#1959

# A Study of Estimation in a Normal Heteroscedastic Regression Model with Type I Censored Data

Marie Davidian*

August 1989

*Marie Davidian is Assistant Professor, Department of Statistics, North Carolina State University, Raleigh, North Carolina 27695-8203.

KEYWORDS: EM algorithm; Generalized least squares; Maximum likelihood; Small sigma asymptotics; Variance function estimation

## Abstract

Much interest has focused on the regression problem with censored data, which has usefulness in biological, pharmaceutical, industrial, survival, and failure time experiments. In these last three applications, the usual assumption is that data after a suitable transformation follow a normal, extreme value, or other distribution. Although usual methods of analysis have traditionally assumed constant dispersion for all responses, some recent interest has focused on censored regression models that accomodate heterogeneity of dispersion by modeling it as a function of location parameters, known variables, and possibly unknown additional parameters. The model can accomodate nonconstant dispersion or serve as the basis for formal assessment. For uncensored data, there is a large literature for heteroscedastic regression that has its roots in normal theory. One key issue is the choice between full normal maximum likelihood or generalized least squares methods which are inefficient but offer robustness properties and relative ease and stability of computation not enjoyed by the former method. In this article, a similar issue is considered for a normal heteroscedastic regression model with Type I censored data, and an estimation method analogous to generalized least squares is developed that, although again inefficient, may be computed by an intuitive procedure with standard software and compares favorably to maximum likelihood. An interesting application of the normal heteroscedastic regression model is to the study of a detection limit problem for assay data. These data are approximately normal, heteroscedastic, and often such that variance is small relative to the range of mean response. Often, there is a detection limit such that the response can not be observed if sufficiently small, and it is common to ignore these censored observations in the analysis. "Small sigma" asymptotic theory used by in Davidian and Carroll (1987) and others is used to show that, theoretically, this practice may not be unreasonable, although to be safe accomdation of censoring may be wise. Furthermore, under this theory, deviations from the assumed normal distribution may also be tolerated.

# 1. INTRODUCTION

Regression problems with censored response arise in industrial (Hamada and Wu 1988), biological, chemical, and pharmaceutical (Carroll and Ruppert 1988 p. 225), and survival and reliability (Anderson 1989; Lawless 1982) experiments. Much work (e.g., Mehortra and Bhattacharyya 1987; Nelson and Hahn 1972, 1973; Schmee and Hahn 1979) has assumed a normal homoscedastic linear model, focusing on analysis of log failure times assumed normal with constant variance. Other techniques (e.g., Buckley and James 1979; Cox 1972; Miller 1976; Miller and Halpern 1982) also assumed constant dispersion. Situations with censored data can arise, however, in which the constant dispersion assumption may be suspect or known to be unreaonable.

Although data transformation such as that for failure times may induce normality or other distribution at each design point, it may not always induce constant dispersion, as noted by Schneider (1986, pp. 154-8), who cautioned that ignoring heteroscedasticity in the normal model leads to bias in standard estimators. He exhibited need for a formal means for assessment of heteroscedasticity in an example in which he used an ad hoc method to investigate it (p. 164). Lu and Pantula (1989) described error heteroscedasticity in transformed normal censored degradation data. Censored data given by Scott and Jones (1985) appear normal after transformation but with some systematic increase of variability with the response. Since a goal is often prediction, it is essential that variability on the scale of the fit be investigated and adequately described (Carroll and Ruppert 1988, pp. 51-54). These authors remarked (p. 161) that a single transformation sometimes cannot simulataneously induce normality (or other distribution) and constant dispersion. Indeed, dispersion heterogeneity may be due in part to exogenous factors or covariates whose influence transformation may not affect. These problems suggest consideration of normal models in which dispersion as well as location is modeled, in part for use as a diagostic tool. For the accelerated failure time model, Anderson (1989) and Glaser (1984) modeled dispersion after transformation as a function of covariates or location and unknown parameters and estimated all parameters by maximum likelihood under the Weibull distribution and provided examples.

Measurement procedures sometimes cannot detect response levels below a known limit, so that data are Type I left censored. This is common for assay data which are often approximately normal but markedly heteroscedastic with nonlinear regression. Private communication with statisticians in the pharmaceutical industry suggests that censored observations are commonly ignored. This article arose as an attempt to consider theoretically the seriousness this practice. The detection limit problem also arises in industrial (Hamada and Wu 1988) and environmental (Shumway, Azari, and Johnson 1989) applications. These authors proposed assumption of the existence of a transformation to normality and constant variance that is either known or estimated. Since such a transformation may be unattainable, one approach would be to transform to normality and investigate dispersion on the transformed scale, as in Nair and Pregibon (1988). Data may in fact be normal but heteroscedastic on the original scale.

These problems lead us to consider a normal censored heteroscedastic regression model. For uncensored data, it is common to postulate models for mean and variance (Carroll and Ruppert 1988, chaps. 2, 3; McCullagh and Nelder 1983, chap. 8; Smyth 1989). Let $y_{ij}$ be the response at k × 1 covariate $x_i$, i = 1, ... , N, j = 1, ... , $m_i$, $m_i \geq 1$, and

$$E(y_{ij}) = f(x_i, \beta) = \mu_i; \quad Var(y_{ij}) = \sigma^2 g^2(\mu_i, \theta) = \sigma^2 g_i^2 \qquad (1.1)$$

for (p × 1) $\beta$. The variance function g depends on $\theta$ (q × 1) and $\beta$ through $\mu_i$; it may also depend on other known variables, e.g., $x_i$, but this is suppressed for simplicity. The (independent) errors $\epsilon_{ij} = (y_{ij} - \mu_i)/\{\sigma g(\mu_i, \theta)\}$ with $E(\epsilon_{ij}) = 0$, $var(\epsilon_{ij}) = 1$. The forms of f and g are suggested by the situation and empirical evidence; e.g., the power function $g(\mu, \theta) = \mu^\theta$ is common in biological and industrial applications, and

$g(\mu, \theta) = \exp(\theta\mu)$ may be appropriate for failure time data (Anderson 1989). Although the form of g may be specified, $\theta$ may be unknown and estimated along with $\beta$ and $\sigma$.

In this article we study theoretically estimation in (1.1) when the response is Type I censored based on the assumption of normality. In analogy to the uncensored case, which we review in Section 2, in Section 3 we investigate the difference between maximum likelihood estimation and a weighted least squares method that leads to a straightforward computation that can be implemented with standard software. In Section 4, we address the issue of ignoring censored data in the detection limit problem for assays through use of an asymptotic theory relevant to assay data, and find that under certain conditions, this practice may not be unreasonable. We investigate nonnormality and show explicitly the effect of slight misspecification of the variance function. Numerical evidence is given in Section 5. Proofs are in the Appendix.

## 2. A REVIEW OF ISSUES FOR HETEROSCEDASTIC REGRESSION

A common method for fitting (1.1) in the uncensored case is generalized least squares (GLS), in which an initial estimate of $\beta$ is used to estimate $\sigma$ and $\theta$, these are used to estimate $\beta$ by weighted least squares, and the process is iterated (Carroll and Ruppert 1988, chaps. 2, 3). Estimation methods for $\theta$ are investigated by Davidian and Carroll (1987), e.g., "pseudo-likelihood" (PL) is based on maximizing the normal likelihood in $(\sigma, \theta)$ given $\beta$. Estimating $(\beta, \sigma, \theta)$ by joint normal maximum likelihood (ML), is not, in contrast to the homoscedastic case ($g \equiv 1$), equivalent to GLS with PL if g depends on $\beta$. Regardless of true distribution, both are consistent, asymptotically normal and require no replication, and GLS is not efficient if data are normal, but ML estimates of $\beta$ can be difficult to compute reliably and are more sensitive to misspecification of g (Carroll and Ruppert, 1982) and nonnormality (Carroll and Ruppert, 1988, pp. 21–22) The latter authors prefer routine use of GLS over ML.

However, if variance is small relative to the range of the mean, arising in (1.1) when $\sigma$ is "small," in an asymptotic theory in which $\sigma \to 0$ as sample size increases (e.g., Davidian and Carroll 1987), GLS and ML are asymptotically equivalent.

Even with no censoring, determining a reasonable g requires adequate information. With many replicates at a few closely-spaced design points, as for some censored failure time data, it may be difficult to investigate and fit systematic heteroscedasticity, but proper characterization of variability will be crucial for prediction at points far from the design space. This suggests that analyses could be improved by attempts to include more design conditions. In biological, physical, or industrial experiments, a reasonable model for g is often available, and design conditions are more numerous. Modification of graphical detection and modeling methods (Carroll and Ruppert 1988, sec. 2.7) under censoring and appropriate variance models for specific applications deserve further study. Our goal is to investigate estimation given a reasonable variance model, so in the sequel we assume we have in mind a particular g.

## 3. ESTIMATION

Assume normality of the (possibly transformed) response, and thus standard normal $\{\epsilon_{ij}\}$, with doubly Type I censored response, i.e., at $x_i$, $y_{ij}$ is unobserved if $y_{ij} \geq R_i$ or $y_{ij} \leq L_i$ for known $L_i < R_i$ for each i. Define $m_i = m_{0i} + m_{ri} + m_{li}$, $m_{0i} = \#\{L_i < y_{ij} < R_i\}$, $m_{ri} = \#\{y_{ij} \leq L_i\}$, $m_{li} = \#\{y_{ij} \geq R_i\}$, $r_i = (R_i - \mu_i)/(\sigma g_i)$, $l_i = (L_i - \mu_i)/(\sigma g_i)$, and $\delta_{ij} = 1$ (0) if $y_{ij}$ is available (censored). Replication or uncensored observations at each $x_i$ are not required. Let $\phi(\cdot)$ and $\Phi(\cdot)$ be the standard normal density and cumulative distribution functions (cdf), $\Gamma(\cdot) = \phi(\cdot)\{1-\Phi(\cdot)\}^{-1}$ and $\Lambda(\cdot) = \phi(\cdot)\Phi(\cdot)^{-1}$. For right (left) censoring, set $l_i = -\infty$ ($r_i = \infty$). Define $\mu_{\beta i} = \partial\mu_i/\partial\beta$, $\nu_{\gamma i} = \partial(\log g_i)/\partial\gamma$, and $\tau_i = (1, \nu_{\theta i}^T)^T$. Summations and products subscripted by i (j) run from 1 to N ($m_i$), $n = \Sigma_i m_i$, $\xrightarrow{L}$ and $\xrightarrow{p}$ represent convergence in distribution and probability, and $N(d, \Delta)$ is a multivariate normal random variable with mean d and covariance $\Delta$.

### 3.1 Maximum Likelihood and Generalized Least Squares

Differentiation of the normal log-likelihood yields the estimating equations

$$\Sigma_i \Big\{ \sigma \Sigma_j \delta_{ij} \{ (y_{ij}-\mu_i)^2/(\sigma^2 g_i^2) - 1 \} \nu_{\beta i} + \Sigma_j \delta_{ij} \{ (y_{ij}-\mu_i)/(\sigma g_i^2) \} \mu_{\beta i} + \{ m_{ri} \Gamma(r_i) - m_{li} \Lambda(l_i) \} \mu_{\beta i}/g_i + \sigma \{ m_{ri} r_i \Gamma(r_i) - m_{li} l_i \Lambda(l_i) \} \nu_{\beta i} \Big\} = 0, \quad (3.1)$$

$$\Sigma_i \Big\{ \Sigma_j \delta_{ij} \{ (y_{ij}-\mu_i)^2/(\sigma^2 g_i^2) - 1 \} \tau_i + \{ m_{ri} r_i \Gamma(r_i) - m_{li} l_i \Lambda(l_i) \} \tau_i \Big\} = 0. \quad (3.2)$$

Call censored maximum likelihood (CML). The presence of $\epsilon_{ij}^2$ in (3.1), as with no censoring, makes estimation different from a "least squares" approach if g depends on $\beta$.

Consider two types of asymptotic theory: (i) $N \to \infty$, $m_i$ fixed, and (ii) $\min(m_i) \to \infty$ such that $m_i/n \to \lambda_i \in (0,1)$. Case (i) applies in situations such as assays where $\{m_i\}$ may be small. Under normality, $E(\epsilon_{ij}\delta_{ij}) = \phi(l_i) - \phi(r_i)$ and $E\{(\epsilon_{ij}^2 - 1)\delta_{ij}\} = l_i \phi(l_i) - r_i \phi(r_i)$. Thus $N^{-1} \times$ (3.1), (3.2) and $n^{-1} \times$ (3.1), (3.2) are unbiased estimating equations under (i) and (ii) as expected from likelihood theory. However, under (i) or (ii),

$$\Sigma_i \Big\{ \Sigma_j \delta_{ij} \{ (y_{ij}-\mu_i)/(\sigma g_i^2) \} \mu_{\beta i} + \{ m_{ri} \Gamma(r_i) - m_{li} \Lambda(l_i) \} \mu_{\beta i}/g_i \Big\} = 0 \quad (3.3)$$

is also unbiased and has the form of a "least squares" equation in that it depends only on $\{\epsilon_{ij}\}$. Thus, an alternative to CML would solve (3.3), (3.2), which is analogous to replacing ML by GLS with PL in uncensored data, so call this estimator CGLS. As in the uncensored case, CML and CGLS are the same if g does not depend on $\beta$.

We now verify that as with no censoring, CGLS is inefficient relative to CML when all assumptions are valid. Rather than pursue rigorous theory, we assume f and g are smooth and other necessary regularity conditions and obtain asymptotic covariance matrices. Define $J(z) = \Gamma(z) - z$, $K(z) = \Lambda(z) + z$, $R(z) = \{ z J(z) - 1 \} \phi(z)$, $L(z) =$

$\{zK(z)+1\}\phi(z)$, and $Q(z_1,z_2) = \Phi(z_1)-\Phi(z_2)$, and represent $[(\hat\beta-\beta)^T/\sigma, (\hat\sigma-\sigma)/\sigma, (\hat\theta-\theta)^T]^T$ by $\hat\eta-\eta$.

*Theorem 1.* Assume $\{\epsilon_{ij}\}$ are N(0,1), (1.1) is true, and asymptotic framework (i).

(1) Assume $(\hat\beta,\hat\sigma,\hat\theta)$ solving (3.1), (3.2) satisfy $N^{1/2}(\hat\eta-\eta)/\sigma = O_p(1)$. Then $N^{1/2}(\hat\eta-\eta)/\sigma \xrightarrow{L} N(0, I^{-1})$, where $I = \lim_{N\to\infty} N^{-1} I_N$,

$$I_N = \begin{bmatrix} I_{N,11} & I_{N,12} \\ I_{N,21} & I_{N,22} \end{bmatrix}, \tag{3.4}$$

$I_{N,11} = \Sigma_i m_i [Q(r_i,l_i)\{\mu_{\beta i}\mu_{\beta i}^T/g_i^2 + 2\sigma^2\nu_{\beta_i}\nu_{\beta i}^T\} + \sigma^2\{r_i R(r_i)+l_i L(l_i)\}\nu_{\beta i}\nu_{\beta i}^T + \sigma\{R(r_i)+L(l_i)\}(\mu_{\beta i}\nu_{\beta i}^T+\nu_{\beta i}\mu_{\beta i}^T)/g_i + \{\phi(r_i)J(r_i)+\phi(l_i)K(l_i)\}\mu_{\beta i}\mu_{\beta i}^T/g_i^2]$, $I_{N,12} = \Sigma_i m_i[\sigma\{2Q(r_i,l_i)+r_i R(r_i) + l_i L(l_i)\}\nu_{\beta i}\tau_i^T + \{R(r_i) + L(l_i)\}\mu_{\beta i}\tau_i^T/g_i]$, $I_{N,21} = I_{N,12}^T$, and $I_{N,22} = \Sigma_i m_i[\{2Q(r_i,l_i) + r_i R(r_i) + l_i L(l_i)\}\tau_i\tau_i^T]$.

(2) Assume $(\hat\beta,\hat\sigma,\hat\theta)$ solving (3.3), (3.2) satisfy $N^{1/2}(\hat\eta-\eta)/\sigma = O_p(1)$. Then $N^{1/2}(\hat\eta-\eta)/\sigma \xrightarrow{L} N(0, B^{-1}AB^{-1})$, where $A = \lim_{N\to\infty} N^{-1}A_N$, $B = \lim_{N\to\infty} N^{-1}B_N$,

$$A_N = \begin{bmatrix} A_{N,11} & A_{N,12} \\ A_{N,21} & A_{N22} \end{bmatrix}, \quad B_N = \begin{bmatrix} B_{N,11} & B_{N,12} \\ B_{N,21} & B_{N22} \end{bmatrix}, \tag{3.5}$$

$A_{N,11} = \Sigma_i m_i[Q(r_i,l_i)\mu_{\beta i}\mu_{\beta i}^T/g_i^2 + \{\phi(r_i)J(r_i)+\phi(l_i)K(l_i)\}\mu_{\beta i}\mu_{\beta i}^T/g_i^2]$, $A_{N,12} = \Sigma_i m_i[\{R(r_i)+L(l_i)\}\mu_{\beta i}\tau_i^T/g_i]$, $B_{N,11} = \Sigma_i m_i[Q(r_i,l_i)\mu_{\beta i}\mu_{\beta i}^T/g_i^2 + \sigma\{R(r_i)+L(l_i)\}\mu_{\beta i}\nu_{\beta i}^T/g_i + \{\phi(r_i)J(r_i)+\phi(l_i)K(l_i)\}\mu_{\beta i}\mu_{\beta i}^T/g_i^2]$, $A_{N,21} = A_{N,12}^T$, $A_{N,22} = I_{N,22}$, $B_{N,12} = A_{N,12}$, $B_{N,21} = I_{N,21}$, and $B_{N,22} = I_{N,22}$.

*Theorem 2.* Assume $\{\epsilon_{ij}\}$ are N(0,1), (1.1) is true, and asymptotic framework (ii).

(1) Assume $(\hat\beta,\hat\sigma,\hat\theta)$ solving (3.1), (3.2) satisfy $n^{1/2}(\hat\eta-\eta)/\sigma = O_p(1)$. Then $n^{1/2}(\hat\eta-\eta)/\sigma \xrightarrow{L} N(0, I_N^{*-1})$, where $I_N^*$ is given by (3.4) with $\{m_i\}$ replaced by $\{\lambda_i\}$.

(2) Assume $(\hat\beta,\hat\sigma,\hat\theta)$ solving (3.3) and (3.2) satisfy $n^{1/2}(\hat\eta-\eta)/\sigma = O_p(1)$. Then $n^{1/2}(\hat\eta-\eta)/\sigma \xrightarrow{L} N(0,B_N^{*-1}A_N^{*}B_N^{*-1})$, where $A_N^*$ and $B_N^*$ are given by (3.5) with $\{m_i\}$ replaced by $\{\lambda_i\}$.

In either Theorem the covariance expressions are the same if g does not depend on $\beta$, and with no censoring ($r_i = \infty$, $l_i = -\infty$) reduce to those of ML for (1) and GLS for (2). In large samples, the theory may be used to construct tests for model adequacy, heteroscedasticity, and other parameter inference. We investigate seriousness of efficiency loss for CGLS and validity of the theory in small samples in Sections 4 and 5.

### 3.2 Computation

With no censoring, both ML and GLS may be computed using nonlinear regression software as described by Carroll and Ruppert (1988, p. 72) and Giltinan and Ruppert (1989). Through use of the EM algorithm (Dempster, Laird, and Rubin 1977), CGLS enjoys this same feature while CML does not. For simplicity consider left-censoring.

Following Schneider (1986, pp. 140-141), under normality $E(y_{ij}|y_{ij} < L_i) = \mu_i - \sigma g_i \Lambda(l_i)$ and $E(y_{ij}^2|y_{ij} < L_i) = \mu_i^2 + \sigma^2 g_i^2 - 2\sigma g_i\mu_i\Lambda(l_i) - \sigma^2 g_i^2 l_i \Lambda(l_i)$. With $\tilde y_{ij} = \delta_{ij}y_{ij} + (1-\delta_{ij})E(y_{ij}|y_{ij} < L_i)$ and $\tilde r_{ij}^2 = \delta_{ij}(y_{ij}-\mu_i)^2 + (1-\delta_{ij})\{E(y_{ij}^2|y_{ij} < L_i) - 2\mu_i E(y_{ij}|y_{ij}<L_i) + \mu_i^2\}$, solving (3.1), (3.2), and (3.3) are equivalent to solving

$$\Sigma_i\Big\{\sigma\Sigma_j\{\tilde r_{ij}^2/(\sigma^2 g_i^2)-1\}\nu_{\beta i} + \Sigma_j(\tilde y_{ij}-\mu_i)\mu_{\beta i}/(\sigma g_i^2)\Big\} = 0, \quad (3.6)$$

$$\Sigma_i\Sigma_j\{\tilde r_{ij}^2/(\sigma^2 g_i^2)-1\}\tau_i = 0, \quad (3.7)$$

$$\Sigma_i\Sigma_j(\tilde y_{ij}-\mu_i)\mu_{\beta i}/(\sigma g_i^2) = 0, \quad (3.8)$$

which have the form of equations for uncensored ML (3.6, 3.7) and GLS with PL (3.6, 3.8) with censored data replaced by their expected values. For uncensored data, $\tilde r_{ij}^2$ would be replaced by $(y_{ij}-\mu_i)^2$ and $\tilde y_{ij}$ by $y_{ij}$. In this case, from Carroll and Ruppert (1988, p. 72) and Giltinan and Ruppert (1989), solving (3.6), (3.7) is equivalent to minimizing in $\beta$ and $\theta$

$$\Sigma_i \Sigma_j \{ (y_{ij} - \mu_i) \dot{g}^{1/n} / g_i \}^2, \ \dot{g} = \Pi_i g_i^{m_i}, \tag{3.9}$$

accomplished by regressing a 0 dummy variable on the function in braces and estimating $\sigma^2$ by the residual mean square. For GLS with PL, (3.7) may be solved for $\theta$ using (3.9) with $\beta$ fixed from the previous iteration of (3.8). When data are censored, this does not extend easily for CML, but, if $\beta$ is fixed, it may be verified by differentiation treating $\tilde{r}_{ij}$ as fixed that (3.7) may be solved by replacing $(y_{ij} - \mu_i)$ by $\tilde{r}_{ij}$ in (3.9) to estimate $\theta$ and estimating $\sigma^2$ by $n^{-1}\Sigma_i \Sigma_j \tilde{r}_{ij}^2 / g_i^2$, suggesting the following iterative scheme for CGLS:

(1) Obtain initial estimates $\hat{\beta}^{(0)}$, $\hat{\sigma}^{(0)}$, $\hat{\theta}^{(0)}$. This could be accomplished, for example, using only the observed data and GLS. Let k = 0.

(2) Let $\hat{l}_i^{(k)} = (L_i - \hat{\mu}_i^{(k)}) / (\hat{\sigma}^{(k)} \hat{g}_i^{(k)})$, where $\hat{\mu}_i^{(k)} = f(x_i, \hat{\beta}^{(k)})$, $\hat{g}_i^{(k)} = g(\hat{\mu}_i^{(k)}, \hat{\theta}^{(k)})$, and define $\tilde{y}_{ij}^{(k)} = \delta_{ij} y_{ij} + (1 - \delta_{ij}) \{ \hat{\mu}_i^{(k)} - \hat{\sigma}^{(k)} \hat{g}_i^{(k)} \Lambda(\hat{l}_i^{(k)}) \}$, $\tilde{r}_{ij}^{(k)2} = \delta_{ij} (y_{ij} - \hat{\mu}_i^{(k)})^2 + (1 - \delta_{ij}) \hat{\sigma}^{(k)2} \hat{g}_i^{(k)2} \{ 1 - \hat{l}_i^{(k)} \Lambda(\hat{l}_i^{(k)}) \}$.

(3) Obtain $\hat{\theta}^{(k+1)}$ by minimizing $\Sigma_i \Sigma_j \{ \tilde{r}_{ij}^{(k)} (\dot{g}^{(k)})^{1/n} / \hat{g}_i^{(k)} \}^2$, $\dot{g}^{(k)} = \Pi_i (\hat{g}_i^{(k)})^{m_i}$.

(4) Obtain $\hat{\beta}^{(k+1)}$ by minimizing $\Sigma_i \Sigma_j \{ \tilde{y}_{ij}^{(k)} - f(x_i, \hat{\beta}^{(k+1)}) \}^2 / g^2(\hat{\mu}_i^{(k)}, \hat{\theta}^{(k+1)})$ and compute $(\hat{\sigma}^{(k+1)})^2 = n^{-1} \Sigma_i \Sigma_j \tilde{r}_{ij}^{(k)2} / g^2(\hat{\mu}_i^{(k+1)}, \hat{\theta}^{(k+1)})$.

(5) Let k = k + 1 and go to (2) and repeat C−1 more times or until convergence.

Steps (3) and (4) may be done with a nonlinear regression program, although it must be able to pass through the data to compute $\dot{g}$ in (3) unless $g_i = \mu_i^\theta$, $e^{\mu_i \theta}$, or does not depend on $\beta$. If g does not depend on $\beta$, the algorithm computes CML. For homoscedasticity, this is the algorithm for maximum likelihood given by Schneider (1986, p. 142) and others, with (3) unnecessary.

Final estimation of $\sigma$ may be accomplished using the final estimates for $\beta$ and $\theta$. A bias-corrected estimate may be preferred (Aitken 1981; Carroll and Ruppert 1988, p.

73). In the homoscedastic case, Scheider (1986, p. 151) notes that correction may be taken at each iteration or at the final stage only with possibly different results. We have had success with the latter approach using $\hat{\sigma}^{*2} = \hat{\sigma}^2 n_0/(n_0-p-q)$, $n_0 = \Sigma_i m_{0i}$.

### 3.3 Other Methods

For the linear homoscedastic problem, simplified estimators have been proposed (Persson and Rootzen 1977; Schmee and Hahn 1979; Tiku 1978). These can be applied to (1.1), but do not represent much simplification, because even without censoring estimation is complex. Since the issue is ML vs. GLS in the uncensored problem, we do not pursue simplifications but investigate CML and CGLS further in the next section.

## 4. VIOLATION OF ASSUMPTIONS

As in the homoscedastic case (Schmee and Hahn 1979; Schneider 1986, pp. 159-160), from (3.1) – (3.3), consistency of CML and CGLS will not obtain if the data are not normal, censoring is ignored, or if g is misspecified. General guidelines are unlikely; however, we can gain insight by applying arguments used in the uncensored problem.

### 4.1 Misspecification of the Variance Function

Carroll and Ruppert (1982) investigated the effect of small misspecification of g on estimation of $\beta$ in the uncensored normal problem by a clever contiguity argument. They found that GLS is unaffected by small misspecification but ML is asymptotically biased. In the following, we apply their arguments to CGLS and CML estimators $\hat{\beta}_{CGLS}$ and $\hat{\beta}_{CML}$. Consider asymptotic framework (i), although similar arguments could be constructed for (ii), and assume necessary regularity conditions.

*Theorem 3.* Suppose that $\{\epsilon_{ij}\}$ are N(0,1), and assume that (a) $y_{ij} = \mu_i + \sigma g_i \epsilon_{ij}$ (1.1), when in truth (b) $y_{ij} = \mu_i + \sigma g_{i,N} \epsilon_{ij}$, where $g_{i,N} = g_i\{1 + 2B_0 N^{-1/2} h_i(\beta,\theta)\}^{-1/2}$ for functions $h_i$ and constant $B_0$. Then under (i), $N^{1/2}(\hat{\beta}_{CGLS}-\beta)/\sigma \xrightarrow{\mathcal{L}}$

$N(d_{CGLS}, B^{-1}AB^{-1})$ and $N^{1/2}(\hat\beta_{CML} - \beta)/\sigma \xrightarrow{L} N(d_{CML}, I^{-1})$, where I, B, and A are given in Theorem 1, I and B are partitioned in analogy to (3.5) and (3.6), $G = B_{11} - B_{12}B_{22}^{-1}B_{21}$, $M = I_{11} - I_{12}I_{22}^{-1}I_{21}$, $d_{CGLS} = B_0 G^{-1} \lim_{N\to\infty} N^{-1} \Sigma_i m_i h_i \Big\{ B_{12}B_{22}^{-1}\tau_i \{2Q(r_i, l_i) + r_i R(r_i) + l_i L(l_i)\} - \mu_{\beta i}\{R(r_i) + L(l_i)\}/g_i \Big\}$, and $d_{CML} = B_0 M^{-1} \lim_{N\to\infty} N^{-1} \Sigma_i m_i h_i \Big\{ \Big( I_{12} I_{22}^{-1} \tau_i - \sigma \nu_{\beta i} \Big) \{2Q(r_i, l_i) + r_i R(r_i) + l_i L(l_i)\} - \mu_{\beta i}\{R(r_i) + L(l_i)\}/g_i \Big\}$.

As expected, both methods are sensitive to small misspecification of g, and this result reduces to that of Carroll and Ruppert (1982) in the uncensored case. Following these authors, if we wish to estimate the linear combination $\alpha^T\beta$, under (a) of the Theorem, $N\,MSE(\alpha^T\hat\beta_{CML})/\sigma^2 = \alpha^T(B^{-1}AB^{-1})\alpha \geq \alpha^T I^{-1}\alpha = N\,MSE(\alpha^T\hat\beta_{CGLS})/\sigma^2$, however, under (b), $N\,MSE(\alpha^T\hat\beta_{CML})/\sigma^2 = \alpha^T(B^{-1}AB^{-1})\alpha + (\alpha^T d_{CML})^2$, $N\,MSE(\alpha^T\hat\beta_{CGLS})/\sigma^2 = \alpha^T I^{-1}\alpha + (\alpha^T d_{CGLS})^2$, so that if g is misspecified, there is no general ordering for robustness. The theory can be used with $B_0 h_i = N^{1/2}\{(g_i/g_{i,N})^2 - 1\}/2$ to construct examples favoring either estimator, so that one is not preferred to the other for greater protection against misspecification.

### 4.2 The Effect of Ignoring Censored Observations and Nonnormality

In biological and physical applications such as assays, the small $\sigma$ asymptotic theory is often practically relevant (Davidian and Carroll 1987). This justifies use of the theory as a simplifying technical device in investigation of the assay detection limit problem. The following discussion is also relevant to the homoscedastic case $g \equiv 1$.

When variability is small, we might hypothesize that mean and variance relationships might be straightforward to assess, so estimation might benefit little from additional information in censored data, and we might expect CGLS and CML and GLS and ML estimators based on ignoring censored data to behave similarly. To investigate this and the effect of nonnormality, consider (i) with $\sigma \to 0$ and left-censoring. Note that if censored observations were ignored, (3.2)−(3.4) would be replaced by

$$\Sigma_i\Big\{\sigma\,\Sigma_j\,\delta_{ij}\{(y_{ij}-\mu_i)^2/(\sigma g_i^2)-1\}\,\nu_{\beta i}+\Sigma_j\,\delta_{ij}\{(y_{ij}-\mu_i)/(\sigma g_i^2)\}\,\mu_{\beta i}\Big\}=0 \qquad (4.1)$$

$$\Sigma_i\Big\{\Sigma_j\,\delta_{ij}\{(y_{ij}-\mu_i)^2/(\sigma^2 g_i^2)-1\}\,\tau_i\Big\}=0 \qquad (4.2)$$

$$\Sigma_i\Big\{\Sigma_j\,\delta_{ij}\{(y_{ij}-\mu_i)/(\sigma g_i^2)\}\,\mu_{\beta i}\Big\}=0 \qquad (4.3)$$

Suppose $\{\epsilon_{ij}\}$ are i.i.d. with density h and cdf H, so that $E(\delta_{ij}\epsilon_{ij}) = \int \epsilon\, h(\epsilon)\, d\epsilon$, $E\{(\epsilon_{ij}^2-1)\,\delta_{ij}\} = \int (\epsilon^2-1)\, h(\epsilon)\, d\epsilon$, and $E(m_{li}) = H(l_i)$, where all integrals have limits from $l_i$ to $\infty$. H and h may depend on $\sigma$, as when $\{y_{ij}\}$ have a gamma distribution. Assuming $\{\mu_i\}$ and $\{g_i\}$ are such that $\{l_i\}$ are well-behaved and $\mu_i \neq L_i\ \forall$ i, we argue that (3.1)−(3.3) and (4.1)−(4.3) should be unbiased. Take (3.1) as an example; the others are similar. The expectation of $N^{-1} \times$ (3.1) is

$$\begin{aligned} &N^{-1}\Sigma_i\, m_i\{\textstyle\int \epsilon\, h(\epsilon)\, d\epsilon\}\,\mu_{\beta i}/g_i + \sigma N^{-1}\Sigma_i\, m_i\{\textstyle\int(\epsilon^2-1)\, h(\epsilon)\, d\epsilon\}\,\nu_{\beta i} \\ &\quad - N^{-1}\Sigma_i\, m_i H(l_i)\Lambda(l_i)\,\mu_{\beta i}/g_i - \sigma N^{-1}\Sigma_i\, m_i\, l_i\, H(l_i)\,\Lambda(l_i)\,\nu_{\beta i}. \end{aligned} \qquad (4.4)$$

Consider the first term in (4.4). Under regularity conditions, the Taylor series

$$\begin{aligned} N^{-1}\Sigma_i m_i\{\textstyle\int \epsilon h(\epsilon)\, d\epsilon\}\mu_{\beta i}/g_i &= N^{-1}\Sigma_i m_i(\mu_{\beta i}/g_i)\{\textstyle\int \epsilon h(\epsilon)\, d\epsilon\}_{\sigma=0} \\ &\quad + N^{-1}\Sigma_i m_i(\mu_{\beta i}/g_i)\,[\partial/\partial\sigma\,\{\textstyle\int \epsilon h(\epsilon)\, d\epsilon\}]_{\sigma=0}\,\sigma + o(\sigma) \end{aligned} \qquad (4.5)$$

is valid as long as the expression in brackets in the second term is finite. Since $E(\epsilon) =$ 0, the first term in (4.5) is 0 regardless of the sign of $\{l_i\}$ as long as h is well-behaved as $\sigma \to 0$. The second term in (4.4) may be expanded similiarly. Similarly, expanding the third term yields

$$\begin{aligned} N^{-1}\Sigma_i\, m_i H(l_i)\Lambda(l_i)\mu_{\beta i}/g_i &= N^{-1}\Sigma_i m_i(\mu_{\beta i}/g_i)\,\{H(l_i)\Lambda(l_i)\}_{\sigma=0} \\ + N^{-1}\Sigma_i m_i(\mu_{\beta i}/g_i)[\partial/\partial\sigma\,\{H(l_i)\}\Lambda(l_i) &+ l_i\, H(l_i)\{l_i\Lambda(l_i)+\Lambda^2(l_i)\}/\sigma]_{\sigma=0}\,\sigma + o(\sigma), \end{aligned} \qquad (4.6)$$

if the term in brackets in finite. A similar argument can be applied to the last term in (4.4). If $\{\epsilon_{ij}\}$ have a standard normal, standard double exponential or contaminated normal distribution scaled to have variance 1, or if $\{y_{ij}\}$ arise from the gamma or certain t distributions, h and H satisfy the conditions needed. For these H and h, the bracketed expressions in the second terms on the right-hand sides of (4.5) and (4.6) are in fact 0, and the first term on the right-hand side of (4.6) converges to 0 as $N \to \infty$, $\sigma \to 0$, as is the case for expansions of the other two terms of (4.4). Thus, as $N \to 0$, $\sigma \to 0$, (4.4) converges to 0 under regularity conditions so that (3.1) is an unbiased estimating equation for several true distributions other than normal. A similar argument applies to (3.2), (3.3), and (4.1)–(4.3), suggesting that if $\sigma$ is small, for a reasonable class of distributions, estimation based on (3.1)–(3.3), (4.1)–(4.3) should yield consistent estimators, even if the censored observations are ignored. We now assume consistency and regularity conditions and derive the form of the asymptotic distributions.

*Theorem 4.* Assume i.i.d. $\{\epsilon_{ij}\}$ with density h and cdf H possibly depending on $\sigma$ and (i) with $\sigma \to 0$ such that $N^{1/2}\sigma \to \gamma$, $0 \leq \gamma < \infty$. Then if h and H satisfy conditions in the Appendix, and $N^{1/2}(\hat{\eta}-\eta) = O_p(1)$ for $(\hat{\beta}, \hat{\sigma}, \hat{\theta})$ solving any of the systems (3.1)-(3.2), (3.1)-(3.3), (4.1)-(4.2), and (4.1)-(4.3), $N^{1/2}(\hat{\eta}-\eta) \xrightarrow{L} N(0, E^{-1}CE^{-1})$, where

$$C = \lim_{N\to\infty} N^{-1} \Sigma_{i,l_i<0}\, m_i \begin{bmatrix} \mu_{\beta i}\mu_{\beta i}^T/g_i^2 & \zeta\,\mu_{\beta i}\tau_i^T/g_i \\ \zeta\,\tau_i\mu_{\beta i}^T/g_i & (2+\kappa)\,\tau_i\tau_i^T \end{bmatrix}, \tag{4.7}$$

$$E = \lim_{N\to\infty} N^{-1} \Sigma_{i,l_i<0}\, m_i \begin{bmatrix} \mu_{\beta i}\mu_{\beta i}^T/g_i^2 & 0 \\ 0 & 2\,\tau_i\tau_i^T \end{bmatrix},$$

$\mathrm{Var}(\epsilon^2) = 2+\kappa$, $\kappa = 0$ for normality, and $E(\epsilon^3) = \zeta$, $\zeta \equiv 0$ for symmetric distributions.

The conditions on h and H are satisfied by the distributions above. The form of (4.7) is that of the asymptotic covariance matrix for GLS or ML in an uncensored problem with

$\sigma \to 0$ based only on design points at which no censoring occurs. Thus, if $\sigma$ is small, we may expect the usual results if the data contain outliers or are otherwise nonnormal, and/or if we ignore censored observations, so that the practice of ignoring observations that are below the detection limit may not be unreasonable for some assay data. The relevance of this theory in real applications is addressed in the next section.

## 5. NUMERICAL EVIDENCE

To investigate loss in efficiency of CGLS relative to CML with censored normal data, usefulness of the theory for small samples, and relevance of Theorem 4, we performed several small simulation studies. Although we considered examples with nonlinear and linear f and different g with qualitatively similar results to those below, for brevity we report in detail only on two simple linear regression examples, $\mu_i = \beta_0 + \beta_1 x_i$, with power variance $g_i = \mu_i^\theta$. Anderson (1989) commented that this model may not be appropriate for failure time data because it is not invariant under constant multiples of the time scale. All cases used 500 simulated data sets.

The first situation, used to investigate the assay detection limit problem, was based on assay data with N = 9 concentrations x and 5 replicates at each given in Table 1, with $\beta_0 = 0.011$, $\beta_1 = 1.502$, $\sigma = 0.025$, and $\theta = 0.8$, and detection limit $L_i = 0.10$. Parameters were estimated by CML, CGLS, and ML and GLS with PL ignoring censored observations, and standard errors were estimated using Theorem 1. Results are given in Table 1 for two error distibutions: standard normal and contaminated normal with proportion 0.10 of data from a normal distribution with standard deviation 3, and the remainder standard normal, with variance standardized to 1.

The estimators are unbiased and similar regardless of distribution, whether censoring was accomodated or ignored, or whether full maximum likelihood or an iterative generalized least squares method was used. The lack of bias in the intercept is especially interesting since the data are left censored. The standard errors based on Theorem 1 (assuming normality) agree well with the Monte Carlo evidence for $\hat{\beta}_0$ and $\hat{\beta}_1$ for all estimators and are virtually the same regardless of distribution. Those for $\hat{\sigma}$ and $\hat{\theta}$ agree well for the normal data but are too small relative to the Monte Carlo values for the nonnormal data. A calculation using $\kappa = 5.333$ shows that the theoretical standard errors for $\hat{\sigma}$ and $\hat{\theta}$ based on normality are too small by a factor roughly on the order of that suggested by Theorem 4. Mean square error (MSE) ratios based on the Monte Carlo MSEs for $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\sigma}$, and $\hat{\theta}$, respectively for CGLS relative to CML were 1.056, 1.006, 1.001, and 1.000 for normal data, and 1.059, 1.023, 1.000, and 1.001 for nonnormal data, with similar values for GLS relative to ML ignoring censored data. MSE ratios for CGLS relative to GLS with PL ignoring censoring were 0.934, 0.969, 0.958, and 0.920, with similar values for the other cases, so that efficiency loss for ignoring censored data is slight. All results support the relevance of Theorem 4. In a similar study with $\theta = 1$, we considered data generated from a gamma distribution, obtaining results very close to those for standard normal errors, again supporting the theory. Another study based on the example in Davidian, Carroll, and Smith (1988) yielded qualitatively similar results with a "larger" value for $\sigma$ of 0.088.

The second situation had right censoring with N=8, $m_i = 5$, and closely spaced design, as might be expected in failure time studies, given in Table 2, $\beta_0$ = -31.2, $\beta_1$ = 0.92, $\sigma = 0.14$, $\theta = 0.55$, and $R_i = \log_{10}(30)$. Parameters were estimated by CGLS, CML with correct g, and CML with $g \equiv 1$ (ignoring heteroscedasticity), and the analogs of these procedures ignoring censored data. Results in Table 2 are for normal data.

Estimates based on ignoring censoring are biased, so although $\sigma$ is somewhat "small," it is not sufficiently small for the result of Theorem 4. For $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\sigma}$, and $\hat{\theta}$, respectively, MSE ratios for ignoring censoring relative to accomodating it for generalized least squares methods were 0.354, 0.352, 0.756, and 0.745, with similar ratios for maximum likelihood, so taking censoring into account is crucial. For CGLS and CML, standard errors based on Theorem 1 compare reasonably to those obtained by Monte Carlo, so that the theory may be relevant for inference. MSE ratios for using CGLS relative to CML, in the same order as above, were 1.006, 1.006, 1.000, 0.999, so that CGLS, computed with a standard nonlinear regression program, performs well even when taking censoring into account is crucial. This was true for every example we considered, suggesting that the simpler CGLS may in general be a reasonable competitor to CML, as in the uncensored case. Finally, note that the MSE ratios for $\hat{\beta}_0$ and $\hat{\beta}_1$ for CML with $g \equiv 1$ relative to CGLS were 0.926 and 0.935, suggesting that from an efficiency standpoint estimation of $\beta$ may not suffer badly by ignoring heteroscedasticity. This is in part due to the closeness of the design. However, it must be kept in mind that as in the uncensored case, inference based on the unweighted fit may be erroneous if heteroscedasticity is ignored.

We also investigated misspecification of g, taking the true $g_i = 0.25 + 0.90\mu_i - 0.15\mu_i^2$ but fitting the power model. For the range of the mean, the ratio of true variance to that based on the power model was between 0.94 and 1.05, so that this represents a slight misspecification. Theorem 3 predicted that the effect of this misspecification would be virtually indistinguishable for $\hat{\beta}_0$ and $\hat{\beta}_1$, and indeed the results for CML and CGLS were very similar to those in Table 2.

## 6. DISCUSSION

In this article we investigated estimation in a normal censored heteroscedastic regression model. The method CGLS, which is the analog of generalized least squares in uncensored data and is straightforward to compute, is a reasonable competitor to the full maximum likelihood scheme, even though it is not efficient, as we have observed in numerous examples. Although theory suggests that ignoring censored assay data in the detection limit problem may be reasonable, our experience with situations like the second example, for which $\sigma$ seemed "small" enough but in which the theory did not apply, suggests that this should not be blindly accepted and accounting for observations below the detection limit may be worthwhile. The small $\sigma$ results are interesting nonetheless from a theoretical standpoint in that they lend insight into the roles of censoring and distribution in regression.

## APPENDIX: SKETCHES OF PROOFS OF THEOREMS

*Theorems 1 and 2.* The method of proof is standard. We sketch the argument for the results for CGLS; results for CML are similar and follow from likelihood theory. For (i), a Taylor series in (3.2), (3.3) along with regularity conditions yields $0 = N^{-1/2}U_N + N^{-1}V_N N^{1/2}(\hat{\eta}-\eta) + o_p(1)$, where $V_N$ is the matrix of partial derivatives, $N^{-1}V_N - N^{-1}B_N = o_p(1)$ and

$$U_N = \Sigma_i \begin{pmatrix} \Sigma_j \delta_{ij}\epsilon_{ij}\mu_{\beta i}/g_i + \{m_{ri}\Gamma(r_i)-m_{li}\Lambda(l_i)\}\mu_{\beta i}/g_i \\ \Sigma_j \delta_{ij}(\epsilon_{ij}^2-1)\tau_i + \{m_{ri}r_i\Gamma(r_i)-m_{li}l_i\Lambda(l_i)\}\tau_i \end{pmatrix}.$$

By central limit arguments, $N^{-1/2}U_N \xrightarrow{\mathcal{L}} N(0,A)$, so that the result of the Theorem follows. For (ii), replace N by n in the above and follow the same argument.

*Theorem 3.* We informally sketch the argument, which follows that in Carroll and Ruppert (1982). Letting $b_{N,i} = (1 + 2B_0N^{-1/2}h_i)$, $l_{N,i} = l_i b_{N,i}$ and $r_{N,i} = r_i b_{N,i}$, the difference of the log-likelihoods for (b) and (a) is $\log L_N \doteq -\Sigma_i\Sigma_j\delta_{ij}\epsilon_{ij}^2 B_0 h_i N^{-1/2} + \Sigma_i\Sigma_j\delta_{ij}\log b_{N,i} + \Sigma_i m_{ri}\log[\{1-\Phi(r_{N,i})\}/\{1-\Phi(r_i)\}] + \Sigma_i m_{li}\log[\{1-\Phi(l_{N,i})\}/\{1-\Phi(l_i)\}]$. Applying $\log(1+x) \doteq x - x^2/2$ for small x and the continuity of $\Phi$, further Taylor series arguments can be used to show that for large N, $\log L_N \doteq -N^{-1/2}\Sigma_i\Sigma_j B_0h_i\delta_{ij}(\epsilon_{ij}^2-1) - N^{-1/2}\Sigma_i B_0h_i\{m_{ri}r_i\Gamma(r_i)-m_{li}l_i\Lambda(l_i)\} - N^{-1}\Sigma_i\Sigma_j\delta_{ij}B_0^2h_i^2 - \frac{1}{2}N^{-1}\Sigma_i\Sigma_j B_0^2h_i^2[m_{ri}r_i\Gamma(r_i)\{r_iJ(r_i)-1\}+m_{li}l_i\Lambda(l_i)\{l_iK(l_i)+1\}]$. Assuming that $N^{-1}\Sigma_i\Sigma_j\delta_{ij}h_i^2 \xrightarrow{p} \gamma_0 \equiv$

$\lim_{N\to\infty} N^{-1}\Sigma_i h_i^2 m_i Q(r_i,l_i)$ and $N^{-1}\Sigma_i\Sigma_j h_i^2[m_{ri}r_i\Gamma(r_i)\{r_iJ(r_i)-1\} + m_{li}l_i\Lambda(l_i)\{l_iK(l_i)+1\}] \xrightarrow{p} \gamma_1 \equiv \lim_{N\to\infty} N^{-1}\Sigma_i h_i^2 m_i\{r_iR(r_i) + l_iL(l_i)\}$, a central limit argument yields that $\log L_N \xrightarrow{\mathcal{L}} N(-B_0^2\Delta/2, B_0^2\Delta)$, $\Delta = 2\gamma_0 + \gamma_1$, so that (b) is contiguous to (a) (Hajek and Sidak 1967, p. 204). We pursue the argument for CGLS; that for CML is similar. From the proof for Theorem 1, we have $G^{-1}N^{1/2}(\hat\beta-\beta) \doteq N^{-1/2}\Sigma_i\Big(-B_{12}B_{22}^{-1}\tau_i[\Sigma_j\delta_{ij}(\epsilon_{ij}^2-1) + \{m_{ri}r_i\Gamma(r_i)-m_{li}l_i\Lambda(l_i)\}] + \mu_{\beta i}[\Sigma_j\delta_{ij}\epsilon_i + \{m_{ri}\Gamma(r_i)-m_{li}\Lambda(l_i)\}]/g_i\Big)$. A detailed but straightforward central limit argument shows that under (a), $\{N^{1/2}(\hat\beta-\beta)^T, \log L_N\}^T \xrightarrow{\mathcal{L}} N\{(0^T, -B_0^2\Delta/2)^T, D\}$, where

$$D = \begin{bmatrix} B^{-1}AB^{-1} & d_{CGLS} \\ d_{CGLS}^T & B_0^2\Delta \end{bmatrix},$$

so by LeCam's third lemma (Hajek and Sidak 1967, p. 208), the result follows.

*Theorem 4.* We sketch the result for the systems (3.2), (3.3) and (4.2), (4.3), showing that the solutions of the two systems are asymptotically equivalent and satisfy (4.7); the arguments for the other equations are similar. Subscripts represent differentiation with respect to that argument. Under assumptions and regularity conditions, we have by a Taylor series that for (3.2), (3.3), $0 = N^{-1/2}U_{N,1} + N^{-1/2}U_{N,2} - N^{-1/2}U_{N,3} + N^{-1}V_{N,1}N^{1/2}(\hat\eta-\eta) + o_p(1)$ and for (4.2), (4.3) that $0 = N^{-1/2}U_{N,1} + N^{-1/2}U_{N,2} + N^{-1}V_{N,2}N^{1/2}(\hat\eta-\eta) + o_p(1)$, where

$$N^{-1/2}U_{N,1} = N^{-1/2}\Sigma_i\begin{pmatrix} \{\Sigma_j\delta_{ij}\epsilon_{ij} - m_i\int\epsilon h(\epsilon)d\epsilon\}\mu_{\beta i}/g_i \\ \{\Sigma_j\delta_{ij}(\epsilon_{ij}^2-1) - m_i\int(\epsilon^2-1)h(\epsilon)d\epsilon\}\tau_i \end{pmatrix}, \tag{A.1}$$

$$N^{-1/2}U_{N,2} = (N^{1/2}\sigma)N^{-1}\Sigma_i m_i\begin{pmatrix} \{(1/\sigma)\int\epsilon h(\epsilon)d\epsilon\}\mu_{\beta i}/g_i \\ \{(1/\sigma)\int(\epsilon^2-1)h(\epsilon)d\epsilon\}\tau_i \end{pmatrix}, \tag{A.2}$$

$$N^{-1/2}U_{N,3} = (N^{1/2}\sigma)N^{-1}\Sigma_i m_i\begin{pmatrix} \{H(l_i)\Lambda(l_i)/\sigma\}\mu_{\beta i}/g_i \\ \{l_iH(l_i)\Lambda(l_i)/\sigma\}\tau_i \end{pmatrix}, \tag{A.3}$$

$$N^{-1}V_{N,2} = -N^{-1}\Sigma_i m_i \begin{bmatrix} v_{11,i} & v_{12,i} & v_{13,i} \\ v_{21,i} & v_{22,i} & v_{23,i} \end{bmatrix}, \qquad (A.4)$$

$$N^{-1}V_{N,2} = -N^{-1}\Sigma_i m_i \begin{bmatrix} (v_{11,i}+u_{11,i}) & (v_{12,i}+u_{12,i}) & (v_{13,i}+u_{13,i}) \\ (v_{21,i}+u_{21,i}) & (v_{22,i}+u_{22,i}) & (v_{23,i}+u_{23,i}) \end{bmatrix}, \qquad (A.5)$$

$v_{11,i} = \{1-H(l_i)\}\,\mu_{\beta i}\mu_{\beta i}^T/g_i^2 + \{\sigma\int\epsilon h(\epsilon)d\epsilon\}(2\mu_{\beta i}\nu_{\beta i}^T-\mu_{\beta\beta i})/g_i$, $v_{12,i} = \{\int\epsilon h(\epsilon)d\epsilon\}\mu_{\beta i}/g_i$, $v_{13,i} = 2\{\int\epsilon h(\epsilon)d\epsilon\}\mu_{\beta i}\nu_{\theta i}^T/g_i$, $v_{21,i} = 2\{\sigma\int\epsilon^2 h(\epsilon)d\epsilon\}\tau_i\nu_{\beta i}^T + 2\{\int\epsilon h(\epsilon)d\epsilon\}\tau_i\mu_{\beta i}/g_i - \{\sigma\int(\epsilon^2-1)h(\epsilon)d\epsilon\}\tau_{\beta i}$, $v_{22,i} = 2\{\int\epsilon^2 h(\epsilon)d\epsilon\}\,\tau_i$, $v_{23,i} = 2\{\int\epsilon^2 h(\epsilon)d\epsilon\}\,\tau_i\nu_{\theta i}^T - \{\int(\epsilon^2-1)h(\epsilon)d\epsilon\}\,\tau_{\theta i}$, and defining $\Pi(x) = \Lambda(x)K(x)$, $u_{11,i} = \{\sigma H(l_i)\Lambda(l_i)\}(\mu_{\beta\beta i}-\mu_{\beta i}\nu_{\beta i}^T)/g_i + \{H(l_i)\Pi(l_i)\}\,\mu_{\beta i}\mu_{\beta i}^T/g_i^2 + \{\sigma l_i H(l_i)\Pi(l_i)\}\,\mu_{\beta i}\nu_{\beta i}^T/g_i$, $u_{12,i} = \{l_i H(l_i)\Pi(l_i)\}\mu_{\beta i}/g_i$, $u_{13,i} = \{-H(l_i)\Lambda(l_i) + l_i H(l_i)\Pi(l_i)\}\mu_{\beta i}\nu_{\theta i}^T/g_i$, $u_{21,i} = -\{l_i H(l_i)\Lambda(l_i)\}\,\tau_i\mu_{\beta i}^T/g_i + \{\sigma l_i H(l_i)\Lambda(l_i)\}(\tau_{\beta i}-\tau_i\nu_{\beta i}^T) + \{l_i H(l_i)\Pi(l_i)\}\,\tau_i\mu_{\beta i}^T/g_i + \{\sigma l_i H(l_i)\Pi(l_i)\}\,\tau_i\nu_{\beta i}^T$, $u_{22,i} = -\{l_i H(l_i)\Lambda(l_i)\}\tau_i + \{l_i^2 H(l_i)\Pi(l_i)\}\tau_i$, and $u_{23,i} = \{l_i H(l_i)\Lambda(l_i)\}(\tau_{i\theta} - \tau_i\nu_{\theta i}^T) + \{l_i^2 H(l_i)\Pi(l_i)\}\,\tau_i\nu_{\theta i}^T$. In each of (A.2)–(A.5), a Taylor series about $\sigma = 0$ is justified in if each term in braces has finite derivative at $\sigma = 0$, a fact that may be verified for each of the distributions mentioned in the text. Letting $N \to \infty$, $\sigma \to 0$, and using $N^{1/2}\sigma \to \gamma$, it is then possible to show that $N^{-1/2}U_{N,2} \to 0$ and $N^{-1/2}U_{N,3} \to 0$ as long as the terms in braces in (A.2) and (A.3) equal 0 for $\sigma = 0$, which is the case for the distributions mentioned. Similarly, it is possible to show that $N^{-1}V_{N,1} \to E$ and $N^{-1}V_{N,2} \to E$, so that the solutions to the two systems of equations are asymptotically equivalent. By central limit arguments, $N^{-1/2}U_{N_1} \xrightarrow{L} N(0, C^*)$, where

$$C^* = \lim_{N\to\infty,\sigma\to 0} N^{-1}\Sigma_i m_i \begin{bmatrix} C^*_{11,i} & C^*_{12,i} \\ C^*_{21,i} & C^*_{22,i} \end{bmatrix},$$

$C^*_{11,i} = [\int\epsilon^2 h(\epsilon)\,d\epsilon - \{\int\epsilon h(\epsilon)\,d\epsilon\}^2]\,\mu_{\beta i}\mu_{\beta i}^T/g_i^2$, $C^*_{12,i} = [\int\epsilon(\epsilon^2-1)h(\epsilon)\,d\epsilon - \{\int\epsilon h(\epsilon)\,d\epsilon\}\{\int(\epsilon^2-1)h(\epsilon)\,d\epsilon\}]\,\mu_{\beta i}\tau_i^T/g_i]$, $C^*_{21,i} = C^{*T}_{12,i}$, and $C^*_{22,i} = [\int(\epsilon^2-1)^2 h(\epsilon)\,d\epsilon - \{\int(\epsilon^2-1)h(\epsilon)\,d\epsilon\}^2]\,\tau_i\tau_i^T$. Again, we may take a Taylor series about $\sigma = 0$ for each element of $C^*$ as long as the expressions in brackets have finite derivatives at $\sigma = 0$, which is true for the distributions mentioned. Then taking the limit $\sigma \to 0$ shows $C^* = C$, completing the proof. Note: If $\{y_{ij}\}$ have a gamma distribution, then $\zeta = O(\sigma)$, $\kappa = O(\sigma^2)$, so that the asymptotic distribution is as if $\{\epsilon_{ij}\}$ were standard normal. For the other distributions mentioned, $\zeta \equiv 0$, so properties of the asymptotic distribution for $\hat{\beta}$ are the same for all distributions, while those for $\hat{\sigma}$ and $\hat{\theta}$ depend on the distribution.

## REFERENCES

Aitken, M. (1981), "A Note on the Regression Analysis of Censored Data," *Technometrics*, 23, 161-163.

Anderson, K.M. (1989), "Heterogeneous Dispersion in the Accelerated Failure Time Model," unpublished report.

Buckley, J., and James, I. (1979), "Linear Regression With Censored Data," *Biometrika*, 66, 429-436.

Carroll, R.J., and Ruppert, D. (1982), "A Comparison Between Maximum Likelihood and Generalized Least Squares in a Heteroscedastic Linear Model," *Journal of the American Statistical Association*, 77, 878-882.

----- (1988), *Transformation and Weighting in Regression*, New York: Chaprman & Hall.

Cox, D.R. (1972), "Regression Models and Life-Tables," *Journal of the Royal Statistical Society*, Ser. B, 34, 187-202.

Davidian, M., and Carroll, R.J. (1987), "Variance Function Estimation," *Journal of the American Statistical Association*, 82, 1079-1091.

Davidian, M., Carroll, R.J., and Smith, W. (1988), "Variance Functions and the Minimum Detectable Concentration in Assays," *Biometrika*, 75, 549-556.

Dempster, A.P., Laird, N.M., and Rubin, D.B. (1977), "Maximum Likelihood From Incomplete Data via the EM Algorithm (with discussion)," *Journal of the Royal Statistical Society,* Ser. B, 39, 1-38.

Giltinan, D.M., and Ruppert, D. (1989), "Fitting Heteroscedastic Regression Models to Individual Pharmacokinetic Data Using Standard Statistical Software," *Journal of Pharmacokinetics and Biopharmaceutics*, to appear.

Glaser, R.E. (1984), "Estimation for a Weibull Accelerated Life Testing Model," *Naval Research Logisitics Quarterly*, 31, 559-570.

Hajek, J., and Sidak, Z. (1967), *Theory of Rank Tests*, New York: Academic Press.

Hamada, M., and Wu. C.F.J., (1988), "Analysis of Incomplete Data From Highly Fractionated Experiments," Report No. 29, Center for Quality and Productivity Improvement, Univeristy of Wisconsin-Madison.

Lawless, J.F. (1982), *Statistical Models and Methods for Lifetime Data*, New York: John Wiley & Sons.

Lu, J.C., and Pantula, S.G. (1989), "A Repeated-Measurements Model for Over-Stressed Degradation Data," Institute of Statistics Mimeo Series Report #1958, North Carolina State University, Raleigh, North Carolina.

McCullagh, P., and Nelder, J.A. (1983), *Generalized Linear Models*, 1st Ed., New York: Chapman & Hall.

Mehrotra, K.G., and Bhattacharyya, G.K. (1987), "Estimation in a Gaussian Linear Regression Model Under Type II Censoring," *Journal of Statistical Planning and Inference*, 17, 241-250.

Miller, R.G. (1976), "Least Squares Regression With Censored Data," *Biometrika*, 63, 449-64.

Miller, R.G., and Halpern, J. (1982), "Regression With Censored Data," *Biometrika*, 69, 521-531.

Nair, V.N., and Pregibon, D. (1988), "Analyzing Dispersion Effects From Replicated Factorial Experiments," *Technometrics*, 30, 247-257.

Nelson, W., and Hahn, G.J. (1972), "Linear Estimation of a Regression Relationship From Censored Data, Part I–Simple Methods and Their Application," *Technometrics*, 14, 247-269.

――――― (1973), "Linear Estimation of a Regression Relationship From Censored

Data, Part II–Best Linear Unbiased Estimation and Theory," *Technometrics*, 15, 133-150.

Persson, T., and Rootzen, H. (1977), "Simple and Highly Efficient Estimators for a Type I Censored Normal Sample," *Biometrika*, 64, 123-128.

Schmee, J., and Hahn, G.J. (1979), "A Simple Method For Regression Analysis With Censored Data," *Technometrics*, 21, 417-432.

Schneider, H. (1986), *Truncated and Censored Samples From Normal Populations*, New York: Marcel Dekker.

Scott, S.J., and Jones, R.A. (1985), "Quantifying Seed Germimation Responses to Low Temperatures: Variation Among *Lycopersicon* SPP.," *Environmental and Experimental Botany*, 25, 129-137.

Shumway, R.H., Azari, A.S., and Johnson, P. (1989), "Estimating Mean Concentration Under Transformation for Environmental Data With Detection Limits," *Technometrics*, 31, 347-356.

Smyth, G.K. (1989), "Generalized Linear Models With Varying Dispersion," *Journal of the Royal Statistical Society,* Ser. B, 51, 47-60.

Tiku, M.L. (1978), "Linear Regression Model With Censored Observations," *Communications in Statistics, Theory and Methods*, A7(13), 1219-1232.

Table 1. Simulation Results for the Assay Detection Limit Example Based on Concentrations x = 0.0000, 0.0625, 0.1250, 0.2500, 0.5000, 1.0000, 2.5000, 5.0000, 10.0000, N = 9, $m_i = 5$, $\beta_0 = 0.011$, $\beta_1 = 1.502$, $\sigma = 0.025$, $\theta = 0.8$, and Detection Limit $L_i = 0.10$.

| | Normal $\{\epsilon_{ij}\}$ | | | | Contaminated Normal $\{\epsilon_{ij}\}$ | | | |
|---|---|---|---|---|---|---|---|---|
| | $\hat\beta_0$ | $\hat\beta_1$ | $\hat\sigma$ | $\hat\theta$ | $\hat\beta_0$ | $\hat\beta_1$ | $\hat\sigma$ | $\hat\theta$ |
| CGLS | .011 | 1.502 | .025 | .802 | .011 | 1.502 | .024 | .811 |
| | .165 | .590 | .303 | 7.734 | .159 | .595 | .467 | 12.227 |
| | .166 | .574 | .283 | 6.965 | .159 | .556 | .270 | 6.942 |
| | .035 | .091 | .036 | .108 | .047 | .147 | .055 | .119 |
| CML | .011 | 1.502 | .025 | .802 | .011 | 1.502 | .024 | .811 |
| | .169 | .592 | .303 | 7.736 | .163 | .602 | .467 | 12.223 |
| | .166 | .574 | .283 | 6.969 | .159 | .556 | .270 | 6.945 |
| | .031 | .091 | .036 | .112 | .147 | .147 | .055 | .122 |
| GLS | .012 | 1.501 | .024 | .823 | .011 | 1.502 | .023 | .824 |
| | .156 | .595 | .296 | 7.726 | .154 | .598 | .471 | 12.476 |
| | .163 | .579 | .274 | 6.920 | .157 | .558 | .264 | 6.859 |
| | .031 | 0.091 | .034 | .161 | .047 | .146 | .053 | .130 |
| ML | .012 | 1.501 | .024 | .823 | .011 | 1.502 | .023 | .823 |
| | .157 | .595 | .297 | 7.726 | .155 | .598 | .471 | 12.474 |
| | .163 | .580 | .274 | 6.920 | .157 | .558 | .264 | 6.858 |
| | .031 | .091 | .034 | .161 | .047 | .146 | .053 | .130 |

NOTE: 4 rows for each entry are: (1) estimate (Monte Carlo mean); (2) 100 x Monte Carlo SE; (3) 100 x estimate of SE based on Theorem 1 (Monte Carlo mean); (4) 100 x Monte Carlo SE for (3).

Table 2. Simulation Results for the Right-Censored Example with Normal Data Based on x = 34.229, 34.347, 34.584, 34.825, 35.069, 35.317, 35.442, 35.568, N = 8, $m_i$ = 5, $\beta_0$ = -31.2, $\beta_1$ = 0.92, $\sigma$ = 0.14, $\theta$ = 0.55, and $R_i = \log_{10}(30)$.

| | $\hat{\beta}_0$ | $\hat{\beta}_1^*$ | $\hat{\sigma}^*$ | $\hat{\theta}$ |
|---|---|---|---|---|
| CML | -30.760 | .970 | .127 | – |
| g≡1 | 1.453 | .419 | .176 | – |
| | 1.551 | .444 | .153 | – |
| | .2157 | .062 | .022 | – |
| CGLS | -31.207 | .920 | .140 | .588 |
| | 1.461 | .422 | .206 | .235 |
| | 1.468 | .424 | .194 | .217 |
| | .203 | .059 | .030 | .011 |
| CML | -31.216 | .921 | .140 | .588 |
| | 1.465 | .423 | .206 | .235 |
| | 1.468 | .424 | .194 | .217 |
| | .203 | .059 | .030 | .011 |
| ML | -28.621 | .845 | .116 | – |
| g≡1 | 1.421 | .405 | .161 | – |
| (LS) | 1.550 | .445 | .140 | – |
| | .228 | .066 | .020 | – |
| GLS | -29.210 | .862 | .128 | .462 |
| | 1.439 | .415 | .205 | .262 |
| | 1.485 | .429 | .183 | .229 |
| | .215 | .062 | .032 | .015 |
| ML | -29.232 | .863 | .128 | .463 |
| | 1.443 | .416 | .205 | .262 |
| | 1.485 | .429 | .183 | .228 |
| | .215 | .062 | .032 | .015 |

NOTE: 4 rows for each entry are: (1) estimate (Monte Carlo mean); (2) Monte Carlo SE; (3) estimate of SE based on Theorem 1 (Monte Carlo mean); (4) Monte Carlo SE for (3). * (2)-(4) are multiplied by 10.