Generalized Linear Models and Extensions

Fourth Edition

James W. Hardin
Department of Epidemiology and Biostatistics
University of South Carolina

Joseph M. Hilbe
Statistics, School of Social and Family Dynamics
Arizona State University

A Stata Press Publication
StataCorp LLC
College Station, Texas

First edition 2001
Second edition 2007
Third edition 2012
Fourth edition 2018

Published by Stata Press, 4905 Lakeway Drive, College Station, Texas 77845
Typeset in LaTeX 2_ε
Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

Print ISBN-10: 1-59718-225-7
Print ISBN-13: 978-1-59718-225-6
ePub ISBN-10: 1-59718-226-5
ePub ISBN-13: 978-1-59718-226-3
Mobi ISBN-10: 1-59718-227-3
Mobi ISBN-13: 978-1-59718-227-0

Library of Congress Control Number: 2018937959

(Pages omitted)

Contents

(Pages omitted)

Preface

We have added several new models to the discussion of extended generalized linear models (GLMs). We have included new software and discussion of extensions to negative binomial regression because of Waring and Famoye. We have also added discussion of heaped data and bias-corrected GLMs because of Firth. There are two new chapters on multivariate outcomes and Bayes GLMs. In addition, we have expanded the clustered data discussion to cover more of the commands available in Stata.

We now include even more examples using synthetically created models to illustrate estimation results, and we illustrate to readers how to construct synthetic Monte Carlo models for binomial and major count models. Code for creating synthetic Poisson, negative binomial, zero-inflated, hurdle, and finite mixture models is provided and further explained. We have enhanced discussion of marginal effects and discrete change for GLMs.

This fourth edition of *Generalized Linear Models and Extensions* is written for the active researcher as well as for the theoretical statistician. Our goal has been to clarify the nature and scope of GLMs and to demonstrate how all the families, links, and variations of GLMs fit together in an understandable whole.

In a step-by-step manner, we detail the foundations and provide working algorithms that readers can use to construct and better understand models that they wish to develop. In a sense, we offer readers a workbook or handbook of how to deal with data using GLM and GLM extensions.

This text is intended as a textbook on GLMs and as a handbook of advice for researchers. We continue to use this book as the required text for a web-based short course through *Statistics.com* (also known as the *Institute for Statistical Education*); see http://www.statistics.com. The students of this six-week course include university professors and active researchers from hospitals, government agencies, research institutes, educational concerns, and other institutions across the world. This latest edition reflects the experiences we have had in communicating to our readers and students the relevant materials over the past decade.

Many people have contributed to the ideas presented in the new edition of this book. John Nelder has been the foremost influence. Other important and influential people include Peter Bruce, David Collett, David Hosmer, Stanley Lemeshow, James Lindsey, J. Scott Long, Roger Newson, Scott Zeger, Kung-Yee Liang, Raymond J. Carroll, H. Joseph Newton, Henrik Schmiediche, Norman Breslow, Berwin Turlach, Gordon Johnston, Thomas Lumley, Bill Sribney, Vince Wiggins, Mario Cleves, William

Greene, Andrew Robinson, Heather Presnal, and others. Specifically, for this edition, we thank Tammy Cummings, Chelsea Deroche, Xinling Xu, Roy Bower, Julie Royer, James Hussey, Alex McLain, Rebecca Wardrop, Gelareh Rahimi, Michael G. Smith, Marco Geraci, Bo Cai, and Feifei Xiao.

As always, we thank William Gould, president of StataCorp, for his encouragement in this project. His statistical computing expertise and his contributions to statistical modeling have had a deep impact on this book.

We are grateful to StataCorp's editorial staff for their equanimity in reading and editing our manuscript, especially to Patricia Branton and Lisa Gilmore for their insightful and patient contributions in this area. Finally, we thank Kristin MacDonald and Isabel Canette-Fernandez, Stata statisticians at StataCorp, for their expert assistance on various programming issues, and Nikolay Balov, Senior Statistician and Software Developer at StataCorp, for his helpful assistance with chapter 20 on Bayesian GLMs. We would also like to thank Rose Medeiros, Senior Statistician at StataCorp, for her assistance in the final passes of this edition.

Stata Press allowed us to dictate some of the style of this text. In writing this material in other forms for short courses, we have always included equation numbers for all equations rather than only for those equations mentioned in text. Although this is not the standard editorial style for textbooks, we enjoy the benefits of students being able to communicate questions and comments more easily (and efficiently). We hope that readers find this practice as beneficial as our short-course participants have found it.

Errata, datasets, and supporting Stata programs (do-files and ado-files) may be found at the publisher's site http://www.stata-press.com/books/generalized-linear-models-and-extensions/. We also maintain these materials on the author sites at http://www.thirdwaystat.com/jameshardin/ and at https://works.bepress.com/joseph_hilbe/. We are very pleased to be able to produce this newest edition. Working on this text from the first edition in 2001 over the past 17 years has been a tremendously satisfying experience.

James W. Hardin
Joseph M. Hilbe

March 2018

(Pages omitted)

2 GLMs

Contents

Nelder and Wedderburn (1972) introduced the theory of GLMs. The authors derived an underlying unity for an entire class of regression models. This class consisted of models whose single response variable, the variable that the model is to explain, is hypothesized to have the variance that is reflected by a member of the single-parameter exponential family of probability distributions. This family of distributions includes the Gaussian or normal, binomial, Poisson, gamma, inverse Gaussian, geometric, and negative binomial.

To establish a basis, we begin discussion of GLMs by initially recalling important results on linear models, specifically those results for linear regression. The standard linear regression model relies on several assumptions, among which are the following:

1. Each observation of the response variable is characterized by the normal or Gaussian distribution; $y_i \sim N(\mu_i, \sigma_i^2)$.
2. The distributions for all observations have a common variance; $\sigma_i^2 = \sigma^2$ for all i.
3. There is a direct or "identical" relationship between the linear predictor (linear combination of covariate values and associated parameters) and the expected values of the model; $E(\mathbf{x}_i\boldsymbol{\beta}) = \mu_i$.

The purpose of GLMs, and the linear models that they generalize, is to specify the relationship between the observed response variable and some number of covariates. The outcome variable is viewed as a realization from a random variable.

Nelder and Wedderburn showed that general models could be developed by relaxing the assumptions of the linear model. By restructuring the relationship between the linear predictor and the fit, we can "linearize" relationships that initially seem to be nonlinear. Nelder and Wedderburn accordingly dubbed these models "generalized linear models".

Most models that were placed under the original GLM framework were well established and popular—some more than others. However, these models had historically been fit using maximum likelihood (ML) algorithms specific to each model. ML algorithms, as we will call them, can be hard to implement. Starting or initial estimates for parameters must be provided, and considerable work is required to derive model-specific quantities to ultimately obtain parameter estimates and their standard errors. In the next chapter, we show much effort is involved.

Ordinary least squares (OLS) extends ML linear regression such that the properties of OLS estimates depend only on the assumptions of constant variance and independence. ML linear regression carries the more restrictive distributional assumption of normality. Similarly, although we may derive likelihoods from specific distributions in the exponential family, the second-order properties of our estimates are shown to depend only on the assumed mean–variance relationship and on the independence of the observations rather than on a more restrictive assumption that observations follow a particular distribution.

The classical linear model assumes that the observations that our dependent variable y represents are independent normal variates with constant variance σ^2. Also covariates are related to the expected value of the independent variable such that

$$E(y) = \mu \tag{2.1}$$

$$\mu = \mathbf{X}\boldsymbol{\beta} \tag{2.2}$$

This last equation shows the "identical" or identity relationship between the linear predictor $\mathbf{X}\boldsymbol{\beta}$ and the mean μ.

Whereas the linear model conceptualizes the outcome y as the sum of its mean μ and a random variable ϵ, Nelder and Wedderburn linearized each GLM family member by means of a link function. They then altered a previously used algorithm called *iterative weighted least squares*, which was used in fitting weighted least-squares regression models. Aside from introducing the link function relating the linear predictor to the fitted values, they also introduced the variance function as an element in the weighting of the regression. The iterations of the algorithm updates parameter estimates to produce appropriate linear predictors, fitted values, and standard errors. We will clarify exactly how all this falls together in the section on the iteratively reweighted least-squares (IRLS) algorithm.

The estimation algorithm allowed researchers to easily fit many models previously considered to be nonlinear by restructuring them into GLMs. Later, it was discovered that an even more general class of linear models results from more relaxations of assumptions for GLMs.

However, even though the historical roots of GLMs are based on IRLS methodology, many generalizations to the linear model still require Newton–Raphson techniques common to ML methods. We take the position here that GLMs should not be constrained to those models first discussed by Nelder and Wedderburn but rather that they encompass all such linear generalizations to the standard model.

Many other books and journal articles followed the cornerstone article by Nelder and Wedderburn (1972) as well as the text by McCullagh and Nelder (1989) (the original text was published in 1983). Lindsey (1997) illustrates the application of GLMs to biostatistics, most notably focusing on survival models. Hilbe (1994) gives an overview of the GLM and its support from various software packages. Software was developed early on. In fact, Nelder was instrumental in developing the first statistical program based entirely on GLM principles—generalized linear interactive modeling (GLIM). Published by the Numerical Algorithms Group (NAG), the software package has been widely used since the mid-1970s. Other vendors began offering GLM capabilities in the 1980s, including GENSTAT and S-Plus. Stata and SAS included it in their software offerings in 1993 and 1994, respectively.

This text covers much of the same foundation material as other books. What distinguishes our presentation of the material is twofold. First, we focus on the estimation of various models via the estimation technique. Second, we present our derivation of the methods of estimation in a more accessible manner than which is presented in other sources. In fact, where possible, we present complete algebraic derivations that include nearly every step in the illustrations. Pedagogically, we have found that this manner of exposition imparts a more solid understanding and "feel" of the area than do other approaches. The idea is this: if you can write your own GLM, then you are probably more able to know how it works, when and why it does not work, and how it is to be evaluated. Of course, we also discuss methods of fit assessment and testing. To model data without subjecting them to evaluation is like taking a test without checking the answers. Hence, we will spend considerable time dealing with model evaluation as well as algorithm construction.

2.1 Components

Cited in various places such as Hilbe (1993b) and Francis, Green, and Payne (1993), GLMs are characterized by an expanded itemized list given by the following:

1. A random component for the response, $\mathbf{y}$, which has the characteristic variance of a distribution that belongs to the exponential family.
2. A linear systematic component relating the linear predictor, $\eta = \mathbf{X}\boldsymbol{\beta}$, to the product of the design matrix $\mathbf{X}$ and the parameters $\boldsymbol{\beta}$.
3. A known monotonic, one-to-one, differentiable link function $g(\cdot)$ relating the linear predictor to the fitted values. Because the function is one-to-one, there is an inverse function relating the mean expected response, $E(y) = \mu$, to the linear predictor such that $\mu = g^{-1}(\eta) = E(y)$.
4. The variance may change with the covariates only as a function of the mean.
5. There is one IRLS algorithm that suffices to fit all members of the class.

Item 5 is of special interest. The traditional formulation of the theory certainly supposed that there was one algorithm that could fit all GLMs. We will see later how this was implemented. However, there have been extensions to this traditional viewpoint. Adjustments to the weight function have been added to match the usual Newton–Raphson algorithms more closely and so that more appropriate standard errors may be calculated for noncanonical link models. Such features as scaling and robust variance estimators have also been added to the basic algorithm. More importantly, sometimes a traditional GLM must be restructured and fit using a model-specific Newton–Raphson algorithm. Of course, one may simply define a GLM as a model requiring only the standard approach but doing so would severely limit the range of possible models. We prefer to think of a GLM as a model that is ultimately based on the probability function belonging to the exponential family of distributions, but with the proviso that this criterion may be relaxed to include quasilikelihood models as well as certain types of multinomial, truncated, censored, and inflated models. Most of the latter type require a Newton–Raphson approach rather than the traditional IRLS algorithm.

Early GLM software development constrained GLMs to those models that could be fit using the originally described estimation algorithm. As we will illustrate, the traditional algorithm is relatively simple to implement and requires little computing power. In the days when RAM was scarce and expensive, this was an optimal production strategy for software development. Because this is no longer the case, a wider range of GLMs can more easily be fit using a variety of algorithms. We will discuss these implementation details at length.

In the classical linear model, the observations of the dependent variable $\mathbf{y}$ are independent normal variates with constant variance σ^2. We assume that the mean value of $\mathbf{y}$ may depend on other quantities (predictors) denoted by the column vectors $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_{p-1}$. In the simplest situation, we assume that this dependency is linear and write

$$E(\mathbf{y}) = \beta_0 + \beta_1 \mathbf{X}_1 + \cdots + \beta_{p-1} \mathbf{X}_{p-1} \tag{2.3}$$

and attempt to estimate the vector $\boldsymbol{\beta}$.

GLMs specify a relationship between the mean of the random variable $\mathbf{y}$ and a function of the linear combination of the predictors. This generalization admits a model specification allowing for continuous or discrete outcomes and allows a description of the variance as a function of the mean.

2.2 Assumptions

The link function relates the mean $\mu = E(y)$ to the linear predictor $\mathbf{X}\boldsymbol{\beta}$, and the variance function relates the variance as a function of the mean $V(y) = a(\phi)v(\mu)$, where $a(\phi)$ is the scale factor. For the Poisson, binomial, and negative binomial variance models, $a(\phi) = 1$.

Breslow (1996) points out that the critical assumptions in the GLM framework may be stated as follows:

1. Statistical independence of the n observations.
2. The variance function $v(\mu)$ is correctly specified.
3. $a(\phi)$ is correctly specified (1 for Poisson, binomial, and negative binomial).
4. The link function is correctly specified.
5. Explanatory variables are of the correct form.
6. There is no undue influence of the individual observations on the fit.

As a simple illustration, in table 2.1 we demonstrate the effect of the assumed variance function on the model and fitted values of a simple GLM.

Table 2.1. Predicted values for various choices of variance function

Observed (y)	1.00	2.00	9.00	
Predicted [Normal: $v(\mu) = \phi$]	0.00	4.00	8.00	$\widehat{y} = -4.00 + 4.00x$
Predicted [Poisson: $v(\mu) = \mu$]	0.80	4.00	7.20	$\widehat{y} = -2.40 + 3.20x$
Predicted [Gamma: $v(\mu) = \phi\mu^2$]	0.94	3.69	6.43	$\widehat{y} = -1.80 + 2.74x$
Predicted [Inverse Gaussian: $v(\mu) = \phi\mu^3$]	0.98	3.33	5.69	$\widehat{y} = -1.37 + 2.35x$

Note: The models are all fit using the identity link, and the data consist of three observations $(y, x) = \{(1, 1), (2, 2), (9, 3)\}$. The fitted models are included in the last column.

2.3 Exponential family

GLMs are traditionally formulated within the framework of the exponential family of distributions. In the associated representation, we can derive a general model that may be fit using the scoring process (IRLS) detailed in section 3.3. Many people confuse the estimation method with the class of GLMs. This is a mistake because there are many estimation methods. Some software implementations allow specification of more diverse models than others. We will point this out throughout the text.

The exponential family is usually (there are other algebraically equivalent forms in the literature) written as

$$f_y(y; \theta, \phi) = \exp\left\{\frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi)\right\} \tag{2.4}$$

where θ is the canonical (natural) parameter of location and ϕ is the parameter of scale. The location parameter (also known as the canonical link function) relates to the means, and the scalar parameter relates to the variances for members of the exponential family of distributions including Gaussian, gamma, inverse Gaussian, and others. Using the notation of the exponential family provides a means to specify models for continuous, discrete, proportional, count, and binary outcomes.

In the exponential family presentation, we construe each of the y_i observations as being defined in terms of the parameters θ. Because the observations are independent, the joint density of the sample of observations y_i, given parameters θ and ϕ, is defined by the product of the density over the individual observations (review section 2.2). Interested readers can review Barndorff-Nielsen (1976) for the theoretical justification that allows this factorization:

$$f_{y_1,y_2,\ldots,y_n}(y_1,y_2,\ldots,y_n;\theta,\phi) = \prod_{i=1}^{n} \exp\left\{\frac{y_i\theta_i - b(\theta_i)}{a(\phi)} + c(y_i,\phi)\right\} \tag{2.5}$$

Conveniently, the joint probability density function may be expressed as a function of θ and ϕ given the observations y_i. This function is called the likelihood, L, and is written as

$$L(\theta,\phi;y_1,y_2,\ldots,y_n) = \prod_{i=1}^{n} \exp\left\{\frac{y_i\theta_i - b(\theta_i)}{a(\phi)} + c(y_i,\phi)\right\} \tag{2.6}$$

We wish to obtain estimates of (θ,ϕ) that maximize the likelihood function. Given the product in the likelihood, it is more convenient to work with the log likelihood,

$$\mathcal{L}(\theta,\phi;y_1,y_2,\ldots,y_n) = \sum_{i=1}^{n} \left\{\frac{y_i\theta_i - b(\theta_i)}{a(\phi)} + c(y_i,\phi)\right\} \tag{2.7}$$

because the values that maximize the likelihood are the same values that maximize the log likelihood.

Throughout the text, we will derive each distributional family member from the exponential family notation so that the components are clearly illustrated. The log likelihood for the exponential family is in a relatively basic form, admitting simple calculations of first and second derivatives for ML estimation. The IRLS algorithm takes advantage of this form of the log likelihood.

First, we generalize the log likelihood to include an offset to the linear predictor. This generalization will allow us to investigate simple equality constraints on the parameters.

The idea of an offset is simple. To fit models with covariates, we specify that θ is a function of specified covariates, $\mathbf{X}$, and their associated coefficients, $\boldsymbol{\beta}$. Within the linear combination of the covariates and their coefficients $\mathbf{X}\boldsymbol{\beta}$, we may further wish to constrain a particular subset of the coefficients β_i to particular values. For example, we may know or wish to test that $\beta_3 = 2$ in a model with a constant, X_0, and three covariates X_1, X_2, and X_3. If we wish to enforce the $\beta_3 = 2$ restriction on the estimation, then we will want the optimization process to calculate the linear predictor as

$$\eta = \widehat{\beta}_0 + \widehat{\beta}_1 X_1 + \widehat{\beta}_2 X_2 + 2X_3 \tag{2.8}$$

at each step. We know (or wish to enforce) that the linear predictor is composed of a linear combination of the unrestricted parameters plus two times the X_3 covariate. If we consider that the linear predictor is generally written as

$$\eta = \mathbf{X}\boldsymbol{\beta} + \text{offset} \tag{2.9}$$

then we can appreciate the implementation of a program that allows an offset. We could generate a new variable equal to two times the variable containing the X_3 observations and specify that generated variable as the offset. By considering this issue from the outset, we can include an offset in our derivations, which will allow us to write programs that include this functionality.

The offset is a given (nonstochastic) component in the estimation problem. By including the offset, we gain the ability to fit (equality) restricted models without adding unnecessary complexity to the model; the offset plays no role in derivative calculations. If we do not include an offset in our derivations and subsequent programs, we can still fit restricted models, but the justification is less clear; see the arguments of Nyquist (1991) for obtaining restricted estimates in a GLM.

2.4 Example: Using an offset in a GLM

In subsequent chapters (especially chapter 3), we illustrate the two main components of the specification of a GLM. The first component of a GLM specification is a function of the linear predictor, which substitutes for the location (mean) parameter of the exponential family. This function is called the link function because it links the expected value of the outcome to the linear predictor comprising the regression coefficients; we specify this function with the `link()` option. The second component of a GLM specification is the variance as a scaled function of the mean. In Stata, this function is specified using the name of a particular member distribution of the exponential family; we specify this function with the `family()` option. The example below highlights a log-link Poisson GLM.

For this example, it is important to note the treatment of the offset in the linear predictor. The particular choices for the link and variance functions are not relevant to the utility of the offset.

Below, we illustrate the use of an offset with Stata's `glm` command. From an analysis presented in chapter 12, consider the output of the following model:

```
. use http://www.stata-press.com/data/hh4/medpar

. glm los hmo white type2 type3, family(poisson) link(log) nolog

Generalized linear models                         No. of obs      =      1,495
Optimization     : ML                             Residual df     =      1,490
                                                  Scale parameter =          1
Deviance         =  8142.666001                   (1/df) Deviance =   5.464877
Pearson          =  9327.983215                   (1/df) Pearson  =   6.260391

Variance function: V(u) = u                       [Poisson]
Link function    : g(u) = ln(u)                   [Log]

                                                  AIC             =   9.276131
Log likelihood   = -6928.907786                   BIC             =  -2749.057

------------------------------------------------------------------------------
             |                 OIM
         los |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         hmo |  -.0715493    .023944    -2.99   0.003    -.1184786    -.02462
       white |   -.153871   .0274128    -5.61   0.000    -.2075991   -.100143
       type2 |   .2216518   .0210519    10.53   0.000     .1803908   .2629127
       type3 |   .7094767    .026136    27.15   0.000     .6582512   .7607022
       _cons |   2.332933   .0272082    85.74   0.000     2.279606    2.38626
------------------------------------------------------------------------------
```

We would like to test whether the coefficient on `white` is equal to -0.20. We could use Stata's `test` command to obtain a Wald test

```
. test white=-.20

 ( 1)  [los]white = -.2

           chi2(  1) =    2.83
         Prob > chi2 =    0.0924
```

which indicates that -0.15 (coefficient on `white`) is not significantly different at a 5% level from -0.20. However, we want to use a likelihood-ratio test, which is usually a more reliable test of parameter estimate significance. Stata provides a command that stores the likelihood from the unrestricted model (above) and then compares it with a restricted model. Having fit the unrestricted model, our attention now turns to fitting a model satisfying our specific set of constraints. Our constraint is that the coefficient on `white` be restricted to the constant value -0.20.

First, we store the log-likelihood value from the unrestricted model, and then we generate a variable indicative of our constraint. This new variable contains the restrictions that we will then supply to the software for fitting the restricted model. In short, the software will add our restriction any time that it calculates the linear predictor $\mathbf{x}_i\beta$. Because we envision a model for which the coefficient of `white` is equal to -0.20, we need to generate a variable that is equal to -0.20 times the variable `white`.

```
. estimates store Unconstrained
. generate offvar = -.20*white
. glm los hmo type2 type3, family(poisson) link(log) offset(offvar) nolog
Generalized linear models                          No. of obs      =      1,495
Optimization     : ML                              Residual df     =      1,491
                                                   Scale parameter =          1
Deviance         =  8145.531652                    (1/df) Deviance =   5.463133
Pearson          =  9334.640731                    (1/df) Pearson  =   6.260658
Variance function: V(u) = u                        [Poisson]
Link function    : g(u) = ln(u)                    [Log]
                                                   AIC             =    9.27671
Log likelihood   = -6930.340612                    BIC             =  -2753.502
------------------------------------------------------------------------------
             |                 OIM
         los |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         hmo |  -.0696133   .0239174    -2.91   0.004    -.1164906    -.022736
       type2 |    .218131    .020951    10.41   0.000     .1770677    .2591942
       type3 |   .7079687   .0261214    27.10   0.000     .6567717    .7591658
       _cons |   2.374881   .0107841   220.22   0.000     2.353744    2.396017
      offvar |          1  (offset)
------------------------------------------------------------------------------
. lrtest Unconstrained
Likelihood-ratio test                                 LR chi2(1)  =       2.87
(Assumption: . nested in Unconstrained)               Prob > chi2 =     0.0905
```

Because we restricted one coefficient from our full model, the likelihood-ratio statistic is distributed as a chi-squared random variable with one degree of freedom. We fail to reject the hypothesis that the coefficient on `white` is equal to -0.20 at the 5% level.

Restricting coefficients for likelihood-ratio tests is just one use for offsets. Later, we discuss how to use offsets to account for exposure in count-data models.

2.5 Summary

The class of GLMs extends traditional linear models so that a linear predictor is mapped through a link function to model the mean of a response characterized by any member of the exponential family of distributions. Because we are able to develop one algorithm to fit the entire class of models, we can support estimation of such useful statistical models as logit, probit, and Poisson.

The traditional linear model is not appropriate when it is unreasonable to assume that data are normally distributed or if the response variable has a limited outcome set. Furthermore, in many instances in which homoskedasticity is an untenable requirement, the linear model is inappropriate. The GLM allows these extensions to the linear model.

A GLM is constructed by first selecting explanatory variables for the response variable of interest. A probability distribution that is a member of the exponential family is selected, and an appropriate link function is specified, for which the mapped range of values supports the implied variance function of the distribution.