5. QUASI LIKELIHOOD MODELS

5.1 Introduction :

In chapter 2, the method of obtaining least square estimates of the parameters in case of classical linear model was discussed. Least square method of estimation does not require form of the distribution of response variate, but it only assumes that variance of the response variate does not depend on on its mean and the systematic effects are linearly related. In chapter 3, fitting generalised linear model to the data is discussed. While fitting generalised linear model it is assumed that, distributional form of the response variate is known and is a member of one parameter natural exponential family.

When distributional form of the response variate is not known and variance depends on the mean or when we are not sure about the linearity of the systematic effects, we are unable to fit either classical linear model or generalised linear model by using the procedures discussed in chapter 2 and chapter 3. In such situations, when there is insufficient information to construct likelihood function, but the relationship between mean and variance is known, a new class of models called 'quasi likelihood models' (QLMs) can be fitted to the data. This new class of models was introduced by Wedderburn(1974). This chapter is devoted to the discussion of this model. Purpose of this chapter is to

(1) define the quasi likelihood function and study its properties,

(2) describe the model and the procedure of model fitting,

(3) study the method of obtaining least absolute deviations estimates of the model parameters in quasi likelihood model,

(4) define extended quasi likelihood function,

(5) describe general quasi likelihood model and model fitting procedure,

and

(6) model fitting for over dispersed and underdispersed grouped binary data.

5.2 : Quasi likelihood function :

The introduction of quasi likelihood function by Wedderburn (1974) widened the scope of generalised linear models. To define the quasi likelihood function some assumptions are needed.

Suppose that observations $Y_i (i=1,2,\dots,n)$ on the response variate Y are independent with mean μ_i and variance $\phi V_{ii}(\mu_i)$, where the variance function $V_{ii}(\mu_i)$ is some known function which depends on μ_i only and ϕ is the nuisance parameter. In addition to this the following two assumptions are made.

(i) μ_i is some known function of the parameters β in the model;

(ii) $V_{ii}(.)$ are identical functions, though their arguments and hence their values are different,

(iii) the dispersion parameter ϕ is constant for all the observations.

To make the meaning of first two assumptions more clear, below one illustration is given.

<u>Illustration</u>: Consider n independent observations $Y_i (i=1,2,\dots,n)$ on the binary response variate. The logistic link function corresponding to $Y_i (i=1,2,\dots,n)$ is,

$$T_i = ln(\mu_i / (1-\mu_i)). \tag{1}$$

Equation (1) along with the definition of the linear predictor $\underline{T}$

gives,

$$\mu_i = \frac{\exp(\beta_o + \sum_j X_{ij}\beta_j)}{1+\exp(\beta_o + \sum_j X_{ij}\beta_j)}, \qquad (2)$$

where the notations have their usual meaning. From equation (2) it is clear that $\mu_i(i=1,2,\ldots,n)$ is a function of β. Further, as pointed out in table (3.2), we have

$$V_{ii}(\mu_i) = \mu_i(1-\mu_i). \qquad (3)$$

Then, the variance functions are identical, though their arguments and hence their values are different. Further discussion in this section is made by taking into account all the three assumptions. Now we define the quasi likelihood function.

<u>Definition-1</u>: <u>Quasi likelihood function</u> (<u>Wedderburn (1974)</u>): If there are n independent observations Y_i $(i=1,2,\ldots,n)$, on the response variate, the quasi likelihood function for i^{th} observation Y_i is defined as,

$$Q_i(\mu_i, Y_i) = \int_{Y_i}^{\mu_i} \{(y_i - t)/[\phi V_{ii}(t)]\}\, dt + h(y_i), \qquad (4)$$

where h(y) is some function of y alone. Equivalently equation (4) can be written as

$$\frac{\partial\, Q_i(\mu_i, Y_i)}{\partial \mu_i} = \frac{(Y_i - \mu_i)}{\phi V_{ii}(\mu_i)}. \qquad (5)$$

Moris (1982) has shown that if there is a distribution from natural exponential family with the same variance function, there exists an equivalent log likelihood (ℓ) corresponding to the quasi likelihood (Q) based on single observation. To verify this result, some illustrations are given below. In all the illustrations, a single observation on the response variate is denoted by Y.

__Illustration__ 1. __Normal distribution__

Suppose the response variate has $N(\mu,\sigma^2)$ distribution. Hence the log likelihood of (μ,σ^2) is given by,

$$\ell(\mu,\sigma^2;Y) = -\ln(2\Pi\sigma^2)/2 - (Y-\mu)^2/(2\sigma^2). \tag{6}$$

For $N(\mu,\sigma^2)$ distribution, we know

$$\left.\begin{array}{l} \phi = \sigma^2 \\ V(\mu) = 1 \end{array}\right\} . \tag{7}$$

Hence equation (5) gives,

$$\frac{\partial Q(\mu,Y)}{\partial\mu} = \frac{(Y-\mu)}{\sigma^2} . \tag{8}$$

Differentiating equation (6) w.r.t. μ implies,

$$\frac{\partial\ell}{\partial\mu} = \frac{(Y-\mu)}{\sigma^2} . \tag{9}$$

Comparison of the equations (8) and (9) show that,

$$\frac{\partial\ell}{\partial\mu} = \frac{\partial Q}{\partial\mu}. \tag{10}$$

One can obtain an expression for quasi likelihood function (Q) by using equation (4). For example, in case of $N(\mu,\sigma^2)$ distribution,

$$Q(\mu,Y) = \int_Y^\mu \{(y - t)/\sigma^2\}\, dt + h(y),$$

$$= -(Y-\mu)/(2\sigma^2). \tag{11}$$

Equation (11) is obtained by taking $h(Y) = 0$. Similarly in other cases also quasi likelihood functions can be obtained. The table below gives quasi likelihood functions for the distributions in table 2.2.

TABLE 5.1

Distribution	Quasi likelihood function $Q(\mu;Y)$
Binary (Grouped)	$m^{*}\{Y\ln(\mu/(1-\mu)) + \ln(1-\mu)\}$
Poisson	$Y\ln(\mu) - \mu$
Normal	$-(y-\mu)^{2}/(2\sigma^{2})$
Exponential	$-(Y/\mu) - \ln(\mu)$
Gamma	$\nu[-(Y/\mu) - \ln(\mu)]$

Now, as n observations Y_i $(i=1,2,\dots,n)$ on the response variate, are assumed to be independent, the quasi likelihood function $Q(\underline{\mu};\underline{Y})$ for the complete data set is given by,

$$Q(\underline{\mu};\underline{Y}) = \sum_i Q_i(\mu_i;Y_i). \tag{12}$$

Quasi likelihood function (Q) has many properties in common with the log likelihood function. These common properties along with some other important properties are stated below in the form of theorems. The detailed proofs are also given.

5.2.1 <u>Properties of quasi likelihood function</u> :

We club some properties of Q which are in common with the log likelihood function (l) in the form of the theorem and then give entire proof of the theorem.Now onwards whenever convenient the subscirlpt 'i' will be dropped for simplicity.

Theorem 5.1: Let Y be the single observation on the response variate, and Q be as defined in equation (4). Then under the above mentioned assumptions, Q has the following properties.

(i) $E(\partial Q/\partial\mu) = 0$,

(ii) $E(\partial Q/\partial\beta_j) = 0$,

(iii) $E[(\partial Q/\partial\mu)^2] = -E(\partial^2 Q/\partial\mu^2) = 1/[\phi V(\mu)]$,

(iv) $E[(\partial Q/\partial\beta_j)(\partial Q/\partial\beta_l)] = -E\{\partial^2 Q/(\partial\beta_j\partial\beta_l)\}$,

and

$$E[(\partial Q/\partial\beta_j)(\partial Q/\partial\beta_l)] = (1/[\phi V(\mu)])[(\partial\mu/\partial\beta_j)(\partial\mu/\partial\beta_l)].$$

Proof : From equation (5) we have,

$$E\left\{\frac{\partial Q}{\partial\mu}\right\} = E\left\{\frac{(Y-\mu)}{\phi V(\mu)}\right\}$$

$$= [\phi V(\mu)]^{-1}\{E(Y)-\mu\}$$

$$= 0. \tag{13}$$

Secondly consider,

$$E\left\{\frac{\partial Q}{\partial\beta_j}\right\} = E\left\{\frac{\partial Q}{\partial\mu}\cdot\frac{\partial\mu}{\partial\beta_j}\right\}$$

$$= \left[\frac{\partial\mu}{\partial\beta_j}\right] E\left\{\frac{\partial Q}{\partial\mu}\right\}$$

$$= 0. \tag{14}$$

Further property (iii) can be treated as special case of the property (iv). Now to prove (iv), consider

$$E[(\partial Q/\partial\beta_j)(\partial Q/\partial\beta_l)] = E[(\partial Q/\partial\mu)^2][(\partial\mu/\partial\beta_j)(\partial\mu/\partial\beta_l)] \tag{15}$$

$$= E\left[\frac{(Y - \mu)^2}{(\phi V(\mu))^2} \right][(\partial\mu/\partial\beta_j)(\partial\mu/\partial\beta_l)]$$

$$= (1/[\phi V(\mu)])[(\partial\mu/\partial\beta_j)(\partial\mu/\partial\beta_l)] \, . \quad (16)$$

Finally,

$$-E[\partial^2 Q/(\partial\beta_j \partial Q/\partial\beta_l)] = -E\left\{ \frac{\partial}{\partial\beta_l}\left[\frac{(Y - \mu)}{(\phi V(\mu))} \frac{\partial\mu}{\partial\beta_j} \right] \right\}$$

$$= -E\left\{ (Y - \mu) \frac{\partial}{\partial\beta_l}\left[\frac{1}{(\phi V(\mu))} \frac{\partial\mu}{\partial\beta_j} \right] \right.$$

$$\left. - \frac{\partial\mu}{\partial\beta_j}\left[\frac{1}{(\phi V(\mu))} \frac{\partial\mu}{\partial\beta_l} \right] \right\}$$

$$= \left[\frac{1}{(\phi V(\mu))} \frac{\partial\mu}{\partial\beta_i} \frac{\partial\mu}{\partial\beta_j} \right] \qquad (17)$$

Equations (16) and (17) combinedly proves the property (iv). This property along with the equation (15) gives the proof of the property (iii).

To see another important property of the quasi likelihood, we prone the next theorem.

Theorem 5.2 : Suppose 'Q' and 'ℓ' denote respectively the quasi likelihood and log likelihood functions based on single observation Y of the response variate. Then

$$\frac{\partial \ell}{\partial\mu} = \frac{\partial Q}{\partial\mu}$$

holds iff the distribution of Y belongs to one parameter natural exponential family.

Proof :- First we prove that the condition is necessary. Then it is shown that the condition is sufficient.

<u>Necessary</u> : Suppose,

$$\frac{\partial \ell}{\partial \mu} = \frac{\partial Q}{\partial \mu} \tag{18}$$

Hence by substituting value of $(\partial Q/\partial \mu)$ from equation (5), above equation becomes,

$$\frac{\partial \ell}{\partial \mu} = \frac{(Y-\mu)}{\phi V(\mu)} . \tag{19}$$

Integrating equation (19) w.r.t. μ and putting

$$\int \frac{1}{V(\mu)} \, d\mu = \theta,$$

we obtain

$$\ell(\theta,\phi;Y) = (1/\phi)[\, Y\theta - g(\theta)] + \beta(\phi;Y)$$

$$= \alpha(\phi) \, [\, Y\theta - g(\theta)] + \beta(\phi;Y), \tag{20}$$

where $\alpha(\phi) = (1/\phi)$ and $\beta(\phi;Y)$ is the function of ϕ and Y only. The log likelihood in the equation (20) corresponds to the distribution which belongs to the natural exponential family. Thus result (18) is true only when Y comes from natural exponential family.

<u>Sufficient</u> : Suppose, the observation Y on the response variate comes from one parameter natural exponential family, so that equation (20) holds.

Differentiation of the equation (20) w.r.t. μ gives,

$$\frac{\partial \ell}{\partial \mu} = \alpha(\phi)[\,Y - g'(\theta)]\{d\theta/d\mu\}.$$

Now, since $g'(\theta) = \mu$, $\alpha(\phi) = \phi$ and $(d\mu/d\theta) = V(\mu)$, above equation can be written as,

$$\frac{\partial \ell}{\partial \mu} = \frac{(Y-\mu)}{\phi V(\mu)}$$

$$= \frac{\partial Q}{\partial \mu}.$$

Thus, if Y comes from one parameter natural exponential family, result (18) holds.

After studying the properties of quasi likelihood function (Q), in the next section we discuss the quasi likelihood model and the model fitting procedure.

5.3 : Quasi likelihood model :

<u>Definition-2</u> : <u>Quasi likelihood model</u> : Suppose Y_i (i=1,2,...,n) are independent observations on the response variate. Under this assumption, Wedderburn(1974) defined the quasi likeihood model as below.

(i) $E(Y_i) = \mu_i$; $i=1,2,\ldots,n$,

(ii) i^{th} $(i=1,2,\ldots,n)$ component of the linear predictor $\underline{T}$ is,

$$T_i = \beta_0 + \sum_j X_{ij}\beta_j \;,$$

(iii) $Var(Y_i) = \phi\, V_{ii}(\mu_i)$; $i=1,2,\ldots,n$,

and

(iv) approximate link between $\underline{\mu}$ and $\underline{T}$ is given by

$$\underline{T} = m(E(\underline{Y})),$$

where the dispersion parameter ϕ is assumed to be constant for all obsevations and m(.) is strictly monotonic differtiable function.

Fitting a quasi likelihood model (QLM) :

Fitting a quasi likelihood model means to estimate the model parameters β with the help of quasi likelihood function (Q). Wedderburn (1974) called these estimates of β as quasi likelihood estimates.

The theorem below gives the direction of computing quasi likelihood estimates.

Theorem 5.3 : The quasi likelihood estimates are same as weighted least square estimates.

Proof :- Proof of this theorem is very much similar to that of theorem (2.1). Let Y be the single observation on response variate. From theorem (5.1) we have the following two results.

$$E(\partial Q/\partial \mu) = 0, \qquad (1)$$

and

$$E(\partial Q/\partial \mu)^2 = -E(\partial^2 Q/\partial \mu^2) = 1/[\phi\ V(\mu)]. \qquad (2)$$

Using chain rule in differential calculus we can write

$$(\partial Q/\partial \beta_j) = (\partial Q/\partial \mu)(d\mu/dT)(\partial T/\partial \beta_j).$$

Substituting the value of $(\partial Q/\partial \mu)$ from equation (5.2-5) in the above equation, we have

$$\frac{\partial Q}{\partial \beta_j} = \frac{(Y-\mu)}{\phi\ V(\mu)}\ \frac{d\mu}{dT}\ X_j. \qquad (3)$$

As per equation (5.2-17),

$$-E\left\{\frac{\partial^2 Q}{\partial\beta_j\partial\beta_l}\right\} = \frac{1}{\phi V(\mu)}\left[\frac{\partial\mu}{\partial\beta_j}\right]\left[\frac{\partial\mu}{\partial\beta_l}\right]$$

$$= \frac{1}{\phi V(\mu)}\left[\frac{d\mu}{dT}\right]^2 X_j X_l$$

$$= \alpha(\phi) W\, X_j X_l. \tag{4}$$

here W is the weight given by

$$W = (V(\mu))^{-1}\left\{\frac{d\mu}{dT}\right\}^2$$

Equation (3) can also be written in terms of W as,

$$\frac{\partial Q}{\partial\beta_j} = \frac{(Y-\mu)}{\phi\, V(\mu)}\,\frac{d\mu}{dT}\, X_j.$$

Equation (3) can also be written in terms of W as,

$$\frac{\partial Q}{\partial\beta_j} = \left\{\frac{\alpha(\phi) W\, [X_j(Y-\mu)]}{(d\mu/dT)}\right\}. \tag{5}$$

The remaining proof goes in very similar way to that of obtaining maximum likelihood estimates by using some numerical methods in case of generalised linear models. Hence we repeat few main steps of the remainder proof.

Suppose $Y_i(i=1,2,\ldots,n)$ are n independent observations on the response variate Y. Hence, similar to the equation (3) we get

$$\frac{\partial Q(\underline{\mu};\underline{Y})}{\partial\beta_j} = \left\{\frac{\alpha(\phi)[X_{ij}(Y_i-\mu_i)]}{V_{ii}(\mu_i)}\right\}\frac{d\mu_i}{dT_i}$$

$$= T_j^* \text{ (say)}. \qquad (6)$$

Generally, the equation $T_j^* = 0$; $(j=0,1,\ldots,k)$ are non linear in β_j, and hence these equations are solved by iteration using Newton-Raphson method. This method gives m^{th} approximation of the estimate of $\underline{\beta}$ as,

$$\hat{\underline{\beta}}^{(m)} = \hat{\underline{\beta}}^{(m-1)} - \left[\underline{T}^{*(m-1)}\right]_{\underline{\beta}=\hat{\underline{\beta}}^{(m-1)}} \cdot \left[\frac{\partial^2 Q}{\partial\beta_j \partial\beta_l}\right]^{-1}_{\beta=\hat{\underline{\beta}}^{(m-1)}}.$$

If the Fisher's scoring method is chosen instead of Newton-Raphson-method to estimate $\underline{\beta}$, the equation (7) becomes

$$\hat{\underline{\beta}}^{(m)} = \hat{\underline{\beta}}^{(m-1)} + \left[\sum_i \left\{\frac{\alpha(\phi)[x_{ij}(Y_i-\mu_i)]}{V_{ii}(\mu_i)}\right\}.\frac{d\mu_i}{dT_i}\right]_{\beta=\hat{\underline{\beta}}^{(m-1)}}$$

$$\cdot\left[\alpha(\phi)\sum_i (W_{ii}x_{ij}x_{il})\right]^{-1}_{\underline{\beta}=\hat{\underline{\beta}}^{(m-1)}} \qquad (8)$$

Equation (8) can be written in matrix form as,

$$\left[(X'W\,X)\right]_{\underline{\beta}=\hat{\underline{\beta}}^{(m-1)}}.\{\hat{\underline{\beta}}^{(m)}\} = X'W\,\hat{\underline{Z}}^{(m-1)},$$

or in other way,

$$\hat{\underline{\beta}}^{(m)} = \left[(X'W\,X)\right]^{-1}_{\underline{\beta}=\hat{\underline{\beta}}^{(m-1)}}.\left\{X'W\,\hat{\underline{Z}}^{(m-1)}\right\}, \qquad (9)$$

where,

$$\underline{Z} = X\underline{\beta} + \left\{F^{-1}(\underline{Y}-\underline{\mu})\right\} . \tag{10}$$

Thus estimates of $\underline{\beta}$ can be obtained by using equations (9) and iterative weighted least square technique.

Note 1. Here it is necessary to note that, the standard errors of the quasi likelihood estimates and those for weighted least square estimates may be different.

2. The deviance function $D(\underline{Y};\underline{\mu})$ can also be written in terms of quasi likelihood function as,

$$D(\underline{Y};\underline{\mu}) = -2\phi\ [\,Q(\underline{\mu};\underline{Y}) - Q(\underline{Y};\underline{Y})] . \tag{11}$$

The values of deviance functions under generalised linear model and quasi likelihood model may be different. This is because, in case of generalised linear models the value of dispersion parameter (ϕ) is assumed to be known, whereas in case of quasi likelihood models, ϕ is unknown nuisance parameter which is to be estimated. If the estimate of ϕ is close to its true value, Wedderburn (1974) suggested estimate of ϕ as,

$$\hat{\phi} = \left\{\frac{\Sigma\,(Y_i-\mu_i)^2}{V_{ii}(\mu_i)}\right\} \frac{1}{(n-k-1)}$$

$$= \frac{X^2}{(n-k-1)} . \tag{12}$$

Below we give one example to illustrate the fitting of quasi likelihood model.

Example 5.1 : This example is taken from Collett (1991). (For complete description about the data see pages 2-3.) For this exmple the linear logistic model is fitted to the data as all the required information is available. In this example, suppose the responses are proportions of the plants survived, instad of the the number of plants survived. Further assume that, sample size is common, but unknown. In the example, two explanatory variates are time of planting and length of cutting respectively. Note that. responses are also binary. Thus, a quasi likelihood model to be fitted can be described as,

(i) $E(Y_i) = \mu_i$; $i=1,2,3,4$.

(ii) i^{th} $(i=1,2,3,4)$ component of the linear predictor $\underline{T}$ is,

$$T_i = \beta_0 + \sum_j X_{ij}\beta_j ,$$

and

(iii) $Var(Y_i) = \phi\, \mu_i(1-\mu_i)$; $i=1,2,3,4$.

Now, if the canonical link

$$T_i = \left\{\frac{\mu_i}{1-\mu_i}\right\}, \quad i=1,2,3,4;$$

is chosen then execution of the program given in Appendix-2(A) with minor changes gives following results under the model

$$\operatorname{logit}(Y_i) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} ;$$
$$i=1,2,3,4.$$

TABLE-5.2

Sr. No.(i)	Parameter estimate	Standard error	Parameter
0	-0.3039205	0.17654	INTERCEPT
1	-1.4275420	0.22069	TIME
2	1.0176910	0.21921	LENGTH

Deviance = 1.010284

Note : This data can be looked as a under dispersed binomial data.

In the same way quasi likelihood model described above can be fitted by using different link functions like complementory log-log function and log-log function. Execution of the program in Appendix-2(A) give the results corresponding to these two link functions also.

Bing (1993) gave some more discussion on the deviance function for quasi likelihood models. Due to the time facing we have not studied it explicitly.

There are some other methods of estimating the model parameters in quasi likelihood model. Below we discuss one of them.

As discussed in section (2.3), 'least absolute deviation' approach to estimate the parameters was introduced by Boscovich in the year 1757, nearly forty years before the introduction of least square approach. Hence, it is naturally quite interesting to see whether least absolute deviation approach can be used in quasi likelihood models instead of usual least square approach to estimate the parameters. Morgenthaler (1992) discussed how least absolute deviation method can be used to fit quasi likelihood model. The next section is devoted to study it explicitly.

5.4 Least absolute deviation method for fitting quasi likelihood model :

Suppose $Y_i (i=1,2,\dots,n)$ are independent observations on the response variate Y with m_i as the median corresponding to the distribution of Y_i. Let $\underline{x}_j$ $(j=1,2,\dots,k)$ be the vectors of known values of the stimulus variates $X_j (j=1,2,\dots,k)$. Then the model here is given by,

$$T_i = G(m_i), \qquad i=1,2,\dots,n, \tag{1}$$

where

(1) G(.) is strictly monotonic differentiable function and is a link function between m_i and β;

(2) $m_i (i=1,2,\dots,n)$ is median of the distribution of Y_i;

(3) $S_i(m_i)$ is the user supplied measure of variation for median m_i.

As discussed earlier in section (2.3), in this method the parameters are estimated so as to minimise,

$$\sum_i \left\{ \frac{|Y_i - m_i|}{[S_i(m_i)]^{(1/2)}} \right\}, \tag{2}$$

While fitting model (1) the question is whether it is suitable for the data to be analysed. When the responses are having discrete distributions, model (1) is not suitable. The reason behind this is, for discrete distributions, median is not a good function to use. Hence this model is appropriate only when the responses are having continuous distribution.

Consider gradient corresponding to the quantity in (2). It is given by,

$$[S(\underline{m})]^{-(1/2)}\left\{Sgn(\underline{y}-\underline{m})\right\}.$$

Hence the estimating equations to estimate β under L_1-fit by using least absolute deviation principle are

$$X'(W^{*})^{-(1/2)}\left\{Sgn(\underline{y}-\underline{m})\right\} = \underline{0}\ , \tag{3}$$

where,

$$W^{*}_{ii} = [\,S_i(m_i)]^{-(1/2)}\left[\frac{dm_i}{dT_i}\right]^{2}. \tag{4}$$

Note that, as for any arbitrary response Y_i

$$E\left\{Sgn(Y_i-m_i)\right\} = 0,$$

estimates using equations (3) are consistent. Morgenthaler (1992) proceeds further to compute the variance of the estimates under the additional regularity conditions.

Since this approach is suitable only for continuous responses and we are more interested in discrete distributions, we terminate this discussion.

For Wedderburn's (1974) original quasi likelihood function (Q) , tests like likelihood ratio tests (LR tests), score tests are useful for testing various hypotheses about the model parameters and link functions as in case of generalised linear models. But, none of these methods is useful for the comparison of different variance functions. Nelder & Pregibon (1987) introduced the new term, 'extended quasi likelihood function (Q^{+})'. This function is useful for the comparison of different variance functions also.

5.5 Extended quasi likelihood function :

Suppose there is a single observation Y on the response variate. Nelder & Pregibon (1987) constructed the extended quasi likelihood function (Q^+) such that, for known dispersion parameter ϕ, it is same as original quasi likelihood (Q). Hence Q^+ must be of the form,

$$Q^+(\mu,\phi;y) = Q(\mu,\phi;y) + \xi^*(\phi;y)$$

$$= -\left[\frac{D(y;\mu)}{2\phi}\right] + \xi^*(\phi;y). \qquad (1)$$

Nelder & Pregibon chosen the function $\xi^*(\phi;y)$ as,

$$\xi^*(\phi;y) = -(1/2)\ln\{2\Pi\phi V(y)\}, \qquad (2)$$

where V(Y) is the function of Y, obtained by replacing 'μ' by 'Y' in the variance function $V(\mu)$. Thus extended quasi likelihood function (Q^+) is given by,

$$Q^+(\mu,\phi;y) = -(1/2)\ln\left\{2\Pi\phi V(y)\right\} - \left[\frac{D(y;\mu)}{2\phi}\right]. \qquad (3)$$

It is easy to see that the estimates of μ obtained by maximising the extended quasi likelihood function (Q^+) coinsides with the usual quasi likelihood estimates of μ. This is because the extended quasi likelihood function (Q^+) is a linear function of the quasi likelihood function (Q).

As stated earlier, Mori's (1982) has shown that, if there is a distribution from natural exponential family with the same variance function, there exists an equivalent log likelihood (ℓ) corresponding to the quasi likelihood (Q) based on single observation. This result remains true for Q^+ in case of normal and inverse Gaussian distributions. For gamma distribution, Q^+ differs from log likelihood by a term depending on ϕ only. For discrete distributions like Poisson, binomial we can show that Q^+ is approximately equal to the log likelihood. This fact will be explicitly justified below. For this suppose Y is the single observation on the dependent variable.

Illustration 1. Normal distribution :

Suppose Y has $N(\mu,\sigma^2)$ distribution. The log likelihood of (μ,σ^2) becomes,

$$\ell(\mu,\sigma^2;y) = -\ ln(2\Pi\sigma^2)/2 - (y-\mu)^2/(2\sigma^2). \quad (4)$$

For $N(\mu,\sigma^2)$ distribution, table (3.2) gives

$$\phi = \sigma^2,\ V(\mu) = 1 \text{ and } D(y;\mu) = (y-\mu)^2. \quad (5)$$

Using results (5), equation (4) becomes

$$\ell(\mu,\phi;y) = -(1/2)ln\left\{2\Pi\phi V(y)\right\} - \left[\frac{D(y;\mu)}{2\phi}\right], \quad (6)$$

which is same as $Q^+(\mu,\phi;y)$. Hence for $N(\mu,\sigma^2)$ distribution,

$$\ell(\mu,\phi;y) = Q^+(\mu,\phi;y). \quad (7)$$

Illustration 2. Gamma distribution :

Suppose Y has gamma distribution with density

$$f(y;\mu,\nu) = \frac{(\nu/\mu)^{\nu}\{\exp(-\nu y/\mu)\}\ y^{\nu-1}}{\Gamma(\nu)}\ I_{(0,\infty)}(y),$$

for $(\mu,\nu > 0)$. Hence, the log likelihood function becomes

$$\ell(\mu,\nu;y) = -ln(\Gamma(\nu)) + \nu ln(\nu) - \nu ln(\mu) - (\nu y/\mu) + (\nu-1)ln(y)$$

$$= -(1/2)ln(2\Pi y^2/\nu) - \nu[-ln(y/\mu) + (y-\mu)/\mu] + g_1^*(\nu), \tag{8}$$

where,

$$g_1^*(\nu) = -\nu + (1/2)\, ln(2\Pi/\nu) - ln(\Gamma(\nu)) + \nu ln(\nu) \tag{9}$$

From table (2.2) we have,

$$\phi = 1/\nu,\ V(\mu) = \mu^2 \text{ and } D(y;\mu) = 2\{-ln(y/\mu) + (y-\mu)/\mu\}. \tag{10}$$

With the help of (10), equation (8) can be written as

$$\ell(\mu,\phi;y) = Q^+(\mu,\phi;y) + g_1^*(1/\phi). \tag{11}$$

Thus in case of gamma distribution, extended quasi likelihood (Q^+) differs from log likelihood by a term depending on ϕ only.

Illustration 3. Poisson distribution :

Let the observation Y be having Poisson distribution with mean μ. Therefore the log likelihood of μ is

$$\ell(\mu;y) = -\mu + y ln(\mu) - ln(y!). \tag{12}$$

Using Stirling's approximation, namely,

$$n! \cong (2\Pi n)^{(1/2)}(n^n)\exp\{-n\}, \tag{13}$$

in equation (12) we get,

$$\ell(\mu;y) \cong -\mu + y ln(\mu) - ln\{(2\Pi y)^{(1/2)} y^{(y)} \exp(-y)\}$$

$$\cong -\mu + y ln(\mu) - (1/2)ln(2\Pi y) - y ln(y) + y$$

$$\cong -(1/2)ln(2\Pi y) - \{y ln(y/\mu) - (y-\mu)\}. \tag{14}$$

For Poisson distribution, the table (2.2) gives

$$\phi = 1,\ V(\mu) = \mu,\ D(y;\mu) = 2\{y\ln(y/\mu) - (y-\mu)\}. \qquad (15)$$

From equations (14) and (15) it can be seen that, the quantity on r.h.s. of equation (14) is same as $Q^{+}(\mu,\phi;y)$. Thus for Poisson distribution,

$$\ell(\mu,\phi;y) = Q^{+}(\mu,\phi;y). \qquad (16)$$

<u>Illustration</u> 4. <u>Binomial</u> <u>distribution</u> :

Assume that Y has $B(m^{*},p)$ distribution. Then the log likelihood based on single observation can be written as

$$\ell(m^{*},p;y) = \ln\left\{{}^{m^{*}}C_{y}\right\} + y\ln(p) + (m^{*}-y)\ln(1-p). \qquad (17)$$

Using Stirling's approximation and readjusting the terms, equation (17) can be written as

$$\ell(m^{*},\mu;y) \cong -(1/2)\left[\ln(2\Pi) + \ln\left\{\frac{y(m^{*}-y)}{m^{*}}\right\}\right] + m^{*}\ln(m^{*}) - y\ln(y)$$

$$- (m^{*}-y)\ln(m^{*}-y) + y\ln(\mu/m^{*}) + (m^{*}-y)\ln\left\{\frac{m^{*}-\mu}{m^{*}}\right\}$$

$$\cong -(1/2)\ln\left\{\frac{2\Pi y(m^{*}-y)}{m^{*}}\right\} - \left\{y\ln(y/\mu) + (m^{*}-y)\ln\left[\frac{m^{*}-y}{m^{*}-\mu}\right]\right\}, \qquad (18)$$

with $\mu = m^{*}p$. For $B(m^{*},p)$ distribution it can be shown that,

$$\left.\begin{aligned} \phi = 1, \qquad & V(\mu) = \mu(m^{*}-\mu) \\ D(y;\mu) = 2[\, & y\ln(y/\mu) + (m^{*}-y)\ln\{(m^{*}-y)/(m^{*}-\mu)\}] \end{aligned}\right\} \qquad (19)$$

Therefore, equations (18), (19) along with the definition of Q^{+} imply,

$$\ell(m,\mu;y) = Q^{+}(m,\mu;y) \qquad (20)$$

Below in table (5.3), expressions for Q^{+} corresponding to above four distributions are written collectively.

TABLE-5.3

Dist.	Expressions for extended quasi likelihood in case of single observation
Binomial	$-(1/2)\ln(2\Pi y(m^*-y)/m^*)$ $-[y\ln(y/\mu)+(m^*-y)\ln((m^*-y)/(m^*-\mu))]$
Poisson	$-(1/2)\ln(2\Pi y)-\{y\ln(y/\mu)-(y-\mu)\}$
Normal	$-(1/2)\ln(2\Pi\sigma^2)-(y-\mu)^2/(2\sigma^2)$
Gamma	$-(1/2)\ln(2\Pi y^2/\nu)-\nu\{-\ln(y/\mu)+[(y-\mu)/\mu]\}+g_1^*(\nu)$

Note : While using Stirling's approximation to replace factorial notations (y!, say), one has to remember that as y approaches to zero, Stirling's approximation for factorial approaches to zero instead of unity. Hence it is better to use the following modified form in place of usual Stirling's approximation. The new formula is,

$$y! \cong \{2\Pi(y+c)\}^{(1/2)}(y^y)\exp(-y).$$

Nelder & Pregibon (1987) have mentioned that this approximation is better than Stirling's approximation.

Nelder & Pregibon (1987) have discussed the application of extended quasi likelihood (Q^+) in estimating non linear parameters affecting the variance function. Here by 'non linear parameter' we mean those parameters not included in the linear predictor.

Quasi likelihood models introduced by Wedderburn (1974) requires variance function in the form Var(Y) = ϕ V(μ). This requirement can be relaxed by using Q^+. Suppose variance function is of the form,

$$\mathrm{Var}(y) = \phi\, V_\theta(\mu). \tag{21}$$

With this new variance function we have,

$$Q_\theta^{+}(\mu,\phi;y) = -(1/2)\ln\left\{2\Pi\phi V_\theta(y)\right\} - \left[\frac{D_\theta(y;\mu)}{2\phi}\right], \tag{22}$$

where,

$$D_\theta(y;\mu) = -2\int \frac{(y-u)}{V_\theta(u)}\, du. \tag{23}$$

Nelder & Pregibon (1987) considered the family of variance function as

$$V_\theta(\mu) = \mu^\theta \;;\theta \geq 0. \tag{24}$$

It is easy to see that, for θ=0,1,2 the family (24) of variance functions gives variance functions corresponding to the normal, Poisson and gamma distributions respectively. Thus for these values of θ, form of the deviance function is well known. For other values of θ, an expression for deviance function can be obtained as below.

Using the form of variance function given in (24), equation (23) can be written as,

$$D_\theta(y;\mu) = -2\int \frac{(y-u)}{(\mu^\theta)}\, du$$

$$= -2\left[\frac{yu^{(1-\theta)}}{(1-\theta)} - \frac{u^{(2-\theta)}}{(2-\theta)}\right]_y^\mu$$

$$=2\left\{\left[y^{(2-\theta)} - (2-\theta)y\mu^{(1-\theta)} + (1-\theta)\mu^{(2-\theta)}\right] / \left[(1-\theta)(2-\theta)\right]\right\} \tag{25}$$

Thus for the family (24) of variance functions, formulae for variance functions can be summarised as below.

$$D_{\theta}(y;\mu) = \begin{cases} (y-\mu)^2 \ ; \text{ for } \theta = 0 \\ 2\{y\ln(y/\mu)-(y-\mu)\} \ ; \text{ for } \theta = 1 \\ 2\{(y-\mu)/\mu-\ln(y/\mu)\} \ ; \text{ for } \theta = 2 \\ 2[y^{(2-\theta)}-(2-\theta)y\mu^{(1-\theta)}+(1-\theta)\mu^{(2-\theta)}]/\{(1-\theta)(2-\theta)\} ; \text{ o.w.} \end{cases}$$

To fit an extended quasi likelihood model by considering the family (24) of variance functions, we must be sure about existance of the distribution which is a member of exponential family for each possible value of θ. Tweedie (1981) has discussed the distribution of variance function. He proved that for non negative value of θ, exponential family exists. Thus family (24) is one of the proper family of variance functions.

After discussing extended quasi likelihood function (Q^+), the next step is to describe the respective model and the model fitting procedure.

5.6 Extended quasi likelihood model :

5.6.1 Defining a model :

This model can be defined as follows.

Definition-4 : Extended quasi likelihood model : For n independent observations Y_i (i=1,2,...,n) on the response variate Y, an extended quasi likelihood model is same as joint model for mean and dispersion. Hence, extended quasi likelihood model has two parts, namely, model for mean and model for dispersion. Thus extended quasi likelihood model can be described as below.

(A) For the i^{th} $(i=1,2,\ldots,n)$ mean, model specification is,

(i) $E(Y_i) = \mu_i$,

(ii) i^{th} $(i=1,2,\ldots,n)$ component of the linear predictor $\underline{T}$ is given by,

$$T_i = \beta_o + \sum_j x_{ij}\beta_j = m(\mu_i)$$

and

(iii) $Var(Y_i) = \phi_i V_{ii}(\mu_i)$.

Thus dispersion parameter changes. Changes in the dispersion parameter are taken into account by the model for dispersion.

(B) For the i^{th} $(i=1,2,\ldots,n)$ dispersion, model specification is,

(i) $E(d_i) = \phi_i$,

(ii) i^{th} $(i=1,2,\ldots,n)$ component of the linear predictor (η) for dispersion is given by,

$$\eta_i = \gamma_o + \sum_{j=1}^{K^*} u_{ij}\gamma_j = h^*(\phi_i),$$

and

(iii) $Var(d_i) = \theta\, V_D(\phi_i)$,

where $V_D(\phi_i)$ is the variance function corresponding to d_i.

In the model for dispersion, d_i $(i=1,2,\ldots,n)$ is a proper measure of dispersion, $h^*(.)$ is the link function and η is the dispersion linear predictor. γ_j $(j=0,1,\ldots,k^*)$ are the model parameters and U_j $(j=1,2,\ldots,k^*)$ are the stimulus variates for dispersion model. Generally, the set U_j $(j=1,2,\ldots,k^*)$ is the subset of X_j $(j=1,2,\ldots,k)$.

5.6.2 Fitting a model :

For proper fitting of extended quasi likelihood model, a

choice of dispersion variance function is very important. In dispersion model, the square of Pearson's residual (r_P) or of deviance residual (r_D) is taken as a measure of dispersion. As discussed in section (2.5) it is clear that, it is more suitable to use deviance residual among the two. In some applications Pearson's residual is used because of computational simplicity.
Now from equations (3.7-1) and (3.7-3) we have,

$$(r_P^2)_i = \left\{\frac{(y_i-\mu_i)^2}{V_{ii}(\mu_i)}\right\} \tag{1}$$

and

$$(r_D^2)_i = D_i(y_i;\mu_i). \tag{2}$$

To fit an extended quasi likelihood model, expressions for mean and variance of the dispersion responses (d_i) are required. So we obtain these expressions.

It is easy to see that for any of the two forms of dispersion responses described in equations (1) and (2), $E(d_i)=0$, (for $i=1,2,\ldots,n$). This fact can be justified as below.

Justification :

<u>part-1</u> : Assume that the dispersion responses are of the form given in equation (1). Then for any component $(r_P^2)_i$ $(i=1,2,\ldots,n)$ we have,

$$\begin{aligned} E\,[(r_P^2)_i] &= E\left\{\frac{(y_i-\mu_i)^2}{V_{ii}(\mu_i)}\right\} \\ &= Var(Y_i)/V_{ii}(\mu_i) \\ &= \phi_i. \end{aligned} \tag{3}$$

<u>Part-2</u> : Suppose, the dispersion responses are taken as deviance residuals $(r_D^2)_i$ $(i=1,2,\ldots,n)$ as given in equation (2).

Now to avoid confusions consider a single observation case. Note that to behave Q^+ like log likelihood w.r.t. ϕ, we must have

$$E(\partial Q^+/\partial\phi) = 0, \text{ and } E(\partial^2 Q^+/\partial\phi^2) = -E(\partial Q^+/\partial\phi)^2. \qquad (4)$$

Equations (5.4-3) and (4) combinedly imply

$$E\left\{ -\left(\frac{1}{2\phi}\right) + \frac{D(Y;\mu)}{2\phi^2} \right\} = 0,$$

which gives

$$E[\,D(Y;\mu)] = \phi\ . \qquad (5)$$

Hence for n observations case we get for i^{th} dispersion response,

$$E[\,D(Y_i;\mu_i)] = \phi_i, \text{ for } i=1,2,\dots,n. \qquad (6)$$

Since

$$[\,D(Y_i;\mu_i)] = (r_D^2)_i \text{ for } i=1,2,\dots,n,$$

the justification is copleted.

To obtain expressions for variance, from equations (1) and (2) it can be observed that, for normal responses dispersion variable d_i has $\phi_i\chi_1^2$ distribution. Hence, $Var(d_i) = 2\phi_i^2$. For non normal distributions this may not be the exact situation. Thus if the responses are having non normal distribution, some adjustment in responses (d_i) of the dispersion model is necessary. If d_i's are taken as r_P^2, adjustment is necessary for additional variability in r_P^2. In case of r_D^2 we show that no adjustment is necessary. Pregibon (1988) has discussed some adjustments. The same adjustments are discussed below in detail.

First we discuss about adjustment for additional variability in r_P^2. Though we are taking

$$Var(d_i) = 2\phi_i^2\ ,\ i=1,2,\dots,n,$$

frequently the variance exceeds this value. To compute correct variance, consider

$$\mathrm{Var}\left\{(r_p^2)_i\right\} = (V_{ii}(\mu_i))^{-2}\left[\mathrm{Var}(Y_i-\mu_i)^2\right]$$

$$= (V_{ii}(\mu_i))^{-2}\ [\,K_4 + 2K_2^2\,]$$

$$= (V_{ii}(\mu_i))^{-2}\ (2K_2^2)[\,1 + \rho_4/2\,], \qquad (7)$$

where

(i) $\rho_4 = K_4/K_2^2$, is the standardised fourth cumulant,

(ii) $K_r(r=1,2,\ldots,)$ denote the respective cumulants corresponding to the response variate.

Since $K_2 = \mathrm{Var}(Y_i)$, the equation (7) reduces to

$$\mathrm{Var}\left\{(r_p^2)_i\right\} = 2\phi_i^2(1+\rho_4/2). \qquad (8)$$

For the variance in r_p^2, the value of ρ_4 is necessary. For the overdispersed binomial and Poisson distributions the adjustment can be made if the fourth cumulant of Y has some particular relation with the second cumulant. If the condition

$$K_{(r+1)} = K_r' K_2,\ \text{for } r \geq 2 \qquad (9)$$

holds upto fourth cumulant, then using

$$K_2 = \phi\ V. \qquad (10)$$

we have

$$K_3 = \left\{\frac{\partial}{\partial\mu}[\,K_2]\right\}$$

$$= \phi^2 V(\partial V/\partial\mu)$$

$$= \phi^2 V(\partial V/\partial\theta)(d\theta/d\mu)$$

$$= \phi^2(\partial V/\partial\theta). \tag{11}$$

Similarly, from the equation (9) we get,

$$K_4 = K_3'K_2$$

$$= \phi^3\left\{\frac{\partial}{\partial\mu}(\partial V/\partial\theta)\right\}V$$

$$= \phi^3 V\left\{\frac{\partial}{\partial\theta}\left[\frac{d\theta}{d\mu}(\partial V/\partial\theta)\right]\right\}$$

$$= \phi^3 V\left\{\frac{\partial}{\partial\theta}\left[\frac{1}{V}(\partial V/\partial\theta)\right]\right\}$$

$$= \phi^3(\partial^2 V/\partial\theta^2). \tag{12}$$

Using the equations (9) to (12), the third and fourth standardised cumulants (ρ_3 and ρ_4) can be obtained as below.

$$\rho_3 = \left\{\frac{K_3^2}{K_2^3}\right\}^{(1/2)}$$

$$= \phi^{1/2}(\partial V/\partial\theta)/(V^{3/2})$$

$$= \phi^{1/2}(V'(\mu))/(V^{1/2}), \tag{13}$$

and

$$\rho_4 = \phi(\partial^2 V/\partial\theta^2)/V^2$$

$$= \frac{\phi}{V^2}\left\{\frac{\partial}{\partial\theta}\left[\frac{\partial}{\partial\mu}(V^2)\right]\right\}$$

$$= \frac{2\phi}{V^2}\left\{\frac{\partial}{\partial\theta}\left[VV'(\mu)\right]\right\}$$

$$= \frac{2\phi}{V^2}\left\{\frac{\partial}{\partial\mu}\left[V^2V'(\mu)\right]\right\}$$

$$= 2\phi V''(\mu) + 4\rho_3^2, \qquad (14)$$

where $V'(\mu)$ and $V''(\mu)$ are repectively the partial derivatives of $V(\mu)$ w.r.t. μ.

Below we obtain expressions for adjustment factor $(1+\rho_4/2)$, corresponding to some well known distributions.

Illustration 1. Normal distribution : Suppose the response variate Y has $N(\mu,\sigma^2)$ distribution. From table (3.2) we have,

$$\phi = \sigma^2 \text{ and } V(\mu) = 1. \qquad (15)$$

Therefore from equations (13) and (14) we see that, $\rho_4 = 0$. This implies,

$$(1+\rho_4/2) = 1. \qquad (16)$$

Illustration 2. Over dispersed binomial distribution : Let the response variate Y is over dispersed $B(m^*,p)$ variate. Then the variance of Y is given by,

$$\text{Var}(Y) = (\phi^*/m^*)\{\mu(m^*-\mu)\}. \qquad (17)$$

It is clear from equation (17) that,

$$\phi = \phi^*/m^*, \text{ and } V(\mu) = \{\mu(m^*-\mu)\}. \qquad (18)$$

Differentiating the variance function $V(\mu)$ in the equation (18) w.r.t. μ successively, we obtain

$$\left.\begin{aligned} V'(\mu) &= (m^* - 2\mu) \\ V''(\mu) &= -2 \end{aligned}\right\} \qquad (19)$$

Hence from the equations (13), (18) and (19), we get

$$\rho_3^2 = \left\{\frac{\phi^*}{(m^*)}\left[\frac{(m^*-2\mu)^2}{\mu(m^*-\mu)}\right]\right\}, \tag{20}$$

which with thw help of equations (14) gives

$$\rho_4 = -4(\phi^*/m^*)\left\{1 + \frac{(m^*-2\mu)^2}{\mu(m^*-\mu)}\right\}.$$

Hence the value of $(1+\rho_4/2)$ becomes,

$$(1+\rho_4/2) = 1 + \frac{2\phi^*}{m^*}\left\{\frac{1-5p(1-p)}{p(1-p)}\right\}. \tag{21}$$

The variance of deviance residual (r_D^2) can be obtained as follows.

By taking differtiation twice of the extended quasi likelihood (Q^+) given in the equation (5.4-3) and taking expectation using equation, we get

$$E(\partial^2 Q^+/\partial\phi^2) = E\left\{\frac{\partial}{\partial\phi}\left[-(\frac{1}{2\phi}) + \frac{D(Y;\mu)}{2\phi^2}\right]\right\}$$

$$= \left\{-\left[\frac{E(D(Y;\mu))}{\phi^2}\right] + \frac{1}{2\phi^2}\right\}$$

$$= -(1/2\phi^2). \tag{22}$$

Consider

$$E(\partial Q/\partial \phi)^2 = E\left\{-\left(-\frac{1}{2\phi}\right) + \frac{D(Y;\mu)}{2\phi^2}\right\}$$

$$= E\left\{\frac{D^2(Y;\mu)}{4\phi^4}\right\} + (1/4\phi^4) - (1/2\phi^2). \quad (23)$$

Now, according to the result (4), we have

$$E(\partial^2 Q/\partial \phi^2) = - E(\partial Q/\partial \phi)^2. \quad (24)$$

Hence, from the equations (22) to (24) we obtain,

$$E(D^2(Y;\mu)) = 3\phi^2. \quad (25)$$

Thus the equations (5) and (25), it is clear that

$$Var(D(Y;\mu)) = 2\phi^2. \quad (26)$$

Thus gamma distribution with scale factor '2' is appropriate for dispersion model. By assuming this distribution to dispersion variates $d_i(i=1,2,\ldots,n)$, the estimating equations for β and γ are respectively;

$$\sum_i \left\{\frac{(y_i-\mu_i)}{\phi_i V(\mu_i)} \cdot \frac{\partial \mu_i}{\partial \beta_j}\right\}, \text{ for } j=0,1,\ldots,k; \quad (27)$$

$$\sum_i \left\{\frac{(d_i-\phi_i)}{\phi_i^2} \cdot \frac{\partial \phi_i}{\partial \gamma_j}\right\}, \text{ for } j=0,1,\ldots,k^*; \quad (28)$$

Equations (27) and (28) are the estimating equations, provided $E(d_i) = \phi_i$ and atleast $V_D(\phi_i) = 2\phi_i^2$, regardless of the distribution of responses Y_i. For the dispersion model, most common link functions are identity and log link functions.

Algorithm for fitting extended quasi likelihood model is same as that given for fitting generalised linear model with varying dispersion in section (3.10). As mentioned earlier, actual programming for the same is quite complicated. The well known software package 'GLIM' is available to fit this model.

As stated in section (4.3), for over dispersed or under dispersed grouped binary responses quasi likelihood model can be fitted. Below we discuss the initial steps of this model fitting.

5.7 Model fitting for over dispersed and under dispersed grouped binary data :

While fitting a grouped binary data, there are many causes giving rise to the problem of over or under dispersion. The main reason is, the model taken has fewer terms than requirement in its systematic part. For example, in case of 'factorial experiments', deletion of significant interactions from the model may indicate that the model is inadequate. Another reason is improper scale of the covariate. Final and very important cause is the dependence of observations. Some times presence of few outliers in the data may indicate that the data is over or under dispersed. The under dispersion case occurs rarely. Hence we consider only the case of over dispersion.

5.7.1 Fitting a model to over dispersed grouped binary data :

Suppose the unobservable random variable $P_i(i=1,2,\ldots,n)$ are independently distributed over the interval $[0,1]$ with mean and variance of $P_i(i=1,2,\ldots,n)$ are respectively,

$$E(P_i) = \Pi_i \;\&\; Var(P_i) = \theta^* \Pi_i (1-\Pi_i), \qquad (1)$$

Assume that for given $P_i(=p_i$, say), Y_i^* $(i=1,2,\ldots,n)$ has $B(m_i^*, p_i)$ distribution. For fitting a model, the unconditional variance of

Y_i^* $(i=1,2,\ldots,n)$ is necessary. To compute the unconditional variance following well known results in probability are to be used (see e.g. Rohatgi (1988) page-170).

<u>Result-1</u>: If X and Y are two random variables, then

$$\left.\begin{aligned} &E(Y) = E\{E(Y|X)\} \\ \text{and} \quad &\\ &Var(Y) = E\{Var(Y|X)\} + Var\{E(Y|X)\}\end{aligned}\right\} \qquad (2)$$

Using the above result in equations (2), we have

$$Var(Y_i^*) = E\{Var(Y_i^*|P_i)\} + Var\{E(Y_i^*|P_i)\} \qquad (3)$$

Consider,

$$\begin{aligned} E\{Var(Y_i^*|P_i)\} &= E\{m_i^* P_i(1-P_i)\} \\ &= m_i^*\{E(P_i)-E(P_i^2)\} \\ &= m_i^*\{\Pi_i-[\theta^*\Pi_i(1-\Pi_i)+\Pi_i^2]\} \\ &= m_i^*\Pi_i\{1-\theta^*(1-\Pi_i)-\Pi_i]\} \\ &= m_i^*\Pi_i(1-\Pi_i)(1-\theta^*), \end{aligned} \qquad (4)$$

and

$$\begin{aligned} Var\{E(Y_i^*|P_i)\} &= Var(m_i^* P_i) \\ &= m_i^{*2}\theta^*\Pi_i(1-\Pi_i). \end{aligned} \qquad (5)$$

With the help of equations (4) and (5), equation (2) can be written as,

$$\begin{aligned} Var(Y_i^*) &= m_i^*\Pi_i(1-\Pi_i)\{(1-\theta^*)+m_i^*\theta^*\} \\ &= m_i^*\Pi_i(1-\Pi_i)\{1+\theta^*(m_i^*-1)\}. \end{aligned} \qquad (6)$$

Hence,

$$\phi_i^* = \{1+\theta^*(m_i^*-1)\}. \qquad (7)$$
$$i=1,2,\ldots n$$

<u>Note</u> : If m_i^* ($i=1,2,\ldots,n$) are equal and the common value is m_o^*, then from equation (7) we have,

$$\phi_i^* = \phi^* = \{1+\theta^*(m_o^*-1)\}. \qquad (8)$$
$$i=1,2,\ldots n$$

Thus in such situations extended quasi likelihood model reduces to a quasi likelihood model given by Wedderburn (1974). Thus quasi likelihood model can be fitted to such type of data.

In the same fashion, the case of over dispersed Poisson distribution can be handled by assuming unobsevable random variables λ_i ($i=1,2,\ldots,n$) having independent gamma distribution with density function given in the equation (2.2-11) and for given λ_i the responses Y_i($i=1,2,\ldots,n$) are independent Poisson variates with respective parameters λ_i. Then it can be shown that the unconditional variance of Y_i is,

$$\mathrm{Var}(Y_i) = (\mu_i^2/\nu_i) + \mu_i. \qquad (9)$$

In the next chapter, which is concluding chapter of the dissertation we try to explain the procedure of data analysis by fitting a model to it by using the available information.