Actuarial Applications of Distance-Based Generalized Linear Models

Eva Boj[1], Josep Fortiana[2], Anna Esteve[3], M. M. Claramunt[1], T. Costa[1]

[1] Departament de Matemàtica Econòmica, Financera i Actuarial, Universitat de Barcelona
Diagonal 690, 08034 Barcelona, Spain.
E_mails: evaboj@ub.edu mmclaramunt@ub.edu tcosta@ub.edu

[2] Departament de Probabilitat, Lògica i Estadística, Universitat de Barcelona
Gran Via 585, 08007 Barcelona, Spain. E_mail: fortiana@ub.edu

[3] Centre d'Estudis Epidemiològics sobre les Infeccions de Transmissió Sexual i Sida de Catalunya (CEEISCAT), Hospital Universitari Germans Trias i Pujol.
Ctra. de Canyet, s/n. 08916 Badalona, Spain. E_mail: aeg.ceescat.germanstrias@gencat.cat

Abstract

In this presentation we propose the use of Distance-Based Generalized Linear Models (DB-GLM) in the solution of actuarial problems for which a GLM is adequate. Distance-Based Regression (DBR), a particular instance of these models where the GLM linear predictor is obtained from the observed explanatory variables through a distance function, has already been successfully applied to the problem of automobile *a priori* rate-making.

DBR extends Ordinary Least Squares (OLS) linear regression, allowing for general type explanatory variables. The *distance-based* name is due to the fact that predictors and response are related by means of distances between observations. OLS is included as a particular case when predictors are continuous and the distance function is Euclidean. DB-GLM generalizes DBR just as GLM does with OLS: flexibilizing the response variable distribution and allowing a link function to insert the response expectation in the model.

We illustrate the usage of this new model with a real portfolio of a well-known data set. We fit Binomial and Poisson DB-GLM models for *claim frequency*. Computations have been programmed with the R language.

Keywords: Distance-based prediction, Generalized Linear Model, Non-Life Insurance, Pricing, Reserving, R language.

1. Introduction

In this presentation we propose the use of DB-GLM in the solution of actuarial problems for which a GLM can be applied, such as the rate-making and the claim reserving processes [see e.g. Anderson *et al.* 2002; Boj *et al.* 2004; Brockman and Wright 1992; de Jong and Heller 2008; Haberman and Renshaw 1998; Kaas *et al.* 2008; Nelder and Verral 1997; Ohlsson and Johansson 2010; Renshaw 1994; Venter 2007; Zehnwirth 1994 among others]

DBR, a particular instance of these models, has already been successfully applied to the problem of automobile *a priori* rate-making [see Boj *et al.* (2000, 2001, 2004, 2007)]. DBR extends Ordinary Least Squares (OLS) linear regression, allowing for mixed explanatory variables (quantitative and qualitative) or of a more general type, such as functional variables. The *distance-based* name is due to the fact that predictors and response are related by means of distances between observations. OLS is included as a particular case when predictors are continuous and the distance function is Euclidean. DB-GLM generalizes DBR just as GLM does with OLS: flexibilizing the response variable distribution and allowing a link function to insert the response expectation in the model.

In order to analyze *claim frequency* a standard practice is to assume a Binomial or a Poisson distribution and, for *claim amount*, continuous distributions such as Gamma or Inverse Gaussian. All these distributions belong to exponential dispersion families, which allows GLM modelling, hence also DB-GLM, with the added benefit of a wider repertoire of predictors.

The presentation is organized as follows: In section 2 we make a brief description of the new model, the DB-GLM, which is described with detail in Boj *et al.* 2011; in section 3 we illustrate the usage of DB-GLM with an academic example using the R program for computations.

2. Distance-based generalized linear models

In this section, first we remember the main characteristics of the DBR and of the GLM. Then we indicate how to construct the DB-GLM.

2.1. The distance-based linear model

Distance-Based Regression was introduced by Cuadras (1989) and has been developed in Cuadras *et al.* (1990, 1996) and in Boj *et al.* (2007, 2010, 2011). A sketchy description of DBR is as follows: We have the value y_i of a real-valued response and a known, constant positive weight $w_i \in (0,1)$ for each *i*-th individual Ω_i in a given set $\Omega = \{\Omega_1,\dots,\Omega_n\}$, randomly drawn from a population. The $n \times 1$ weight vector $w = \{w_1,\dots,w_n\}$ is standardized to unit sum, i.e., $\mathbf{1}'w = 1$, where $\mathbf{1}$ is the $n \times 1$ vector of ones. We assume that the $n \times 1$ response vector $y = (y_i)$ is *w*-centered, i.e., $w'y = 0$. Individuals in Ω are described by a set *X* of variables, henceforth observed predictors, possibly including both quantitative and qualitative measurements or, possibly, other nonstandard quantities, such as character strings or functions. A distance (metric or semi-metric) $\delta(\cdot,\cdot)$ is defined in Ω, a function of the *X* variables. We denote by Δ the $n \times n$ matrix, whose entries are the squared distances $\delta^2(\Omega_i, \Omega_j)$.

The $n \times n$ inner-products matrix is defined as:

$$G_w = -\frac{1}{2} J_w \Delta J_w', \tag{2.1}$$

where J_w is the *w*-centering matrix, defined as $J_w = Id - \mathbf{1}w'$. We denote by g_w the $1 \times n$ row vector containing the (necessarily nonnegative) diagonal entries of G_w.

Any $n \times k$ matrix X_w such that $G_w = X_w X_w'$ is called a *Euclidean configuration* of Δ. $k \geq r \equiv rank\ G_w$ and $w'X_w = 0$. The DBR of response *y* with weights *w* and predictor matrix Δ, an $n \times n$ square distances matrix, is defined as the Weighted Least Squares (WLS) regression of *y* on a *w*-centered Euclidean configuration of Δ, X_w, with weights *w*.

Assume a new case Ω_{n+1} is available, and we are given the $1 \times n$ vector δ_{n+1} of squared distances from Ω_{n+1} to the *n* previously known individuals. Ω_{n+1} can be represented as a *k*-vector x_{n+1} in the row space of X_w. Then, the predicted *Y* for Ω_{n+1} is $x_{n+1}\hat{\beta}$, where $\hat{\beta}$ is the vector of estimated regression coefficients.

DBR does not depend on a specific X_w, since the final quantities are obtained directly from the distances. Usually such a configuration needs not be made explicit, and neither do $\hat{\beta}$ or x_{n+1}. In DBR, the hat matrix is

$$H_w = G_w \left(D_w^{1/2} F_w^{+} D_w^{1/2} \right) \tag{2.2}$$

where $D_w = diag(w)$ is the diagonal matrix whose diagonal entries are the weights *w*, $F_w = D_w^{1/2} G_w D_w^{1/2}$ and F_w^+ is the *Moore-Penrose* pseudo-inverse of F_w. Then, $\hat{y} = H_w y$.

Finally, the predicted *Y* for a new case Ω_{n+1}, given its δ_{n+1} vector is:

$$\hat{y}_{n+1} = \frac{1}{2}\left(g_w - \delta_{n+1}\right)\left(D_w^{1/2} F_w^+ D_w^{1/2}\right) y \, . \tag{2.3}$$

Thus, the estimations obtained with (2.2) and (2.3) are intrinsic quantities, meaning that they can be expressed directly as a function of the distances or, equivalently, the inner products (2.1).

DBR reproduces the results of WLS: if we start from a $n \times r$ *w*-centered matrix X_w of *r* continuous predictors corresponding to *n* individuals and we define Δ as the matrix of squared Euclidean distance between rows of X_w, then X_w is trivially a Euclidean configuration for Δ, hence the DBR hat matrix, response and predictions coincide with the corresponding WLS quantities.

2.2. The generalized linear model

In this section we review the basic concepts and notations of GLM, for the sake of an easy reference. As is well-known [see eg. McCullagh and Nelder 1989], in a GLM we have a linear predictor $\eta = X\beta$, which is related to the response variable *Y* by means of a link function $g(\bullet)$, $\eta = g(\mu)$, then

$$\mu = g^{-1}(\eta) \, . \tag{2.4}$$

In a GLM it is assumed that each component of the response has a distribution in the exponential family, taking the form:

$$f_Y(y;\theta,\phi) = \exp\left\{\left(y\theta - b(\theta)\right)/a(\phi) + c(y,\phi)\right\} \tag{2.5}$$

for some specific functions $a(\bullet)$, $b(\bullet)$, $c(\bullet)$. If ϕ is known, this is an exponential-family model with canonical parameter θ .

The log-likelihood function for a GLM is $l(\theta;y)=\left(y\theta-b(\theta)\right)/a(\phi)+c(y,\phi)$, and the mean and the variance of *Y*, (2.6), can be derived easily from the relations $E\left(\frac{\partial l}{\partial\theta}\right)=0$ and $E\left(\frac{\partial^2 l}{\partial\theta^2}\right)+E\left(\frac{\partial l}{\partial\theta}\right)^2=0$. From (2.5) we have that $\frac{\partial l}{\partial\theta}=\left\{y-b'(\theta)\right\}/a(\phi)$ and $\frac{\partial^2 l}{\partial\theta^2}=-b''(\theta)/a(\phi)$ and then,

$$E(Y)=\mu=b'(\theta) \quad \text{and} \quad \operatorname{var}(Y)=b''(\theta)a(\phi) \tag{2.6}$$

The variance of *Y* is the product of two functions; one, $b''(\theta)$, depends on the canonical parameter only (and hence on the mean) and will be called the *variance function*, while the other is independent of θ and depends only on ϕ. The variance function as a function of μ will be written $V(\mu)=b''(\theta)$. The function $a(\phi)$ is commonly of the form $a(\phi)=\phi/w$, where ϕ is called the dispersion parameter and is constant over observations. Respect w it is a known *prior weight* that varies from observations to observation. If we have *n* independent readings $w=n$. Finally, we can write

$$\operatorname{var}(Y)=V(\mu)\frac{\phi}{w}. \tag{2.7}$$

<u>An algorithm for fitting GLM</u>

In a GLM the maximum-likelihood estimates of the parameters β in the linear predictor η can be obtained by iterative weighted least squares (IWLS) [see McCullagh and Nelder 1989 for a more detailed description and justification of the algorithm]. In the IWLS the dependent variable of the regression is not *y* but *z*, a linearized form of the link function applied to *y*, and the weights are functions of the fitted values $\hat{\mu}$. The process is iterative because both the adjusted dependent variable *z* and the weight *W* depend on the fitted values, for which only current estimates are available. The procedure underlying the iteration is as follows. Let $\hat{\eta}_0$ be the current estimate of the linear predictor, with corresponding fitted value $\hat{\mu}_0$ derived from the link function $\eta=g(\mu)$. Form the adjusted dependent variate with typical value

$$z_0=\hat{\eta}_0+(y-\hat{\mu}_0)\left(\frac{d\eta}{d\mu}\right)_0 \tag{2.8}$$

where the derivate of the link is evaluated at $\hat{\mu}_0$. The quadratic weight is defned by:

$$W_0^{-1} = \left(\frac{d\eta}{d\mu}\right)_0^2 \frac{V_0}{w} \tag{2.9}$$

where V_0 is the variance function evaluated at $\hat{\mu}_0$. Now regress z_0 on the covariates with weight W_0 to give new estimates of $\hat{\beta}_1$ of the parameters; from these form a new estimate $\hat{\eta}_1$, of the linear predictor. Repeat until changes are sufficiently small.

Note that z is just a linearized form of the link function applied to the data, for, to first order, $g(y) \simeq g(\mu) + (y-\mu) g'(\mu)$. The variance of z is just W^{-1} (ignoring the dispersion parameter), assuming that η and μ are fixed and known.

2.3. Construction of the distance-based generalized linear model

As explained in Boj *et al.* (2011), for both DB-GLM and DBR we have the same elements: a set of individuals with associated standardized to unit sum weights w, for which we have observed the w-centered response variable and a set of predictors X. From predictors we calculate the $n \times n$ distances matrix Δ. But, in this new model, the DB-GLM, we have two extensions respect to the DBR:

1) We assume the responses distribution is in an exponential dispersion family (2.5), as in any GLM,
2) and we consider that the relation between the linear predictor $\eta = X_w \beta$ (obtained from the latent Euclidean configuration X_w) and the response y is given by a link function $g(\bullet)$ as in (2.4).

In Boj *et al.* (2011) it is shown that the proposed IWLS estimation process for DB-GLM does not depend on a specific X_w, since the final quantities are obtained directly from distances. In the first IWLS step one needs an initial $\hat{\mu}_0$. An easy choice is the DBR estimate (i.e. a DB-GLM with a Gaussian distribution for response and Identity link). Then one calculates $\hat{\eta}_0$ and $\left(\frac{d\eta}{d\mu}\right)_0$ as in a GLM. These two elements only depend on the link function. Finally, one calculates V_0, the function (2.7) evaluated at $\hat{\mu}_0$, which only depends on the fitted vales $\hat{\mu}$ in each step.

3. Example

As an academic illustration we fit DB-GLM to the well-known data set on Swedish third-party motor insurance in 1977 described in Hallin and Ingenbleek (1983). Data for factor Zone =1 can be found in Andrews and Herzberg 1985 pages 413-421. These data correspond to the cities of Stockholm, Göteburg and Malmo, and were obtained from a committee study of risk premiums in motor insurance. The file can be downloaded electronically from:

http://www.statsci.org/data/general/motorins.html.

All the computations of this example have been made with the R language. Various authors of this presentation are now elaborating (joint with Adrià Caballé and Pedro Delicado of the Universitat Politècnica de Catalunya) a Distance-based library which includes the DB-GLM. It is foreseen to submit it to the Cran R Project for Statistical Computing during this year 2011. In that way it will become free software for actuaries and other applicants. Respect to the academic example in this section we have used the related file which is incorporated in package “faraway” of R named “motorins”.

The total number of observations (for Zone=1) is n = 295 corresponding to different non-empty risk groups. For each group, Y is the number of claims suffered by the automobile insured in the exposure w, which is the number of insured in policy-years. The factors that are thought to be important in modeling the occurrence of claims are three: Distance, Bonus and Make. The number of levels of each factor are 5, 7 and 9 respectively. Distance and Bonus are continuous numerical predictors and we have coded numerically versions of them as follows:

We have represented each state of Distance by a class mark. Central classes are represented by the interval average, whereas class marks for the extreme classes are reasonably representative values. The codes are:

"<1000 Km per year "	------------->	750 Kilometers Travelled per year
"1000-15000 Km per year "	------------->	8000 Kilometers Travelled per year
"15000-20000 Km per year "	------------->	17500 Kilometers Travelled per year
"20000-25000 Km per year "	------------->	22500 Kilometers Travelled per year
">25000 Km per year "	------------->	40000 Kilometers Travelled per year

Bonus is represented by the (arbitrary) numerical code, 1 to 7. Insured starts in the class 1 and is moved up one class (to a maximum of 7) each year there is no claim.

Make will be considered as a nominal categorical variable in Gower's formula (3.1). It is coded numerically (as 1 to 9) just as a programming convenience. It represents 9 specified car makes.

The first step in the treatment of these data by DB-GLM is the choice of a suitable metric. In principle it is possible to tailor a metric to reflect specific information on predictors and on how their proximity relates to the particular prediction under study. Here it is sufficient to utilize an omnibus metric function which satisfies the Euclidean condition. One very popular such metric for mixtures of numerical continuous, categorical and binary predictor variables is the one based on Gower's general similarity coefficient (see Gower 1971):

$$s_{ij} = \frac{\sum_{h=1}^{p_1}\left(1-\left|x_{ih}-x_{jh}\right|/G_h\right)+a+\alpha_{ij}}{p_1+\left(p_2-d\right)+p_3} \qquad (3.1)$$

where p_1 is the number of continuous variables, a and d are the number of positive and negative matches, respectively, for the p_2 binary variables, and α_{ij} is the number of matches for the p_3 multi-state categorical variables. G_h is the range of the *h*-th continuous variable. The squared distance is computed as: $\delta_{ij}^2 = 1 - s_{ij}$. Gower (1971) proves that (3.1) satisfies the Euclidean condition. In our example, $p_1 = 2$, $p_2 = 0$ and $p_3 = 1$.

For GLM, we use 11 parameters: 2 for Distance and Bonus with the mark classes defined above and 9 binary variables for each nominal class of Make, taking into account an intercept term.

In both, GLM and DB-GLM, we have supposed the Binomial and the Poisson distributions for *claim frequency*, combined with its associated canonical links. The weights of regressions are the exposures *w*.

We fit three DB-GLM in the cases when:

- ✓ rel.gvar = 1, i.e., we take into account for the model all the dimensions of the latent Euclidean configurations
- ✓ method = ”GCV”, i.e., we take for the model the effective rank which minimizes a generalized cross-validation (leave-one-out) statistic
- ✓ rel.gvar = 0.90, i.e., we take into account for the model the 90% of explained geometric variability. In both cases (Poisson and Binomial) coincide with the choice of taking into account an effective rank of 10, which is the number of parameters in the fitted GLM (without counting the intercept term).

We obtain in both cases (the Poisson and the Binomial ones), lower residual deviances with the distance-based treatment of the GLM than those obtained with the classical GLM (see Table 1 and 2). More details can be found in Annex 1.

Poisson / Logarithmic	**Residual Deviance**	**Eff.rank**
DB-GLM (rel.gvar = 1)	453.9998	18
DB-GLM (method = ”GCV”)	485.6362	13
DB-GLM (rel.gvar = 0.90)	539.4509	10
GLM	779.4	10

Table 1. Results of the fitting for the Poisson model with the Logarithmic link.

Binomial / Logit	**Residual Deviance**	**Eff.rank**
DB-GLM (rel.gvar = 1)	498.3561	18
DB-GLM (method = ”GCV”)	538.1463	13
DB-GLM (rel.gvar = 0.90)	595.8747	10
GLM	889.1	10

Table 2. Results of the fitting for the Binomial model with the link Logit.

Acknowledgments

Work supported in part by the Spanish Ministerio de Educación y Ciencia and the Agència de Gestió d’Ajuts Universitaris i de Recerca de la Generalitat de Catalunya, grants: MTM2010-17323, ECO2010-22065-C03-03, 2009SGR970 and 2009SGR932.

References

Anderson, D., Feldblum, S., Modlin, C., Schirmacher, D., Schirmacher, E. and N. Thandi (2002). A Practitioner's Guide to Generalized Linear Models (Third edition). *A CAS Study Note*.

Andrews, D. F., and A. M. Herzberg (1985). *Data. A collection of problems from many fields for the student and reseach worker*. Springer, New York.

Boj, E., Claramunt, M. M. and J. Fortiana (2000). Una alternativa en la selección de los factores de riesgo a utilizar en el cálculo de primas. *Anales del Instituto de Actuarios Españoles*, Tercera Época 6, pp. 11–35.

Boj, E., Claramunt, M. M. and J. Fortiana (2001). Herramientas estadísticas para el estudio de perfiles de riesgo. *Anales del Instituto de Actuarios Españoles*, Tercera Época 7, pp. 59–89.

Boj, E., Claramunt, M. M. and J. Fortiana (2004). *Análisis multivariante aplicado a la selección de factores de riesgo en la tarificación.* Cuadernos de la Fundación MAPFRE, 88. Fundación MAPFRE Estudios, Madrid.

Boj, E., M. M. Claramunt, and J. Fortiana (2007). Selection of predictors in distance-based regression. *Communications in Statistics A. Theory and Methods* 36, 87–98.

Boj, E., P. Delicado, and J. Fortiana (2010). Local linear functional regression based on weighted distance-based regression. *Computational Statistics and Data Analysis* 54, 429–437.

Boj, E., P. Delicado, Fortiana, J. and A. Esteve (2011). Local Distance-Based Generalized Linear Models. *Working Paper*.

Brockman, M. J., and Wright, T. S. (1992). Statistical model rating: making effective use of your data. *Journal of the Institute of Actuaries* 119, 457.

Cuadras, C. M., C. Arenas, and J. Fortiana (1996). Some computational aspects of a distance-based model for prediction. *Communications in Statistics B. Simulation and Computation* 25, 593–609.

Cuadras, C. and C. Arenas (1990). A distance-based regression model for prediction with mixed data. *Communications in Statistics A. Theory and Methods* 19, 2261–2279.

Cuadras, C. M. (1989). Distance analysis in discrimination and classification using both continuous and categorical variables. In: Y. Dodge (Ed.), *Statistical Data Analysis and* Inference, Amsterdam, The Netherlands, pp. 459–473. North-Holland Publishing Co.

de Jong P. and G. Z. Heller (2008). *Generalized Linear Models for Insurance Data*. New York: Cambridge University Press.

Gower, J. C. (1971). A general coefficient of similarity and some of its properties. *Biometrics* 27, 857–874.

Haberman, S., and Renshaw, A. E. (1998). Actuarial applications of generalized linear models. In: *Statistics in Finance*, D. J. Hand and S. D. Jacka (eds), Arnold, London.

Hallin, M. and J. F. Ingenbleek. (1983). The Swedish automobile portfolio in 1977. A statistical study. *Scandinavian Actuarial Journal* 49-64.

Kaas, R., Goovaerts, M., Dhaene, J. and M. Denuit (2008). *Modern Actuarial Risk Theory -- Using R*. Second edition. Heidelberg: Springer-Verlag.

McCullagh, P. and J. A. Nelder (1989). *Generalized Linear Models* (2nd ed). London: Chapman and Hall.

Nelder, J. A., and Verral, R. J. (1997). Credibility theory and generalized linear models. *ASTIN Bulletin* 27, 71-82.

Ohlsson, E. and B. Johansson (2010). *Non-Life Insurance Pricing with Generalized Linear Models* (EAA Series). New York: Springer.

Renshaw, A. E. (1994). Modelling the claims process in the presence of covariates. *ASTIN Bulletin* 24, 265-286.

Venter, G. G. (2007). Generalized Linear Models beyond the Exponential Family with Loss Reserve Applications. *ASTIN Bulletin* 37:2, 345-364.

Zehnwirth, B. (1994). Ratemaking: From Bailey and Simon (1960) to Generalized Linear Regression Models. *Casualty Actuarial Society Forum.* Virginia 1994: Winter, 615-659.

Annex 1. Results of the fitting.

Results for the Poisson (Logarithmic link):

Call: dbglm.default(y = y, x = D2, family = poisson(link = "log"), weights = w, rel.gvar = 1)
family: poisson
Degrees of Freedom: 294 Total (i.e. Null); 276 Residual [eff..rank = 18]
Null Deviance: 6978
Residual Deviance: 453.9998

Call: dbglm.default(y = y, x = D2, family = poisson(link = "log"), weights = w, method = "GCV")
family: poisson
Degrees of Freedom: 294 Total (i.e. Null); 281 Residual [eff..rank = 13]
Null Deviance: 6978
Residual Deviance: 485.6362

Call: dbglm.default(y = y, x = D2, family = poisson(link = "log"), weights = w, rel.gvar = 0.9)
family: poisson
Degrees of Freedom: 294 Total (i.e. Null); 284 Residual [eff..rank = 10]
Null Deviance: 6978
Residual Deviance: 539.4509

Call: glm(formula = frequency ~ KmC + BonC + factor(MakeC), family = poisson(link = "log"), data = Motor3, weights = Motor3$weights)
Coefficients:

```
   (Intercept)          KmC          BonC  factor(MakeC)2  factor(MakeC)3
    -1.640e+00    1.431e-05    -2.165e-01      1.282e-01     -2.140e-01
factor(MakeC)4  factor(MakeC)5  factor(MakeC)6  factor(MakeC)7  factor(MakeC)8
    -5.162e-01      1.270e-01     -3.976e-01     -1.320e-01      1.396e-01
factor(MakeC)9
    -3.079e-02
```

Degrees of Freedom: 294 Total (i.e. Null); 284 Residual [eff..rank = 10]
Null Deviance: 6978
Residual Deviance: 779.4

Results for the Binomial (Link Logit):

Call: dbglm.default(y = y, x = D2, family = binomial(link = "logit"), weights = w, rel.gvar = 1)
family: binomial
Degrees of Freedom: 294 Total (i.e. Null); 276 Residual [eff.rank=18]
Null Deviance: 7666
Residual Deviance: 498.3561

Call: dbglm.default(y = y, x = D2, family = binomial(link = "logit"), weights = w, method = "GCV")

```
family: binomial
Degrees of Freedom: 294 Total (i.e. Null);  281 Residual  [eff.rank=13]
Null Deviance:      7666
Residual Deviance: 538.1463
```

```
Call:  dbglm.default(y = y, x = D2, family = binomial(link = "logit"), weights=w, rel.gvar = 0.9)
family: binomial
Degrees of Freedom: 294 Total (i.e. Null);  284 Residual [eff.rank=10]
Null Deviance:      7666
Residual Deviance: 595.8747
```

```
Call:   glm(formula = frequency ~ KmC + BonC + factor(MakeC), family = binomial(link = "logit"), data = Motor3, weights = Motor3$weights)

Coefficients:
   (Intercept)            KmC           BonC  factor(MakeC)2  factor(MakeC)3
    -1.451e+00       1.586e-05      -2.391e-01       1.415e-01      -2.326e-01
factor(MakeC)4  factor(MakeC)5  factor(MakeC)6  factor(MakeC)7  factor(MakeC)8
    -5.664e-01       1.406e-01      -4.311e-01      -1.431e-01       1.538e-01
factor(MakeC)9
    -3.316e-02

Degrees of Freedom: 294 Total (i.e. Null);  284 Residual [eff.rank=10]
Null Deviance:      7666
Residual Deviance: 889.1
```
