N \to \infty$, $\sigma \to 0$

$$\frac{K_N}{\sigma^2} \xrightarrow{p} 0.$$

If (3.17) holds for all $\sigma$, then we should also have as $N \to \infty$, $\sigma \to 0$ that

$$\frac{K_N}{\sigma^4} \xrightarrow{p} 0. \tag{3.18}$$

It can be shown by detailed calculations that (3.18) does not hold in general. These calculations involve expansions of (3.10) and (3.11) to terms of order $\sigma^4$.

The implication of the above discussion is that the extended quasi-likelihood method may not even admit in general a consistent estimator for $\theta$, but in situations where the small $\sigma$ asymptotics are relevant the method does provide consistent estimates of the parameters. We now show that in the small $\sigma$ asymptotics, the extended quasi-likelihood estimator of $\theta$ is in fact asymptotically equivalent to the maximum likelihood estimator and the estimators of Section 2.5.1.

<u>Theorem 3.3</u>. Under regularity conditions, if, as $N \to \infty$ and $\sigma \to 0$, $N^{1/2}\sigma \to \lambda \geq 0$, and either $\lambda = 0$ or the $\{\epsilon_i\}$ are symmetrically distributed, then the extended quasi-likelihood estimator admits the expansion

$$B_{1,N}\, N^{1/2} \begin{bmatrix} \hat{\eta} - \eta \\ \hat{\theta} - \theta \end{bmatrix} = 2\, N^{-1/2}\, \Sigma_{i=1}^{N}\, (\epsilon_i^2 - 1)\, \tau(i,\beta,\theta) + o_p(1).$$

<u>Proof</u>: A Taylor series in (3.8) using consistency, a Taylor series in $\sigma$ about 0 using Lemma 3.2 and laws of large numbers yield after simplification

$$\begin{bmatrix} G_{1,N} & 0 \\ 0 & \frac{1}{2} B_{1,N} \end{bmatrix} N^{1/2} \begin{bmatrix} \hat{\beta} - \beta \\ \hat{\eta} - \eta \\ \hat{\theta} - \theta \end{bmatrix}$$

(3.19)

$$= \begin{bmatrix} \sigma\, N^{-1/2}\, \Sigma_{i=1}^{N}\, \epsilon_i\, \rho(1,\beta,\theta) \\ N^{-1/2}\, \Sigma_{i=1}^{N}\, (\epsilon_i^2 - 1)\, \tau(i,\beta,\theta) \end{bmatrix} + \mathcal{O}_p(N^{1/2}\sigma).$$

From Lemma 3.2, (3.10) and (3.11), the $\mathcal{O}_p(N^{1/2}\sigma)$ term will be of the form

$$(3.20) \qquad (N^{1/2}\sigma)\, N^{-1}\Sigma_{i=1}^{N}\, s_i\, \epsilon_i^3$$

for some constants $\{s_i\}$, thus, if the $\{\epsilon_i\}$ are symmetric, we only require $N^{1/2}\sigma = \mathcal{O}(1)$ in order that the remainder term be $o_p(1)$, while if the $\{\epsilon_i\}$ are not necessarily symmetric we require $N^{1/2}\sigma \to 0$. (3.19) implies that

$$N^{1/2}\, (\hat{\beta} - \beta) = \sigma\, N^{-1/2}\, \Sigma_{i=1}^{N}\, \epsilon_i\, G_{1,N}^{-1}\, \rho(i,\beta,\theta) + o_p(1),$$

which shows upon comparison with (3.5) that the extended quasi-likelihood estimator of $\beta$ is asymptotically equivalent to a generalized least squares estimator of $\beta$ as in Theorem A of Chapter 1 when $\sigma \to 0$. (3.19) also implies that, as $N \to \infty$, $\sigma \to 0$,

$$B_{1,N}\, N^{1/2} \begin{bmatrix} \hat{\eta} - \eta \\ \hat{\theta} - \theta \end{bmatrix} = N^{-1/2} \Sigma_{i=1}^{N} C_i + o_p(1),$$

which from (3.4) and (3.6) is the expansion for the maximum likelihood estimator and the estimators of Section 2.5.1 when $\sigma \rightarrow 0$. □

The asymptotic results presented so far suggest that in situations where small $\sigma$ asymptotics are relevant, the extended quasi-likelihood estimator is a reasonable competitor to maximum likelihood and the weighted squared residual methods in terms of asymptotic performance. Note that technically we need the additional requirement on the rates at which $N \rightarrow \infty$ and $\sigma \rightarrow 0$ for this result to hold. While we see that the estimator may not be consistent when these asymptotics do not apply, the seriousness of this result is unclear from our discussion. We now investigate the character of the inconsistency by means of several examples.

For simplicity, assume that the $\{\mu_i\}$ (and hence $\beta$) are known so that our focus is restricted to estimation of $\theta$ and $\sigma$. For definiteness, assume also the power of the mean model

$$(3.20) \qquad g(\mu_i, z_i, \theta) = \mu_i^{\theta}.$$

Note in (1.9) that for (3.20), $g(Y_i, z_i, \theta) = 0$ when $Y_i = 0$ so that $Q^+$ becomes infinite. Nelder and Pregibon (1986) suggest replacing $g(Y,z,\theta)$ by $g(Y+c,z,\theta)$ with $c = 1/6$ in (1.9); this suggestion arises from the fact that when $Q^+$ is an approximation to a discrete distribution such as the Poisson, the problem lies with the Stirling approximation used for the factorials, and the factor $c = 1/6$

represents the proper correction for the first term in the Stirling series. In the investigation described below we use this correction.

Since we assume $\beta$ to be known, we consider only solution of the second two equations of (3.8) and write $Q_{j,N}(\beta,\theta,\eta) = Q_{j,N}(\theta,\sigma)$, j = 2, 3, to emphasize this fact and the fact that we are estimating $\sigma$. Briefly, the theory of M-estimation as in Huber (1981, p. 130-132) implies that under regularity conditions, if $\hat{\theta}$ and $\hat{\sigma}$ are such that

$$N^{1/2} \begin{bmatrix} Q_{2,N}(\hat{\theta},\hat{\sigma}) \\ \\ Q_{3,N}(\hat{\theta},\hat{\sigma}) \end{bmatrix} \xrightarrow{p} 0,$$

and if

$$\lambda(\theta,\sigma) = E \begin{bmatrix} Q_{2,N}(\theta,\sigma) \\ \\ Q_{3,N}(\theta,\sigma) \end{bmatrix}$$

exists for all $\theta$ and $\sigma$ and has a unique 0 at some $(\theta_*,\sigma_*)$, then

$$\hat{\theta} \xrightarrow{p} \theta_* \quad \text{and} \quad \hat{\sigma} \xrightarrow{p} \sigma_*.$$

It is easily seen that under (3.20), $\lambda(\theta,\sigma)$ exists and is finite only if expected values of the form $E\ \{\ Y^{\alpha}\ (\log Y)^{k}\ \}$ exist for various values of $\alpha$ and k; we thus restrict our investigation to distributions for Y and values for $\theta$ and $\sigma$ for which this requirement is met and consider the correction described in the preceding paragraph. This requirement precludes consideration of strictly normal data, for example.

The above theory can be used to determine $\theta_*$; we cite three examples when the $\{\mu_i\}$ take on a finite number of values in equal proportions:

(i) $Y_i$ distributed as Poisson with mean $\mu_i$. In this case, $\theta = 0.5$ and $\sigma = 1$. However, if the $\{\mu_i\}$ take on the values 1 and 4 in equal proportions, then $\theta_* = 0.675$; if the $\{\mu_i\}$ take on the values 1 and 5, then $\theta_* = 0.640$. In these cases, $\sigma$ is large relative to the means. If the $\{\mu_i\}$ take on larger values, such as 30, 40 and 50 or 50, 75 and 100 in equal proportions, however, then $\theta_*$ is very nearly equal to 0.5. This example shows that there are cases in which the extended quasi-likelihood estimator can be badly inconsistent, but also invites the following conjecture. It appears that as the $\{\mu_i\}$ increase, $\hat{\theta}$ is consistent. The Poisson distribution is one for which $Q^+$ is an approximation to the actual distributional likelihood, and for large $\mu_i$ appears to be behaving very nearly like the normal likelihood. This leads one to suspect that the extended quasi-likelihood estimator may have properties similar to those of the maximum likelihood and weighted squared residual estimators when the means are large. We investigate this conjecture below; we now consider examples for which $Q^+$ is not an approximation to the distributional likelihood in this way.

(ii) $Y_i = \mu_i + \sigma\, \mu_i\, \epsilon_i$, where $\epsilon_i = (X_i - \mu_i) / \mu_i^{1/2}$ and the $X_i$ are distributed as Poisson with mean $\mu_i$. In this case the theory is valid as long as $\sigma\, \mu_i^{(\theta-1/2)} \le 1$. If $\sigma = 1$ so that $\mu_i \le 1$, this setting is a perhaps slightly artificial example of a case where $\sigma$ is large relative to the means. If $\sigma$ is considered known so that only $\theta$ need be estimated, and the $\{\mu_i\}$ take on the values shown in equal proportions, we obtain the following results:

| $\theta$ | $\mu_i$ | $\theta_*$ |
|---|---|---|
| 1.0 | 0.3,0.5 | 1.194 |
| 0.8 | 0.3,0.6 | 0.992 |
| 0.8 | 0.1,0.9 | 0.952 |
| 0.6 | 0.3,0.7 | 0.825 |
| 0.6 | 0.1,0.9 | 0.845 |
| 0.55 | 0.1,0.9 | 0.815 |

This example represents a case in which the small $\sigma$ asymptotics are not valid and the extended likelihood estimator may be poor.

(iii) $Y_i = \mu_i + \sigma \mu_i^{\theta} \epsilon_i$ with $\epsilon_i = v_i/\omega^{1/2}$, where $v_i$ is truncated standard normal on $(-a,a)$ and $\omega = 1 - (2/\pi)^{1/2}\{a\exp(-a^2/2)\}/\{2\Phi(a)-1\}$. The theory is valid if $\mu_i^{(1-\theta)} \geq \sigma\ a/\omega^{1/2}$ so that the means must be relatively large compared to $\sigma$. For the following values of the $\{\mu_i\}$, a and $\theta$ with $\sigma = 1$, in virtually every case $\theta_*$ is very nearly equal to $\theta$:

| $\theta$ | $\mu_i$ | a |
|---|---|---|
| 0.2 | 5,10 | 2.5 |
| 0.5 | 10,20 | 2.5 |
| 0.5 | 50,100 | 2.5 |
| 0.5 | 10,20 | 2.0 |
| 0.5 | 50,75 | 2.0 |

It can also be shown that if the $\epsilon_i$ are centered Uniform (a,b) random variables with variance 1, similar results obtain for a variety of values for the means.

The above examples show that while the extended quasi-likelihood estimator may be inconsistent in some instances, this need not always be the case.

We now examine more closely the behavior of the extended quasi-likelihood estimator when the means are large as suggested by the discussion at the end of Example (i). More specifically, we show in the case of (3.20) for fixed $\sigma$, under regularity conditions, that as long as $\theta < 1$, when the means increase the extended quasi-likelihood estimator behaves as it does in the small $\sigma$ asymptotics, thus in some sense verifying the conjecture. For the following discussion, let $\beta$ (and hence the $\{\mu_i\}$) be known so that we focus on estimation of $\theta$ and $\sigma$. Let $\mu_{0,N}$ be the median of the $\{\mu_i\}$ and define

$$\mu_i^* = \mu_i/\mu_{0,N} \qquad \text{and} \qquad Y_i^* = Y_i/\mu_{0,N}$$

so that $\epsilon_i = (Y_i^* - \mu_i^*) \,/\, (\delta\, \mu_i^*)$, where $\delta = \sigma\, \mu_{0,N}^{(\theta-1)}$. Assume further that as $N \to \infty$,

$$\min_{1 \le i \le N} \mu_i \longrightarrow \infty \qquad \text{and} \qquad \mu_{0,N} \longrightarrow \infty$$

in such a way that as $N \to \infty$ the $\{\mu_i^*\}$ and $\{Y_i^*\}$ as well as various sums of functions of the $\{\mu_i\}$ are well-behaved in a sense that will become obvious in the calculations below. Rewrite (3.7) as

$$H_\theta(y,\mu) = \int_Y^{\mu} (y - u)/u^{2\theta}\, du.$$

Note that as long as $\theta < 1$, $\delta \to 0$ as $N \to \infty$. We now show that under the above conditions, results similar to those of Lemma 3.2 hold for $H_\theta(y,\mu)$, so that as $N \to \infty$ the asymptotic calculations parallel those for the small $\sigma$ case.

<u>Lemma 3.4</u>. Under regularity conditions, if $\theta < 1$, then,

(3.21) $$N^{-1}\Sigma_{i=1}^{N} H_\theta(Y_i,\mu_i) = -\frac{\sigma^2}{2} N^{-1}\Sigma_{i=1}^{N}\epsilon_i^2 + \delta\, N^{-1}\Sigma_{i=1}^{N} t_{1,i}\epsilon_i^3 + \mathcal{O}_p(\delta^2);$$

(3.22) $$N^{-1}\Sigma_{i=1}^{N} \partial/\partial\theta\, H_\theta(Y_i,\mu_i) = \sigma^2\, N^{-1}\Sigma_{i=1}^{N}\, \epsilon_i^2 \log \mu_i$$

$$+ \delta\, N^{-1}\Sigma_{i=1}^{N}\, t_{2,i}\, \epsilon_i^3 + \mathcal{O}_p(\delta^2);$$

(3.23) $$N^{-1}\Sigma_{i=1}^{N} \partial^2/\partial\theta^2\, H_\theta(Y_i,\mu_i) = -2\,\sigma^2\, N^{-1}\Sigma_{i=1}^{N}\, \epsilon_i^2\, (\log \mu_i)^2$$

$$+ \delta\, N^{-1}\Sigma_{i=1}^{N}\, t_{3,i}\, \epsilon_i^3 + \mathcal{O}_p(\delta^2).$$

<u>Sketch of proof</u>: Write $\mu_0$ = median of the $\{\mu_i\}$, $\mu^* = \mu/\mu_0$, $y^* = y/\mu_0$ and $\delta = \sigma\, \mu_0^{(\theta-1)}$. Let

$$\epsilon = (y - \mu)/(\sigma\, \mu^\theta) = (y^* - \mu^*)/(\delta\, \mu^{*\theta}).$$

By a change of variables, note that

$$(3.24) \qquad H_\gamma(y,\mu) = -\frac{\sigma^2}{\delta^2}\int_{\mu^*}^{\mu^* + \delta\,\mu^{*\theta}\epsilon} (y^* - w)/w^{2\gamma}\, dw;$$

$$(3.25) \qquad \partial/\partial\gamma\; H_\gamma(y,\mu) = -2\int_y^{\mu} (y - u)\,\log u/u^{2\gamma}\, du$$

$$= -\frac{2\sigma^2}{\delta^2}\int_{\mu^*}^{\mu^* + \delta\,\mu^{*\theta}\epsilon} (y^*-w)(\log z + \log \mu_0)/w^{2\gamma}\, dw;$$

$$(3.26) \qquad \partial^2/\partial\gamma^2\; H_\gamma(y,\mu) = 4\int_y^{\mu} (y - u)\,(\log u)^2/u^{2\theta}\, du$$

$$= -\frac{4\sigma^2}{\delta^2}\int_{\mu^*}^{\mu^* + \sigma\mu^{*\theta}\epsilon} (y^*-w)(\log w + \log \mu_0)^2/w^{2\gamma}\, dw.$$

The integrals on the right-hand sides of (3.24) - (3.26) are of the same form as those of (3.13) - (3.15) with $\mu^*$, $y^*$ and $\delta$ playing the roles of $\mu$, y and $\sigma$, respectively. Thus, by the same calculations used in Lemma 3.2, where the Taylor series are now in $\delta$ about 0, we obtain that for constants $\{t_j\}$, j = 1, 2, 3, depending on $\mu^*$,

$$H_\theta(y,\mu) = -\frac{\sigma^2}{2}\,\epsilon^2 + \delta\, t_1\,\epsilon^3 + \mathcal{O}(\delta^2);$$

$$\partial/\partial\theta\; H_\theta(y,\mu) = \sigma^2\,\epsilon^2\,(\log \mu^* + \log \mu_0) + \delta\, t_2\,\epsilon^3 + \mathcal{O}(\delta^2)$$

$$= \sigma^2\,\epsilon^2 \log \mu + \delta\, t_2\,\epsilon^3 + \mathcal{O}(\delta^2);$$

$$\partial^2/\partial\theta^2 \; H_\theta(y,\mu) = -2\,\sigma^2\,(\log \mu^* + \log \mu_0)^2 + \delta\, t_3\, \epsilon^3 + \mathcal{O}(\delta^2)$$

$$= -2\,\sigma^2\,\epsilon^2\,(\log \mu)^2 + \delta\, t_3\, \epsilon^3 + \mathcal{O}(\delta^2).$$

Applying these calculations to the left-hand sides of (3.21) - (3.23) yields under regularity conditions the results. □

We may now use arguments entirely similar to those in the small $\sigma$ asymptotics to conclude that the extended quasi-likelihood estimator is consistent under these conditions and, by a Taylor series in the second two equations of (3.8) using consistency, a Taylor series in $\delta$ about 0 using Lemma 3.4 and laws of large numbers, that

$$(3.27) \qquad B_{1,N}\, N^{1/2} \begin{bmatrix} \hat{\eta} - \eta \\ \hat{\theta} - \theta \end{bmatrix} = N^{-1/2}\, \Sigma_{i=1}^{N}\, C_i + \mathcal{O}_p(N^{1/2}\delta),$$

where here $\nu_\theta(i,\beta,\theta) = \log \mu_i$ in the definitions of $B_{1,N}$ and $C_i$. As in the proof of Theorem 3.3, in order that the second term on the right-hand side be $o_p(1)$, we require that either the $\{\epsilon_i\}$ be symmetric and $N^{1/2}\delta = \sigma\{N^{1/2}\mu_{0,N}^{(\theta-1)}\} \to \lambda^* \geq 0$ or that $N^{1/2}\delta \to 0$. From the results of Sections 2.5.1 and 3.1, we see that if $\beta$ is known and the means become large in such a way that certain sums of functions of the means remain bounded, the maximum likelihood and weighted squared residual estimators will admit expansion (3.27) with remainder term which is $o_p(1)$.

The implication of the preceding discussion is that the conjecture at the end of Example (i) is valid in the sense that for "large" values of the means, the extended quasi-likelihood estimator will again behave similarly to the maximum likelihood and weighted squared residual estimators. Thus, while the formulation and motivation for this estimator may differ from those of the latter estimators, extended quasi-likelihood can be competitive under certain conditions. Computation of the extended quasi-likelihood estimator may not be as straightforward as for the weighted squared residual estimators, however, since the latter may be computed using standard software while the former requires iterative solution of a "nonstandard" problem which includes the requirement of the correction factor described previously in certain settings.

## 3.3 Modified maximum likelihood

Throughout this section, we use the notation for the case of replication as in equation (1.10). Recall that $N = Mm$ and $N \to \infty$ in such a way that m remains fixed. For modified maximum likelihood, we require that (1.8) holds; thus, we write $g(\mu_i, z_i, \theta)$ for the variance function and $\nu(\mu_i, z_i, \theta) = \log g(\mu_i, z_i, \theta)$. As before, $\nu_\theta$ represents the derivative of $\nu$ with respect to $\theta$; similarly, let $\nu_{\theta\theta} = \{\partial/\partial\theta\ \partial\theta^t\}\ \nu$ and let a superscript " ' " denote differentiation of the quantity once with respect to its first argument. Also as before, a "—" above a quantity represents its mean, e.g.,

$$\bar{\nu}_\theta(\mu, z, \theta) = N^{-1} \Sigma_{i=1}^{N} \nu_\theta(\mu_i, z_i, \theta).$$

Let $\eta$ = log $\sigma$ and write

$$\varsigma(\mu,z,\theta) = N^{-1}\Sigma_{i=1}^{N}\{\nu_\theta(\mu_i,z_i,\theta) - \overline{\nu}_\theta(\mu,z,\theta)\}\{\nu_\theta(\mu_i,z_i,\theta) - \overline{\nu}_\theta(\mu,z,\theta)\}^t.$$

It turns out that it is notationally simpler to derive an expansion for the modified maximum likelihood estimator of $\theta$ directly rather than for the estimators of $\theta$ and $\eta$ simultaneously. In Chapter 2, we note that the estimator of Sadler and Smith (1985) can be shown to be equivalent to the modified maximum likelihood estimator. From Theorem 2.3 and algebra, we can show that the Sadler and Smith estimator satisfies

$$\varsigma(\mu,z,\theta)\, N^{1/2}(\hat{\theta} - \theta)$$

(3.28)

$$= \frac{1}{2} N^{-1/2}\Sigma_{i=1}^{N} (s_i^2 - 1)\, \{\nu_\theta(\mu_i,z_i,\theta) - \overline{\nu}_\theta(\mu,z,\theta)\} + \mathcal{O}_p(N^{1/2}\sigma),$$

where the $\mathcal{O}_p(N^{1/2}\sigma)$ term is $o_p(1)$ if $N^{1/2}\sigma \to \lambda \geq 0$ and either the $\{\epsilon_{ij}\}$ are symmetric or $\lambda = 0$. We now show that the modified maximum likelihood estimator of Raab admits expansion (3.28).

For the following, let $\hat{\theta}$, $\hat{\eta}$ and $\hat{\mu}_1$, ..., $\hat{\mu}_M$ denote the joint modified maximum likelihood estimators of $\theta$, $\eta$ and $\mu_1$, ..., $\mu_M$. Define the following conditions:

(i) $N^{1/2}\sigma \to \lambda$, $0 \leq \lambda < \infty$;

(ii) The $\{\epsilon_{ij}\}$ are independent and identically distributed random variables. Either the $\{\epsilon_{ij}\}$ are symmetric or $\lambda = 0$;

(iii) Under (i) and (ii), $N^{1/2}(\hat{\theta} - \theta) = \mathcal{O}_p(1)$ and $N^{1/2}(\hat{\eta} - \eta) = \mathcal{O}_p(1)$;

(iv) The $\hat{\mu}_i$ are uniformly consistent for the $\mu_i$;

(v) For every $\gamma > 0$, $P\{ \max_{1 \le i \le N} \frac{| \hat{\mu}_i - \mu_i |^k}{\sigma^j} > \gamma \} \longrightarrow 0$ for $k = 4$, $j = 3$, and $k = 3$, $j = 2$.

Before we state the result, we comment on (iii) and (iv) above. As for the general estimator of Chapter 2, to prove (iii) would be very detailed and essentially noninformative; as we show in Chapter 4, under the assumption of (iii), the modified maximum likelihood estimator is not competitive with several other important estimators so to prove (iii) would be a laborious task resulting in little gain. While we have not verified (iv) and (v) for the modified maximum likelihood estimator itself, under reasonable conditions these assertions hold for $\overline{Y}_{i\cdot}$ replacing $\hat{\mu}_i$ and for a one-step estimator of $\mu_i$ starting from $\overline{Y}_{i\cdot}$ and the Sadler and Smith estimator of $\theta$. Thus, it is reasonable to expect that conditions such as (iv) and (v) hold for the $\{\hat{\mu}_i\}$ themselves. Since our major aim is to obtain insight mainly for comparative purposes we do not pursue these points further.

We now state the main result.

Theorem 3.5. Suppose that as $N \to \infty$, $\sigma \to 0$ simultaneously, (i) - (v) hold. Under these and further regularity conditions, the modified maximum likelihood estimator of $\theta$ satisfies

$$\varsigma(\mu,z,\theta)\ N^{1/2}(\hat{\theta} - \theta)$$

$$= \frac{1}{2} N^{-1/2} \Sigma_{i=1}^{N} (s_i^2 - 1) \{\nu_\theta(\mu_i,z_i,\theta) - \bar{\nu}_\theta(\mu,z,\theta)\} + o_p(1).$$

Sketch of proof: We present a heuristic argument and, as the algebra involved is lengthy, only briefly summarize the major steps. From (1.10), $\hat{\theta}$, $\hat{\eta}$ and $\hat{\mu}_1, \ldots, \hat{\mu}_M$ solve

$$\Sigma_{j=1}^{m} \frac{(Y_{ij} - \hat{\mu}_i)}{\exp(2\hat{\eta})\ g(\hat{\mu}_i, z_i, \hat{\theta})}$$

(3.29)

$$+ \Sigma_{j=1}^{m} \left\{ \frac{(Y_{ij} - \hat{\mu}_i)^2}{\exp(2\hat{\eta})\ g^2(\hat{\mu}_i, z_i, \hat{\theta})} - \frac{(m-1)}{m} \right\} \nu'(\hat{\mu}_i, z_i, \hat{\theta}) = 0,$$

$$i = 1, \ldots, M;$$

(3.30) $$N^{-1}\Sigma_{i=1}^{M}\Sigma_{j=1}^{m} \frac{(Y_{ij} - \hat{\mu}_i)^2}{g^2(\hat{\mu}_i, z_i, \hat{\theta})} \{ \nu_\theta(\hat{\mu}_i, z_i, \hat{\theta}) - \bar{\nu}_\theta(\hat{\mu}, z, \hat{\theta}) \} = 0.$$

By the Mean Value Theorem and (iii), (3.30) implies that

(3.31) $$H_{1,N}\ N^{1/2}(\hat{\theta} - \theta) = H_{2,N} + o_p(1),$$

where

$$(3.32)\qquad H_{1,N} = \frac{N^{-1}}{\sigma^2}\,\Sigma_{i=1}^{M}\,\Sigma_{j=1}^{m}\,\frac{(Y_{ij}-\hat{\mu}_i)^2}{g^2(\hat{\mu}_i,z_i,\hat{\theta})}\,\xi(\hat{\mu}_i,z_i,\hat{\theta}),$$

$$(3.33)\qquad H_{2,N} = \frac{N^{-1/2}}{\sigma^2}\Sigma_{i=1}^{M}\Sigma_{j=1}^{m}\,\frac{(Y_{ij}-\hat{\mu}_i)^2}{g^2(\hat{\mu}_i,z_i,\hat{\theta})}\,\{\,\nu_\theta(\hat{\mu}_i,z_i,\hat{\theta}) - \bar{\nu}_\theta(\hat{\mu},z,\hat{\theta})\,\}$$

and

$$\xi(\mu_i,z_i,\theta) = 2\,\{\,\nu_\theta(\mu_i,z_i,\theta) - \bar{\nu}(\mu,z,\theta)\,\}\,\nu_\theta^t(\mu_i,z_i,\theta).$$

$$-\,\{\,\nu_{\theta\theta}(\mu_i,z_i,\theta) - \bar{\nu}_{\theta\theta}(\mu,z,\theta)\,\}.$$

A Taylor series of (3.32) in $\hat{\mu}_i$ about $\mu_i$ using (iv) and (v) yields after much tedious algebra that

$$\begin{aligned} H_{1,N} &= N^{-1}\,\Sigma_{i=1}^{M}\Sigma_{j=1}^{m}\,\epsilon_{ij}^2\,\xi(\mu_i,z_i,\theta) \\ &+ N^{-1}\Sigma_{i=1}^{M}\Sigma_{j=1}^{m}\,[\,-2\,\epsilon_{ij}\,\xi(\mu_i,z_i,\theta)/\{\sigma\, g(\mu_i,z_i,\theta)\} \\ &\qquad + h_{1,i}\,\epsilon_{ij}^2\,](\hat{\mu}_i - \mu_i) \\ (3.34)\quad & \\ &+ N^{-1}\Sigma_{i=1}^{M}\Sigma_{j=1}^{m}\,[\,2\,\xi(\mu_i,z_i,\theta)\,/\,\{\sigma^2\, g^2(\mu_i,z_i,\theta)\} \\ &\qquad + h_{2,i}\epsilon_{ij}/\sigma + h_{3,i}\epsilon_{ij}^2\,]\,(\hat{\mu}_i - \mu_i)^2 + o_p(1) \end{aligned}$$

for constants $\{h_{k,i}\}$, $k = 1, 2, 3$. Similarly, after much simplification a Taylor series of (3.33) yields

$$H_{2,N} = N^{-1/2}\Sigma_{i=1}^{M} \Sigma_{j=1}^{m} \epsilon_{ij}^{2} \{ \nu_\theta(\mu_i, z_i, \theta) - \overline{\nu}_\theta(\mu, z, \theta) \}$$

$$+ N^{-1/2}\Sigma_{i=1}^{M}\Sigma_{j=1}^{m}[-2\epsilon_{ij}\{\nu_\theta(\mu_i, z_i, \theta) - \overline{\nu}_\theta(\mu, z, \theta)\}/\{\sigma\ g(\mu_i, z_i, \theta)\}$$

$$+ h_{1,i}^{*} \epsilon_{ij}^{2} ]\ (\hat{\mu}_i - \mu_i)$$

(3.35)

$$+ N^{-1/2}\Sigma_{i=1}^{M}\Sigma_{j=1}^{m}[\ 2\ \{\nu_\theta(\mu_i, z_i, \theta) - \overline{\nu}_\theta(\mu, z, \theta)\}/\{\sigma^2 g^2(\mu_i, z_i, \theta)\}$$

$$+ h_{2,i}^{*} \epsilon_{ij}/\sigma + h_{3,i}^{*} \epsilon_{ij}^{2} ]\ (\hat{\mu}_i - \mu_i)^2$$

$$+ N^{-1/2}\ \Sigma_{i=1}^{M}\Sigma_{j=1}^{m}\ [-2\{\nu_\theta(\mu_i, z_i, \theta) - \overline{\nu}_\theta(\mu, z, \theta)\}/\{\sigma^2 g^2(\mu_i, z_i, \theta)\}$$

$$+ \{\nu_\theta'(\mu_i, z_i, \theta) - \overline{\nu}_\theta'(\mu, z, \theta)\}/\{\sigma^2 g^2(\mu_i, z_i, \theta)\}$$

$$+ h_{4,i}^{*} \epsilon_{ij}/\sigma + h_{5,i}^{*} \epsilon_{ij}^{2} ]\ (\hat{\mu}_i - \mu_i)^3 + o_p(1)$$

for constants $\{h_{k,i}^{*}\}$, $k = 1, \ldots, 5$.

To obtain an expression for $(\hat{\mu}_i - \mu_i)$ we focus on (3.29). Note that using (iii) and (iv) we have that

$$(\hat{\mu}_i - \mu_i) \doteq \frac{\sigma m \bar{\epsilon}_{i.}}{A_i\ g(\mu_i, z_i, \theta)} + \frac{\sigma^2 \nu'(\mu_i, z_i, \theta)}{A_i} \Sigma_{j=1}^{m} \{\epsilon_{ij}^2 - \frac{(m-1)}{m}\}$$

(3.36)

$$- \frac{2\ (N^{-1/2}\sigma)\ m\ \bar{\epsilon}_{i.}}{A_i\ g(\mu_i, z_i, \theta)} \{\ 1 + \nu_\theta'(\mu_i, z_i, \theta)\ \}$$

$$+ \frac{N^{-1/2}\sigma^2}{A_i} [\ \Sigma_{j=1}^{m} \{\ \epsilon_{ij}^2 - \frac{(m-1)}{m}\}\ \nu_\theta'(\mu_i, z_i, \theta)$$

$$- 2\ \Sigma_{j=1}^{m} \epsilon_{ij}^2\ \nu'(\mu_i, z_i, \theta)\ \{\ 1 + \nu_\theta(\mu_i, z_i, \theta)\ \}],$$

where

$$A_i = \{m/g^2(\mu_i, z_i, \theta)\} + (m-1)\ \nu''(\mu_i, z_i, \theta)\ \sigma^2$$

$$+ \{4\ \sigma m\ \nu'(\mu_i, z_i, \theta)\ \bar{\epsilon}_{i.}/g(\mu_i, z_i, \theta)\}$$

$$+ \sigma^2[\nu''(\mu_i, z_i, \theta) - 2\{\nu'(\mu_i, z_i, \theta)\}^2]\ \Sigma_{j=1}^{m} \epsilon_{ij}^2.$$

A Taylor series in $\sigma$ about 0 in (3.36) then gives that

$$(\hat{\mu}_i - \mu_i) \doteq \sigma\ g(\mu_i, z_i, \theta)\ \bar{\epsilon}_{i.}$$

$$- 2(N^{-1/2}\sigma)\ g(\mu_i, z_i, \theta)\ \{1 + \nu_\theta(\mu_i, z_i, \theta)\}\ \bar{\epsilon}_{i.}$$

$$+ \sigma^2 g^2(\mu_i, z_i, \theta)\nu'(\mu_i, z_i, \theta)[-4\bar{\epsilon}_{i.} + m^{-1}\Sigma_{j=1}^{m}\{\epsilon_{ij}^2 - \frac{(m-1)}{m}\}]$$

(3.37)

$$+ (N^{-1/2}\sigma^2)g^2(\mu_i, z_i, \theta)\nu'(\mu_i, z_i, \theta)\ [8\{1 + \nu_\theta(\mu_i, z_i, \theta)\}\bar{\epsilon}_{i.}^2$$

$$+ \ \{\nu'_{\theta}(\mu_i, z_i, \theta)/\nu'(\mu_i, z_i, \theta)\} \ \Sigma^{m}_{j=1}\{\epsilon^{2}_{ij} - \frac{(m-1)}{m}\}$$

$$- \ 2 \ \{1 + \nu_{\theta}(\mu_i, z_i, \theta) \ m^{-1}\Sigma^{m}_{j=1} \ \epsilon^{2}_{ij} \ ].$$

Using (3.37) in (3.34) and (3.35) yields, after much tedious algebra and repeated use of (i), that

$$(3.38) \qquad H_{1,N} = N^{-1}\Sigma^{M}_{i=1}\Sigma^{m}_{j=1} \ \xi(\mu_i, z_i, \theta) \ (\epsilon_{ij} - \bar{\epsilon}_{i.})^{2} + o_p(1)$$

and

$$(3.39) \qquad \begin{aligned} H_{2,N} &= N^{-1/2}\Sigma^{M}_{i=1}\Sigma^{m}_{j=1} \ \{ \ \nu_{\theta}(\mu_i, z_i, \theta) - \bar{\nu}_{\theta}(\mu, z, \theta) \ \} \ (\epsilon_{ij} - \bar{\epsilon}_{i.})^{2} \\ &+ \ (N^{1/2}\sigma) \ N^{-1}\Sigma^{M}_{i=1}\Sigma^{m}_{j=1} \ b_{ij} + o_p(1), \end{aligned}$$

where the $\{b_{ij}\}$ are functions of various constants and the $\{\epsilon_{ij}\}$ only through third powers of the $\{\epsilon_{ij}\}$. From (i) and (ii) and the nature of the $\{b_{ij}\}$, the second term on the right-hand side of (3.38) is $o_p(1)$. Multiplying (3.38) and (3.39) by $m/(m-1)$ and abusing summation notation slightly to facilitate the comparison with (3.28), we obtain that

$$\frac{m}{(m-1)} H_{2,N} = N^{-1/2}\Sigma^{N}_{i=1} \ (s^{2}_{i} - 1)\{ \ \nu_{\theta}(\mu_i, z_i, \theta) - \bar{\nu}_{\theta}(\mu, z, \theta) \ \} + o_p(1)$$

and, by laws of large numbers and the fact that $E \ s^{2}_{i} = 1$, that

$$\frac{m}{(m-1)} H_{1,N} = \zeta(\mu, z, \theta) + o_p(1).$$

Using the above in (3.31) now yields the result. □

Theorem 3.5 shows that under the small $\sigma$ asymptotics of Theorem 2.3, the modified maximum likelihood estimator of Raab is asymptotically equivalent to the estimator of Sadler and Smith, thus supporting their empirical observations and the discussion at the end of Section 2.3.

# CHAPTER IV

# COMPARISON AND DISCUSSION

## 4.0 Introduction

In this Chapter we use the results of Chapters 2 and 3 to offer some theoretical comparisons based on asymptotic efficiency of the variance function estimators of Section 1.3. Our findings, along with the developments in previous chapters, allow us to make several general conclusions regarding efficient and appropriate variance function estimation.

## 4.1 Comparison of methods based on residuals

In order to make simple comparisons among the methods of pseudo-likelihood, restricted maximum likelihood or weighted residuals, maximum likelihood, extended quasi-likelihood, the logarithm method and weighted absolute residuals, we assume that the errors are symmetric and independent and identically distributed and that either g does not depend on $\beta$ or $\sigma$ is small. Recall from Chapter 3 that maximum likelihood and the weighted squared residual methods are thus asymptotically equivalent here and that the same methods are asymptotically equivalent to extended quasi-likelihood when $\sigma \rightarrow 0$.

From (2.6) and (2.11), we see that the asymptotic relative efficiency of the weighted absolute residual method with respect to pseudo-likelihood and the other methods of Section 2.5.1 is

$$(4.1) \qquad \{(2+\kappa)(1-\delta)\} \,/\, (4\delta).$$

This is the asymptotic relative efficiency of the mean absolute deviation with respect to the sample variance for a single sample, thus the problem of comparing these two estimation methods is identical to that of the Eddington-Fisher dispute, see Huber (1981, page 3). For normal errors, using absolute residuals results in a 12% loss in efficiency while for standard double exponential errors there is a 25% gain in efficiency for using absolute residuals.

From (2.6) and (2.10), the asymptotic relative efficiency of the logarithm method with respect to those based on squared residuals is given by

$$(4.2) \qquad (2+\kappa)\ [\ \mathrm{var}\ \{\log\ (\epsilon^2)\}\ ]^{-1}.$$

For normal errors, the logarithm method represents a 59% loss of efficiency with respect to pseudo-likelihood.

Huber (1981, page 3) presents a table of asymptotic relative efficiencies for mean absolute deviation with respect to mean square deviation for various contaminated normal distributions. Here, we adapt part of that table as our Table 4.1, augmenting it to include the result (4.2). The table shows that while at normality neither the absolute residuals nor the logarithm methods are efficient, a very slight fraction of "bad" observations is enough to offset the superiority of squared residuals in a dramatic fashion. For example, just two bad observations in 1000 negate the superiority of squared residuals. If 1% or 5% of the data are "bad," absolute residuals and the logarithm method, respectively, show substantial gains over squared residuals. The implication is that while it is commonly perceived that methods based on squared residuals are to be preferred in general, these methods can be highly non-robust. Our formulation includes this result for maximum likelihood, showing its inadequacy under slight departures from the assumed distributional structure.

### 4.2 Methods based on sample standard deviations

Assume that $m \geq 2$ replicate observations are available at each design point. It is often the case in practice that m is small, usually no more than 4 and most often 2, see Raab (1981). We now compare using absolute residuals to using sample standard deviations in the estimators of Section 1.3.1. For simplicity, assume that the errors are independent and identically and symmetrically distributed and that either g does not depend on $\beta$ or $\sigma$ is small. If the errors are not symmetric and $\sigma$ is not small or the variance depends on $\beta$, using sample standard deviations will be more efficient than suggested in the discussion below.

Let $s_m^2$ be the sample variance of m errors $\{\epsilon_1,\ldots,\epsilon_m\}$. It is easily shown by calculations analogous to those of section 4.1 that replacing absolute residuals by sample standard deviations has the effect of changing the asymptotic covariance matrix (2.6) to

(4.3) $\{(2 + \kappa) + 2/(m - 1)\} \{4N \varsigma(\beta,\theta)\}^{-1}$

so that the asymptotic relative efficiency of using sample standard deviations to weighted squared residuals is given by

(4.4) $\{(2 + \kappa)(m - 1)\} / \{(2 + \kappa)(m - 1) + 2\}$.

From the discussion in Sections 2.3 and 3.3 we see that under the additional conditions presented there, (4.4) represents the asymptotic relative efficiency of the Raab estimator to the weighted squared residual methods.

By calculations similar to those in Section 2.5.2, we find that replacing absolute residuals by sample standard deviations in the logarithm method changes (2.10) to

(4.5) $m \text{ var } \{ \log (s_m^2) \} \{4N \varsigma(\beta,\theta)\}^{-1}$.

Note that this is thus the asymptotic covariance for the Rodbard and Frazier estimator. When the errors are normally distributed, it can be shown routinely that the asymptotic relative efficiency of using sample standard deviations to absolute residuals is given by

(4.6) $\pi^2 / [2m \Psi'\{(m - 1)/2\}]$,

where $\Psi'$ is the trigamma function, see Abramowitz and Stegun (1972, chapter 6).

Replacing absolute residuals by sample standard deviations in the weighted absolute residual method of Section 2.5.3 can be shown to change the asymptotic covariance matrix of this estimator from (2.11) to

$$\{m\,\delta_* \;/\; (1 - \delta_*)\}\; \{N\, \varsigma(\beta,\theta)\}^{-1},$$

where $\delta_* = \mathrm{var}\,(s_m)$. The asymptotic relative efficiency of using sample standard deviations to absolute residuals in this method is thus

$$\{\delta(1 - \delta_*)\} \;/\; \{m\,\delta_*(1 - \delta)\}. \tag{4.7}$$

Table 4.2 contains the asymptotic relative efficiencies (4.4), (4.6) and (4.7) for various values of m when the errors are standard normal. The values in the table for $H_1(x) = x^2$ and x indicate that if the data are approximately normally distributed, using sample standard deviations can entail a loss in efficiency with respect to using residuals if m is small. For substantial replication ($m \geq 10$), using sample standard deviations produces a slight edge in efficiency with respect to residuals for $H_1 = x$. It is interesting to note from Table 4.1 that for normal data, the asymptotic relative efficiency of weighted squared residuals with respect to absolute residuals is the reciprocal of the limit of the asymptotic relative efficiency of weighted absolute residuals with respect to weighted absolute sample standard deviations as $m \to \infty$. This implies that for large m, using sample standard deviations for $H_1(x) = x$ is the same as using weighted squared residuals. For the degree of replication likely to be encountered in practice ($m \leq 4$), however, the implication is that residuals are to be preferred to sample standard deviations in both methods when the data are approximately normal.

The second column of Table 4.2 shows that for the logarithm method, using sample standard deviations surpasses using residuals in terms of efficiency except when m = 2 and is more than twice as efficient for large m. In its raw form, log $|r_i|$ is very unstable because, at least occasionally, $|r_i| \approx 0$, producing a wild "outlier" in the regression. The effect of using sample standard deviations is to decrease the possibility of such outliers; the sample standard deviations will be likely more uniform, especially as m increases. The implication is that the logarithm method should not be based on residuals unless remedial measures are taken if the data are approximately normally distributed. The suggestion to trim a few of the smallest absolute residuals before using this method is clearly supported by the theory; presumably, such trimming would reduce or negate the theoretical superiority of using sample standard deviations.

Table 4.3 contains the asymptotic relative efficiencies of weighted squared sample standard deviations and logarithms of these to weighted squared residuals under normality of the errors. The first column is thus the efficiency of Raab's method to pseudo-likelihood, and the second column is the efficiency of the Rodbard and Frazier method to pseudo-likelihood. Dividing the second column by the first yields the asymptotic relative efficiency of the Rodbard and Frazier method to that of Raab; these numbers agree quite well with Monte Carlo efficiencies for m = 2, 3, and 4 of 39%, 62% and 77% and 39%, 64% and 74% reported by Raab (1981) and Sadler and Smith (1985), respectively. The results of the table imply that using the Raab and Rodbard and Frazier methods, which are popular in the analysis of radioimmunoassay data, can entail a dramatic loss of efficiency when compared to methods based on weighted squared residuals when the data are approximately normally distributed. In Chapter 5 we exhibit a specific implication of this result for an important application in the analysis of assay data.

Note from (4.4) that the squared residual methods will always be more efficient than Raab's method, regardless of the underlying distribution of the data. When the distribution is not normal, the superiority of the methods of Section 2.5.1 to that of Rodbard and Frazier is not as clear cut. When the $\{\epsilon_i\}$ are independent and identically distributed double exponential random variables such that $E\epsilon_i = 0$ and $E\epsilon_i^2 = 1$ and $m = 2$, the asymptotic relative efficiency of the Rodbard and Frazier estimator with respect to weighted squared residual methods is 0.448, which is more than twice that given for normal errors. For some highly contaminated normal distributions, it can be shown that in some instances the Rodbard and Frazier estimator can be preferable to weighted squared residual methods in terms of relative efficiency. When $m = 2$, 1% contamination by observations which are normal with standard deviation six times that of standard normal yields an asymptotic relative efficiency of 2.203.

## 4.3 Discussion and conclusions

In Chapter 2 we constructed a general theory of regression-type estimation for $\theta$ in the heteroscedastic model (1.1) and (1.2). This theory includes as special cases common methods described in Section 1.3.1 and allows for the regression to be based on absolute residuals from the current regression fit as well as sample standard deviations in the event of replication at each design point. In Chapter 3 we showed that the implications of the general theory are valid under certain conditions for other estimators not fitting strictly into this class. Under various restrictions such as symmetry or small $\sigma$, when the variance function g does not depend on $\beta$, we showed in that we can draw simple general conclusions about this class of estimators as well as estimators which do not fall into this class and make simple comparisons among the various methods.

Our conclusions apply strictly only to symmetric error distributions, but they are fairly definitive and one is unlikely to be too successful ignoring them in practice. Our results indicate first that robustness plays a great role in the efficiency of variance function estimation. Squared residual methods are preferable for approximately normally distributed data, but this preference is tenuous, as methods based on these can be highly nonrobust under only slight departures from normality. Methods based on logarithms or the absolute residuals themselves exhibit relatively more robust behavior.

A second conclusion is that when employing methods based on residuals, one should weight the residuals appropriately. Further, because asymptotic efficiency of the variance function estimators is an increasing function of the current regression fit, one should use an iterative weighted fitting so that the estimate of $\theta$ is based on generalized least squares residuals.

Another conclusion concerns the use of residuals versus that of sample standard deviations. For the small amount of replication found in practice, using sample standard deviations rather than residuals can entail a large loss of efficiency if estimation is based on the squares of these quantities or the quantities themselves. For the logarithm method of Harvey based on residuals, trimming the smallest few absolute residuals is essential, since for normal data using sample standard deviations is almost always more efficient than using residuals, even for a small number of replicates. Popular methods in applications such as radioimmunoassay based on sample means and standard deviations can be less efficient than methods based on weighted squared residuals.

Efficient variance function estimation in heteroscedastic regression analysis is an important problem in its own right. There are important differences in estimators for variance when it is modeled parametrically.

## Table 4.1

Asymptotic relative efficiency with respect to weighted squared residuals for contaminated normal distributions with distribution function $F(x) = (1 - \alpha)\Phi(x) + \alpha\Phi(x/3)$.

| contamination fraction $\alpha$ | weighted absolute residuals | logarithms of absolute residuals |
|---|---|---|
| 0.000 | 0.876 | 0.405 |
| 0.001 | 0.948 | 0.440 |
| 0.002 | 1.016 | 0.480 |
| 0.010 | 1.439 | 0.720 |
| 0.050 | 2.035 | 1.220 |

## Table 4.2

Asymptotic relative efficiency of using sample standard deviations to using absolute residuals under normality for $H_1(x)$ (weighted methods).

| | | $H_1(x)$ | |
|---|---|---|---|
| $m$ | $x^2$ | $\log x$ | $x$ |
| 2 | 0.500 | 0.500 | 0.500 |
| 3 | 0.667 | 1.000 | 0.696 |
| 4 | 0.750 | 1.320 | 0.801 |
| ⋮ | ⋮ | ⋮ | ⋮ |
| 9 | 0.889 | 1.932 | 0.986 |
| 10 | 0.900 | 1.984 | 1.001 |
| $\infty$ | 1.000 | 2.467 | 1.142 |

## Table 4.3

Asymptotic relative efficiency of using sample standard deviations to weighted squared residuals under normal errors for $H_1(x)$.

| | $H_1(x)$ | |
|---|---|---|
| $m$ | $x^2$ | log $x$ |
| 2 | 0.500 | 0.203 |
| 3 | 0.667 | 0.405 |
| 4 | 0.750 | 0.535 |
| 5 | 0.800 | 0.620 |
| 6 | 0.833 | 0.680 |
| 7 | 0.857 | 0.723 |
| 8 | 0.875 | 0.757 |
| 9 | 0.889 | 0.783 |
| 10 | 0.900 | 0.804 |

# CHAPTER V

## APPLICATION - THE ROLE OF $\hat{\theta}$ IN THE PROPERTIES OF THE ESTIMATOR FOR MINIMUM DETECTABLE CONCENTRATION

### 5.0 Introduction

In this chapter we investigate a specific setting in which the choice of the estimator for $\theta$ plays an important role in determining the theoretical properties of estimators for other important quantities. In particular, we study the theoretical properties of an estimator for the minimum detectable concentration, an important calibration quantity in the analysis of assay data, and show how the properties of this estimator are controlled by those of the estimator for $\theta$. We present a Monte Carlo study and an example which not only support the asymptotic results of this chapter but our results on iterating the generalized least squares algorithm from Chapter 2.

### 5.1 Analysis of assay data and the minimum detectable concentration

The analysis of assay data has long been an important problem in clinical chemistry and the biological sciences; see, for example, Finney (1964) and Oppenheimer, et al. (1983). The most common method of analysis is to fit a nonlinear regression model to the data. Much recent work suggests that these data can be markedly heteroscedastic; in radioimmunoassay, for example, this characteristic has been observed repeatedly and incorporated into the analysis as discussed by Finney (1976), Rodbard (1978), Tiede and Pagano (1979) and Raab (1981). Such analyses are for the most part special cases of the general heteroscedastic nonlinear regression model (1.1) and (1.2). Specifically, we observe independent counts $Y_{ij}$ at concentrations $x_i$ for $i = 1,\ldots,M$ and $j = 1,\ldots m_i$ with mean and variances given by (1.1) and (1.2). Since replication is involved we use this notation throughout for convenience; for our analysis we adopt the simplifying assumption that $m_i = m \ \forall i$. A standard model for the mean in a radioimmunoassay is the four parameter logistic model

$$(5.1) \qquad f(x,\beta) = \beta_1 + (\beta_2 - \beta_1)/[1 + \exp\{\beta_4(\log x - \beta_3)\}].$$

Almost without exception, the variances have been modeled as functions of the mean response as in (1.8), usually either as a quadratic function or as a power of the mean, i.e.,

$$(5.2) \qquad \sigma g(z_i,\beta,\theta) = \sigma f(x_i,\beta)^{\theta}.$$

The assay problem does not always stop with estimating $\beta$, but also addresses issues of calibration. These issues include confidence intervals for a true $x_*$ given a new $Y_*$, the classic calibration problem. Also of interest is determination of the sensitivity of the assay using such concepts as the minimum detectable concentration of Rodbard (1978) and the critical level, detection level and determination limit of Oppenheimer, et al. (1983). In this chapter we show that a unique feature of these calibration problems is that the efficiency of estimation is essentially determined by how well one estimates the variance parameter $\theta$. We show that far from being a nuisance parameter, $\theta$ plays a central role in these problems.

For definiteness, we focus on the determination of minimum detectable concentration. There is no unique definition of this concept; we use the following based on the definition of Rodbard (1978).

Definition. Let $\overline{Y}(x,m)$ be the mean response based on m replicates at concentration level x, taken independently of the calibration data set $\{Y_{ij}\}$. Let $f(0,\beta)$ be the expected response at zero concentration based on the calibration data set. The minimum detectable concentration $x_c$ at level $(1-\alpha)$ is the smallest concentration x for which

$$(5.3) \qquad \Pr\{\overline{Y}(x,m) \geq f(0,\beta)\} > 1 - \alpha .$$

If $t(\alpha,N-p)$ is the $(1-\alpha)^{th}$ percentile of the t-distribution with N-p degrees of freedom, where $N = Mm$ is the total sample size and M is the number of concentrations, the usual estimate $\hat{x}_c$ of $x_c$ satisfies

$$(5.4) \quad \{f(\hat{x}_c,\hat{\beta}) - f(0,\hat{\beta})\}^2 = \{t(\alpha,N-p)\}^2\{\hat{\sigma}^2 g^2(\hat{x}_c,\hat{\beta},\hat{\theta})/m + \widehat{\mathrm{var}}[f(0,\hat{\beta})]\},$$

where $\widehat{\mathrm{var}}[f(0,\hat{\beta})]$ is an estimate of the variance of $f(0,\hat{\beta})$ and $\hat{\sigma}^2$ is the usual mean squared error from the weighted fit:

$$\hat{\sigma}^2 = (N-p)^{-1}\Sigma_{i=1}^{M}\Sigma_{j=1}^{m}\{Y_{ij} - f(x_i,\hat{\beta})\}^2 f^{-2\hat{\theta}}(x_i,\hat{\beta}).$$

In application, practitioners involved in the analysis of radioimmunoassay data commonly use the methods of Rodbard and Frazier and Raab to estimate $\theta$ and then perform generalized least squares with

$$\hat{\sigma}_i = \bar{Y}_{i\cdot}^{\hat{\theta}}.$$

Thus, in our investigation, we will first investigate the properties of $\hat{x}_c$ in general and then for illumination use the general results to examine the behavior of $\hat{x}_c$ when $\theta$ is estimated by the methods of Rodbard and Frazier, Raab and any of the weighted squared residual methods; we will refer in particular to the latter methods as pseudo-likelihood for brevity.

## 5.2 Asymptotic theory

In this section we outline an asymptotic theory for the estimator of minimum detectable concentration. As mentioned in Section 2.4, values of $\sigma$ which are small relative to the means are typical of assay data. Furthermore, as in Chapters 2 and 3, the assumption of small $\sigma$ is needed from a technical standpoint for study of common estimators for $\theta$ used in this application such as the Rodbard and Frazier and Raab estimators. Thus, the asymptotic theory we develop here is based on the assumption that $N \to \infty$ and $\sigma \to 0$ simultaneously in such a way that m remains fixed. For simplicity of presentation and immediate application, we express our results in terms of the power of the mean model (5.2).

Throughout, assume that the errors $\{\epsilon_{ij}\}$ are independent and identically distributed. Define

$$v_i = \log f(x_i,\beta),$$

$$s_m^2 = \{(m-1)^{-1}\Sigma_{j=1}^{m}(\epsilon_{ij} - \bar{\epsilon}_{i.})^2\}, \text{ and}$$

$$\sigma_v^2 = \lim_{N\to\infty} (M-1)^{-1}\Sigma_{i=1}^{M}(v_i - \bar{v})^2.$$

For the minimum detectable concentration, note that in (5.4) the term $\hat{var}\{f(0,\hat{\beta})\}$ is of the order $N^{-1}$ and is hence small relative to all the other terms. It turns out that in our asymptotic framework, the solution to (5.4) is estimating the quantity $x_c^*$, where

$$(5.5) \qquad 0 = \{z(\alpha)\}^2\sigma^2 f^{2\theta}(x_c^*,\beta)/m - \{f(x_c^*,\beta) - f(0,\beta)\}^2,$$

where $z(\alpha)$ is the $(1 - \alpha)^{th}$ percentile point of the standard normal distribution.

Here are the major results, the technical details for which are given in Section 5.4. Define

$$(5.6) \qquad d_0 = \log f(0,\beta) - \lim_{N\to\infty} M^{-1}\Sigma_{i=1}^{M} \log f(x_i,\beta).$$

The main theorem follows from the results of Chapter 2.

<u>Theorem 5.1.</u> Let $\hat{x}_c(RF)$, $\hat{x}_c(MML)$ and $\hat{x}_c(PL)$ denote the estimated minimum detectable concentrations using the Rodbard and Frazier estimate of $\theta$, the modified maximum likelihood estimate of Raab and the pseudo-likelihood estimate ,respectively. Then there is a constant $A_0$ and a sequence $b_N$ for which

$$(5.7)\qquad b_N A_0 N^{1/2}(\hat{x}_c(RF) - x_c^*)/\sigma \xrightarrow{\mathscr{L}} N(0,\ (2 + \kappa) + \mathrm{var}(\log s_m^2)d_0^2 m/\sigma_v^2\ ),$$

$$(5.8)\qquad b_N A_0 N^{1/2}(\hat{x}_c(MML) - x_c^*)/\sigma \xrightarrow{\mathscr{L}} N(0,\ (2 + \kappa) + \mathrm{var}(s_m^2)d_0^2 m/\sigma_v^2\ ),$$

$$(5.9)\qquad b_N A_0 N^{1/2}(\hat{x}_c(PL) - x_c^*)/\sigma \xrightarrow{\mathscr{L}} N(0,\ (2 + \kappa)\{1 + d_0^2/\sigma_v^2\}\ ).\ \square$$

From (4.3), the asymptotic relative efficiency of the modified maximum likelihood estimate of minimum detectable concentration relative to the pseudo-likelihood estimate is

$$(5.10)\qquad (2 + \kappa)(\sigma_v^2 + d_0^2)/\{(2 + \kappa)(\sigma_v^2 + d_0^2) + 2d_0^2/(m - 1)\}$$

which is less than 1 for all m regardless of the value of $\kappa$. Similarly, the asymptotic relative efficiency of the log-linearized estimate of minimum detectable concentration to the pseudo-likelihood estimate is

$$(5.11)\qquad (2 + \kappa)(\sigma_v^2 + d_0^2)/\{\sigma_v^2(2 + \kappa) + md_0^2\mathrm{var}(\log s_m^2\}.$$

It follows from Theorem 5.1 and (5.10) and (5.11) that the ordering in efficiency of estimated minimum detectable concentration is the same as the ordering for estimating $\theta$ and thus from Chapter 4 will favor pseudo-likelihood for normally distributed data in the case of the Rodbard and Frazier estimator and for all distributions in the case of the modified maximum likelihood method. The numerical efficiencies depend on the logarithm of the true means through $d_0^2$ and $\sigma_v^2$. For example, in the simulation discussed in the next section, the asymptotic relative efficiency of the Rodbard and Frazier estimator is 27% for m = 2 and 63% for m = 4.

The asymptotic theory thus suggests that inefficiencies in estimating the variance parameter $\theta$ translate into inefficiencies for estimating the minimum detectable concentration.

5.3 A simulation

To check the qualitative nature of the asymptotic theory, we ran a small simulation. We restrict our focus here to the Rodbard and Frazier method and weighted squared residual methods, in particular, pseudo-likelihood iterated with number of cycles $\mathscr{C}$.

The responses $Y_{ij}$ were normally distributed with mean and variance satisfying (5.1), (5.2), where $\beta_1 = 29.5274$, $\beta_2 = 1.8864$, $\beta_3 = 1.5793$, $\beta_4 = 1.0022$, $\theta = 0.7$ and $\sigma = .0872$. The 23 concentrations chosen are given in Table 5.1. The parameters and the concentrations were chosen to represent some assays we have observed. We studied the case m = 2 or duplicates and m = 4 or quadruplicates. For each situation, there were 500 simulated data sets. A limited second simulation was run with the larger value $\sigma = 0.17$, but there did not appear to be significant qualitative differences from the case reported here.

The estimators chosen were unweighted least squares for $\beta$, the Rodbard and Frazier method and the pseudo-likelihood/generalized least squares combination which we report only for $\mathscr{C}$ = 1 and 2 cycles of the algorithm. The methods of estimating the minimum detectable concentration are as discussed in Section 5.1. The estimates of $\theta$ were constrained to lie in the interval $0 \leq \theta \leq 1.50$.

In Table 5.2, we compare the estimators of $\theta$ on the basis of bias and variance. The biases are large relative to the standard error, so that mean-squared error comparisons are artificial and dramatic. The bias in the pseudo-likelihood estimate of $\theta$ when doing only $\mathscr{C}$ = 1 cycles of the algorithm has been previously observed as mentioned in the discussion of Chapter 1. One sees here that the effect of doubling the replicates from two to four for a given set of concentrations improves the Monte-Carlo efficiency of the Rodbard and Frazier estimate of $\theta$ from 0.351 to 0.585, compared to the theoretical asymptotic increase from 0.203 to 0.535, supporting the theory of Chapter 4. This example indicates that pseudo-likelihood estimation of $\theta$ can in some circumstances be a considerable improvement over the method of Rodbard and Frazier.

In Table 5.3 we consider the mean-squared errors for estimating the regression parameters $(\beta_1, \beta_2, \beta_3, \beta_4)$. The value of $\sigma$ is very small here. Larger values of $\sigma$ would presumably cause greater differences. As expected, the unweighted least squares estimate is unacceptably inefficient.

Finally, we turn to the minimum detectable concentration. We chose $\alpha$ = 0.05. For all of the methods used in the study, the probability requirement (5.3) was easily satisfied; rather than 95% exceedance probability, every case was more than 97%. The mean values of the minimum detectable concentrations are reported in Table 5.4, with variances given in Table 5.5. Note that in both of these tables, we give results for the case that $\theta$ is known as well as estimated. The relatively poor behavior of unweighted least squares is evident. To quote from Oppenheimer, et al. (1983): "Rather dramatic differences have been observed depending on whether a valid weighted or inappropriate unweighted analysis is used." When the variance parameter $\theta$ is known, there is little difference between any of the weighted methods.

When $\theta$ is unknown, there are rather large proportional differences. The figures in Table 5.4 show that the mean minimum detectable concentration for the Rodbard and Frazier method is 10% larger than for the pseudo-likelihood method based on $\mathcal{C} = 2$ cycles; whether the raw numerical difference is of any practical consequence will depend on the context.

For $m = 2$ replicates, the pseudo-likelihood estimate of minimum detectable concentration with unknown $\theta$ has mean $3.934 \times 10^{-2}$ and standard deviation $0.05 \times 10^{-4}$; the corresponding figures for $M = 4$ are $2.722 \times 10^{-2}$ and $0.028 \times 10^{-4}$. Proportionately, when $\theta$ is unknown, the method of estimating it seems to have important consequences for the estimate of minimum detectable concentration, particularly in the variability of the estimate. For the case of duplicates, the Monte-Carlo variance of pseudo-likelihood is only 0.368 as large as that based on the Rodbard and Frazier estimate, while the asymptotics suggest 0.273, increasing to 0.709 and 0.629 respectively for quadruplicates.

## 5.4. An example

Differences among the three estimators of $\theta$ and the subsequent estimators of minimum detectable concentration which are reminiscent of the qualitative implications of the asymptotic theory and simulation can be seen in the following example. The data are from a radioimmunoassay based on 4 replicates at each of the 23 concentrations given in Table 5.1 and are presented in Table 5.6. The analysis presented here is for illustrative purposes only; we do not claim to be analyzing these data fully. Our aim is to exhibit the fact that the three methods of estimation of $\theta$ considered in this chapter can lead to nontrivially different results.

We assumed in all cases the model (5.1) and (5.2). For the full data set and reduced data sets considering all possible permutations of duplicates (except one set for which an $s_i^2 = 0$, complicating the application of the Rodbard and Frazier method), we computed the estimates of $\theta$, $\sigma^2$ and $x_c$ using the pseudo-likelihood and Rodbard and Frazier methods and the estimate of Sadler and Smith (1985) as described in Section 2.3 in place of the more computationally difficult modified maximum likelihood estimate. We also computed the estimate of minimum detectable concentration based on ordinary least squares. The results are given in Table 5.7.

An investigation of both the full and reduced data sets suggests that there are no massive outliers and that design points 1, 22 and 23 are possible high leverage points. For our purposes of illustration we do not pursue this point; we could investigate the behavior of one of the methods that account for leverage, for example.

The results of Table 5.7 show that the three estimates can vary greatly. As a measure of this, consider the means and standard deviations of $\hat{\theta}$ and $\hat{x}_c$ for the five data sets obtained by considering duplicates (ignoring the fact that these data sets are not strictly independent). Using these, we list "relative efficiencies" for the estimators below:

"Relative efficiencies" for estimators of $\theta$ and $x_c$

for data in example when m =2

| | RF to PL | MML to PL | RF to MML |
|---|---|---|---|
| $\hat{\theta}$ | .222 | .351 | .632 |
| $\hat{x}_c$ | .529 | .659 | .802 |

Qualitatively, the estimates exhibit the type of behavior predicted by the asymptotic theory; quantitatively, the values compare favorably with what the theory would predict given the crudity of the comparison.

This example shows that there can be wide differences among the various estimation methods for $\theta$ and minimum detectable concentration in application and that the qualitative way in which the differences manifest themselves is predicted by the asymptotic theory of Chapter 2 and Section 5.2.

## 5.5 Summary

The point made in this chapter is that the relative efficiency of the estimated minimum detectable concentration can be affected by the method used to estimate the variance parameter $\theta$. While for estimation of $\beta$ the effect of how one estimates the variance function is only second order, for estimation of other quantities, the effect can be first order. While we have shown this result in the specific instance of the minimum detectable concentration, we conjecture that it is probably a much more general result for calibration quantities in general. Development of a unified theory for the estimation of calibration quantities would be a worthwhile endeavor.

## 5.6 Proofs

The analysis of the estimator for minimum detectable concentration is complicated by the behavior of the derivative of $f(x,\beta)$ with respect to $\beta$ at x=0, especially for the standard model (5.1). We will write $f(x,\beta) = h(\xi,\beta)$, where $\xi = \ell(x,\beta)$, $\xi_c^* = \ell(x_c^*,\beta)$ and $\hat{\xi}_c = \ell(\hat{x}_c,\hat{\beta})$. In the model (5.1), for example, $\xi = \ell(x,\beta) = \exp(\beta_4 \log x)$. We assume throughout that $f(0,\beta) > 0$ and that all functions are sufficiently smooth. Assume further that

(5.12) $\ell(0,\beta) = 0$ ;

(5.13) $\partial/\partial\xi \; h(0,\beta) \neq 0$ ;

(5.14) $\partial/\partial\beta \; \ell(0,\beta) = 0$ ;

(5.15) If $w \to 0$ and $v$ is a random variable such that

$\ell(v,\beta)/\ell(w,\beta) \xrightarrow{p} 1$, then

$$\sup\{\, 0 \le \alpha \le 1 : \left| \{\partial/\partial\xi \; \ell(\alpha v+(1-\alpha)w,\beta)\}/\{\partial/\partial\xi \; \ell(w,\beta) - 1\} \right| \,\} \xrightarrow{p} 0.$$

These assumptions are satisfied for the model (5.1) if $\beta_4 > 0$.

We need the following results. The proofs of Lemmas 5.4 and 5.5 will be at the end of the section. Define $c = \{z(\alpha)\}^2$.

<u>Lemma 5.2</u> As $\sigma \to 0$ for $f(0,\beta) > 0$,

$$\xi_c^* = \sigma a_c + \mathcal{O}(\sigma^2), \qquad a_c = (c/m)^{1/2} f^{\theta}(0,\beta)\{\partial/\partial\xi \; h(0,\beta)\}^{-1}.$$

<u>Proof</u>: A Taylor series expansion of (5.5) in $\xi_c^*$ and around zero. □

<u>Lemma 5.3</u> Assume that as $N \to \infty$, $\sigma \to 0$, $(\hat{\xi}_c - \xi_c^*) = \mathcal{O}_p(\sigma N^{1/2})$. Define

$$A_0 = 2ma_c\{\partial/\partial\xi \; h(0,\beta)\}^2/\{cf^{2\theta}(0,\beta)\}.$$

Then as $N \to \infty$, $\sigma \to 0$, if $N^{1/2}(\hat{\theta} - \theta) = \mathcal{O}_p(1)$, we have the asymptotic expansion

$$(5.16) \qquad A_0 N^{1/2}(\hat{\xi}_c - \xi_c^*)/\sigma$$

$$= N^{1/2}(\hat{\sigma}^2 - \sigma^2)/\sigma^2 + 2\{\log f(0,\beta) N^{1/2}(\hat{\theta} - \theta) + o_p(1). \quad □$$

<u>Lemma 5.4</u> Consider Lemma 5.3. Then

$$N^{1/2}(\hat{\sigma}^2 - \sigma^2)/\sigma^2$$

$$= N^{-1/2}\Sigma_{i=1}^{M}\Sigma_{j=1}^{m}(\epsilon_{ij}^{2} - 1) - 2\bar{v}N^{1/2}(\hat{\theta} - \theta) + o_p(1)$$

so that

$$A_0 N^{1/2}(\hat{\xi}_c - \xi_c^{*})/\sigma$$

$$= N^{-1/2}\Sigma_{i=1}^{M}\Sigma_{j=1}^{m}(\epsilon_{ij}^{2} - 1) + 2d_0 N^{1/2}(\hat{\theta}-\theta) + o_p(1),$$

where $d_0$ is defined in (5.6). □

<u>Proposition 1</u> : The limit results (5.7)-(5.9) hold for $A_0 N^{1/2}(\hat{\xi}_c - \xi_c^{*})/\sigma$, where (in obvious notation) $\hat{\xi}_c = \hat{\xi}_c(RF)$, $\hat{\xi}_c(MML)$ or $\hat{\xi}_c(PL)$.

<u>Proof of Proposition 1</u> : From the results of Chapter 2, we have that (in obvious notation)

$$N^{1/2}(\hat{\theta}_{RF} - \theta) = (1/2)\ N^{-1/2} m\ \Sigma_{i=1}^{M}\ \{\log q_i^2 - E(\log q_i^2)\}(v_i - \bar{v}) + o_p(1);$$

$$N^{1/2}(\hat{\theta}_{MML} - \theta) = (1/2)\ N^{-1/2} m\ \Sigma_{i=1}^{M}\ \{q_i^2 - E(q_i^2)\}\ (v_i - \bar{v}) + o_p(1)\ ;$$

$$N^{1/2}(\hat{\theta}_{PL} - \theta) = (1/2)\ N^{-1/2}\ m^{-1/2}\ \Sigma_{i=1}^{M}\ \Sigma_{j=1}^{m}(\epsilon_{ij}^{2} - 1)\ (v_i - \bar{v}) + o_p(1),$$

so that by Lemmas 5.2 - 5.4 and equation (5.16), we have

$$A_0 N^{1/2}(\hat{\xi}_c(PL) - \xi_c^*)/\sigma$$

$$= N^{-1/2} \Sigma_{i=1}^M \Sigma_{j=1}^m (\epsilon_{ij}^2 - 1)\{1 + d_0(v_i - \bar{v})/\sigma_v^2\} + o_p(1),$$

which with the central limit theorem gives the same limit distribution as in (5.9). We also have that

$$A_0 N^{1/2}(\hat{\xi}_c(RF) - \xi_c^*)/\sigma = N^{-1/2} \Sigma_{i=1}^M \Sigma_{j=1}^m (\epsilon_{ij}^2 - 1)$$

$$+ d_0 \, mN^{-1/2} \Sigma_{i=1}^M (v_i - \bar{v}) \log q_i^2/\sigma_v^2 + o_p(1)$$

$$A_0 N^{1/2}(\hat{\xi}_c(MML) - \xi_c^*)/\sigma = N^{-1/2} \Sigma_{i=1}^M \Sigma_{j=1}^m (\epsilon_{ij}^2 - 1)$$

$$+ d_0 \, mN^{-1/2} \Sigma_{i=1}^M (v_i - \bar{v}) q_i^2/\sigma_v^2 + o_p(1).$$

Simple central limit theorem calculations yield the same limit distribution as in (5.7) and (5.8). □

Remark: Result (5.8) is based on $\hat{\sigma}$ obtained from the residuals of the final fit of the mean response function as for the log-linearized method and pseudo-likelihood, so that Lemma 5.4 holds. The modified maximum likelihood method also provides a joint estimate of $\sigma$ along with the estimate of $\theta$. If one considers this estimator in place of $\hat{\sigma}$ in Lemma 5.4, it can be shown that the resulting estimator of minimal detectable concentration has even larger asymptotic variance than that in (5.8). This follows from calculations using the results of Section 3.3 and has interesting implications for practice.

Proof of Theorem 5.1 : By (5.16), for any of the estimators $\hat{x}_c$, since $\xi = \ell(x,\beta)$, we have the limit result that for some $\Delta$,

$$(5.17) \qquad N^{1/2} \{ \ell(\hat{x}_c,\beta) - \ell(x_c^*,\beta) \}/\sigma \xrightarrow{\mathscr{L}} N(0,\Delta).$$

Thus, for $\gamma_c$ between $\hat{x}_c$ and $x_c^*$, defining

$$W_N = N^{1/2} \ell_x(x_c^*,\beta) \ (\hat{x}_c - x_c^*) \ / \ \sigma, \text{ we have}$$

$$W_N \ \ell_x(\gamma_c,\beta)\} \ / \ \ell_x(x_c^*,\beta) \xrightarrow{\mathscr{L}} N(0,\Delta),$$

where $\ell_x(v,\beta)$ is the derivative with respect to the first argument of $\ell(v,\beta)$. It thus suffices through (5.15) to prove that

$$\ell(\gamma_c,\beta) \ / \ \ell(x_c^*,\beta) \xrightarrow{p} 1.$$

But this follows from (5.17) since $\xi_c^*/\sigma = \ell(x_c^*,\beta)/\sigma \to a_c$, see Lemma 5.2. □

Proof of Lemma 5.3: By a series of Taylor expansions and using Lemma 5.2,

$$(5.18)\qquad N^{1/2}\{h(\hat{\xi}_c,\hat{\beta}) - h(0,\hat{\beta})\}^2/\sigma^2$$

$$= N^{1/2}\{h(\hat{\xi}_c,\beta) - h(0,\beta)\}^2/\sigma^2 + o_p(N^{-1/2})$$

$$= N^{1/2}\{h(\xi_c^*,\beta) - h(0,\beta)\}^2/\sigma^2$$

$$+ 2\{h(\xi_c^*,\beta) - h(0,\beta)\}\{\partial/\partial\xi\ h(\xi_c^*,\beta)\}N^{1/2}(\hat{\xi}_c-\xi_c^*)/\sigma^2$$

$$+ o_p(N^{-1/2})$$

$$= N^{1/2}\{h(\xi_c^*,\beta) - h(0,\beta)\}^2/\sigma^2$$

$$+ 2\xi_c^*\{\partial/\partial\xi\ h(0,\beta)\}^2N^{1/2}(\hat{\xi}_c - \xi_c^*)/\sigma^2 + o_p(1)$$

$$= N^{1/2}\{h(\xi_c^*,\beta) - h(0,\beta)\}^2/\sigma^2$$

$$+ 2a_c\{\partial/\partial\xi\ h(0,\beta)\}^2N^{1/2}(\hat{\xi}_c - \xi_c^*)/\sigma + o_p(1).$$

Similar calculations taking into account that $N^{1/2}(\hat{\beta} - \beta) = O_p(\sigma)$ and $\xi_c^* \to 0$ yield

$$(5.19)\qquad N^{1/2}h^{2\hat{\theta}}(\hat{\xi}_c,\hat{\beta})\hat{\sigma}^2/(m\sigma^2)$$

$$= N^{1/2}h^{2\theta}(\xi_c^*,\beta)/m + h^{2\theta}(\xi_c^*,\beta)(N^{1/2}/m)(\hat{\sigma}^2 - \sigma^2)/\sigma^2$$

$$+ (2/m)h^{2\theta}(\xi_c^*,\beta)\ \{\log h(\xi_c^*,\beta)\}\ N^{1/2}(\hat{\theta} - \theta)$$

$$+ (2\theta/m)h^{2\theta-1}(\xi_c^*,\beta)h_\beta(\xi_c^*,\beta)\ N^{1/2}(\hat{\beta} - \beta) + o_p(1)$$

$$= \{h^{2\theta}(\xi_c^*,\beta)/m\}N^{1/2} + (c/m)\ h^{2\theta}(0,\beta)\ N^{1/2}(\hat{\sigma}^2 - \sigma^2)/\sigma^2$$

$$+ (2c/m)\ h^{2\theta}(0,\beta)\ \{\log h(0,\beta)\}\ N^{1/2}(\hat{\theta} - \theta) + o_p(1).$$

Combining (5.4), (5.5), (5.18) and (5.19) yields (5.16). □

<u>Proof of Lemma 5.4</u>: Define

$$\hat{\sigma}_0^2 = N^{-1}\Sigma_{i=1}^{M}\Sigma_{j=1}^{m}[\{Y_{ij} - f(x_i,\beta)\}/f^{\theta}(x_i,\beta)]^2 = N^{-1}\sigma^2\Sigma_{i=1}^{M}\Sigma_{j=1}^{m}\epsilon_{ij}^2.$$

Then, since $N^{1/2}(\hat{\beta} - \beta) = O_p(\sigma)$,

$$N^{1/2}(\hat{\sigma}^2 - \hat{\sigma}_0^2)/\sigma^2 = N^{-1/2}\Sigma_{i=1}^{M}\Sigma_{j=1}^{m}\ [\{Y_{ij} - f(x_i,\hat{\beta})\}^2/f^{2\hat{\theta}}(x_i,\hat{\beta})$$

$$- \{Y_{ij} - f(x_i,\beta)\}^2/f^{2\theta}(x_i,\beta)]\sigma^{-2}$$

$$= N^{-1/2}\Sigma_{i=1}^{M}\Sigma_{j=1}^{m}\ [\{Y_{ij} - f(x_i,\beta)\}^2/f^{2\hat{\theta}}(x_i,\beta)$$

$$-\{Y_{ij} - f(x_i,\beta)\}^2/f^{2\theta}(x_i,\beta)]\sigma^{-2} + o_p(1)$$

$$-2\bar{v}\ N^{1/2}(\hat{\theta} - \theta) + o_p(1),$$

completing the proof. □

**Table 5.1**

Concentration Levels Used in the Simulation

| *X* | *X* |
|---|---|
| 0.00 | 2.50 |
| 0.075 | 3.25 |
| 0.1025 | 4.50 |
| 0.135 | 6.00 |
| 0.185 | 8.25 |
| 0.25 | 11.25 |
| 0.40 | 15.00 |
| 0.55 | 20.25 |
| 0.75 | 27.50 |
| 1.00 | 37.00 |
| 1.375 | 50.00 |
| 1.85 | |

## Table 5.2

Three Estimates of the Variance Parameter $\theta$

| | m = 2 Replicates | | m = 4 Replicates | |
|---|---|---|---|---|
| Bias | Monte-Carlo | Asymptotic | Monte-Carlo | Asymptotic |
| Rodbard and Frazier | 0.15 | - | 0.001 | - |
| Pseudo-likelihood | | | | |
| $\mathscr{C} = 1$ | 0.045 | - | 0.022 | - |
| $\mathscr{C} = 2$ | 0.000 | - | 0.001 | - |
| $\mathscr{C} = 3$ | 0.004 | - | 0.000 | - |
| Variance of Pseudo -likelihood with $\mathscr{C} = 2$ Relative to Variance of | | | | |
| Rodbard and Frazier | 0.351 | 0.203 | 0.585 | 0.535 |
| Pseudo-likelihood | | | | |
| $\mathscr{C} = 1$ | 1.010 | - | 1.010 | - |
| $\mathscr{C} = 3$ | 1.000 | - | 1.000 | - |

**Table 5.3**

Mean Squared Error Ratios for Estimating the Regression Parameter

Pseudo-likelihood with $\mathscr{C}$ = 2 Steps Relative to

| | m = 2 Replicates | | | | m = 4 Replicates | | | |
|---|---|---|---|---|---|---|---|---|
| | $\beta_1$ | $\beta_2$ | $\beta_3$ | $\beta_4$ | $\beta_1$ | $\beta_2$ | $\beta_3$ | $\beta_4$ |
| Least Squares | 0.621 | 0.330 | 0.704 | 0.424 | 0.694 | 0.383 | 0.725 | 0.418 |
| Rodbard and Frazier | 0.901 | 0.917 | 0.901 | 0.893 | 1.000 | 0.971 | 1.000 | 0.990 |
| Pseudo-likelihood | | | | | | | | |
| $\mathscr{C}$ = 1 | 0.980 | 0.961 | 0.990 | 0.971 | 1.000 | 1.000 | 0.990 | 0.990 |
| $\mathscr{C}$ = 3 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |

**Table 5.4**

100 × Mean Minimum Detectable Concentrations

| | m = 2 Replicates | | m = 4 Replicates | |
|---|---|---|---|---|
| | $\theta$ Known | $\theta$ Estimated | $\theta$ Known | $\theta$ Estimated |
| Unweighted Least Squares | 13.106 | 13.106 | 9.173 | 9.173 |
| Rodbard and Frazier | 3.937 | 4.346 | 2.718 | 2.785 |
| Pseudo-likelihood | | | | |
| $\mathscr{C}$ = 1 | – | 4.216 | – | 2.809 |
| $\mathscr{C}$ = 2 | 3.927 | 3.934 | 2.715 | 2.722 |

**Table 5.5**

Ratio of Monte-Carlo Variance of the Estimate of Minimum Detectable Concentration -- Pseudo-likelihood with $\mathscr{C}$ = 2 Cycles Relative to

| | m = 2 Replicates | | m = 4 Replicates | |
|---|---|---|---|---|
| | $\theta$ Known | $\theta$ Estimated | $\theta$ Known | $\theta$ Estimated |
| Unweighted Least Squares | 0.067 | 0.118 | 0.053 | 0.073 |
| Rodbard and Frazier | 0.980 | 0.368 | 1.000 | 0.709 |
| Pseudo-likelihood | | | | |
| $\mathscr{C}$ = 1 | – | 0.840 | – | 0.910 |

**Table 5.6**

Data for Example of Section 5.4

| Concentration (x) | Response (Y) |
|---|---|
| 0.000 | 1.700, 1.660, 1.950, 2.070 |
| 0.075 | 1.910, 2.270, 2.110, 2.390 |
| 0.1025 | 2.220, 2.250, 3.260, 2.920 |
| 0.135 | 2.800, 2.940, 2.380, 2.700 |
| 0.185 | 2.780, 2.640, 2.710, 2.850 |
| 0.250 | 3.540, 2.860, 3.150, 3.320 |
| 0.400 | 3.910, 3.830, 4.880, 4.210 |
| 0.550 | 4.540, 4.470, 4.790, 5.680 |
| 0.750 | 6.060, 5.070, 5.000, 5.980 |
| 1.000 | 5.840, 5.790, 6.100, 7.810 |
| 1.375 | 7.310, 7.080, 7.060, 6.870 |
| 1.850 | 9.880, 10.120, 9.220, 9.960 |
| 2.500 | 11.040, 10.460, 10.880, 11.650 |
| 3.250 | 13.510, 15.470, 14.210, 13.920 |
| 4.500 | 16.070, 14.670, 14.780, 15.210 |
| 6.000 | 17.340, 16.850, 16.740, 16.870 |
| 8.250 | 18.980, 19.850, 18.750, 18.510 |
| 11.250 | 21.666, 21.218, 19.790, 22.669 |
| 15.000 | 23.206, 22.239, 22.436, 22.597 |
| 20.250 | 23.922, 24.871, 23.815, 24.871 |
| 27.500 | 25.748, 25.874, 24.907, 24.871 |
| 37.000 | 24.441, 25.874, 25.748, 27.270 |
| 50.000 | 29.580, 26.698, 26.536, 27.181 |

## Table 5.7

Estimates of $\theta$, $\sigma$ and $x_c$ based on example of Section 5.4

| | Least Squares | | Pseudo-likelihood $\mathscr{C}$=2 | | | Rodbard and Frazier | | | Modified Max. Likelihood | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\hat{x}_c$ | $\hat{\sigma}^2$ | $\hat{x}_c$ | $\hat{\theta}$ | $\hat{\sigma}^2$ | $\hat{x}_c$ | $\hat{\theta}$ | $\hat{\sigma}^2$ | $\hat{x}_c$ | $\hat{\theta}$ | $\hat{\sigma}^2$ |
| Full | .1554 | .4721 | .0790 | .4750 | .0487 | .0793 | .4757 | .0485 | .0822 | .4500 | .0487 |
| Duplicates | | | | | | | | | | | |
| 1 & 2 | .2230 | .5200 | .0728 | .7000 | .0063 | .0476 | .9404 | .0158 | .0659 | .7500 | .0063 |
| 2 & 3 | .2385 | .4013 | .1555 | .3500 | .0809 | .1870 | .1950 | .1603 | .1739 | .2500 | .1253 |
| 3 & 4 | .2513 | .5361 | .1324 | .5750 | .0356 | .1112 | .6940 | .0197 | .1104 | .7000 | .0186 |
| 1 & 4 | .1593 | .4737 | .0612 | .5500 | .0302 | .0601 | .5931 | .0252 | .0695 | .5000 | .0377 |
| 1 & 3 | .1859 | .4692 | .0938 | .4500 | .0500 | .0981 | .4233 | .0564 | .0909 | .4750 | .0452 |
| Mean | .2116 | .4800 | .1031 | .5250 | .0406 | .1008 | .5692 | .0554 | .1021 | .5350 | .0466 |
| SD | .0763 | .0471 | .0357 | .1183 | .0246 | .0491 | .2511 | .0544 | .0439 | .1997 | .0466 |

Note: Means and SDs are based only on the five reduced permutations of the data with duplicates.

## REFERENCES


Abramowitz, M. and Stegun, I. A. (1972). Handbook of Mathematical Functions. Dover Publications, New York.

Amemiya, T. (1977). A note on a heteroscedastic model. Journal of Econometrics 6, 365-370 and corrigenda 8, 265.

Bates, D. M., Wolf, D. A. and Watts, J. A. (1985). Nonlinear least squares and first order kinetics. Unpublished manuscript.

Beal, S. L. and Sheiner, L. B. (1985). Heteroscedastic nonlinear regression with pharmokinetic type data. Preprint.

Box, G. E. P. and Hill, W. J. (1974). Correcting inhomogeneity of variance with power transformation weighting. Technometrics 16, 385-389.

Box, G. E. P. and Meyer, R. D. (1986). Dispersion effects from fractional designs. Technometrics 28, 19-28.

Carroll, R. J. (1982). Adapting for heteroscedasticity in linear models, Annals of Statistics 10, 1224-1233.

Carroll, R. J., and Ruppert, D. (1982b). A comparison between maximum likelihood and generalized least squares in a heteroscedastic linear model. Journal of the American Statistical Association 77, 878-882.

Carroll, R. J., and Ruppert, D. (1982a). Robust estimation in heteroscedastic linear models, Annals of Statistics 10, 429-441.

Carroll, R. J., and Ruppert, D. (1985). Power transformations when fitting theoretical models to data. Journal of the American Statistical Association 79, 321-328.

Carroll, R. J., Ruppert, D., and Wu, C. F. J. (1986). Variance expansion and the bootstrap in generalized least squares. Preprint.

Cook, R. D. and Weisberg, S. (1982). Residuals and Influence in Regression. New York: Chapman and Hall.

Efron, B. (1985). Double exponential families and their use in generalized linear regression. Technical Report #107, Division of Biostatistics, Stanford University.

Finney, D. J. (1964). Statistical Methods in Biological Assay. London: Griffin.

Finney, D. J. (1976). Radioligand assay. Biometrics 32, 721-740.

Fuller, W. A. and Rao, J. N. K. (1978). Estimation for a linear regression model with unknown diagonal covariance matrix. Annals of Statistics 6, 1149-1158.

Giltinan, D. M., Carroll, R. J. and Ruppert, D. (1986). Some new methods for weighted regression when there are possible outliers. Technometrics 28, 219-230.

Glejser, H. (1969). A new test for heteroscedasticity. Journal of the American Statistical Association 64, 316-323.

Goldfeld, S. M. and Quandt, R. E. (1972). Nonlinear Methods in Econometrics. Amsterdam: North-Holland.

Gong, G. and Samaniego, F. J. (1981). Pseudo maximum likelihood estimation: theory and applications. Annals of Statistics 9, 861-869.

Harville, D. (1977). Maximum likelihood approaches to variance component estimation and to related problems. Journal of the American Statistical Association 79, 302-308.

Harvey, A. C. (1976). Estimating regression models with multiplicative heteroscedasticity. Econometrics 44, 461-465.

Hildreth, C. and Houck, J. P. (1968). Some estimators for a linear model with random coefficients. Journal of the American Statistical Association 63, 584-595.

Huber, P. J. (1981). Robust Statistics. New York: John Wiley and Sons.

Jacquez, J. A., Mather, F. J. and Crawford, C. R. (1968). Linear regression with non-constant, unknown error variances: sampling experiments with least squares and maximum likelihood estimators. Biometrics 24, 607-626.

Jacquez, J. A. and Norusis, M. (1973). Sampling experiments on the estimation of parameters in heteroscedastic linear regression. Biometrics 29, 771-780.

Jennrich, R. I. (1969). Asymptotics properties of nonlinear least squares estimators. Annals of Mathematical Statistics 40, 633-643.

Jobson, J. D. and Fuller, W. A. (1980). Least squares estimation when the covariance matrix and parameter vector are functionally related. Journal of the American Statistical Association 75, 176-181.

Matloff, N., Rose, R. and Tai, R. (1985). A comparison of two methods for estimating optimal weights in regression. Unpublished report.

McCullagh, P. (1983). Quasi-likelihood functions. Annals of Statistics 11, 59-67.

McCullagh, P. and Nelder, J. A. (1983). Generalized Linear Models. Chapman & Hall, New York.

Nel, D. G. (1980). On matrix differentiation in statistics. South African Statistical Journal 14,87-101.

Nelder, J. A. and Pregibon, D. (1986). An extended quasi-likelihood function. Preprint.

Oppenheimer, L., Capizzi, T. P., Weppelman, R. M. and Mehto, H. (1983). Determining the lowest limit of reliable assay measurement. Analytical Chemistry 55, 638-643.

Patterson, H. D. and Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal. Biometrika 58, 545-554.

Pritchard, D. J., Downie, J. and Bacon, D. W. (1977). Further consideration of heteroscedasticity in fitting kinetic models. Technometrics 19, 227-236.

Raab, G. M. (1981). Estimation of a variance function, with application to radioimmunoassay. Applied Statistics 30, 32-40.

Rodbard D. and Frazier, G. R. (1975). Statistical analysis of radioligand assay data. Methods of Enzymology 37, 3-22.

Rothenberg, T. J. (1984). Approximate normality of generalized least squares estimates. Econometrica 52, 811-825.

Ruppert, D. and Carroll, R. J. (1980). Trimmed least squares estimation in the linear model. Journal of the American Statistical Association 77, 828-838.

Sadler, W. A. and Smith, M. H. (1985). Estimation of the response-error relationship in immunoassay. Clinical Chemistry 31/11, 1802-1805

Serfling, R. J. (1980). Approximation Theorems of Mathematical Statistics. New York: John Wiley and Sons.

Stefanski, L. A. and Carroll, R. J. (1985). Covariate measurement error in logistic regression. Annals of Statistics 13, 1335-1351.

Taguchi, G. and Wu. Y. (1980). Introduction to Off-line Quality Control. Nagoya: Central Japan Quality Control Association.

Theil, H. (1971). Principles of Econometrics, New York: John Wiley and Sons.

Tiede, J. J. and Pagano, M. (1979). The application of robust calibration to radioimmunoassay. Biometrics 35, 567-574.

Wedderburn, R. W. M. (1974). Quasi-likelihood functions, generalized linear models and the Gauss-Newton method. Biometrika 61, 439-447

Williams, J. S. (1975). Lower bounds on convergence rates of weighted least squares to best linear unbiased estimators. In A Survey of Statistical Design and Linear Models, J. N. Srivastava, editor. Amsterdam: North Holland.

Wolter, K. M., and Fuller, W. A. (1982). Estimation of nonlinear errors-in-variables models. Annals of Statistics 10, 539-548.