

TR/03/86

February 1986

Improved Estimators for
The Shape Parameter in Gamma Regression

by

A.M. Al-Abood

S.T. Bakir

D.H. Young

w9259254

Improved Estimators For The Shape Parameter In Gamma Regression

by

A.M. Al-Abood,	Brunel University
S.T. Bakir,	Kuwait University
D.H. Young,	Brunel University

Summary

A regression model is considered in which the response variables have gamma distributions with a common shape parameter. A review is given of existing estimators for the shape parameter. Bias expressions for the maximum likelihood estimates of the regression coefficients and the shape parameter are developed. A new estimator for the shape parameter based on bias correction for the maximum likelihood estimator is shown to have markedly better variance and mean square error properties in small to moderate sized samples. Approximations to the low order moments of the Pearson statistic are derived for gamma regression models with general link functions. These are used for the case of a logarithmic link to develop new estimators for the shape parameter which have better moment properties than the estimators based on the Pearson statistic which have been used previously. Finally, the small sample variance and mean square error efficiencies of the estimators relative to the maximum likelihood estimator are evaluated by simulation for the case of a single explanatory variable and a logarithmic link, for a range of sample sizes less than or equal to 100.

Contents

1. Introduction.
2. Bias Approximations For The ML Estimators.
3. Bias Corrected Estimators For The Shape Parameter.
4. Approximations To The Expectation And Variance Of The Pearson Statistic.
5. Estimators For θ Based On The Pearson Statistic.
6. Monte Carlo Results.

1. INTRODUCTION

We consider a regression model in which the response variables Y_1, \dots, Y_n are independent gamma r. v. 's where Y_i has density

$$f_i(y) = \frac{1}{\Gamma(\theta)} \frac{\theta}{\mu_i} y^{\theta-1} \exp -\frac{\theta y}{\mu_i}, \quad y > 0 \quad (1.1)$$

where $\theta > 0, \mu_i > 0$. We have $E(Y_i) = \mu_i$ and if $\mu_v(Y_i)$ denotes the v th central moment of Y_i ,

$$\mu_2(Y_i) = \frac{\mu_i^2}{\theta}, \quad \mu_3(Y_i) = \frac{2\mu_i^3}{\theta^2}, \quad \mu_4(Y_i) = \frac{3\mu_i^4(\theta+2)}{\theta^3}, \quad \mu_4(Y_i) = \frac{4\mu_i^5(5\theta+6)}{\theta^4}. \quad (1.2)$$

The skewness and kurtosis coefficients are $\gamma_1 = 2\theta^{-\frac{1}{2}}$, and $\gamma_2 = 6\theta^{-1}$, respectively. When $\theta = 1$ the densities are exponential, and as $\theta \rightarrow \infty$ the densities approach the normal.

Since the shape parameter θ is assumed to be constant for all observations, each observation has the same coefficient of variation equal to $\theta^{-\frac{1}{2}}$. This contrasts with the classical linear model which assumes that the variance of the response is a constant independent of the mean.

For the i th individual, we let x_{i1}, \dots, x_{ik} denote values on k non-random regressor variates. The regression model for the mean response is written as

$$g(\mu_i) = \tilde{X}_i' \tilde{\beta}, \quad i=1, \dots, n \quad (1.3)$$

where $\tilde{X}_i' = (1, X_{i1}, \dots, X_{ik})$, $\tilde{\beta}' = (\beta_0, \beta_1, \dots, \beta_k)$ is a vector of unknown regression coefficients and $g(\bullet)$ is the link function. Two functions that are commonly used for gamma regression are the reciprocal link $\mu_i^{-1} = \tilde{X}_i' \tilde{\beta}$ and the logarithmic $\log \mu_i = \tilde{X}_i' \tilde{\beta}$. The first function provides the canonical link which yields sufficient statistics which are linear functions of the observations (McCullagh and Nelder (1983)), but has the disadvantage that restrictions on $\tilde{\beta}$ must be imposed to ensure that $\mu_i > 0$. The second link assumes that the effects of the

regressor variates are multiplicative and it maps the range of μ_i on to the whole real line.

Estimates for β and $\mu_i = g^{-1}(\underline{X}'_i \beta)$, $i = 1, \dots, n$ are usually found by maximum likelihood (ML). The ML estimator $\hat{\beta}$ is the solution of the likelihood equations

$$\sum_i \frac{1}{\hat{\mu}_i} \frac{y_i}{\hat{\mu}_i} - 1 \frac{\partial \mu_i}{\partial \beta_r} \Big|_{\beta=\hat{\beta}} = 0, \quad r = 0, 1, \dots, k \quad (1.4)$$

for any β . The covariance matrix of the parameter estimates is approximated by

$$\text{cov}_a(\hat{\beta}) \approx \theta^{-1} \sum_i \frac{1}{\mu_i^2} \frac{\partial \mu_i}{\partial \beta_r} \frac{\partial \mu_i}{\partial \beta_s}^{-1}. \quad (1.5)$$

Inference for β is usually based on a standard large sample ML approach, taking $\hat{\beta}$ to have an approximate multivariate normal distribution with mean β and covariance matrix $\text{cov}_a(\hat{\beta})$. Usually θ is

unknown and must be estimated. McCullagh and Nelder (1983) consider a number of estimators. The first is the ML estimator $\hat{\theta}$ which is the solution of

$$2n\{\log \hat{\theta} - \psi(\hat{\theta})\} = D$$

where $\psi(X) = d \log \Gamma(X) / dx$ is the digamma function and D is the deviance statistic given by

$$D = 2 \sum_i \log \frac{\hat{\mu}_i}{Y_i} + \frac{Y_i}{\hat{\mu}_i} - 1. \quad (1.7)$$

The deviance D is proportional to twice the difference between the maximum attainable value of the log-likelihood when no model is imposed on the $\{\mu_i\}$ and the log-likelihood under the fitted gamma regression model, when β is treated as known. Nelder and Wedderburn (1972) show that D simplifies to $2 \sum_i \log(\hat{\mu}_i / Y_i)$ under the power link $\mu_i^\lambda = \underline{X}'_i \beta$ and the logarithmic link.

An exact solution for $\hat{\theta}$ satisfying (1.6) has to be found iteratively. Putting $D_1 = D/(2n)$, Greenwood and Durand (1960) give the approximations

$$\begin{aligned} \hat{\theta} &= D_1^{-1}(0.5000876 + 0.1648852D_1 - 0.0544274D_1^2), \quad 0 < D_1 < 0.5772 \\ \hat{\theta} &= \frac{8.898919 + 9.059950D_1 + 0.9775373D_1^2}{D_1(17.79728 + 11.968477D_1 + D_1^2)}, \quad D_1 < 0.5772. \end{aligned} \quad (1.8)$$

The maximum errors in these approximations are claimed to be 0.0088% and 0.0054%, respectively, which for all practical purposes are negligible.

Using the asymptotic formula

$$\psi(X) = \log X - \frac{1}{2X} - \frac{1}{12X^2} + \frac{1}{210X^4} + 0 \frac{1}{X^6}, \quad (1.9)$$

if $\hat{\theta}$ is sufficiently large and terms of order $\hat{\theta}^{-2}$ are ignored, an estimator providing a simple approximation to $\hat{\theta}$ is

$$\hat{\theta}_1 = nD^{-1} \quad (1.10)$$

From Cordeiro (1983), the expectation of the deviance statistic is

$$E(D) = 2n\{\log \theta - \psi(\theta)\} - (k+1)\theta^{-1} + 0(n^{-1}) \quad (1.11)$$

where the term of order n^{-1} depends on the link function and the x -configuration. Cordeiro gives an explicit representation for this term for the power family link and the logarithmic link. Equating D to its expected value correct to $0(1)$, McCullagh and Nelder (1983) suggest that an improvement to the ML procedure is to use the estimator $\hat{\theta}_2$ given by the solution of

$$2n\{\log \hat{\theta}_2 - \psi(\hat{\theta}_2)\} - (k+1)\hat{\theta}_2^{-1} = D. \quad (1.12)$$

If terms of order $\hat{\theta}_2^{-2}$ are ignored, the estimator

$$\hat{\theta}_3 = (n-k-1)D^{-1} \quad (1.13)$$

provides an approximation to $\hat{\theta}_2$ and removes the need for iteration.

The final estimator proposed by McCullagh and Nelder is the moment estimator

$$\hat{\theta}_4 = (n - k - 1)T^{-1} \quad (1.14)$$

where $T = \sum_i \{(Y_i - \hat{\mu}_i)/\hat{\mu}_i\}^2$ is the Pearson statistic for the gamma

regression model. This estimator has the advantage of being much less sensitive to very small observations for the response variable than the estimators based on the deviance which is infinite if any observation is zero.

In this report, we propose a number of alternative estimators for the shape parameter and compare their moment properties with those of the commonly used estimators $\hat{\theta}$, $\hat{\theta}_3$ and $\hat{\theta}_4$. In section 2, bias approximations to the ML estimators of the regression coefficient vector β and are given for the gamma regression model with a general link function. In section 3 the biases are used to provide bias corrected estimators for which are shown to have much better bias, variance and mean square error properties than the ordinary ML estimator. Expressions for the mean and variance of the Pearson statistic to $O(1)$ and $O(n)$ respectively, are derived in section 4. The results are used in section 5 to examine the properties of a class of estimators $(n-a)T^{-1}$ where a is a constant. It is shown that simple modifications to the estimators based on bias correction leads to estimators with markedly improved bias, variance and mean square error properties. Finally in section 6, we report the findings of a large scale simulation investigation into the small sample properties of the estimators for the shape parameter under the logarithmic link.

2. BIAS APPROXIMATIONS FOR THE ML ESTIMATORS

Bias approximations for the ML estimators to order n^{-1} can be found applying general results given by Cox and Snell (1968). Define

$$U_r^{(i)} = \frac{\partial \log f_i(Y_i)}{\partial \beta_r}, \quad V_{rs}^{(i)} = \frac{\partial^2 \log f_i(Y_i)}{\partial \beta_r \partial \beta_s}, \quad W_{rst}^{(i)} = \frac{\partial^3 \log f_i(Y_i)}{\partial \beta_r \partial \beta_s \partial \beta_t} \quad (2.1)$$

where for notational convenience we set $\beta_\theta = \theta$. A straightforward calculation gives

$$U_r^{(i)} = \frac{\theta}{\mu_i} \frac{\partial \mu_i}{\partial \beta_r} \frac{Y_i}{\mu_i} - 1 \quad (2.2)$$

$$U_r^{(i)} = \log \theta - \psi(\theta) + \log \frac{Y_i}{\mu_i} - \frac{Y_i}{\mu_i} - 1 \quad (2.3)$$

$$V_{rs}^{(i)} = \theta \frac{1}{\mu_i} \frac{\partial^2 \mu_i}{\partial \beta_r \partial \beta_s} \frac{Y_i}{\mu_i} - 1 - \frac{1}{\mu_i^2} \frac{\partial \mu_i}{\partial \beta_r} \frac{\partial \mu_i}{\partial \beta_s} 2 \frac{Y_i}{\mu_i} - 1 \quad (2.4)$$

$$V_{r\theta}^{(i)} = \frac{1}{\mu_i} \frac{\partial \mu_i}{\partial \beta_r} \frac{Y_i}{\mu_i} - 1 \quad (2.5)$$

$$V_{\theta\theta}^{(i)} = \theta^{-1} - \psi^{(1)}(\theta) \quad (2.6)$$

$$W_{rst}^{(i)} = \frac{\theta Y_i}{\mu_i^3} \frac{4}{\mu_i} \frac{\partial \mu_i}{\partial \beta_r} \frac{\partial \mu_i}{\partial \beta_s} \frac{\partial \mu_i}{\partial \beta_t} - \frac{\partial \mu_i}{\partial \beta_r} \frac{\partial^2 \mu_i}{\partial \beta_s \partial \beta_t} - \frac{\partial \mu_i}{\partial \beta_s} \frac{\partial^2 \mu_i}{\partial \beta_r \partial \beta_t} - \frac{\partial \mu_i}{\partial \beta_t} \frac{\partial^2 \mu_i}{\partial \beta_r \partial \beta_s} \\ + \theta \frac{Y_i}{\mu_i} - 1 \frac{\partial^2}{\partial \beta_t} \frac{1}{\mu_i} \frac{\partial^2 \mu_i}{\partial \beta_r \partial \beta_s} - \frac{1}{\mu_i^2} \frac{\partial \mu_i}{\partial \beta_r} \frac{\partial \mu_i}{\partial \beta_s} \quad (2.7)$$

$$W_{rs\theta}^{(i)} = \frac{1}{\mu_i} \frac{\partial^2 \mu_i}{\partial \beta_r \partial \beta_s} \frac{Y_i}{\mu_i} - 1 - \frac{1}{\mu_i^2} \frac{\partial \mu_i}{\partial \beta_r} \frac{\partial \mu_i}{\partial \beta_s} 2 \frac{Y_i}{\mu_i} - 1 \quad (2.8)$$

$$W_{r\theta\theta}^{(i)} = 0, \quad W_{\theta\theta\theta}^{(i)} = -\{\theta^{-2} + \beta \psi^{(2)}(\theta)\} \quad (2.9)$$

for $r, s, t = 0, 1, \dots, k$, where $\psi^{(v)}(\cdot)$ is the v th derivative of the digamma function. Set

$$I_{rs} = E \sum_i V_{rs}^{(i)}, \quad K_{rst} = E \sum_i W_{rst}^{(i)}, \quad J_{r,st} = E \sum_i U_r^{(i)} V_{st}^{(i)}. \quad (2.10)$$

The elements in the information matrix are

$$I_{rs} = \theta \sum_i \frac{1}{\mu_i^2} \frac{\partial \mu_i}{\partial \beta_r} \frac{\partial \mu_i}{\partial \beta_s}, \quad I_{r\theta} = 0, \quad I_{\theta\theta} = n\{\psi^{(1)}(\theta) - \theta^{-1}\} \quad (2.11)$$

for $r, s = 0, 1, \dots, k$. It is seen that the estimates of the regression

coefficients are asymptotically independent of $\hat{\theta}$. Also to order n^{-1} , we have

$$\text{var}(\hat{\theta}) = n^{-1} \{ \psi^{(1)}(\theta) - \theta^{-1} \}^{-1} \quad (2.12)$$

which is independent of the link function for μ_i . Our later results show that (2.12) gives a serious underestimation of the variances even for moderate sized samples, under the logarithmic link. For the K functions we have

$$K_{rst} = \theta \sum_i \frac{1}{\mu_i^2} \frac{4}{\mu_i} \frac{\partial \mu_i}{\partial \beta_r} \frac{\partial \mu_i}{\partial \beta_s} \frac{\partial \mu_i}{\partial \beta_t} - \frac{\partial \mu_i}{\partial \beta_t} \frac{\partial^2 \mu_i}{\partial \beta_s \partial \beta_t} - \frac{\partial \mu_i}{\partial \beta_s} \frac{\partial^2 \mu_i}{\partial \beta_r \partial \beta_t} - \frac{\partial \mu_i}{\partial \beta_t} \frac{\partial^2 \mu_i}{\partial \beta_r \partial \beta_s} \quad (2.13)$$

$$K_{rs\theta} = - \sum_i \frac{1}{\mu_i^2} \frac{\partial \mu_i}{\partial \beta_r} \frac{\partial \mu_i}{\partial \beta_s}, \quad K_{r\theta\theta} = 0, \quad K_{\theta\theta\theta} = -n \{ \theta^{-2} + \psi^{(2)}(\theta) \} \quad (2.14)$$

for $r, s, t = 0, 1, \dots, k$. For evaluation of the J functions we use the result that

$$E \frac{Y_i}{\mu_i} \log \frac{Y_i}{\mu_i} = \{ \Gamma(\theta) \theta^V \}^{-1} \{ d\Gamma(\theta + V) / d\theta - \log \theta \Gamma(\theta + V) \} \quad (2.15)$$

and obtain

$$J_{r,st} = \theta \sum_i \frac{1}{\mu_i^2} \frac{\partial \mu_i}{\partial \beta_r} \frac{\partial^2 \mu_i}{\partial \beta_s \partial \beta_t} - \frac{2}{\mu_i} \frac{\partial \mu_i}{\partial \beta_s} \frac{\partial \mu_i}{\partial \beta_t} \quad (2.16)$$

$$J_{r,s\theta} = \sum_i \frac{1}{\mu_i^2} \frac{\partial \mu_i}{\partial \beta_r} \frac{\partial \mu_i}{\partial \beta_s}, \quad (2.17)$$

$$J_{r,\theta\theta} = J_{\theta,st} = J_{\theta,s\theta} = J_{\theta,\theta\theta} = 0. \quad (2.18)$$

From Cox and Snell (1968), we have to order n^{-1}

$$b_{\hat{\theta}} = \frac{1}{2} \sum_s \sum_t \sum_u I^{\theta s} I^{tu} (K_{stu} + 2J_{t,su}) \quad (2.19)$$

where I^{rs} denotes the element corresponding to r and s in the inverse of the information matrix. Each summation in (2.19) is over the values $0, 1, \dots, k$. Since $I^r = 0$ for $r = 0, 1, \dots, k$ and $J_{t,u} = -K_{tu} = -I_{tu}$, we obtain

$$\begin{aligned} b_{\hat{\theta}} &= \frac{1}{2} I^{\theta\theta} \theta^{-1} \sum_{t=0}^k \sum_{u=0}^k I^{tu} I_{tu} + I^{\theta\theta} K_{\theta\theta\theta} \\ &= \frac{1}{2} I^{\theta\theta} \left\{ (k+1)\theta^{-1} + I^{\theta\theta} K_{\theta\theta\theta} \right\} \\ &= \frac{1}{2n\{\theta\psi^{(1)}(\theta) - 1\}} \left[k+1 - \frac{\theta^2 \psi^{(2)}(\theta) + 1}{\theta\psi^{(1)}(\theta) - 1} \right]. \end{aligned} \quad (2.20)$$

For the estimates of the regression coefficients the biases $b_r = E(\hat{\beta}_r) - \beta_r$ to order n^{-1} are given by

$$b_r = \frac{1}{2} \sum_{s=0}^k \sum_{t=0}^k \sum_{u=0}^k I^{rs} I^{tu} (K_{stu} + 2J_{t,su}), \quad r = 0, 1, \dots, k. \quad (2.21)$$

For the logarithmic link, these biases take on particularly simple forms if it is assumed that the x 's satisfy the centring conditions

$$\sum_i X_{ir} = 0 \quad \text{for } r = 1, \dots, k. \quad (2.22)$$

In this case $I^{\theta\theta} = 0$ for $r = 1, \dots, k$ and $K_{stu} = -j_{t,su} = \theta \sum_i x_{is} x_{it} x_{iu}$

and we obtain

$$b_0 = -\frac{(k+1)}{2n\theta}, \quad b_r = -\frac{1}{2\theta} \sum_{s=1}^k \sum_{t=1}^k \sum_{u=1}^k M^{rs} M^{tu} \sum_i x_{is} x_{it} x_{iu} \quad (2.23)$$

for $r = 1, \dots, k$, where M^{rs} denotes the (r, s) th element in the inverse of the sum of cross products matrix $\tilde{M} = ((\sum_i x_{ir} x_{is}))$. In

the case of a single regressor variate we have

$$b_1 = -(2\theta)^{-1} \sum_i x_i^3 / (\sum_i x_i^2)^2 \quad \text{which is zero if the } x \text{ values are equally}$$

spaced.

3. BIAS CORRECTED ESTIMATORS FOR THE SHAPE PARAMETER

From (2.20) and (2.12), the bias and variance of the ML estimator of the shape parameter to order n^{-1} may be written as

$$b_{\hat{\theta}} = n^{-1}h_1(\theta), \quad \text{var}(\hat{\theta}) = n^{-1}h_2(\theta) \quad (3.1)$$

where

$$h_1(\theta) = \frac{1}{2\{\theta\psi^{(1)}(\theta) - 1\}} \{k+1 - \frac{\theta^2\psi^{(2)}(\theta)+1}{\theta\psi^{(1)}(\theta) - 1}\} \quad (3.2)$$

$$h_2(\theta) = \{\psi^{(1)}(\theta) - \theta^{-1}\}^{-1}. \quad (3.3)$$

The approximating bias depends only on n , and the number of regressor variates k and is independent of the link function for the mean and the values of the regressor variates. If n is large, then using the expansions

$$\theta\psi^{(1)}(\theta) - 1 = (2\theta)^{-1} + (6\theta^2)^{-1} + O(\theta^{-4}) \quad (3.4)$$

$$\theta\psi^{(2)}(\theta) + 1 = -\theta^{-1} - (2\theta^2)^{-1} + O(\theta^{-4}) \quad (3.5)$$

and neglecting terms $O(n^{-2})$, we obtain

$$b_{\hat{\theta}} = n^{-1} \left\{ (k+3)\theta - \frac{k+2}{3} + \frac{k+1}{9\theta} \right\}. \quad (3.6)$$

When $k = 0$, this bias reduces to that given by Bowman and Shenton (1968) for the case when no regressor variates are present.

Making a direct correction for the bias of $\hat{\theta}$, we are led to consider the estimator

$$\hat{\theta}_5 = \hat{\theta} - n^{-1}h_1(\hat{\theta}) \quad (3.7)$$

or neglecting terms of $O(n^{-2})$, the estimator

$$\hat{\theta}_6 = \hat{\theta} \left\{ 1 - \frac{k+3}{n} + \frac{k+2}{3n} - \frac{(k+1)}{9n\hat{\theta}} \right\}. \quad (3.8)$$

The bias of $\hat{\theta}_5$ is $O(n^{-2})$.

We now compare the variance and mean square error properties of the estimators $\hat{\theta}_5$ and $\hat{\theta}_6$ with those of the ML estimator. To do this it is convenient to consider a general estimator of the form

$$\hat{\theta}^* = \hat{\theta} + n^{-1}a(\hat{\theta}) \quad (3.9)$$

where $a(\hat{\theta})$ is a function of $\hat{\theta}$ which is independent of n . We have

$$b_{\hat{\theta}^*} = b_{\hat{\theta}} + n^{-1}a(\theta) + 0(n^{-2}). \quad (3.10)$$

Since $\text{var}(\hat{\theta}^*) = \text{var}(\hat{\theta}) + 2n^{-1}\text{cov}(\hat{\theta}, a(\hat{\theta})) + 0(n^{-3})$ and $\text{cov}(\hat{\theta}, a(\hat{\theta})) = a'(\theta)\text{var}(\hat{\theta}) + 0(n^{-2})$, we obtain

$$\text{var}(\hat{\theta}) - (\text{var}(\hat{\theta}^*)) = -2n^{-2}h_2(\theta)a'(\theta) + 0(n^{-3}). \quad (3.11)$$

The proportionate reduction in the variance to $0(n^{-1})$ is

$$R_v(\theta) = -2n^{-1}a'(\theta). \quad (3.12)$$

Using (3.10) and (3.11) we obtain

$$\text{mse}(\hat{\theta}) - \text{mse}(\hat{\theta}^*) = -n^{-2}[2h_2(\theta) + a(\theta)\{2h_1(\theta) + a(\theta)\}] + 0(n^{-3}). \quad (3.13)$$

The proportionate reduction in the mean square error to $0(n^{-1})$ is

$$R_m(\theta) = -n^{-1}[2a'(\theta) + a(\theta)\{2h_1(\theta) + a(\theta)\}h_2^{-1}(\theta)] \quad (3.14)$$

Values of $nR_v(\)$ and $nR_m(\)$ given by (3.12) and (3.14), respectively are shown in table 1 for the estimator $\hat{\theta}_5$, putting $a(\) = -h_1(\)$ for $\theta = 0.25(0.25)1.00(0.50)3.00(1.00)6.00$ and $k = 1, 2, 3, 4$. The corresponding values for $\hat{\theta}_6$, putting

$a(\theta) = -(k+3)\theta + \frac{1}{3}(k+2) - \frac{k+1}{9\theta}$, are shown in Table 2.

Table 1Values of $nR_v(\cdot)$ and $nR_m(\cdot)$ for the estimator $\hat{\theta}_5$.

	$nR_v(\theta)$				$nR_m(\theta)$			
	k=1	k=2	k=3	k=4	k=1	k=2	k=3	k=4
0.25	5.86	7.25	8.64	10.03	11.07	15.27	20.09	25.51
0.50	6.80	8.42	10.03	11.65	12.77	17.58	23.07	29.24
0.75	10.08	12.25	14.42	16.58	17.65	23.68	30.51	38.13
1.00	7.55	9.37	11.20	13.02	14.31	19.76	25.99	32.99
1.50	7.78	9.69	11.59	13.50	14.92	20.68	27.27	34.69
2.00	7.88	9.82	11.76	13.71	15.22	21.15	27.95	35.61
2.50	7.92	9.88	11.85	13.81	15.39	21.44	28.36	36.18
3.00	7.95	9.92	11.89	13.87	15.51	21.62	28.64	36.56
4.00	7.97	9.96	11.94	13.93	15.64	21.85	28.99	37.05
5.00	7.98	9.97	11.96	13.95	15.72	21.98	29.19	37.34
6.00	7.99	9.98	11.98	13.97	15.77	22.07	29.33	37.53

Table 2Values of $nR_v(\cdot)$ and $nR_m(\cdot)$ for the estimator $\hat{\theta}_6$.

	$nR_v(\theta)$				$nR_m(\theta)$			
	k=1	k=2	k=3	k=4	k=1	k=2	k=3	k=4
0.25	0.89	-0.67	-2.22	-3.78	5.21	4.45	3.13	1.27
0.50	6.22	7.33	8.44	9.56	12.19	16.48	21.44	27.07
0.75	7.21	8.81	10.42	12.02	14.67	20.12	26.36	33.39
1.00	7.56	9.33	11.11	12.89	14.32	19.72	25.90	32.86
1.50	7.80	9.70	11.60	13.51	14.94	20.69	27.28	34.69
2.00	7.89	9.83	11.78	13.72	15.23	21.17	27.96	35.62
2.50	7.93	9.89	11.86	13.82	15.40	21.45	28.38	36.19
3.00	7.95	9.93	11.90	13.88	15.51	21.63	28.65	36.57
4.00	7.97	9.96	11.94	13.93	15.64	21.85	28.99	37.05
5.00	7.98	9.97	11.96	13.96	15.72	21.98	29.19	37.34
6.00	7.99	9.98	11.98	13.97	15.77	22.07	29.33	37.53

The results in tables 1 and 2 have to be used with caution, particularly if n is not large, since terms $O(n^{-2})$ will not be negligible. They are clearly not applicable if $nR_v(\cdot) > n$ or $nR_m(\cdot) > n$. However, simulation results given in section 6 for the case $k=1$ with a logarithmic link function, suggest that they provide a useful guide for $n \geq 50$.

The results indicate that the difference in variance and mean square error performance between the bias reduction estimators $\hat{\theta}_5$ and $\hat{\theta}_6$ will be negligible for $\lambda \geq 1.0$. The estimator $\hat{\theta}_5$ will have a markedly better performance than the uncorrected ML estimator $\hat{\theta}$ for all values of λ . The same is true of $\hat{\theta}_6$ except at very small values of λ . For $n = 50$, $k = 1$ and $\lambda = 1$, proportionate reduction in variance is approximately 16% and the proportionate reduction in mean square error is approximately 30%. With $k = 4$, these percentages rise to 27% and 70%, respectively.

Finally, we consider the bias and variance properties of the estimators $\hat{\theta}_1 = nD^{-1}$ and $\hat{\theta}_3 = (n-k-1)D^{-1}$. The expectation of D to $O(1)$ is given by (1.1). To order n^{-1} , $\text{var}\{\log \hat{\theta} - \psi(\hat{\theta}) = nh_2(\hat{\theta})\}^{-1}$. Use of (1.6) gives to order n^{-1}

$$\text{var}(D) = 4n/h_2(\theta). \quad (3.15)$$

Using the approximation

$$E(\hat{\theta}_1) = n\{E(D)\}^{-1}\{1 + \text{var}(D)/E^2(D)\} \quad (3.16)$$

the bias of $\hat{\theta}_1$ to order n^{-1} is

$$b_{\hat{\theta}_1} = \frac{1}{2}\eta(\theta) - \theta + \frac{\eta^2(\theta)}{2n} - \frac{k+1}{2\theta} + \frac{\eta(\theta)}{h_2(\theta)} \quad (3.17)$$

where $\eta(\theta) = \{\log \theta - \psi(\theta)\}^{-1}$. A similar approach gives the bias of $\hat{\theta}_3$ to order n^{-1} as

$$b_{\hat{\theta}_3} = \frac{1}{2}\eta(\theta) - \theta + \frac{\eta^2(\theta)}{2n} - \frac{k+1}{2\theta} + \frac{\eta(\theta)}{h_2(\theta)} - \frac{(k+1)\eta(\theta)}{2n}. \quad (3.18)$$

To order n^{-1} ,

$$\text{var}(\hat{\theta}_1) = \text{var}(\hat{\theta}_3) = \eta^4(\theta)/\{4nh_2(\theta)\}. \quad (3.19)$$

From (3.17) and (3.18), it is seen that the estimators $\hat{\theta}_1$ and $\hat{\theta}_3$ are not asymptotically unbiased, the biases approaching $\frac{1}{2}\eta(\theta) - \theta$ as $n \rightarrow \infty$. Use of (3.19) and (3.1) shows that the asymptotic variance efficiency of $\hat{\theta}_1$ and $\hat{\theta}_3$ relative to the ML estimator $\hat{\theta}$, which is asymptotically unbiased, is

$$E_{\theta}^{(1)} = 4h_2^2(\theta)/\eta^4(\theta). \tag{3.20}$$

Values of the asymptotic biases of $\hat{\theta}_1$ and $\hat{\theta}_3$ are shown in table 3 and values of the asymptotic variance efficiencies $E_{\theta}^{(1)}$ are given in table 4 for $\theta = 0.25(0.25)1.00, 1.50, 2.00, 3.00, 4.00(2.00)10.00$. The asymptotic bias is negative for all θ and is approximately -0.16 for $\theta = 2$. The asymptotic variance efficiency $E_{\theta}^{(1)}$ is a decreasing function of θ and approaches 1 fairly rapidly.

Table 3

Asymptotic biases of estimators $\hat{\theta}_1$ and $\hat{\theta}_3$

θ :	0.25	0.50	0.75	1.00	1.50	2.00
$\frac{1}{2}\eta(\theta) - \theta$:	-0.074	-0.106	-0.124	-0.134	-0.145	-0.151
θ :	3.00	4.00	6.00	8.00	10.00	
$\frac{1}{2}\eta(\theta) - \theta$:	-0.156	-0.159	-0.161	-0.162	-0.164	

Table 4

Asymptotic variance efficiency of $\hat{\theta}_1$ and $\hat{\theta}_3$ relative to $\hat{\theta}$

θ :	0.25	0.50	0.75	1.00	1.50	2.00	3.00	4.00	6.00	8.00	10.00
$E_{\theta}^{(1)}$:	1.496	1.210	1.298	1.068	1.031	1.017	1.007	1.004	1.002	1.001	1.001

4. APPROXIMATIONS TO THE EXPECTATION AND VARIANCE OF THE PEARSON STATISTIC

In this section, we obtain expressions for $E(T)$ to $O(1)$ and $\text{var}(T)$ to $O(n)$ which will be used in the next section to examine moment properties of estimators for β based on T . The expressions are derived applying general results given by Cox and Snell (1968).

Define

$$R_i = \frac{Y_i}{\hat{\mu}_i} - 1 = h_i(Y_i, \hat{\beta}), \quad \varepsilon_i = \frac{Y_i}{\mu_i} - 1 = h_i(Y_i, \beta) \quad (4.1)$$

and let

$$H_r^{(i)} = \frac{\partial h_i(Y_i, \hat{\beta})}{\partial \beta_r}, \quad H_{rs}^{(i)} = \frac{\partial^2 h_i(Y_i, \beta)}{\partial \beta_r \partial \beta_s}. \quad (4.2)$$

Since the $\{\varepsilon_i\}$ are independently and identically distributed, we have to $O(n^{-1})$

$$E(R_i) = E(\varepsilon_i) + \sum_{r=0}^k b_r E(H_r^{(i)}) + \sum_{r=0}^k \sum_{s=0}^k I^{rs} E(H_r^{(i)} U_s^{(i)} + \frac{1}{2} H_{rs}^{(i)}) \quad (4.3)$$

where b_r is the bias of $\hat{\beta}_r$ correct to $O(n^{-1})$. We have

$$H_r^{(i)} = -\frac{2(Y_i - \mu_i)Y_i}{\mu_i^3} \frac{\partial \mu_i}{\partial \beta_r} \quad (4.4)$$

$$H_{rs}^{(i)} = \frac{2Y_i(3Y_i - 2\mu_i)}{\mu_i^4} \frac{\partial \mu_i}{\partial \beta_r} \frac{\partial \mu_i}{\partial \beta_s} - \frac{2Y_i(Y_i - \mu_i)}{\mu_i^3} \frac{\partial^2 \mu_i}{\partial \beta_r \partial \beta_s} \quad (4.5)$$

$$U_s^{(i)} = \frac{\theta(Y_i - \mu_i)}{\mu_i^2} \frac{\partial \mu_i}{\partial \beta_s} \quad (4.6)$$

A straightforward calculation yields

$$E(H_r^{(i)}) = -\frac{2}{\theta \mu_i} \frac{\partial \mu_i}{\partial \beta_r}, \quad (H_r^{(i)} U_s^{(i)}) = -\frac{2(2+\theta)}{\theta \mu_i^2} \frac{\partial \mu_i}{\partial \beta_r} \frac{\partial \mu_i}{\partial \beta_s} \quad (4.7)$$

$$E(H_{rs}^{(i)}) = \frac{2(3+\theta)}{\theta \mu_i^2} \frac{\partial \mu_i}{\partial \beta_r} \frac{\partial \mu_i}{\partial \beta_s} - \frac{2}{\theta \mu_i} \frac{\partial^2 \mu_i}{\partial \beta_r \partial \beta_s}. \quad (4.8)$$

Using (4.3), we have to $O(n^{-1})$

$$E(R_i) = \frac{1}{\mu_i} - \frac{2}{\mu_i^2} \sum_{r=0}^k \frac{b_r}{\mu_i} \frac{\partial \mu_i}{\partial r} - \sum_{r=0}^k \sum_{s=0}^k I_{rs} \frac{(1+r)}{\mu_i^2} \frac{\partial \mu_i}{\partial s} + \frac{1}{\mu_i} \frac{\partial^2 \mu_i}{\partial r \partial s} \quad (4.9)$$

Since $T = \sum_{i=1}^n R_i$ and using the result

$$\sum_{r=0}^k \sum_{s=0}^k I_{rs} \sum_i \frac{1}{\mu_i^2} \frac{\partial \mu_i}{\partial r} \frac{\partial \mu_i}{\partial s} = -1 \sum_{r=0}^k \sum_{\text{all } s} I_{rs} I_{rs} = -1(k+1)$$

we obtain to $O(1)$,

$$E(T) = \frac{n}{\mu} - \frac{(k+1)(k+1)}{2} - \frac{2}{\mu} \sum_{r=0}^k b_r \sum_i \frac{1}{\mu_i} \frac{\partial \mu_i}{\partial r} - \frac{1}{\mu} \sum_{r=0}^k \sum_{s=0}^k r^{rs} \sum_i \frac{1}{\mu_i} \frac{\partial^2 \mu_i}{\partial r \partial s}. \quad (4.10)$$

we now consider two special cases

a) Logarithmic Link

For the link $\log \mu_i = x_{i\sim}$, we have

$$\mu_i^{-1} \frac{\partial \mu_i}{\partial r} = x_{ir}, \quad \mu_i^{-1} \frac{\partial^2 \mu_i}{\partial r \partial s} = x_{ir} x_{is}, \quad I_{rs} = \sum_i x_{ir} x_{is}$$

giving

$$E(T) = \frac{n}{\mu} - \frac{(k+1)(k+1)}{2} - \frac{2}{\mu} \sum_{r=0}^k b_r \sum_i x_{ir}. \quad (4.11)$$

if $\sum_i x_{ir} = 0$ for $r = 1, \dots, k$ then $b_0 = -(k+1)/(2n)$

$$E(T) = \frac{n}{\mu} - \frac{(k+1)(k+1)}{2} \quad (4.12)$$

to $O(1)$

b) Power Link

For the power link $\mu_i = x_{i\sim}^2$, we have

$$\mu_i^{-1} \frac{\partial \mu_i}{\partial r} = \frac{x_{ir}}{\mu_i}, \quad \mu_i^{-1} \frac{\partial^2 \mu_i}{\partial r \partial s} = -\frac{(-1)x_{ir}x_{is}}{2\mu_i^2}, \quad I_{rs} = \sum_i \frac{x_{ir}x_{is}}{2\mu_i^2}$$

giving

$$E(T) = \frac{n}{2} - \frac{(k+1)(n+2)}{2} - \frac{2}{n} \sum_{r=0}^k b_r \sum_i \frac{x_{ir}}{\mu_i}. \quad (4.13)$$

To find the leading term of $O(n)$ in $\text{var}(T)$, we write $E(R_i) = n^{-1} + a_i$ where a_i is $O(n^{-1})$ and given by (4.9). To $O(1)$ we have

$$\text{var}(R_i) = \text{var}(U_i) = 3(n+2)^{-3} \quad (4.14)$$

and to $O(n^{-1})$ we have

$$\text{cov}(R_i, R_j) = \sum_{r=0}^k \sum_{s=0}^k I^{rs} E\{ U_r^{(j)} U_s^{(i)} + U_r^{(i)} U_s^{(j)} + U_r^{(i)} U_s^{(j)} \} \quad (4.15)$$

A straightforward calculation gives

$$E(U_r^{(j)} U_s^{(i)}) = -\frac{48^{-2}}{\mu_i \mu_j} \frac{\partial \mu_j}{\partial r} \frac{\partial \mu_i}{\partial s}, \quad E(U_r^{(i)} U_s^{(j)}) = \frac{48^{-2}}{\mu_i \mu_j} \frac{\partial \mu_i}{\partial r} \frac{\partial \mu_j}{\partial s}$$

and hence

$$\text{cov}(R_i, R_j) = -\frac{4}{n^2} \sum_r \sum_s \frac{r^{rs}}{\mu_i \mu_j} \frac{\partial \mu_j}{\partial r} \frac{\partial \mu_i}{\partial s} + O(n^{-2}). \quad (4.16)$$

Hence

$$\begin{aligned} \text{var}(T) &= \sum_i \text{var}(R_i) + \sum_{i \neq j} \text{cov}(R_i, R_j) \\ &= 2n \frac{(n+3)}{3} - \frac{4}{n^2} \sum_{r=0}^k \sum_{s=0}^k I^{rs} \sum_{i=1}^n \frac{1}{\mu_i} \frac{\partial \mu_i}{\partial r} \sum_{i=1}^n \frac{1}{\mu_i} \frac{\partial \mu_i}{\partial s}. \end{aligned} \quad (4.17)$$

to $O(n)$.

For the logarithmic link function with the x 's centred such that

$\sum_i x_{ir} = 0$ for $r = 1, \dots, k$, we obtain

$$\text{var}(T) = \frac{2n}{3} (n+1) + O(1) \quad (4.18)$$

No useful simplification occurs for the power link.

5. ESTIMATORS FOR θ BASED ON THE PEARSON STATISTIC

In this section, we shall restrict attention to the gamma regression model with a logarithmic link and assume, without loss of generality, that the regressor variables are centred such that $\sum_i x_{ir} = 0$ for $r = 1, \dots, k$. We consider a class of estimators which include the McCullagh-Nelder estimator \hat{a}_4 as a special case. It is shown that simple bias adjustments to the estimators leads to estimators with markedly better bias, variance and mean square error properties in small to moderate sized samples.

Consider the estimator

$$\hat{a}(a) = (n-a)T^{-1} \quad (5.1)$$

where a is a constant which is small compared with n . Using the standard approximations

$$E\{\hat{a}(a)\} \approx (n-a)\{E(T)\}^{-1}\{1 + \text{var}(T)/E^2(T)\} \quad (5.2)$$

$$\text{var}\{\hat{a}(a)\} \approx (n-a)^2 \text{var}(T)/E^4(T), \quad (5.3)$$

then from (4.12) and (4.18) we obtain to $O(n^{-1})$,

$$E\{\hat{a}(a)\} = [1 + n^{-1}\{(k+3)(\theta+1)^{-1} - a\}] \quad (5.4)$$

$$\text{var}\{\hat{\theta}(a)\} = 2\theta(\theta+1)/n \quad (5.5)$$

Use of (2.12) and (5.5) shows that the asymptotic variance efficiency of $\hat{\theta}(a)$ relative to the ML estimator $\hat{\theta}$ is

$$E_{\theta}^{(2)} = [2(\theta+1)\{\theta\psi^{(1)}(\theta) - 1\}]^{-1}. \quad (5.6)$$

The efficiency is independent of the number of regressor variables and their values and so is equal to the efficiency of the method of moments estimator relative to the ML estimator when no regressor variables are present.

Values of E_{θ} are shown in table 5 for $\theta = 0.25(0.25)1.00, 1.50, 2.00, 3.00, 4.00(2.00)10.00$. The results show that the efficiency is low for small values of θ and increases relatively slowly as θ increases.

Table 5

Asymptotic variance efficiency of $\hat{\theta}(a)$ relative to $\hat{\theta}$

θ	0.25	0.5	0.75	1.00	1.5	2.00	3.00	4.00	6.00	8.00	10.00
$E_{\theta}^{(2)}$	0.12	0.23	0.34	0.39	0.50	0.57	0.68	0.74	0.81	0.85	0.88

The variance and mean square error properties of the estimator $\hat{\theta}(a)$ can be substantially improved in small samples by adjusting for bias which to $O(n^{-1})$ is

$$b_{\hat{\theta}(a)} = n^{-1}\{(k+3-a)\theta + k+3\}. \quad (5.7)$$

The bias corrected estimator is

$$\hat{\theta}^*(a) = \hat{\theta}(a)\{1 - n^{-1}(k+3-a)\} - n^{-1}(k+3) \quad (5.8)$$

with

$$b_{\hat{\theta}^*(a)} = b_{\hat{\theta}(a)} - n^{-1}\{(k+3-a)\theta + k+3\} \quad (5.9)$$

which is $O(n^{-2})$ and

$$\text{var}\{\hat{\theta}^*(a)\} = \{1 - n^{-1}(k+3-a)\}^2 \text{var}\{\hat{\theta}(a)\} \quad (5.10)$$

The correction leads to an increase in variance when $a > k+3$.

Assuming that $0 \leq a \leq k+3$, the proportionate decrease in variance is

$$R_v(a) = n^{-1}(k+3-a)\{2 - n^{-1}(k+3-a)\} \quad (5.11)$$

For the McCullagh-Nelder estimator $\hat{\theta}_4, R_v(k+1) = 4n^{-1}(1 - n^{-1})$ showing a 19% decrease in variance when $n = 20$ and a 7.8% decrease when $n = 50$.

When $a = 0$, the bias corrected estimator is

$$\hat{\theta}_7 = (n - k - 3)T^{-1} - (k + 3)n^{-1}. \quad (5.12)$$

The proportionate decrease in variance using $\hat{\theta}_7$ instead of the uncorrected estimator n/T is

$$R_v(0) = n^{-1}(k + 3)\{2 - n^{-1}(k + 3)\} \quad (5.13)$$

When $k = 1$ there is a 76% decrease in variance when $n = 20$ and a 15% decrease when $n = 50$. The percentage decreases become substantially larger as k increases.

The estimator $\hat{\theta}_7$ provides a good approximation to the minimum variance estimator within the class of bias corrected estimators $\{\hat{a}^*(a)\}$. This is seen by writing

$$\{\hat{a}^*(a)\} = c_a(n, k)\hat{\theta}_4 - (k + 3)n^{-1}. \quad (5.14)$$

where

$$c_a(n, k) = \frac{\{(n - a)(n - k - 3 + a)\}}{\{n(n - k - 1)\}} \left[1 - \frac{2}{n - k - 1} - \frac{a(a - k - 3)}{n(n - k - 1)} \right]. \quad (5.15)$$

To order n^{-2} , $c_a(n, k)$ is independent of a and $\hat{a}^*(a) \approx \hat{\theta}_7$.

The proportionate decrease in variance using $\hat{\theta}_7$ instead of the McCullagh-Nelder estimator $\hat{\theta}_4$ is

$$R_v(n, k) = 4(n - k - 2)/(n - k - 1)^2. \quad (5.16)$$

Table 6 gives values of $R_v(n, k)$ for $n = 10(10)40(20)100$ and $k = 1(1)4$.

To order n^{-1} , $R_v(n, k) = 4n^{-1}$ and it is seen that this approximation works well for $n \geq 20$.

Table 6

Values of the variance reduction factor $R_v(n,k)$ for comparing $\hat{\gamma}_7$ with $\hat{\gamma}_4$

k \ n	10	20	30	40	60	80	100
1	0.438	0.210	0.138	0.102	0.068	0.051	0.040
2	0.490	0.221	0.143	0.105	0.069	0.051	0.041
3	0.556	0.234	0.148	0.108	0.07	0.052	0.041
4	0.640	0.249	0.154	0.111	0.071	0.053	0.042

The proportionate reduction in mean square error through using $\hat{\gamma}_7$ instead of $\hat{\gamma}_4$ is to order n^{-1} ,

$$R_m(n,k) = 4n^{-1} \left[1 + (20+k+3)^2 / \{80(n+1)\} \right] \quad (5.17)$$

Values of $nR_m(n,k)$ are shown in table 7 for $\gamma = 0.25(0.25)1.00(0.50)3.00(1.00)6.00$ and $k = 1, 2, 3, 4$. When $k = 1$, $n = 50$, the values of $R_m(n,k)$ are 0.41, 0.26, 0.15 for $\gamma = 0.5, 1.0, 4.0$, respectively. When $k=3$, the corresponding values rise to 0.73, 0.40 and 0.18, respectively. The results indicate the marked gains to be had from using $\hat{\gamma}_7$ instead of the estimator $\hat{\gamma}_4$.

Table 7

Values of the mean square error reduction factor $nR_m(n,k)$

k \ \gamma	0.25	0.50	0.75	1.00	1.50	2.00	2.50	3.00	4.00	5.00	6.00
1	36.40	20.67	15.52	13.00	10.53	9.33	8.63	8.17	7.60	7.27	7.05
2	52.40	28.00	20.10	16.25	12.53	10.75	9.71	9.04	8.23	7.75	7.44
3	71.60	36.67	25.43	20.00	14.80	12.33	10.91	10.00	8.90	8.27	7.86
4	94.00	46.67	31.52	24.25	17.33	14.08	12.23	11.04	9.63	8.82	8.30

6. MONTE CARLO RESULTS

In the previous sections, we have considered several alternative estimators for the shape parameter θ under the logarithmic link function. Theoretical approximations to their biases and variances were developed and used to compare the variance and mean square error properties of the estimators.

In order to assess the adequacy of the theoretical approximations, a large Monte—Carlo study was made for the case of a single explanatory variable with equally spaced x values, the model for the means being

$$\mu_i = \exp(\beta_0 + \beta_1 x_i), \quad i = 1, \dots, n \quad (6.1)$$

with $x_i = i - \frac{1}{2}(n+1)$, $(\sum_i x_i = 0)$. Since the distributions of

$\hat{\beta} - \beta$ and $\hat{\theta}$ are independent of $\tilde{\beta}$, values $\beta_0 = \beta_1 = 0$ were used without

loss of generality. Values $\theta = 0.5(0.5)2.0, 3.0, 4.0$ were used for the shape parameter. For integer θ with $\mu_i = 1$, the density of Y_i

is the special Erlangian distribution and the observation on Y_i can be generated as the sum of θ independent standard exponential observations.

For the half integer values of θ , the observation on Y_i was generated as the scaled sum of squares of 2θ $N(0,1)$

observations, using the result that $Y_i \sim \chi_{2\theta}^2 / (2\theta)$. Sample sizes

$n = 10, 20, 30, 50$ and 100 were used in the investigation. The run size was 2000 in each case and calculations were performed using the statistical package GLIM.

Table 8 shows the values of $\tilde{b}_\theta / \{n^{-1}h_1(\theta)\}$, where \tilde{b}_θ is the simulation estimate of the bias of $\hat{\theta}$ and $n^{-1}h_1(\theta)$ is the approximating bias given by (3.2). The results show that the actual biases are considerably larger than the approximate biases given by (2.20).

For $n=10$, the biases obtained by simulation were more than 50% higher than the approximate biases. With increasing n , the agreement between the approximate biases and biases obtained by simulation improved rapidly and the results suggest that (2.20) may be safely used when $n \geq 50$.

Table 8Values of the ratio $n \tilde{b} / h_1(\theta)$

\backslash n	10	20	30	50	100
0.5	1.89	1.34	1.08	1.10	1.21
1.0	1.71	1.19	1.20	1.01	0.94
1.5	1.73	1.17	1.20	1.04	1.00
2.0	1.56	1.20	1.17	0.99	1.00
3.0	1.66	1.23	1.21	1.07	1.10
4.0	1.56	1.29	1.14	1.12	1.11

Table 9 shows the values of $\tilde{\text{var}}(\hat{\theta}) / \{n^{-1}h_2(\theta)\}$, where $\tilde{\text{var}}(\hat{\theta})$

is the simulation estimate of $\text{var}(\hat{\theta})$ and $n^{-1}h_2(\theta)$ is the approximating variance given by (3.3). The results show that the first order approximation to the variance of the ML estimate given by (2.12) seriously underestimates the variance in small to moderate sized samples. When $n = 20$, the variances obtained by simulation are more than double the approximating variance. For $n = 100$, the values are approximately 15% higher. These results indicate that second order approximations to $\text{var}(\hat{\theta})$ as well as bias correction for $\hat{\theta}$ are needed for the approximating inference procedures for $\hat{\theta}$ to be satisfactory in small samples.

Table 9Values of the ratio $n \tilde{\text{var}}(\hat{\theta}) / h_2(\theta)$,

\backslash n	10	20	30	50	100
0.5	4.42	2.03	1.35	1.26	1.14
1.0	5.97	2.18	1.66	1.25	1.14
1.5	6.37	2.26	1.84	1.29	1.12
2.0	5.66	2.25	1.58	1.26	1.11
3.0	5.87	2.14	1.72	1.36	1.16
4.0	5.99	2.34	1.71	1.33	1.15

The estimators $\hat{\theta}, \hat{\theta}_1 = nD^{-1}, \hat{\theta}_3 = (n-k-1)D^{-1}, \hat{\theta}_5 = \hat{\theta} - n^{-1}h_1(\hat{\theta})$
 And $\hat{\theta}_6 = \hat{\theta} \{1 - n^{-1}(k+3)\} + (3n)^{-1} \{k+2 - (k+1)/30\}$ form a class of
 estimators as they are all functions of the deviance statistic. We
 let

$$E_j^{(v)}(n) = \frac{\text{var}(\hat{\theta}_j)}{\text{var}(\hat{\theta})}, \quad E_j^{(m)}(n) = \frac{\text{mse}(\hat{\theta}_j)}{\text{mse}(\hat{\theta})}, \quad (6.2)$$

denote the variance and mean square error efficiencies of $\hat{\theta}_j$
 relative to the ML estimator $\hat{\theta}$. Simulation estimates of the values
 of these efficiencies are shown in tables 10, 11, 12 and 13 for
 $\hat{\theta}_1, \hat{\theta}_3, \hat{\theta}_5$ and $\hat{\theta}_6$ respectively. The broad findings are as follows

(i) The estimator $\hat{\theta}_1$ has only a slightly better variance per-
 formance than $\hat{\theta}$. Even though its mean square error efficiency
 approaches 0 as $n \rightarrow \infty$ the estimator had a better mean square error
 performance than $\hat{\theta}$ for $n \leq 50$.

(ii) Use of the estimator $\hat{\theta}_3$ gives a worthwhile improvement in
 variance and mean square error performance compared with $\hat{\theta}$, in
 small to moderate size samples. For $n > 50$ and small values of k ,
 the mean square efficiency of $\hat{\theta}_3$ is less than one.

(iii) The bias corrected estimator $\hat{\theta}_5$ has much better variance
 and mean square error properties than $\hat{\theta}$ in small to moderate sized
 samples. For $n = 20$, the mean square error efficiency was only a
 little less than 2 and was still greater than 1.5 when $n = 50$.

(iv) The performance of the estimator $\hat{\theta}_6$ was not unexpectedly
 very similar to that of the estimator $\hat{\theta}_5$.

Table 10Simulation estimates of variance and me efficiencies of $\hat{\theta}_1$ relative to $\hat{\theta}$

θ	n	$E_1^{(v)}(n)$					$E_1^{(m)}(n)$				
		10	20	30	50	100	10	20	30	50	100
0.5		1.07	1.03	1.17	1.18	1.19	1.36	1.42	1.27	0.73	0.37
1.0		1.02	1.04	1.05	1.06	1.06	1.14	1.24	1.25	1.04	0.68
1.5		1.01	1.02	1.02	1.03	1.03	1.09	1.17	1.19	1.13	0.90
2.0		1.00	1.01	1.01	1.01	1.01	1.07	1.13	1.17	1.13	1.00
3.0		1.00	1.00	1.00	1.01	1.01	1.05	1.10	1.12	1.12	1.08
4.0		1.00	1.00	1.00	1.00	1.00	1.03	1.07	1.09	1.10	1.09

Table11Simulation estimates of variance and use efficiencies of $\hat{\theta}_3$ relative to $\hat{\theta}$

θ	n	$E_3^{(v)}(n)$					$E_3^{(m)}(n)$				
		10	20	30	50	100	10	20	30	50	100
0.5		1.67	1.40	1.34	1.28	1.24	2.46	1.51	0.94	0.62	0.33
1.0		1.59	1.28	1.20	1.15	1.11	2.06	1.53	1.34	0.95	0.60
1.5		1.57	1.26	1.17	1.11	1.07	1.99	1.53	1.37	1.13	0.83
2.0		1.57	1.25	1.16	1.10	1.06	1.96	1.53	1.41	1.16	0.95
3.0		1.57	1.24	1.15	1.09	1.05	1.94	1.54	1.39	1.23	1.08
4.0		1.56	1.24	1.15	1.09	1.04	1.90	1.52	1.36	1.25	1.11

Table 12Simulation estimates of variance and use efficiencies of $\hat{\theta}_5$ relative to $\hat{\theta}$

θ	n	$E_5^{(v)}(n)$					$E_5^{(m)}(n)$				
		10	20	30	50	100	10	20	30	50	100
0.5		2.42	1.43	1.26	1.15	1.07	3.50	1.80	1.48	1.28	1.14
1.0		2.73	1.50	1.28	1.15	1.07	3.62	1.82	1.52	1.27	1.13
1.5		2.82	1.58	1.34	1.17	1.07	3.74	1.92	1.58	1.31	1.14
2.0		2.81	1.59	1.36	1.21	1.10	3.70	1.97	1.65	1.34	1.18
3.0		2.78	1.56	1.33	1.18	1.08	3.76	1.98	1.61	1.33	1.17
4.0		2.78	1.56	1.33	1.18	1.08	3.64	1.99	1.59	1.35	1.17

Table 13

Simulation estimates of variance and use efficiencies of $\hat{\theta}_6$ relative to $\hat{\theta}_6$

	n	$E_6^{(v)}(n)$					$E_6^{(m)}(n)$				
		10	20	30	50	100	10	20	30	50	100
0.5		2.55	1.46	1.26	1.14	1.07	3.74	1.84	1.48	1.27	1.14
1.0		2.72	1.54	1.32	1.17	1.08	3.61	1.87	1.57	1.30	1.14
1.5		2.75	1.55	1.32	1.18	1.08	3.67	1.89	1.57	1.32	1.15
2.0		2.76	1.56	1.33	1.18	1.08	3.63	1.92	1.61	1.33	1.16
3.0		2.77	1.56	1.33	1.18	1.08	3.75	1.98	1.61	1.33	1.17
4.0		2.77	1.56	1.33	1.18	1.08	3.64	1.99	1.59	1.35	1.17

We now turn to the properties of the estimators based on the Pearson statistic, namely $\hat{\theta}_4 = (n - k - 1)T^{-1}$, $\hat{\theta}_7 = (n - k - 3)T^{-1} - (k + 3)n^{-1}$ and the bias corrected form of $\hat{\theta}_4$ given by

$$\hat{\theta}_8 = \hat{\theta}_4(1 - 2n^{-1}) - (k + 3)n^{-1}. \quad (6.3)$$

Values of the simulation estimates of the variance efficiencies $E_j^{(v)}$ and the mean square error efficiencies $E_j^{(m)}(n)$ of $\hat{\theta}_j$ relative to the ML estimator $\hat{\theta}$ are shown in tables 14, 15 and 16 for $\hat{\theta}_4, \hat{\theta}_7$ and $\hat{\theta}_8$, respectively. The broad findings from the results are as follows

(i) The McCullagh-Nelder estimator $\hat{\theta}_4$ has a better performance than the ML estimator $\hat{\theta}$ in very small samples. However, its use cannot be recommended when $n > 20$ unless there are good grounds for suspecting the ML estimate.

(ii) The performance of the bias corrected estimator $\hat{\theta}_7$ compared with $\hat{\theta}$ is good for sample sizes less than 50, except when $\hat{\theta}$ is very small.

(iii) The performance of $\hat{\theta}_8$, the bias corrected form of the McCullagh-Nelder estimator is very similar to that of $\hat{\theta}_7$ except in very small samples.

References

Bowman, K.O. and Shenton, L.R. (1968). Properties of estimators for the gamma distribution, Report CTC-1, Union Carbide Corp., Oak Ridge, Tennessee.

Cordeiro, G.M. (1983). Improved likelihood ratio statistics for generalised linear models. *J.R. Statist. Soc. B*, 45, 404-413.

Cox, D.R. and Snell, E.J. (1968). A general definition of residuals. *J.R. Statist. Soc. B*, 30, 248-265.

Greenwood, J.A. and Durand, D. (1960). Aids for fitting the gamma distribution by maximum likelihood. *Technometrics*, 2, 55-65.

McCullagh, P. and Nelder, J.A. (1983). Generalised Linear Models. London, Chapman and Hall.

Nelder, J.A. and Wedderburn, R.W.M. (1972). Generalised linear models. *J.R. Statist. Soc. A*, 135, 370-384.

