STATISTICS

Dalarna University

D-level Master's Thesis 2007

Which estimator of the dispersion parameter for the Gamma family generalized linear models is to be chosen?

Submitted by: Juan Du

Registration Number: 810926-T084

Supervisor: Md.Moudud Alam

Date: June 2007

ABSTRACT

For the Gamma family generalized linear models, the dispersion parameter is contained in the variance of the model parameter estimator. So it will affect the results of statistical inference or any kinds of tests that refer to variance. This paper reviewed several existing estimators of dispersion parameter via the Monte Carlo experiments to see which one is to be preferred when the sample size is different. The simulation results show that the bias corrected maximum likelihood estimator performs the best in comparison with the other methods when the sample size is small; in large sample size all the estimate methods perform similar.

Keywords: Generalized linear models; Gamma distribution; Dispersion parameter; Inference on model parameters.

1 Introduction

Normality and constancy of variance are no longer required in Generalized Linear Models (GLIM) [7]. Abandon these strict assumptions modeling become realistic and have more general fields of application. Exponential family is a class of distribution family that GLIM deals with.

Gamma distribution, which belongs to exponential family, is one of the commonly used distributions in GLIM. It is assumed to deal with responses which are positive and continuous. In general, we also suppose that these variables have constant coefficient variation.

Gamma distribution is applied in many fields, for example, it is commonly used in meteorology and climatology to represent variations in precipitation amount [10]. It also has been found practical application as a model in studies relating to life, fatigue, and reliability characteristics of industrial product [9].

1.1 Member of exponential family

The conventional probability density function of Gamma distribution can be expressed as:

$$f\left(y;\alpha,\beta\right)=\frac{1}{\Gamma\left(\alpha\right)\beta^{\alpha}}y^{\alpha-1}\exp(-\frac{y}{\beta}),\qquad y\geq 0,\qquad \alpha,\beta>0 \qquad (1)$$

Where, α and β are called the shape and the scale parameters respectively.

A different parameterization of Gamma distribution is used in generalized linear models for some convenience. The form of Gamma distribution used in GLIM can be presented as:

$$f(y;\mu,\nu)=\frac{1}{\Gamma(\nu)}\left(\frac{\nu}{\mu}\right)^{\nu}\exp[-\frac{\nu y}{\mu}]y^{\nu-1}; y\geq 0,\quad \nu>0,\quad \mu>0 \qquad (2)$$

Compare the two expressions above we have $\alpha=\nu$, $\beta=\mu/\nu$. In this form the parameter ν is called shape parameter which determines the shape of the distribution. It can be shown that the above Gamma distribution belongs to the exponential family of distributions (see Appendix A for detailed derivation). The canonical link is the reciprocal form of the expectation i.e. $g(\mu) = \frac{1}{\mu} = E(Y)$ and the dispersion parameter has the form $a(\phi) = \frac{1}{v}$.

1.2 Role of the dispersion parameter

According to the asymptotic theory, the maximum likelihood estimator of β follows the normal distribution: $\hat{\beta} \sim N(\beta, I^{-1})$, where I is the Fisher's information matrix. The corresponding variance is the diagonal elements of the inverse Fisher's information matrix, which is a function of the dispersion parameter (see Appendix C).

From the calculation (see Appendix C) we see that estimate of the model parameter does not depend on the dispersion parameter. But the effect of the dispersion parameter is obvious when we perform any kinds of statistical inference about the model parameter. Since the expression of standard error of the model parameter contains the dispersion parameter, $Cov(\hat{\beta}) = a(\phi)(X^T W X)^{-1}$, where X is the design matrix and W is the diagonal weight matrix[7].

In addition, We can see that $V(y) = \frac{\mu^2}{v}$, $E(y) = \mu$. So $V(y) = \frac{E^2(y)}{v}$ or $V(y) = a(\phi)E^2(y)$. Thus, the variance of the random variable is also affected by the dispersion parameter. We can denote the dispersion parameter as $\sigma^2 = \frac{1}{v}$.

In application, different statistical packages provide different default settings for the dispersion parameter estimator in the generalized linear models' procedure. For example, SAS provides MLE as the default while R provides method of the moment estimator as the default. Therefore, it is important for a data analyst to know in which case the default setting of the respective software package is good enough and in which case it does not and in that case which alternative is to be preferred.

This paper summarizes different existing methods of estimating the dispersion parameter and compares these estimators in terms of their preciseness and applicability in producing valid inference about model parameters by means of Monte-Carlo experiments. This paper is organized in the following way: section 2 lists the different estimation methods after literature review; section 3 compares the simulation results and section 4 draws the conclusions from the comparison of the results.

2 Literature Review

In order to draw a valid statistical inference about the model parameters, first we have to obtain a good estimate of the dispersion parameter. There are many suggestions in literatures on Gamma distribution and its parameters' estimation. Reviewed those related literatures we found four methods to estimate the dispersion parameter.

2.1 Maximum Likelihood Estimator (MLE)

The method of Maximum likelihood estimate is one of the most useful methods to estimate model parameters. McCullagh and Nelder (1989) apply this method to estimate the dispersion parameter of the gamma distribution. This estimator is denoted as

$$\sigma^2_{MLE} = \frac{1}{\hat{\nu}} \approx \frac{\bar{D}(6+\bar{D})}{6+2\bar{D}} \qquad (3)$$

Where, $\bar{D} = D(y;\hat{\mu})/n$ and $D(y;\hat{\mu})$ is the deviance of the model.

This estimator is based on deviance. It is extremely sensitive to rounding errors[1] in very small observations and in fact deviance is infinite if any component of y is zero. If gamma assumption is false, $\hat{\nu}^{-1/2}$ does not consistently estimate the coefficient of

[1] Is the difference between the calculated approximation of a number and its exact mathematical value.

variation [7]. It is well known that maximum likelihood estimates maybe biased when the sample size *n* or the total Fisher's information is small. The bias is usually ignored in practice, the justification being that it is negligible compared with the standard errors [2].

2.2 Bias Corrected Maximum Likelihood Estimator (BMLE)

In small or moderate size samples, however, a bias correction can be appreciable [2]. Bias corrected maximum likelihood estimator is obtained by including the term of order n^{-1} in the expected deviance. McCullagh and Nelder (1989) presented the form of the bias corrected maximum likelihood estimator of the dispersion parameter as,

$$\sigma^2_{BMLE} = \frac{1}{\hat{v}} \approx \tilde{D}\frac{6(n-k)+n\tilde{D}}{6(n-k)+2n\tilde{D}} \quad (4)$$

Where, $\tilde{D} = D(y;\hat{\mu})/\left(n-k\right)$ and n is the sample size and k is the number of parameters.

2.3 Moment Estimator (ME)

The method of Moment is another most commonly used way to estimate the unknown parameters. The moment estimator of the dispersion parameter is given as,

$$\sigma^2_{ME} = \frac{1}{\hat{v}} = \sum\left[\left(y-\hat{\mu}\right)/\hat{\mu}\right]^2/(n-k) = \frac{\chi^2}{n-k} \quad (5)$$

Where, χ^2 is the Pearson's Chi-square statistic, n is the sample size and k is the number of parameters.

Here, σ^2_{ME} is consistent for σ^2, if β has been consistently estimated. But it is inefficient, particularly for small values of the shape parameter [10], this kind of objection plagues moment method's application. In addition, unlike the normal model, the method of moment estimator of the dispersion parameter is not unbiased for the

Gamma models.

2.4 Quasi-maximum Likelihood Estimator (QMLE)

Stacy (1973) presented a set of estimators for the parameters of Gamma distribution using the method of quasi-likelihood based on the complete sample size[2]. The estimator of the shape parameter is appeared as the form of the inverse. We know the inverse form is equal to the dispersion parameter ($\alpha = \nu, \sigma^2 = \frac{1}{\nu}$). The estimator that he gave is demoted as,

$$\sigma^2_{QMLE}\,\frac{1}{\hat{\nu}} = \frac{n}{n-1}\sum_{i=1}^{n}\left(z_i - n^{-1}\right)\ln z_i \qquad (6)$$

Where, $z_i = y_i / n\bar{y}$, $\bar{y}$ is the sample mean.

In addition, Stacy (1973) has three constrains on the random sample data:

1) n>2
2) $y_i \neq y_j$ for some values of i and j.
3) $y_i > 0,\ for \quad i = 1, \ldots, n$

In his estimation, first he gave the log-likelihood of the generalized Gamma distribution, and then follows the steps of maximum likelihood method to get the estimators.

A brief note on the mathematical derivation of Stacy's (1973) method is given in Appendix D. For the convenience of comparison according to the equation (a) in maximum likelihood method (see Appendix B), we make little change and have the equation as,

$$\log\nu - \frac{\Gamma'(\nu)}{\Gamma(\nu)} = n^{-1}\sum_{i}\left[\log\frac{\hat{\mu}_i}{y_i}\right] \qquad (7)$$

When we compare equation (c) (see Appendix D) with (7), we can see that they are

[2] In fact, Stacy's Quasi-likelihood is different from Quasi-likelihood used in GLIM. The latter is based on variables' mean and variance to build likelihood. But Stacy's is just a maximum likelihood under generalized form of the Gamma distribution.

the same. Because $\hat{\mu}_i$ in the above equation is the expected response of the model while it is replaced by to the arithmetic mean $\bar{y}_p$ in the quasi-likelihood method. Changing a little bit of the above equation we get,

$$\frac{\Gamma'(\nu)}{\Gamma(\nu)} - \log\nu = n^{-1}\sum_i\left[\log\frac{y_i}{\hat{\mu}_i}\right] \qquad (8)$$

In the beginning of the introduction we have shown that $\alpha = \nu$. Thus, in this progress we notice that though the estimator in Stacy's(1973) paper is different from the method of maximum likelihood, when we see the steps of the calculation or the final equation which is used to estimate the dispersion parameter, it is clear that they are the same. The only difference is that the final form Stacy gives is the unbiased one, nothing new. Thus, we can see that quasi maximum likelihood is same to the maximum likelihood.

So, there are three methods used in this paper to estimate the dispersion parameter.

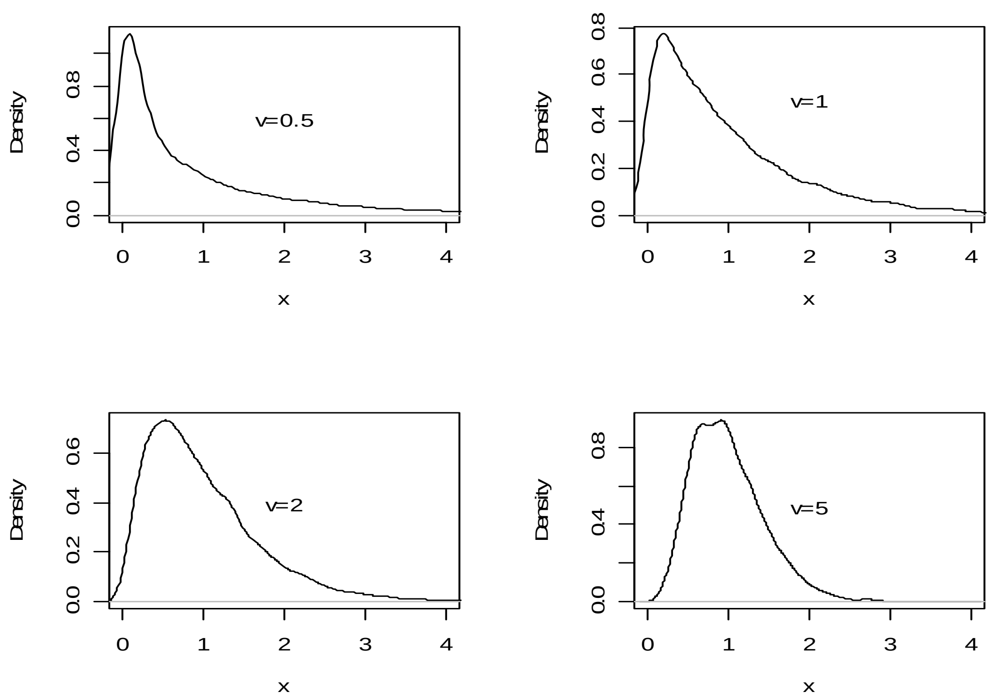


Figure 1 The Gamma distribution with different shape parameter under $\mu = 1$

3 Comparisons of different estimate methods

3.1 Comparison of estimate results

In this section we apply these three methods to estimate the model parameters and the dispersion parameter by Monte Carlo experiments. We assign four different true values of the shape parameter $v = 0.5, 1, 2, 5$, which is the inverse of the dispersion parameter. We choose these values because Gamma distribution with these shape parameters has different shape, which is shown is Figure 1.

Four different sample sizes we use is: $n = 10, 20, 50, 1000$; form of the link function is the inverse link: $g(\mu) = \frac{1}{\mu}$; and the linear predictor is $\eta = \alpha + \beta x$, where $\alpha = 0.5, \beta = 1$. The Monte-Carlo results are based on 10 thousand iterations. All the simulations (mean of the estimates of dispersion parameter and the corresponding confidence intervals) have been carried out in R 2.4.0 and are tabulated in Table 1.

From Table 1 we can see that when true value of v^{-1} decreases from 2 to 0.2, bias of the simulation results of the dispersion parameter is different (see Appendix E). For example, the maximum likelihood estimator bias is 2.23%, 0.76%, 0.3%, and 0.2% respectively for the different dispersion parameter $v^{-1} = 2, 1, 0.5, 0.2$ when sample size is 1000. In other word this means that the simulation results are affected by the value of dispersion parameter or the shape parameter.

Comparing results on each column of Table 1, we see the method of bias corrected maximum likelihood is the best among these three. This is very obvious when the sample size is small. For example, when n=10 and $v^{-1} = 2$, we can see that the mean of MLE is 1.6359, BMLE is 1.9055, and ME is 1.4851. The corresponding central 95% Monte-Carlo quantile intervals are (0.5573, 3.1022), (0.6692, 3.5976) and (0.5001, 3.2835) respectively. Here the simulated value under the bias corrected maximum likelihood method (1.9055) is the closest to the true dispersion parameter value 2; even though it seems that this estimator is still not unbiased. With the increasing of the sample size the simulation estimator is closer to the true value. When n=20 three different estimators are 1.8092, 1.9435, and 1.6882, the corresponding confidence intervals, of course, become shorter.

Table 1 **Simulation results of dispersion parameter**

1/v	Sample size	$1/\hat{\nu}$ MLE	BMLE	ME
2	10	1.6359(0.5573, 3.1022)	1.9055(0.6692, 3.5976)	1.4851(0.5001, 3.2835)
	20	1.8092(0.9582, 2.8444)	1.9435(1.0379, 3.0484)	1.6882(0.7564, 3.3588)
	50	1.8996(1.3279, 2.5460)	1.9529(1.3675, 2.6154)	1.8589(1.0673, 3.1942)
	1000	1.9554(1.8196, 2.0946)	1.9581(1.8220, 2.0974)	1.9921(1.7209, 2.3230)
1	10	0.8198(0.2540,1.6279)	0.9700(0.3109, 1.8941)	0.8606(0.2723, 1.9321)
	20	0.9074(0.4513, 1.4761)	0.9831(0.4941, 1.5889)	0.9185(0.4194, 1.7912)
	50	0.9600(0.6497, 1.3184)	0.9904(0.6719, 1.3578)	0.9643(0.5777, 1.5748)
	1000	0.9924(0.9176, 1.0693)	0.9939(0.9191, 1.0709)	0.9982(0.8833, 1.1308)
0.5	10	0.4070(0.1183, 0.8413)	0.4910(0.1463, 0.9973)	0.4643(0.1386, 1.0386)
	20	0.4536(0.2168, 0.7614)	0.4962(0.2390, 0.8277)	0.4795(0.2177, 0.9071)
	50	0.4812(0.3176, 0.6739)	0.4983(0.3295, 0.6968)	0.4906(0.3034, 0.7607)
	1000	0.4985(0.4589, 0.5395)	0.4993(0.4597, 0.5404)	0.4994(0.4489, 0.5555)
0.2	10	0.1613(0.0448, 0.3447)	0.1982(0.0558, 0.4192)	0.1939(0.0547, 0.4318)
	20	0.1806(0.0841, 0.3111)	0.1992(0.0932, 0.3420)	0.1964(0.0898, 0.3554)
	50	0.1923(0.1246, 0.2735)	0.1998(0.1296, 0.2839)	0.1985(0.1254, 0.2954)
	1000	0.1996(0.1830, 0.2168)	0.2000(0.1833, 0.2172)	0.2000(0.1814, 0.2198)

Note: 1. figures in the parenthesis present 95% quantile interval.

2. The estimator tabulated in this table is the mean of the 10 thousands simulations.

This results are true not only in small sample sizes but also in the medium sample size, n=50. We can get the same conclusion that estimator of bias correction maximum likelihood is the most appropriate estimator of the true value of dispersion parameter.

But in large sample size n=1000, when true value is 2 and 1, the simulation results show that ME is better than estimator of other methods. For example, when $\nu^{-1} = 2$, value of MLE, BMLE, and ME is 1.9554, 1.9581 and 1.9921 respectively. And when $\nu^{-1} = 1$ difference between ME and BMLE is 0.0043, which is not very large. But from the length of the 95% Monte-Carlo interval we see that BMLE's is smaller than the other two methods. For the other two values of dispersion parameter $\nu^{-1} = 0.5, 0.2$ the simulation results are similar. Except the case in large sample size, when $\nu^{-1} = 2$, method of BML seems better, no matter what is the true value of the dispersion parameter. This can be seen from the results of Table 1 and Appendix E.

Thus, we conclude that the bias corrected maximum likelihood method is relative better than the other two methods. But ME is better than BMLE when the dispersion parameter is large under the large sample size.

3.2 Evaluation of the hypothesis test

In the other part of the estimation experiment I have compared the estimators on the basis of their performance in the test of significance about the model parameters α and β. We have stated in the above section that the asymptotic distribution of the model parameter is normal. The null hypothesis assigns for α and β are $H_0 : \alpha = 0.5$ and $H_0^{'} : \beta = 1$ respectively. Under these two hypotheses we have,

$$Z_\alpha = \frac{\hat{\alpha} - \alpha}{s.e(\hat{\alpha})} \sim N(0,1),$$

$$Z_\beta = \frac{\hat{\beta} - \beta}{s.e(\hat{\beta})} \sim N(0,1).$$

After Simulation of the model parameters we get the distribution of test statistics

Z (see Figure2, 3, 4). Here we just give figures under dispersion parameter $\nu^{-1} = 2$ and the sample size is $n = 10, 50, 1000$ respectively as an example. From these 3 figures we can see that when sample size is small density plot of Z does not fit with the normal curve. However, there is nearly no large difference among those three estimation methods. From the figures based on Medium sample size n=50 and large sample size n=1000, we can get the same results. And with the increase of the sample size the data fits better.

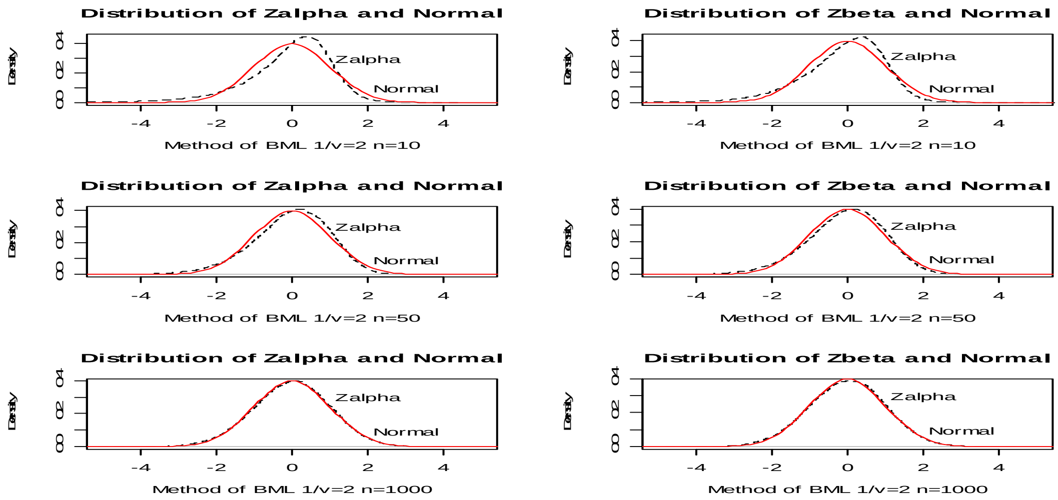


Figure 2 Distribution of statistics Z under method of BML

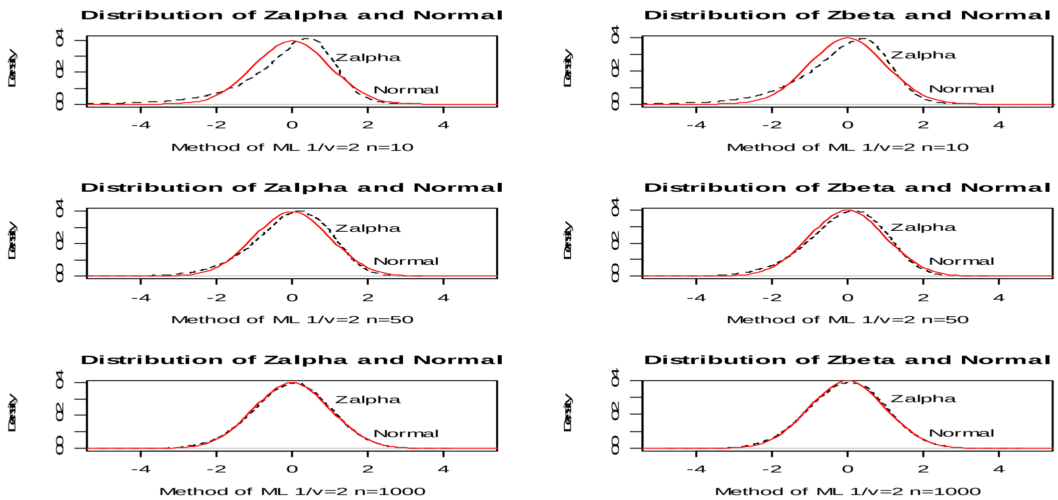
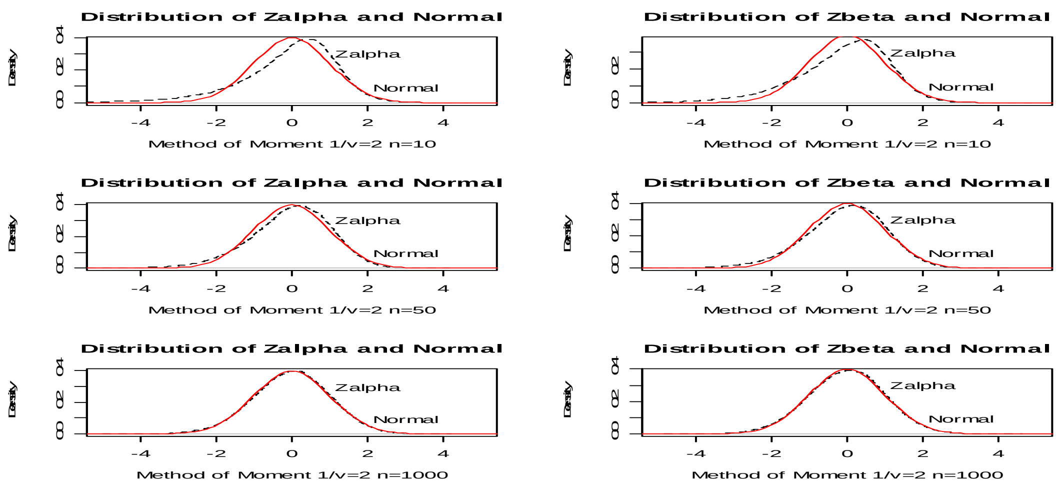


Figure 3 Distribution of statistics Z under method of Maximum Likelihood



Note: The broken line is the density of the test statistics Z.

Figure 4 Distribution of statistics Z under method of Moment

To evaluate a hypothesis test the common method is to evaluate the Type-I error and the Type-II error. In this section we choose the Type-I error to evaluate our hypothesis. Type-I error is the probability that when the null hypothesis is true but the hypothesis test incorrectly decides to reject it. For the null hypothesis in the above we calculate the probability of the Type-I error to see the confidence about the null hypotheses.

We denote probability of the Type-I error as P (I),

P (I) =P (Reject H_0 when H_0 is true)

= (value of test statistic is in rejection region when H_0 is true)

For our null hypothesis $H_0: \alpha = 0.5$ and $H_0^{'}: \beta = 1$, rejection region is absolute value of Z (test statistics for α or β) statistics lager than $Z_{\alpha/2}$, which is 1.96 when significance level takes 0.05.

Thus,

$$P(I) = P(|Z| > 1.96)$$

Applying this definition we get the probability of the Type-I error under different estimate methods and sample sizes. The results are shown in Table 2.

Table 2 **Probability of Type-I error**

		P(I)					
		MLE		BMLE		ME	
v	Sample size	α	β	α	β	α	β
0.5	10	0.09	0.088	0.074	0.071	0.105	0.105
	20	0.071	0.07	0.063	0.062	0.087	0.087
	50	0.059	0.059	0.056	0.056	0.067	0.067
	1000	0.053	0.054	0.053	0.053	0.051*	0.052
1	10	0.101	0.098	0.079	0.077	0.098	0.096
	20	0.074	0.072	0.064	0.062	0.078	0.077
	50	0.06	0.06	0.057	0.056	0.064	0.064
	1000	0.051*	0.05*	0.051*	0.05*	0.051*	0.05*
2	10	0.107	0.108	0.081	0.082	0.091	0.092
	20	0.076	0.076	0.065	0.065	0.073	0.073
	50	0.061	0.06	0.057	0.056	0.062	0.06
	1000	0.051*	0.051*	0.05*	0.051*	0.05*	0.051*
5	10	0.112	0.114	0.083	0.085	0.087	0.089
	20	0.077	0.079	0.065	0.066	0.068	0.07
	50	0.06	0.06	0.056	0.055	0.058	0.058
	1000	0.051*	0.05*	0.051*	0.05*	0.051*	0.05*

Note: * means the probability of Type-I error is non-significantly different from 0.05.

From Table 2 we can see that with the increase of the sample size, the probability of committing Type-I error both for α and β is decreasing. The largest commit probability for α and β is 0.112, 0.114 respectively; the smallest probability for α and β is the same: 0.05, which is equal to the significant level. Small and medium sample size does not have the prefect probability as large sample size, so the statistical hypothesis test does not work well under these kinds of sample sizes for all of the estimation methods. But it works well under every method when sample size is large.

Thus, we conclude that at 5% significant level we can not get a reasonable hypothesis test when sample size is not large enough. This is consistent with the fact that model parameters follow the asymptotic normal distribution.

But when we compare the error probability for different estimate methods, bias corrected maximum likelihood has smaller error commit probability than the other two methods. In other word, it is also relatively the best among these three. This result is coinciding with our first one, which obtained from comparison of the appropriateness of different estimators values (see Table 1).

The Kolmogorov-Smirnov test is used to show whether a sample comes from a special distribution, we apply it here to check whether test statistics Z has a standard normal distribution under the null hypothesis: $\alpha = 0.5, \beta = 1$.

The results of p-value of the KS tests for both α and β are tabulated in Appendix F. It shows that in small and medium sample size p-values are all less than significance level 5%, so we should reject the null hypothesis; when sample size is large, p-values are larger than significance level which means that we should accept null hypothesis. The situation for all these three methods is the same. And this conclusion is similar with that we get from the probability of the Type-I error.

4 Conclusions:

Maximum likelihood is the most commonly used methods in parameter estimation, but it has some disadvantages in some special cases especially with small sample sizes. By adjusting its weak points, for example the biased property, we can get better estimator. At the same time by introducing other methods such as moment method, we can also avoid some kind of weakness of MLE. A comparison of these rival methods for the estimation of the dispersion parameter of Gamma distribution has been studied in this paper via Monte Carlo experiments, which show that in small and medium sample sizes, bias corrected maximum likelihood estimator performs better than the other two estimators. But when we do simulation under large sample size, those three estimators of the dispersion parameter show almost the same performance.

Since dispersion parameter affect the variance of the model parameters, we judge the appropriateness of the dispersion parameter on the asymptotic standard normal distribution under the null hypothesis. Figures of the density of the standardized form of the dispersion parameter show that they fit the normal curve well just under large sample size. No big advantages can be checked from this part, results from those three methods seem very similar. Then, comparison the probabilities of the probability of the Type-I error and the p-values of the Kolmogorov-Smirnov test have the similar results with appropriateness of the estimator simulation.

Thus, depend on the results of the appropriateness of the simulation, the Type-I error and p-values of Kolmogorov-Smirnov test, we conclude that in the case of small or medium sample sizes bias corrected maximum likelihood is to be preferred. In large sample size all those methods are similar.

So, when use statistical software to estimate the dispersion parameter, if the sample size is large there is nearly no big difference between the SAS (default, MLE) and R (default, ME). But when the sample size is not big enough neither of these two methods are good. Then, we should apply bias corrected maximum likelihood method in the statistical software to estimate the dispersion parameter.

Reference:

[1] Choi, S. C. and Wette, R. (1969), Maximum likelihood estimation of the parameters of the Gamma distribution and their bias, *Technometrics* 11(4).

[2] Cordio, G. M. and McCullagh, P. (1991), Bias correction in generalized linear models, *J. R. Stat. Soc.* B 53(3), pp. 629-643.

[3] Davidian, M. and Carrol, R. J. (1987), Variance function estimation, *Journal of the American Statistical Association* 82(400), 1079-1091.

[4] Engelhardt, M. and Bain, L. J. (1977), Uniformly most powerful unbiased test of the scale parameter of a gamma distribution with a nuisance shape parameter, *Technometrics* 19, pp.77-81.

[5] Engelhardt, M. and Bain, L. J. (1978), Construction of optimal unbiased inference procedures for the parameters of the gamma distribution, *Technometrics* 20, pp.485-489.

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

[7] McCullagh, P. and Nelder, J. A. (1989), *Generalized Liner Models*, Chapman and Hall, London.

[8] Olsson, U. (2001), *Generalized Liner Models: an applied approach*, Studentlitteratur, Lund.

[9] Stacy, E. W. (1973), Quasi maximum likelihood estimator for two parameter gamma distribution, *IBM Res*. and *Devp*., pp 115-124.

[10] Wilks, D. L. (1990), Maximum likelihood estimation of the Gamma distribution using data containing zeros, *Journal of Climate* 3, pp 1495-1501.

Appendix A: Detail about the gamma distribution as a member of exponential family

$$f(y;\mu,\nu)=\frac{1}{\Gamma(\nu)}\left(\frac{\nu}{\mu}\right)^{\nu}\exp[-\frac{\nu y}{\mu}]y^{\nu-1};\, y\geq 0,\quad \nu>0,\quad \mu>0$$

$$=\exp\left[-\frac{\nu y}{\mu}+\nu\log(\nu)-\nu\log(\mu)+(\nu-1)\log(y)-\log(\Gamma(\nu))\right]$$

$$=\exp\{\nu\left[y\left(-\frac{1}{\mu}\right)-\log(\mu)\right]+\nu\log(\nu)+(\nu-1)\log(y)-\log(\Gamma(\nu))\}$$

Where,

The canonical parameter: $\theta=-\frac{1}{\mu}$

The cumulant function: $b(\theta)=\log(\mu)=-\log(-\theta)$

The dispersion parameter: $a(\phi)=\frac{1}{\nu}$

The mean of the distribution: $E(y)=b'(\theta)=\mu$

The variance of the distribution: $V(y)=b''(\theta)a(\phi)=\frac{\mu^2}{\nu}$

$C(y;\phi)=\nu\log(\nu)+(\nu-1)\log(y)-\log(\Gamma(\nu))$

The coefficient of the variation is $CV=\frac{\sqrt{\mathrm{var}(y)}}{mean(y)}=\frac{1}{\sqrt{\nu}}$, the parameter $\phi=\frac{1}{\nu}=CV^2$ is the dispersion parameter. Thus the gamma distributions have the same shape parameter then they have the same coefficient of variation.

Appendix B: Maximum Likelihood Estimate

The likelihood function and log-likelihood function of gamma distribution are

$$L = \exp\left\{\sum_i \left[-\frac{\nu y_i}{\mu_i} - \nu \log \mu_i + \nu \log \nu + (\nu - 1)\log y_i - \log \Gamma(\nu) \right]\right\}$$

and

$$\log L = \sum_i \left[-\frac{\nu y_i}{\mu_i} - \nu \log \mu_i + \nu \log \nu + (\nu - 1)\log y_i - \log \Gamma(\nu) \right]$$

The maximum likelihood estimator of ν is

$$\frac{\partial \log L}{\partial \nu} = \sum_i \left[-\frac{y_i}{\mu_i} - \log \mu_i + \log \nu + 1 + \log y_i - \frac{\Gamma'(\nu)}{\Gamma(\nu)} \right]$$

Let this derivative equal to zero we have

$$n\left(\log \nu - \frac{\Gamma'(\nu)}{\Gamma(\nu)} \right) = \sum_i \left[\frac{y_i}{\mu_i} - 1 + \log \mu_i - \log y_i \right] = \sum_i \left[\frac{y_i - \mu_i}{\mu_i} + \log \frac{\mu_i}{y_i} \right]$$

We know that the right hand side is the deviance of the gamma distribution,

$$D(y_i; \hat{\mu}) = 2\left[\log(y, y) - \log(y, \hat{\mu})\right] = 2\sum_i \left[\frac{y_i - \hat{\mu}_i}{\hat{\mu}_i} + \log \frac{\hat{\mu}_i}{y_i} \right]$$

Thus, we have the equation,

$$2n\left(\log \nu - \frac{\Gamma'(\nu)}{\Gamma(\nu)} \right) = D(y_i; \hat{\mu}) \qquad \text{(a)}$$

By solving the equation above we can get the maximum likelihood estimator

$$\sigma^2_{MLE} = \frac{1}{\hat{\nu}} \approx \frac{\bar{D}(6 + \bar{D})}{6 + 2\bar{D}}$$

Appendix C:

For exponential family the density function for a single observation is

$$f(y;\theta,\phi)=\exp\left[\frac{y\theta-b(\theta)}{a(\phi)}+c(y,\phi)\right]$$

The corresponding log-likelihood function is

$$l=\log(L(y;\theta,\phi))=\frac{y\theta-b(\theta)}{a(\phi)}+c(y,\phi)$$

To get the maximum likelihood estimate of parameter we take derivative of l with respect β_j, according to the chain rule we have

$$\frac{\partial l}{\partial\beta_j}=\frac{\partial l}{\partial\theta}\frac{d\theta}{d\mu}\frac{d\mu}{d\eta}\frac{\partial\eta}{\partial\beta_j}$$

We know that

$$b'(\theta)=\mu,\ \ b''(\theta)=V(\mu),\Rightarrow\frac{\partial\mu}{\partial\theta}=V(\mu)$$

$$\eta=\sum_i x_i\beta_i\ ,\ \ \Rightarrow\frac{\partial\eta}{\partial\beta_j}=x_j\ ,\ \ \Rightarrow\frac{\partial\eta}{\partial\beta_k}=x_k$$

Thus,

$$\frac{\partial l}{\partial\beta_j}=\frac{(y-\mu)}{a(\phi)}\frac{1}{V(\mu)}\frac{d\mu}{d\eta}x_j$$

In general, for n observations we have

$$\frac{\partial l}{\partial\beta_j}=\sum_i\frac{(y_i-\mu_i)}{a(\phi)}\frac{1}{V(\mu_i)}\frac{d\mu_i}{d\eta_i}x_{ij}$$

Let the first derivative equal to zero we can get the maximum likelihood estimates of the model parameters, which have no relation with the dispersion parameter.

In addition, according to the relation $g(\mu)=\eta$, we know that μ is the function of models parameter, so when take second derivative we have,

$$\frac{\partial^2 l}{\partial\beta_j\partial\beta_k}=\frac{\partial}{\partial\beta_k}\left[\frac{(y-\mu)}{a(\phi)}\frac{1}{V(\mu)}\frac{d\mu}{d\eta}x_j\right]$$

$$=\frac{\partial}{\partial\mu}\left[\frac{(y-\mu)}{a(\phi)}\frac{1}{V(\mu)}\frac{d\mu}{d\eta}x_j\right]\frac{\partial\mu}{\partial\eta}\frac{\partial\eta}{\partial\beta_k}$$

$$=\left[-\frac{1}{a(\phi)}\frac{1}{V(\mu)}\frac{d\mu}{d\eta}x_j+\frac{(y-\mu)}{a(\phi)}x_j\left(\frac{-1}{V^2(\mu)}V'(\mu)-\frac{1}{\left(g'(\mu)\right)^2}g''(\mu)\right)\right]\frac{1}{g'(\mu)}x_k$$

The information matrix I is the minus expectation of the second derivative, which can be expressed as

$$I=-E\left(\frac{\partial^2 l}{\partial\beta_j\partial\beta_k}\right)=-E\left(-\frac{1}{a(\phi)}\frac{1}{V(\mu)}\frac{d\mu}{d\eta}\frac{1}{g'(\mu)}x_jx_k\right)$$

$$=\frac{1}{a(\phi)}\frac{1}{V(\mu)}\left(\frac{1}{g'(\mu)}\right)^2x_jx_k$$

Where, $E(y-\mu)=0$.

Appendix D (Stacy (1973))

The generalized Gamma distribution is

$$f(y,\alpha,\beta,p)=\frac{|p|\,y^{p\alpha-1}}{\beta^{p\alpha}}\exp[-(y/\beta)^{p}],\qquad y\ge 0$$

Compare with the general form of this distribution we can see that it is got by setting the parameter p=1.

The logarithm likelihood function is

$$l=\log(\frac{|p|\,y^{p\alpha-1}}{\beta^{p\alpha}}\exp[-(y_i/\beta)^{p}])$$

Take the first derivatives and let them equal to zero,

$$-np\alpha+p\sum_{i=1}^{n}(y_i/\beta)^{p}=0,$$

$$-n\Psi(\alpha)+p\sum_{i=1}^{n}\log(y_i/\beta)=0,$$

$$\frac{n}{p}-n\alpha\log(\beta)+\alpha\sum_{i=1}^{n}y_i-\sum_{i=1}^{n}(y_i/\beta)^{p}\log(y_i/\beta)=0.$$

The corresponding results are

$$\beta=(\bar{y}_p\alpha^{-1})^{1/p}\qquad(b)$$

$$\Psi(\alpha)-\log(\alpha)=n^{-1}\sum_{i=1}^{n}\log(t_{ip})\qquad(c)$$

$$\alpha=\left[\sum_{i=1}^{n}\left(z_{ip}-\frac{1}{n}\right)\log(z_{ip})\right]^{-1}\qquad(d)$$

Where

$$\bar{y}_p=\sum_{i=1}^{n}y_i^{p}/n,$$

$$t_{ip}=y_i^{p}/\bar{y}_p\,,$$

$$z_{ip}=t_{ip}/n=y_i^{p}\Big/\sum_{i=1}^{n}y_i^{p}\,.$$

Appendix E: Table of the estimators' bias (%)

		Bias		
1/v	Sample size	MLE	BMLE	ME
2	10	18.205	4.725	25.745
	20	9.54	2.825	15.59
	50	5.02	2.355	7.055
	1000	2.23	2.095	0.395
1	10	18.02	3	13.94
	20	9.26	1.69	8.15
	50	4	0.96	3.57
	1000	0.76	0.601	0.18
0.5	10	18.6	1.8	7.14
	20	9.28	0.76	4.1
	50	3.76	0.34	1.88
	1000	0.3	0.14	0.12
0.2	10	19.35	0.9	3.05
	20	9.7	0.4	1.8
	50	3.85	0.1	0.75
	1000	0.2	0	0

Appendix F: p-value of Kolmogorov-Smirnov Tests

		p-value					
		MLE		BMLE		ME	
v	Sample size	Zalpha	Zbeta	Zalpha	Zbeta	Zalpha	Zbeta
0.5	10	< 2.2e-16	< 2.2e-16	< 2.2e-16	< 2.2e-16	< 2.2e-16	< 2.2e-16
	20	< 2.2e-16	< 2.2e-16	< 2.2e-16	< 2.2e-16	< 2.2e-16	< 2.2e-16
	50	< 2.2e-16	< 2.2e-16	< 2.2e-16	< 2.2e-16	< 2.2e-16	< 2.2e-16
	1000	0.00006715	0.000002582	0.0001008	0.000004176	0.00433	0.000604
1	10	< 2.2e-16	< 2.2e-16	< 2.2e-16	< 2.2e-16	< 2.2e-16	< 2.2e-16
	20	< 2.2e-16	< 2.2e-16	< 2.2e-16	< 2.2e-16	< 2.2e-16	< 2.2e-16
	50	< 2.2e-16	< 2.2e-16	< 2.2e-16	< 2.2e-16	< 2.2e-16	< 2.2e-16
	1000	0.02637	= 0.04125	0.02893	0.05524	0.03054	0.07354
2	10	< 2.2e-16	< 2.2e-16	< 2.2e-16	< 2.2e-16	< 2.2e-16	< 2.2e-16
	20	< 2.2e-16	< 2.2e-16	< 2.2e-16	< 2.2e-16	< 2.2e-16	< 2.2e-16
	50	< 2.2e-16	< 2.2e-16	< 2.2e-16	3.33E-16	< 2.2e-16	< 2.2e-16
	1000	0.05536	0.1927	0.05366	0.2473	0.05076	0.1919
5	10	< 2.2e-16	< 2.2e-16	< 2.2e-16	< 2.2e-16	< 2.2e-16	< 2.2e-16
	20	< 2.2e-16	< 2.2e-16	< 2.2e-16	< 2.2e-16	< 2.2e-16	< 2.2e-16
	50	< 2.2e-16	< 2.2e-16	1.90E-09	3.16E-08	3.39E-11	1.03E-10
	1000	0.3067	0.4076	0.3206	0.3271	0.3357	0.3488