Pak. J. Statist.
2003 Vol.19(1) pp129-141

A CLASS OF METHOD OF MOMENTS ESTIMATORS FOR THE TWO-PARAMETER GAMMA FAMILY

Douglas P. Wiens
Department of Mathematical and Statistical Sciences
University of Alberta, Canada

Julian Cheng
Department of Electrical and Computer Engineering
University of Alberta, Canada

Norman C. Beaulieu
Department of Electrical and Computer Engineering
University of Alberta, Canada

ABSTRACT

We introduce and study a class of estimators, for the two-parameter Gamma family, which are based on three moments of the sample. The members of this family are easy to compute, relative to the maximum likelihood estimator or its commonly used approximation. A limiting member of the family is surprisingly efficient. The other members of the class, while approaching in efficiency this limiting member, possess a property of robustness against "inliers" - observations which are aberrant by virtue of being too small.

KEY WORDS

Asymptotic, Efficiency, Influence function, Inliers, Likelihood, Nakagami distribution, Robustness.

1. INTRODUCTION

In this article we revisit the problem of parameter estimation in the Gamma family $\mathcal{G}(\alpha, \beta)$ of probability distributions. A member of this family has the density

$$p(x; \boldsymbol{\theta}) = \frac{\left(\frac{x}{\beta}\right)^{\alpha-1} e^{-\frac{x}{\beta}}}{\beta\Gamma(\alpha)} I(x > 0), \tag{1}$$

parameterised by $\boldsymbol{\theta} = (\alpha, \beta)^T$, $\alpha, \beta > 0$. We present a new class of easily computed "method of moments" estimators of $\boldsymbol{\theta}$, a limiting member of which is also highly efficient.

Much of the parameter estimation literature for $\mathcal{G}(\alpha, \beta)$ has concentrated on the computation or approximation of the maximum likelihood estimate $\hat{\boldsymbol{\theta}}_{ML}$. Let $X_1, ..., X_n$ be a sample from (1), with average $\bar{X}$. The likelihood equation for β has solution $\hat{\beta}_{ML} = \bar{X}/\hat{\alpha}_{ML}$. In terms of the positive random variable $Y = \ln \bar{X} - \overline{\ln X}$ (here $\overline{\ln X}$ is the average of the logarithms of the data values) and the digamma function $\psi(\alpha) = (d/d\alpha) \ln \Gamma(\alpha)$, $\hat{\alpha}_{ML}$ is then obtained from the equation

$$Y = g(\hat{\alpha}_{ML}) := \ln \hat{\alpha}_{ML} - \psi(\hat{\alpha}_{ML}). \tag{2}$$

Since $g(\alpha)$ is strictly decreasing, from ∞ to 0, on $(0, \infty)$, (2) has a unique solution $\hat{\alpha}_{ML} = g^{-1}(Y)$. The resulting estimates are asymptotically normally distributed and asymptotically efficient:

$$\sqrt{n}\left(\hat{\boldsymbol{\theta}}_{ML} - \boldsymbol{\theta}\right) \xrightarrow{\mathcal{L}} N\left(\mathbf{0}, \mathbf{I}^{-1}(\boldsymbol{\theta})\right), \tag{3}$$

where

$$\mathbf{I}(\boldsymbol{\theta}) = \begin{pmatrix} \psi'(\alpha) & \frac{1}{\beta} \\ \frac{1}{\beta} & \frac{\alpha}{\beta^2} \end{pmatrix} \tag{4}$$

is the Fisher information in one observation.

Computational difficulties associated with the digamma function seem to have motivated a search for approximations to $\hat{\alpha}_{ML}$, for instance for applications in Electrical Engineering calling for rapid and continuous on-line estimation. Thom (1958) proposed an approximation obtained by replacing $\psi(\alpha)$ by its second order asymptotic expansion. A more popular product of the search seems to be the estimator of Greenwood and Durand (1960), who approximated $g^{-1}(y)$ by

$$f_{GD}(y) = \begin{cases} f_1(y), & y < .5772, \\ f_2(y), & .5772 \le y \le 17, \end{cases}$$

where

$$\begin{aligned} f_1(y) &= \frac{.5000876 + .1648852y - .0544274y^2}{y}, \\ f_2(y) &= \frac{8.898919 + 9.059950y - .9775373y^2}{(17.79728 + 11.968477y + y^2)\, y}. \end{aligned}$$

There appears to be no mention in the literature of the significance of the upper bound of 17, and it seems that in practice this bound is ignored. The significance of the .5772 may be that $g(1) = C$ (Catalan's constant) $= .5772...$. Despite being inconsistent, the resulting estimator $\hat{\alpha}_{GD} = f_{GD}(Y)$ is surprisingly accurate in moderately sized samples.

Bowman and Shenton (1988) give a fixed point algorithm to solve (2):

$$\alpha^{(i)} = \frac{\alpha^{(i-1)} g\left(\alpha^{(i-1)}\right)}{Y}, \; i = 1, 2, 3... \; . \tag{5}$$

If convergent, $\alpha^{(i)} \to \hat{\alpha}_{ML}$.

This and other algorithms to obtain $\hat{\alpha}_{ML}$ are often initiated from the "standard" method of moments estimate $\hat{\alpha}_1 = \bar{X}^2/S^2$ (correspondingly, $\hat{\beta}_1 = S^2/\bar{X}$). Here S^2 is the sample variance. These estimates arise from the observation that if $X \sim \mathcal{G}(\alpha, \beta)$ then

$$\alpha = \frac{E[X]^2}{VAR[X]}, \; \beta = \frac{E[X]}{\alpha}. \tag{6}$$

Dusenberry and Bowman (1977) carried out a comparative study of $\hat{\alpha}_1$, $\hat{\alpha}_{GB}$ and Thom's (1958) approximation. Cohen and Whitten (1982) discussed modifications of $\hat{\boldsymbol{\theta}}_1$ in the three-parameter gamma family.

The non-uniqueness of method of moments estimators, and the possibility of improving on $\hat{\boldsymbol{\theta}}_1$, seem not to have been exploited until very recently. Rafiq, Ahmad and Muhammed (1996) develop estimators from two fractional moments $E[X^r]$ and $E[X^s]$, $0 < r, s < 1$. Bickel and Doksum (2001, p. 102) mention that an estimator may be developed from the first and third moments. Cheng and Beaulieu (2002) make a relevant observation, in their study of the Nakagami m-distribution (Nakagami 1960 - a r.v. R has the Nakagami distribution with parameters m and Ω if $mR^2 \sim \mathcal{G}(m, \Omega)$; the "fading parameter" m is of interest in Electrical Engineering). In the current context this observation implies that if $X \sim \mathcal{G}(\alpha, \beta)$ then for $k > 0$,

$$\alpha = \frac{E\left[X^k\right] E[X]}{COV\left[\frac{X^k}{k}, X\right]}. \tag{7}$$

The special case $k = 1$ is (6); otherwise (7) employs three moments. Although Cheng and Beaulieu (2002) did not make the passage to the limit as $k \to 0$, if we do so now we obtain

$$\alpha = \frac{E[X]}{COV[\ln X, X]}. \tag{8}$$

This may be seen by writing $COV\left[k^{-1}X^k, X\right]$ as $COV\left[k^{-1}\left(X^k - 1\right), X\right]$, and then noting that $k^{-1}\left(X^k - 1\right) \to \ln X$ as $k \to 0$. Both (7) and (8) may also be verified by noting that their right hand sides are invariant under scale changes in X, and then using

$$E\left[\left(\frac{X}{\beta}\right)^\delta \ln^j\left(\frac{X}{\beta}\right)\right] = \frac{\frac{d^j}{d\alpha^j}\Gamma(\alpha + \delta)}{\Gamma(\alpha)}, \; \delta > 0, \; j = 0, 1, 2, ... \; .$$

For a sample $X_1, ..., X_n \overset{i.i.d.}{\sim} \mathcal{G}(\alpha, \beta)$ let $\hat{\mu}_k$ denote the k^{th} sample moment, and let $S_{X^k,X}$ be the sample covariance $\sum \left(X_i^k - \hat{\mu}_k\right)(X_i - \hat{\mu}_1)/(n-1)$. Analogously, define $S_{\ln X,X} = \sum \left(\ln X_i - \overline{\ln X}\right)(X_i - \hat{\mu}_1)/(n-1)$. We propose the class of method of moment estimators $\hat{\boldsymbol{\theta}}_k = \left(\hat{\alpha}_k, \hat{\beta}_k\right)^T$, with $\hat{\beta}_k = \hat{\mu}_1/\hat{\alpha}_k$ and

$$\hat{\alpha}_k = \begin{cases} \frac{\hat{\mu}_k\hat{\mu}_1}{k^{-1}S_{X^k,X}}, & k > 0, \\ \frac{\hat{\mu}_1}{S_{\ln X,X}}, & k = 0. \end{cases}$$

In Section 2 we study the asymptotic properties of $\hat{\boldsymbol{\theta}}_k$. Asymptotic and finite sample comparisons with the MLE, and with the estimator of Greenwood and Durand (1960), are made in Section 3. On the basis of these comparisons we claim that the ease of computation of $\hat{\boldsymbol{\theta}}_0$ and its high efficiency, especially when bias as well as variation are taken into account, make it a strong competitor. We also point out a moderate robustness property of $\hat{\boldsymbol{\theta}}_k$ for $k > 0$.

2. ASYMPTOTIC THEORY

Define $\nu_k = \Gamma(\alpha + k)/\Gamma(\alpha)$, so that $E\left[X^k\right] = \beta^k \nu_k$. We shall prove:

Theorem 1 (i) For $k > 0$ the vector $\hat{\boldsymbol{\theta}}_k = \left(\hat{\alpha}_k, \hat{\beta}_k\right)^T$ is asymptotically normally distributed:

$$\sqrt{n}\left(\hat{\boldsymbol{\theta}}_k - \boldsymbol{\theta}\right) \xrightarrow{\mathcal{L}} N\left(\mathbf{0}, \mathbf{C}_k\right) \tag{9}$$

where, with

$$\sigma_k^2 = \alpha^2 \left\{ \frac{\nu_{2k}}{\nu_k^2} + \frac{\nu_{2k+1} - \frac{\nu_{k+1}^2}{\nu_1}}{k^2 \nu_k^2} \right\} \tag{10}$$

the asymptotic covariance matrix is

$$\mathbf{C}_k = \begin{pmatrix} \sigma_k^2 & -\frac{\beta}{\alpha}\sigma_k^2 \\ * & \left(\frac{\beta}{\alpha}\right)^2 \sigma_k^2 + \frac{\beta^2}{\alpha} \end{pmatrix}. \tag{11}$$

(ii) The vector $\hat{\boldsymbol{\theta}}_0$ is asymptotically normally distributed as at (9), with asymptotic covariance matrix $\mathbf{C}_0$ defined by (11) with

$$\sigma_0^2 = \alpha^2 \left[1 + \alpha\psi'(\alpha + 1)\right].$$

Remarks:

1. The numerator $\nu_{2k+1} - \nu_{k+1}^2/\nu_1$ in the second summand of (10) can be seen to equal α times the variance of the r.v. U^k, if $U \sim \mathcal{G}(\alpha+1, 1)$. Thus both summands in (10) are positive, and the second tends to $\alpha VAR[\ln U] = \alpha\psi'(\alpha+1)$ as $k \to 0$.

2. We initially believed that σ_k^2 was minimised at $k = 0$, but further investigation showed this to be not quite true. See Figure 1, where $V_k^2 = \sigma_k^2/\alpha^2$ is plotted against k, in the typical case $\alpha = 1$. The plots were prepared using S-PLUS. In each, the horizontal line is at $V_0^2 = \pi^2/6$. For very small values of k, V_k^2 is quite unstable and can drop below V_0^2, albeit by a practically insignificant amount.

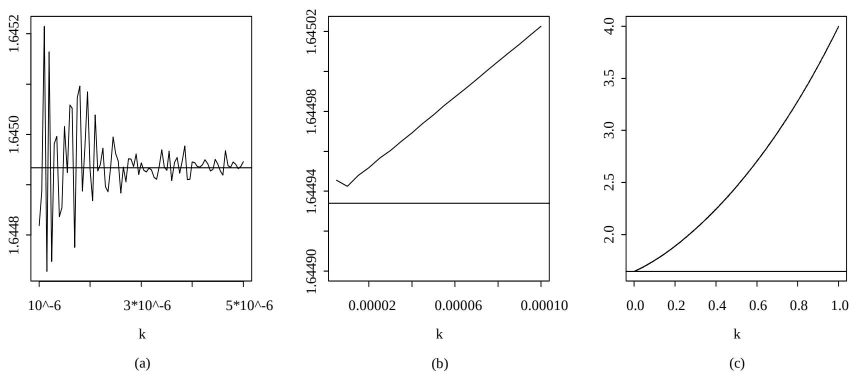

Figure 1: Asymptotic variance V_k^2 of $\ln\hat{\alpha}_k$ vs. k for (a) $10^{-6} \le k \le 5\cdot 10^{-6}$, (b) $5\cdot 10^{-6} \le k \le 10^{-4}$, (c) $10^{-4} \le k \le 1$; all with $\alpha = 1$. Horizontal lines are at $V_0^2 = \pi^2/6$.

3. The variance of $\hat{\alpha}_k$ is $2\alpha^2\left(1 + O\left(\alpha^{-1}\right)\right)$, so that the width of a confidence interval will increase approximately linearly, with α. This inconvenient point can be addressed by applying the delta method to obtain

$$\sqrt{n}\left(\ln\hat{\alpha}_k - \ln\alpha\right) \xrightarrow{\mathcal{L}} N\left(0, V_k^2\right).$$

Confidence intervals on $\ln\alpha$, with asymptotic coverage $100(1-p)\,\%$, are then $\ln\hat{\alpha}_k \pm z_{p/2}\hat{V}_k/\sqrt{n}$, with approximately constant widths. Since V_k is a continuous function of α, a consistent estimate $\hat{V}_k$ may be obtained by replacing all

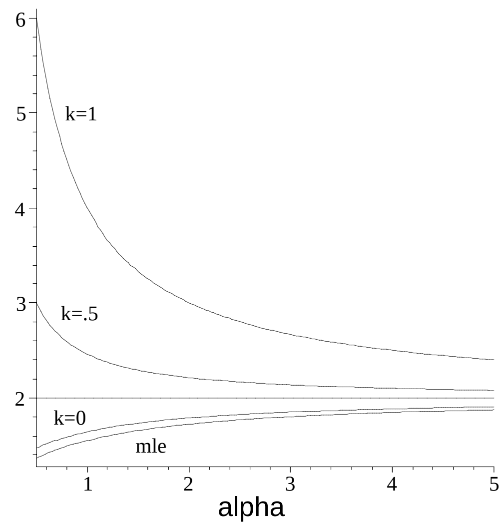

Figure 2: Asymptotic variances of $\ln \hat{\alpha}_1$, $\ln \hat{\alpha}_{.5}$, $\ln \hat{\alpha}_0$ and $\ln \hat{\alpha}_{ML}$ vs. α.

occurrences of α, in V_k, by $\hat{\alpha}_k$. If $k = 0$ an interval which is conservative, but only slightly so for even moderately large α, is obtained by setting $\hat{V}_0 = \sqrt{2}$ - see Figure 2. The log transformation is also recommended since, as seen in §3, it typically hastens the approach to normality.

4. The asymptotic variance V_k^2 of $\ln \hat{\alpha}_k$ is plotted against α, in Figure 2, for $k = 1$, $k = 0$ and for the maximum likelihood estimator $\ln \hat{\alpha}_{ML}$. Using (3) and (4) to obtain the asymptotic variance of $\hat{\alpha}_{ML}$ and dividing by that of $\hat{\alpha}_0$ gives the asymptotic relative efficiency of $\hat{\alpha}_0$ (equivalently, that of $\ln \hat{\alpha}_0$):

$$\begin{aligned} ARE\left(\hat{\alpha}_0\right) &= \left\{\alpha^2\left(1+\alpha\psi'\left(\alpha+1\right)\right)\left(\psi'\left(\alpha\right)-\alpha^{-1}\right)\right\}^{-1} \\ &= \begin{cases} 1-\left(\frac{\pi^2}{6}-1\right)\alpha+O\left(\alpha^2\right), & \text{as } \alpha\to 0, \\ 1-\frac{1}{12\alpha}+O\left(\alpha^{-2}\right), & \text{as } \alpha\to\infty. \end{cases} \end{aligned}$$

The expansions were obtained with the aid of MAPLE, and give a precise sense in which $\hat{\alpha}_0$ is "almost fully efficient". As is evident from Figure 2 the ARE of the standard estimator $\hat{\alpha}_1$ is much smaller; the corresponding expansions are $\alpha/2 + O\left(\alpha^3\right)$ and $1 - 4/3\alpha + O\left(\alpha^{-2}\right)$.

5. The ratio of the determinants of the asymptotic covariance matrices of $\hat{\boldsymbol{\theta}}_{ML}$ and $\hat{\boldsymbol{\theta}}_k$ is the same as that of the asymptotic variances of $\hat{\alpha}_{ML}$ and $\hat{\alpha}_k$; thus

the relative efficiencies of $\hat{\alpha}_k$, given above, are also those of $\hat{\boldsymbol{\theta}}_k$. Similarly, up to terms which are $O\left(\alpha^{\pm 2}\right)$, they are the relative efficiencies of $\hat{\beta}_k$.

Proof of Theorem 1

We prove the theorem for $k > 0$ only, the case $k = 0$ being very similar. We show below that the marginal distribution of $\hat{\alpha}_k$ is as claimed, i.e. that

$$\sqrt{n}\left(\hat{\alpha}_k - \alpha\right) \xrightarrow{\mathcal{L}} N\left(0, \sigma_k^2\right). \tag{12}$$

Since $\hat{\alpha}_k$ is scale invariant it is distributed freely of β and hence, by Basu's Theorem, is independent of $\bar{X}$. This together with (12) and the Central Limit Theorem applied to $\bar{X}$ gives

$$\sqrt{n}\left(\begin{pmatrix}\hat{\alpha}_k \\ \bar{X}\end{pmatrix} - \begin{pmatrix}\alpha \\ \alpha\beta\end{pmatrix}\right) \xrightarrow{\mathcal{L}} N\left(\mathbf{0}, \boldsymbol{\Sigma}_k\right),$$

where $\boldsymbol{\Sigma}_k = diag\left(\sigma_k^2, \alpha\beta^2\right)$. Now define the function $\boldsymbol{\eta}$ by $\boldsymbol{\eta}\left(z_1, z_2\right) = \left(z_1, z_2/z_1\right)$, with Jacobian

$$\mathbf{J}_{\boldsymbol{\eta}}\left(\mathbf{z}\right) = \begin{pmatrix} 1 & 0 \\ -z_2/z_1^2 & 1/z_1 \end{pmatrix}.$$

Then $\hat{\boldsymbol{\theta}}_k = \boldsymbol{\eta}\left(\hat{\alpha}_k, \bar{X}\right)$. With $\mathbf{J}_0$ denoting $\mathbf{J}_{\boldsymbol{\eta}}\left(\mathbf{z}\right)$ evaluated at $\mathbf{z} = \left(\alpha, \alpha\beta\right)^T$, the delta method shows that (9) holds with $\mathbf{C}_k = \mathbf{J}_0 \boldsymbol{\Sigma}_k \mathbf{J}_0^T$. A calculation then yields (11).

It remains to show (12). For this we may take $\beta = 1$ without loss of generality. By the multivariate Central Limit Theorem,

$$\sqrt{n}\begin{pmatrix} \hat{\mu}_1 - \nu_1 \\ \hat{\mu}_k - \nu_k \\ \hat{\mu}_{k+1} - \nu_{k+1} \end{pmatrix} \xrightarrow{\mathcal{L}} N\left(\mathbf{0}, \boldsymbol{\Lambda}\right),$$

the trivariate normal distribution with mean vector $\mathbf{0}$ and (symmetric) covariance matrix

$$\begin{aligned}
\boldsymbol{\Lambda} &= \begin{pmatrix} VAR\left[X\right] & COV\left[X, X^k\right] & COV\left[X, X^{k+1}\right] \\ * & VAR\left[X^k\right] & COV\left[X^k, X^{k+1}\right] \\ * & * & VAR\left[X^{k+1}\right] \end{pmatrix} \\
&= \begin{pmatrix} \nu_2 & \nu_{k+1} & \nu_{k+2} \\ \nu_{k+1} & \nu_{2k} & \nu_{2k+1} \\ \nu_{k+2} & \nu_{2k+1} & \nu_{2k+2} \end{pmatrix} - \begin{pmatrix} \nu_1 \\ \nu_k \\ \nu_{k+1} \end{pmatrix} \begin{pmatrix} \nu_1 & \nu_k & \nu_{k+1} \end{pmatrix}.
\end{aligned} \tag{13}$$

Define a function $\xi(x, y, z) = k\left(\frac{z}{xy} - 1\right)^{-1}$, with gradient

$$\dot{\boldsymbol{\xi}}\left(x, y, z\right) = \frac{\xi^2(x, y, z)}{kxy}\left(\frac{z}{x}, \frac{z}{y}, -1\right)^T.$$

Then $\hat{\alpha}_k = \frac{n-1}{n}\xi(\hat{\mu}_1, \hat{\mu}_k, \hat{\mu}_{k+1})$, $\alpha = \xi(\nu_1, \nu_k, \nu_{k+1})$ and the delta method together with Slutsky's Theorem gives (12) with

$$\sigma_k^2 = \dot{\boldsymbol{\xi}}^T(\nu_1, \nu_k, \nu_{k+1})\,\boldsymbol{\Lambda}\dot{\boldsymbol{\xi}}(\nu_1, \nu_k, \nu_{k+1}). \tag{14}$$

It is convenient to write $\dot{\boldsymbol{\xi}}$ in (14) as

$$\dot{\boldsymbol{\xi}}(\nu_1, \nu_k, \nu_{k+1}) = \frac{\alpha}{k\nu_k}\left(\frac{\nu_{k+1}}{\nu_1}, \frac{\nu_{k+1}}{\nu_k}, -1\right)^T;$$

substituting this and (13) into (14) yields (10), after a calculation. □

3. ASYMPTOTIC AND FINITE SAMPLE COMPARISONS

In this section we compare $\hat{\alpha}_0$, $\hat{\alpha}_1$, $\hat{\alpha}_{GD}$ and $\hat{\alpha}_{ML}$ with respect to various criteria. We first exhibit the asymptotic properties of $\hat{\alpha}_{GD}$, and draw comparisons with $\hat{\alpha}_0$. Recall that $\hat{\alpha}_{GD} = f_{GD}(Y) = f_{GD}(\ln\hat{\alpha}_{ML} - \psi(\hat{\alpha}_{ML}))$, so that an application of the delta method shows that

$$\sqrt{n}\left(\hat{\alpha}_{GD} - f_{GD}(\ln\alpha - \psi(\alpha))\right) \xrightarrow{\mathcal{L}} N\left(0, \sigma_{GD}^2 = \left(f'_{GD}(\ln\alpha - \psi(\alpha))\right)^2\left(\psi'(\alpha) - \alpha^{-1}\right)\right).$$

Since $\hat{\alpha}_{GD}$ is inconsistent, the proper basis of comparison is the ratio of the asymptotic mean squared errors. Figure 3 shows a plot of

$$\frac{MSE(\hat{\alpha}_{GD})}{VAR(\hat{\alpha}_0)} = \frac{\sigma_{GD}^2/n + \left(f_{GD}(\ln\alpha - \psi(\alpha)) - \alpha\right)^2}{\sigma_0^2/n}$$

for $n = 100$ and $\alpha \leq 5$. This ratio exceeds unity for all sufficiently small or large α: $\alpha \leq .003$ and $\alpha \geq 236$ suffice. Of course these bounds coalesce as $n \to \infty$. The cusp at $\alpha = 1$ is a result of the change from $f_{GD} = f_2$ to $f_{GD} = f_1$ at $\ln\alpha - \psi(\alpha)_{|\alpha=1} = .5772$.

How do these estimators compare in finite samples? To answer this question we simulated 2500 values of $\hat{\alpha}_0$, $\hat{\alpha}_1$, $\hat{\alpha}_{GD}$ and $\hat{\alpha}_{ML}$ for each of a range of sample sizes from $n = 20$ to $n = 800$ and for each $\alpha = .5, 1, 4.5$. In the computation of $\hat{\alpha}_{ML}$ we started with $\alpha^{(0)} = \hat{\alpha}_0$ and iterated as at (5) until the maximum (over all 2500 simulations) change between iterates was less than .0001. This generally required $4-5$ iterations. The biases, standard deviations and root mean squared errors, for all four estimators, are plotted for $20 \leq n \leq 250$ in Figure 4. They are plotted against $n \geq 300$ in Figure 5. The conclusion pointed to is that with respect to all three measures of performance, $\hat{\alpha}_0$, $\hat{\alpha}_{GD}$ and $\hat{\alpha}_{ML}$ are almost indistinguishable. In fact, except for the plot of bias, when $\alpha = 4.5$ and $n \geq 300$, the graphs for $\hat{\alpha}_{GD}$ and

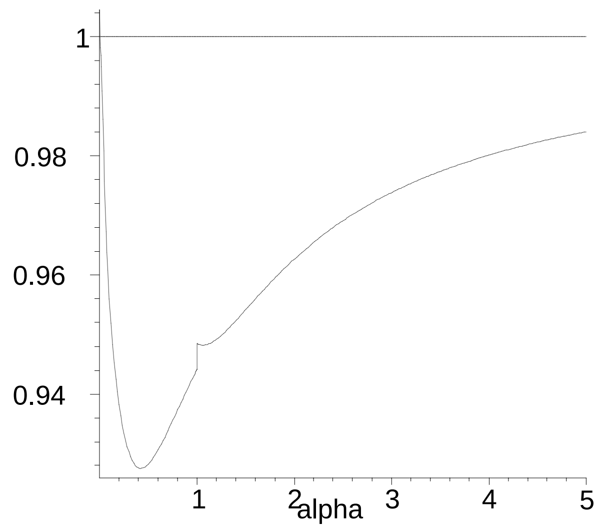

Figure 3: Ratio of asymptotic mean squared error of $\hat{\alpha}_{GD}$ to that of $\hat{\alpha}_0$; $n = 100$.

$\hat{\alpha}_{ML}$ are indistinguishable. In all cases $\hat{\alpha}_0$ has a smaller bias, but this advantage is masked by the larger contribution of variation to rmse. As expected in light of the expansions in Remark 4 of §2 , $\hat{\alpha}_1$ performed less well than the others.

We have also computed empirical measures of the approach to normality of the estimators. For $n = 20$, $n = 100$ and for each value of α we prepared QQ plots (not shown), and computed the correlations between the empirical and normal quantiles. See Table 1. In most cases $\hat{\alpha}_1$ appeared to approach normality slightly more quickly than the others, which were in this respect almost identical. In all cases the approach to normality was improved by the taking of logarithms.

To this point, our discussion will have left the impression that $\hat{\alpha}_k$ for $k > 0$ has little to recommend it. There is however one respect in which it is superior - robustness against "inliers". An inlier is an observation which is aberrant by virtue of being too small. To quantify the robustness we have calculated the influence functions of these estimates. Recall that the influence function $IF(x; \hat{\alpha})$ measures the limiting and suitably normed influence of a single observation x, in an otherwise clean sample, on an estimate $\hat{\alpha}$. If $\hat{\alpha}$ is consistent then typically $\sqrt{n}(\hat{\alpha} - \alpha)$ is asymptotically normally distributed, with a mean of $E[IF(X; \hat{\alpha})] = 0$ and variance of $E\left[IF^2(X; \hat{\alpha})\right]$. See Huber (1981) for more details.

Table 1. Correlations between normal and empirical quantiles of estimates $\hat{\alpha}$. Those of $\ln \hat{\alpha}$ are in parentheses.

α	n	$\hat{\alpha}_{GD}$	$\hat{\alpha}_1$	$\hat{\alpha}_0$	$\hat{\alpha}_{ML}$
.5	20	.962 (.997)	.974 (.998)	.963 (.997)	.962 (.997)
1	20	.952 (.996)	.967 (.999)	.953 (.997)	.952 (.996)
4.5	20	.950 (.997)	.958 (.999)	.950 (.997)	.950 (.997)
.5	100	.990 (.998)	.996 (.994)	.992 (.999)	.990 (.998)
1	100	.991 (.999)	.995 (.999)	.991 (.999)	.991 (.999)
4.5	100	.990 (.999)	.993 (1.00)	.990 (.999)	.990 (.999)

NOTE: The "normal scores" test of the hypothesis of normality rejects for small values of the coefficient of correlation. For n = 20 the critical values, at the 10%, 5% and 1% significance levels, are .960, .951 and .926 respectively. For n = 100 they are .989, .987 and .982.

The influence functions are

$$IF(x;\hat{\alpha}_k) = \begin{cases} \frac{(k+\alpha)\nu_k(x-\alpha)+(k+\alpha)\alpha\left(x^k-\nu_k\right)-\alpha\left(x^{k+1}-\nu_{k+1}\right)}{k\nu_k}, & k>0, \\ (x-\alpha)-\alpha\left\{(x-\alpha)(\ln x-\psi(\alpha))-1\right\}, & k=0, \end{cases}$$

$$IF(x;\hat{\alpha}_{ML}) = \frac{\ln x-\psi(\alpha)-\alpha^{-1}(x-\alpha)}{\psi'(\alpha)-\alpha^{-1}},$$

$$IF(x;\hat{\alpha}_{GD}) = \left\{\frac{d}{d\alpha}f_{GD}(\ln\alpha-\psi(\alpha))\right\}\cdot IF(x;\hat{\alpha}_{ML}).$$

Note that all of these tend to $-\infty$ as $x \to \infty$, implying a lack of robustness against extreme outliers - a single, arbitrarily large observation can drive the estimate to 0. Only $IF(x;\hat{\alpha}_k)$ for $k>0$ remains finite as $x \to 0$: $IF(0;\hat{\alpha}_k) = -\alpha(1+\alpha/k)$.

4. SUMMARY

The "classical" approach to estimating the parameters of the Gamma distribution has largely centred on approximating the solution to the likelihood equation. Here we have instead given a class of moment-based estimators which are easily computed, close to efficient, and which exhibit appropriate asymptotic behaviour. The best and in some respects simplest member of this class - $\hat{\alpha}_0$ - seems to perform as well as the maximum likelihood estimate when bias as well as variation is taken into account. If robustness against abnormally small observations is an issue, then a choice $\hat{\alpha}_k$ for small but positive k will furnish more protection.

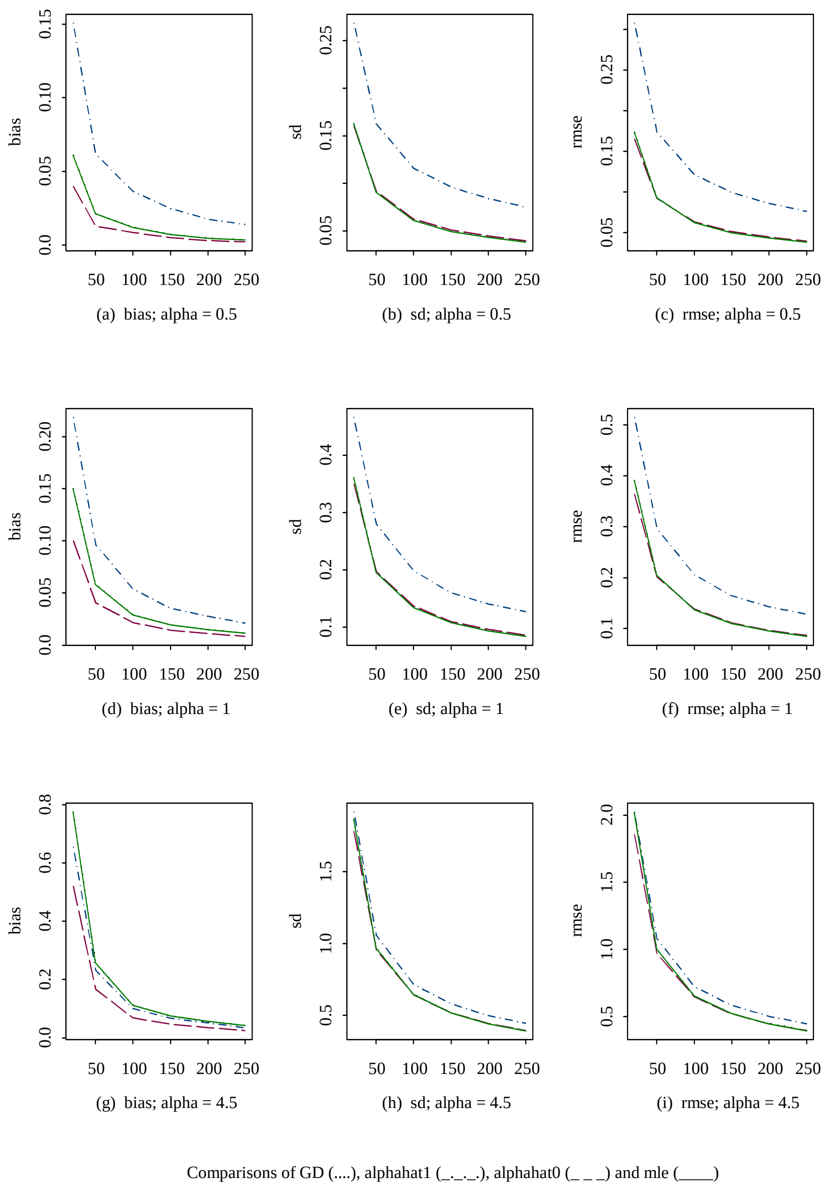

Figure 4: Comparisons of $\hat{\alpha}_{GD}$ $(\cdots)$, $\hat{\alpha}_1$ $(\cdot - \cdot - \cdot)$, $\hat{\alpha}_{ML}$ (——) and $\hat{\alpha}_0$ $(- - -)$. Bias, standard deviation and root mean-squared error vs. sample size n, $20 \leq n \leq 250$.

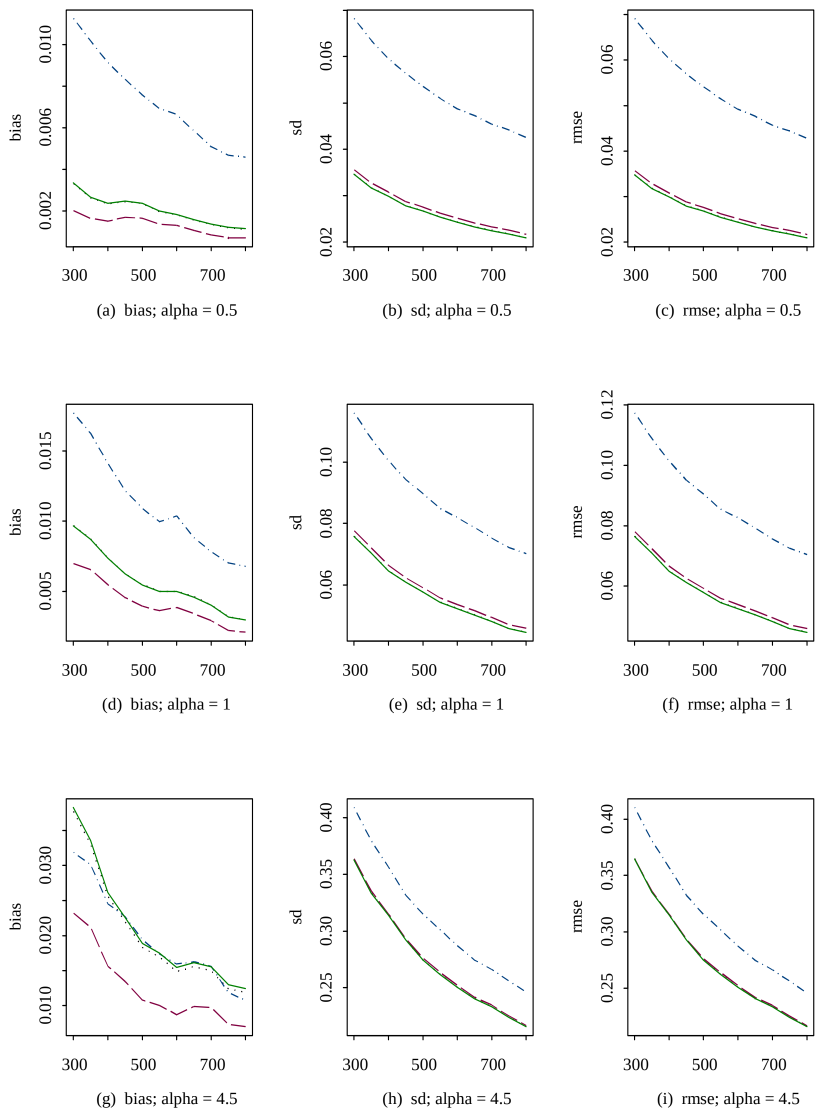

Figure 5: Comparisons of $\hat{\alpha}_{GD}$ $(\cdots)$, $\hat{\alpha}_1$ $(\cdot - \cdot - \cdot)$, $\hat{\alpha}_{ML}$ (——) and $\hat{\alpha}_0$ $(---)$. Bias, standard deviation and root mean-squared error vs. sample size n, $300 \leq n \leq 800$.

REFERENCES

Bickel, P.J. and Doksum, K.A. (2001). *Mathematical Statistics: Basic Ideas and Selected Topics Vol. I*. London: Prentice-Hall.

Bowman, K.O. and Shenton, L.R. (1988). *Properties of Estimators for the Gamma Distribution.* New York: Marcel Dekker.

Cheng, J. and Beaulieu, N. (2002). Generalized Moment Estimators for the Nakagami Fading Parameter. *IEEE Comm. Letters*, 6, 144-146.

Cohen, A.C. and Whitten, B.J. (1982). Modified Moment and Maximum Likelihood Estimators for Parameters of the Three-Parameter Gamma Distribution. *Comm. Statist., Part B - Simulation and Computation*, 11, 197-216.

Dusenberry, W.E. and Bowman, K.O. (1977). The Moment Estimator for the Shape Parameter of the Gamma Distribution. *Comm. Statist., Part B - Simulation and Computation*, 6, 1-20.

Greenwood, J.A. and Durand, D. (1960). Aids for Fitting the Gamma Distribution by Maximum Likelihood. *Technometrics*, 2, 55-65.

Huber, P.J. (1981). *Robust Statistics*. New York: Wiley.

Nakagami, M. (1960). The m-Distribution – A General Formula of Intensity Distribution of Rapid Fading. In *Statistical Methods in Radio Wave Propagation*, ed. W.G. Hoffman, Oxford: Pergamon, pp. 3-36.

Rafiq, M. Ahmad, M. and Muhammed, F. (1996). Estimation of Parameters of the Gamma Distribution by the Method of Fractional Moments. *Pak. J. Statist.*, 12, 265-274.

Thom, H.C. (1958). A Note on the Gamma Distribution. *Monthly Weather Review*, 86, 117-122.