The dbm package
(Version 1.0-0)

Alexios Ghalanos

August 7, 2013

Contents

1 Introduction

Binary response models have a very rich history in the statistical research literature with diverse applications in many fields where the dependent variable takes on a dichotomous value. Early work on these models included the tetrachoric correlation analysis of Pearson (1900) and the standard biometric textbook of Finney (1971), with more recent advances dealing with such issues as full or partial separation in Zorn (2005), small sample bias reduction in Firth (1993) and Heinze and Schemper (2002) and heteroscedasticity (see for instance Keele and Park (2006)). In econometrics, Amemiya (1981) provided an early review of applications, but more recent research has focused on direction of change forecast for stock market returns and recession forecasting which is also the motivation of this author.
The Dynamic Binary Model (`dbm`) package implements an autoregressive binary regression model described in Kauppi and Saikkonen (2008) and further discussed and extended in Nyberg(2010a, 2010b, and 2011). R already includes a large number of packages implementing dynamic binary models, with basic functionality already available in the *glm* function of the `stats` package, and more advanced variations in Zeileis and Windberger (2011)(skew-or generalized-logit), Zeileis et al. (2011) (heteroscedasticy), and Heinze and Ploner (2013) (bias reduction methods), to name but a few. The `dbm` package uniquely implements the autoregressive model of Kauppi and Saikkonen (2008) which is believed to provide a much improved fit to typical problems in the time series domain.
The vignette discusses the model background in Section 2 and its implementation in Section 3. An interesting application, and exposition of the package's functionality is available on my blog (see below). The appendix derives the analytic gradient expressions used in the numeric optimization routines.

The `dbm` package is currently available in the *teatime* package repository on r-forge (`http://cran.r-project.org/web/packages/teatime/index.html`). An example is available on my blog (`http://www.unstarched.net`).

The package is provided AS IS, without any implied warranty as to its accuracy or suitability. A lot of time and effort has gone into the development of this package, and it is offered under the GPL-3 license in the spirit of open knowledge sharing and dissemination. If you do use the model in published work DO remember to cite the package and author (type `citation`("dbm") for the appropriate BibTeX entry) , and if you have used it and found it useful, drop me a note and let me know.

USE THE R-SIG-FINANCE MAILING LIST FOR QUESTIONS.

2 Background

Consider the stochastic process y_t, which is binary valued, and the vector of k explanatory variables, $x_{i,t}$ for $i = 1 \dots k$. Let $\Im_t$ be the information set available at time t, then y_t has a Bernoulli distribution with probability p_t:

$$y_t \left| \Im_{t-1} \sim B\left(p_t\right) \right. \tag{1}$$

The objective is to model p_t through the CDF transformed dynamics[1] of a linear process π_t such that $p_t = \Phi\left(\pi_t\right)$. Formally,

$$E_{t-1}\left(y_t\right) = P_{t-1}\left(y = 1\right) = \Phi\ \left(\pi_t\right) = p_t \tag{2}$$

[1]The CDF transformation is monotonically increasing and guarantees that the resulting transformation will be in the unit interval.

The CDF function, also called the link function, can be one of any number of distributions but has typically been either the Gaussian (probit model), Logistic (logit model) or some skewed variation (scobit model). In the `dbm` package, the Generalized Logistic distribution has been chosen as a skewed choice since it nests the Logistic. In the model of Kauppi and Saikkonen (2008), the dynamics of π_t take on the following form:

$$\pi_t = \omega + \sum_{i=1}^{k} \beta_i x_{i,t-l_i} + \sum_{i=1}^{q} \delta_i y_{t-i} + \sum_{i=1}^{p} \alpha_i \pi_{t-i} \tag{3}$$

where δ_i represents the coefficient on the q autoregressive terms of the binary variable y_t, β_i the coefficient on the i^{th}(of k) explanatory variable x_t with lag ${l_i}^2$, and α_i the coefficient on the p autoregressive terms of the dynamics π_t. The specification without the latter term has already been examined elsewhere, and some results with regards to its asymptotic properties can be found in de Jong and Woutersen (2011). Related literature on general binomial ARMA type models can be found, among others, in Al-Osh and Alzaid (1991) and more recently, with financial/econometric applications, in Rydberg and Shephard (2003) and Startz (2008). Nyberg (2011) introduced a restriction to Equation 3 by setting $\delta_1 = 1 - \alpha_1$, leading to a type of error correction model (ecm) with strong persistence in the autoregressive dynamics parameter α_1 usually observed.

2.1 Maximum Likelihood Estimation

The log-likelihood function (at time t) in the maximization, given the vector of parameters θ, the conditional dynamics $\pi_t(\theta)$ and an appropriate link function Φ, can be represented as:

$$l(\theta) = \sum_{t=1}^{T} \left[y_t \log \Phi(\pi_t(\theta)) + (1 - y_t) \log(1 - \Phi(\pi_t(\theta))) \right] \tag{4}$$

conditional on initial values for the recursion. Appendix 4 contains the details of the gradient for each of the 3 link functions used in the `dbm` package and the recursion initialization method.

2.2 Regularization

Regularization, originally proposed by Tikhonov and Arsenin (1974), enables to avoid overfitting in systems with many parameters or generally ill posed problems. While there are a number of ways to do this, regularization in the `dbm` package is based on the L_2 norm which is rotationally invariant (as compared to the L_1 norm). The likelihood can be represented as:

$$l(\theta) = \sum_{t=1}^{T} \left[y_t \log \Phi(\pi_t(\theta)) + (1 - y_t) \log(1 - \Phi(\pi_t(\theta))) \right] - \frac{C}{2} \sum_{j=1}^{m} \theta_j^2 \tag{5}$$

where it is clear how higher values of the m parameters are penalized, with the cost C determined by the user. In future, cross validation may be implemented internally to enable the determination of this cost function.

[2]The representation used here, which is the one used in the package, is such that the vector $\mathbf{x}_t$ can contain the same explanatory variable but with a different lag thus enabling a great degree of flexibility, albeit at the cost of some redundancy, in the dynamics.

2.3 Forecasting

Unlike other nonlinear models, the binary nature of the model allows explicit multi-period iterated forecasts by enumeration of all the possible binary paths. Following from the Appendix of Kauppi and Saikkonen (2008), and adjusting the notation to use forward times, the h-period ahead forecast can be represented as follows:

$$\begin{aligned} E_t\left(y_{t+h}\right) &= E_t\Phi\left(\alpha^h\pi_t + \sum_{j=1}^{h}\left[\alpha^{j-1}\left(\omega + \delta y_{t+h-j} + \sum_{i=1,(h-l_i)<j}^{p}\beta_i x_{i,t+(h-l_i)+1-j}\right)\right]\right) \\ &= \sum_{y_{t+1}^{t+h-1}\in B_{h-1}} P_t\left(y_{t+1}^{t+h-1}\right)\Phi\left(\alpha^h\pi_t + \sum_{j=1}^{h}\left[\alpha^{j-1}\left(\omega + \delta y_{t+h-j} + \sum_{i=1,L(x_i)\geqslant j}^{l}\beta_i x_{i,t+1-j}\right)\right]\right) \end{aligned} \tag{6}$$

where, $y_{t+1}^{t+h-1} \in B_{h-1}$ indicates the evaluation of all possible binary paths for y up to time $t+h-1$, l_i is the lag of each explanatory variable $x_i, i = 1, \ldots k$, and

$$\begin{aligned} P_t\left(y_{t+1}^{t+h-1}\right) &= \prod_{n=1}^{h-1}\left(p_{t+n}\right)^{y_{t+n}}\left(1-p_{t+n}\right)^{(1-y_{t+n})} \\ p_{t+n} &= \Phi\left(\alpha^n\pi_t + \sum_{j=1}^{n}\alpha^{j-1}\left(\omega + \delta y_{t-1+j} + \sum_{i=1,l_i\geqslant j}^{k}\beta_i x_{\{i,t-l_i+j\}}\right)\right) \end{aligned} \tag{7}$$

Some caution is warranted here since the multi-path forecast, while explicit, starts to grow very fast. Because the implementation in the `dbm` package does not enforce any compact representation (it simply makes use of the *expand.grid* function), memory issues are likely to arise for forecast horizons greater than 15.

3 Implementation

The dbm package estimates the model of Kauppi and Saikkonen (2008) by maximum likelihood using analytic gradient information which is passed to the appropriate solver. The default is to use the Nelder-Mead algorithm from the optim package with the use of a bounding logistic transformation for the parameters which are constrained to ensure either stationarity (autoregressive parameter α) or existence in the real domain (the Generalized Logistic skew parameter $skew[k]$). The result of using such a transformation (given that the optim package does not necessarily include lower and upper bounds for all but one solver), is that the gradient of those parameters are shifted slightly from the unconstrained case and I make no special provision to account for this. Since the transformation is turned off during the calculation of the final hessian and scores, the standard errors and resulting information is not affected, and it is also my experience that estimation is not in the slightest affected by this. If in doubt, there is always the possibility of using any of bound constrained solvers from the nloptr package with analytic gradient or the gosolnp solver from the Rsolnp package with numerical gradient. Finally, both the likelihood and gradient, as well as the main part of the forecast routine are implemented in C and C++ for speed.
The main functionality can be found in the **dbm** function which is used for the ML estimation of a model:

```
> args(dbm)

function(y, x.vars = NULL, x.lags = 1, arp = 1, arq = 0, ecm = FALSE,
        constant = TRUE, link = "gaussian", fixed.pars = NULL,
```

```
solver = "optim", control=list(), parsearch = TRUE, parsim = 5000,
method = "Nelder-Mead", ...)
```

The first argument, y, is a multivariate xts matrix with the first column being the binary dependent variable, and the **names** of the independent explanatory variables (corresponding to the column names in y) are passed via the *x.vars* character vector. The lags for each of the explanatory variables is then passed using the x.lags integer vector. This means that if you want to include multiple lags from the same variable, then the y matrix must include that same variable data as many times as the lags required (with different names). Also note that you must not pass any of the data lagged since that is done by the routine (which is why *x.lags* is used). The *arp* and *arp* options denote the lags for the auto-regression in the dynamics and dependent variable, respectively, but as currently implemented only lag 1 is allowed. The *ecm* option indicates whether to use the error correction restriction of Nyberg (2011), *constant* whether to include an intercept, and *link* the CDF link function with a choice of 'gaussian' (probit), 'logistic' (logit) and 'glogistic' (scobit). The other options are related to the parameter start, fixed values (as a named vector), the choice of solver etc. Standard extraction and inference methods are implemented such as *fitted*, *residuals*,*coef*, *vcov* (with choice of robust), *likelihood*, *deviance* (with null model choice), *plot* and *summary*. Additionally, a score method is also included to extract either the analytic or numeric scores with the option of also recalculating them for a new set of parameters. This is particularly useful when performing a Lagrange Multiplier (LM) test which requires the score of the fixed parameter. An example of this and additional functions are documented in the package help.
Finally, it is up to the user to determine the adequacy of the estimated model. There are a lot of packages on CRAN which implement diagnostic tests, and the `dbm` package returns enough information in the estimated object (including the scores) in order to run most of the available tests. There are a number of important issues to keep in mind when estimating a model, including heteroscedasticity, partial or complete non-separation, and multi-collinearity, all of which are covered in standard textbooks and numerous articles. Particular care should also be taken to ensure that the explanatory variables are stationary.

3.1 Methods

The main methods for working with the returned estimation object after calling the **dbm** function are summarized in Table 1.

Table 1: **dbm** class methods.

Method	Args	Description
logLik	(object, ...)	Returns the log likelihood at the maximized optimal
coef	(object, ...)	Returns the coefficient values (excluding any fixed parameters)
vcov	(object, robust=FALSE, ...)	Returns the variance covariance matrix of the parameters
score	(object, pars=NULL, analytic=TRUE, ...)	Returns the score matrix (or calculates one using a new set of pars)
deviance	(object, null = FALSE, ...)	Extracts the deviance (or optionally the null deviance)
fitted	(object, ...)	Returns the estimated probability
residuals	(object, type, ...)	Returns the residuals (with option for type of residuals)
BIC	(object, ...)	Returns the Bayesian Information Criterion
AIC	(object, ...)	Returns the Akaike Information Criterion
summary	(object, ...)	Returns a summary of the estimated model
print	(x, digits, ...)	Prints the estimated model summary
model.matrix	(object, ...)	Returns the model matrix (including autoregressive data)
hat	(x, ...)	Returns the hat matrix
hatvalues	(model, ...)	Returns the Pregibon leverage
plot	(x, ...)	Plots of the estimated object
predict	(object, newdata = NULL, n.ahead=1, ...)	Prediction (including n.ahead iterated)
mfx	(model, x.mean = TRUE, rev.dum = TRUE, ...)	Returns marginal effects

4 Appendix: Analytic Derivatives

The appendix provides a dry exposition of the analytic expressions for the derivatives used in the numerical optimization routine for the 3 link functions. The subscripts for the parameter lag orders and explanatory variables have been suppressed for ease and compactness of notation, and denote a lag-1 in all variables with one explanatory variable, but the results easily generalize to higher lags and k explanatory variables. The gradient formulae shown are for the basic model and do not include the **ecm** case nor the case of regularization (which are however implemented internally in the package since ver 1.0-1), and in any case require only simple modification.

4.1 Recursion Initialization

The recursion is initialized by setting:

$$\pi_0 = \frac{\omega + \delta\bar{y} + \beta\bar{x}}{1-\alpha} \tag{8}$$

where the bar over the variables denotes their unconditional mean. Therefore, at time t_1, $\pi_1 = \omega + \alpha\pi_0$, and the partial derivatives with respect to π_1 are:

$$\begin{aligned}
\frac{\partial \pi_1}{\partial \omega} &= \frac{\alpha}{1-\alpha} + 1 \\
\frac{\partial \pi_1}{\partial \beta} &= \frac{\alpha\bar{x}}{1-\alpha} \\
\frac{\partial \pi_1}{\partial \delta} &= \frac{\alpha\bar{y}}{1-\alpha} \\
\frac{\partial \pi_1}{\partial \alpha} &= \pi_0 + \frac{\alpha\pi_0}{1-\alpha}
\end{aligned} \tag{9}$$

4.2 Logistic

The logistic link function, which gives rise to the *logit* model, has the following log-likelihood function at time t (maximization):

$$l_t(\theta) = y_t \log\left(\frac{1}{1+e^{-\pi_t}}\right) + (1-y_t)\log\left(1 - \frac{1}{1+e^{-\pi_t}}\right) \tag{10}$$

The partial derivatives with respect to the log-likelihood are defined as follows:

$$\begin{aligned}
\frac{\partial l_t}{\partial \omega} &= \frac{(y_t + y_t \mathrm{e}^{-\pi_t} - 1)}{(1+\mathrm{e}^{-\pi_t})}\left(1 + \frac{\partial \pi_{t-1}(\theta)}{\partial \omega}\right) \\
\frac{\partial l_t}{\partial \beta} &= \frac{(y_t + y_t \mathrm{e}^{-\pi_t} - 1)}{(1+\mathrm{e}^{-\pi_t})}\left(x_{t-1} + \alpha\frac{\partial \pi_{t-1}(\theta)}{\partial x_{t-1}}\right) \\
\frac{\partial l_t}{\partial \delta} &= \frac{(y_t + y_t \mathrm{e}^{-\pi_t} - 1)}{(1+\mathrm{e}^{-\pi_t})}\left(y_{t-1} + \alpha\frac{\partial \pi_{t-1}(\theta)}{\partial \delta_{t-1}}\right) \\
\frac{\partial l_t}{\partial \alpha} &= \frac{(y_t + y_t \mathrm{e}^{-\pi_t} - 1)}{(1+\mathrm{e}^{-\pi_t})}\left(\pi_{t-1} + \alpha\frac{\partial \pi_{t-1}(\theta)}{\partial \alpha}\right)
\end{aligned} \tag{11}$$

The $\partial \pi_{t-1}(\theta)$ is based on the initialization values and their partial derivatives given in Equation 9.

4.3 Gaussian

The Gaussian link function, which gives rise to the *probit* model, has the following log-likelihood function at time t (maximization):

$$l_t(\theta) = y_t \ln\left(\frac{1}{2} + \frac{1}{2}\operatorname{erf}\left(\frac{\pi_t}{\sqrt{2}}\right)\right) + (1 - y_y)\ln\left(\frac{1}{2} - \frac{1}{2}\operatorname{erf}\left(\frac{\pi_t}{\sqrt{2}}\right)\right) \tag{12}$$

where the error function erf is used to approximate the Gaussian distribution. The recursion is initialized as in Equation 8. The partial derivatives with respect to the log-likelihood are defined as follows:

$$\begin{aligned}
\frac{\partial l_t}{\partial \omega} &= \frac{\mathrm{e}^{-0.5\pi_t^2}\sqrt{2}\left(\operatorname{erf}\left(-0.5\sqrt{2}\,\pi_t\right) - 2\,y_t + 1\right)}{\sqrt{\pi}\left(\operatorname{erf}\left(-0.5\sqrt{2}\pi_t\right)^2 - 1\right)}\left(1 + \alpha\frac{\partial \pi_{t-1}(\theta)}{\partial \omega}\right) \\
\frac{\partial l_t}{\partial \beta} &= \frac{\mathrm{e}^{-0.5\pi_t^2}\sqrt{2}\left(\operatorname{erf}\left(-0.5\sqrt{2}\,\pi_t\right) - 2\,y_t + 1\right)}{\sqrt{\pi}\left(\operatorname{erf}\left(-0.5\sqrt{2}\pi_t\right)^2 - 1\right)}\left(x_{t-1} + \alpha\frac{\partial \pi_{t-1}(\theta)}{\partial \beta}\right) \\
\frac{\partial l_t}{\partial \delta} &= \frac{\mathrm{e}^{-0.5\pi_t^2}\sqrt{2}\left(\operatorname{erf}\left(-0.5\sqrt{2}\,\pi_t\right) - 2\,y_t + 1\right)}{\sqrt{\pi}\left(\operatorname{erf}\left(-0.5\sqrt{2}\pi_t\right)^2 - 1\right)}\left(y_{t-1} + \alpha\frac{\partial \pi_{t-1}(\theta)}{\partial \delta}\right) \\
\frac{\partial l_t}{\partial \alpha} &= \frac{\mathrm{e}^{-0.5\pi_t^2}\sqrt{2}\left(\operatorname{erf}\left(-0.5\sqrt{2}\,\pi_t\right) - 2\,y_t + 1\right)}{\sqrt{\pi}\left(\operatorname{erf}\left(-0.5\sqrt{2}\pi_t\right)^2 - 1\right)}\left(\pi_{t-1}(\theta) + \alpha\frac{\partial \pi_{t-1}(\theta)}{\partial \alpha}\right)
\end{aligned} \tag{13}$$

The $\partial \pi_{t-1}(\theta)$ is based on the initialization values and their partial derivatives given in Equation 9.

4.4 Generalized Logistic

The Generalized Logistic distribution allows for the asymmetric impact of the explanatory variables with the following distribution function:

$$\Phi(y_t) = \left(1 + e^{-\pi_t}\right)^{-k}, \tag{14}$$

with $k \in \Re^+$ and π_t representing the dynamics given in Equation 3. This has also been called the *scobit* model[3] since it nests the logistic distribution when $k = 1$. The log-likelihood to be maximized in the binary response model is then given by the following equation at time t:

$$l_t(\theta) = y_t \log\left(\left(1 + e^{-\pi_t}\right)^{-k}\right) + (1 - y_t)\log\left(1 - \left(1 + e^{-\pi_t}\right)^{-k}\right) \tag{15}$$

[3]See for instance Zeileis and Windberger (2011) for *one* implementation.

The partial derivatives with respect to the log-likelihood are defined as follows:

$$
\begin{aligned}
\frac{\partial l_t(\theta)}{\partial \omega} &= \frac{k\mathrm{e}^{-\pi_t}\left(\mathrm{e}^{k\pi_t} - y_t(1+\mathrm{e}^{\pi_t})^k\right)}{(1+\mathrm{e}^{-\pi_t})\left(\mathrm{e}^{k\pi_t} - (1+\mathrm{e}^{\pi_t})^k\right)}\left(1+\alpha\frac{\partial \pi_{t-1}(\theta)}{\partial \omega}\right) \\
\frac{\partial l_t(\theta)}{\partial \beta} &= \frac{k\mathrm{e}^{-\pi_t}\left(\mathrm{e}^{k\pi_t} - y_t(1+\mathrm{e}^{\pi_t})^k\right)}{(1+\mathrm{e}^{-\pi_t})\left(\mathrm{e}^{k\pi_t} - (1+\mathrm{e}^{\pi_t})^k\right)}\left(x_{t-1}+\alpha\frac{\partial \pi_{t-1}(\theta)}{\partial \beta}\right) \\
\frac{\partial l_t(\theta)}{\partial \delta} &= \frac{k\mathrm{e}^{-\pi_t}\left(\mathrm{e}^{k\pi_t} - y_t(1+\mathrm{e}^{\pi_t})^k\right)}{(1+\mathrm{e}^{-\pi_t})\left(\mathrm{e}^{\mathrm{k}}\pi_{\mathrm{t}} \text{ - } (1+\mathrm{e}^{\pi_{\mathrm{t}}})^{\mathrm{k}}\right)}\left(y_{t-1}+\alpha\frac{\partial \pi_{t-1}(\theta)}{\partial \delta}\right) \\
\frac{\partial l_t(\theta)}{\partial \alpha} &= \frac{k\mathrm{e}^{-\pi_t}\left(\mathrm{e}^{k\pi_t} - y_t(1+\mathrm{e}^{\pi_t})^k\right)}{(1+\mathrm{e}^{-\pi_t})\left(\mathrm{e}^{k\pi_t} - (1+\mathrm{e}^{\pi_t})^k\right)}\left(\pi_{t-1}(\theta)+\alpha\frac{\partial \pi_{t-1}(\theta)}{\partial \alpha}\right) \\
\frac{\partial l_t(\theta)}{\partial k} &= -\frac{\log(1+\mathrm{e}^{-\pi_t})\left(\mathrm{e}^{k\pi_t} - y_t(1+\mathrm{e}^{\pi_t})^k\right)}{\mathrm{e}^{k\pi_t} - (1+\mathrm{e}^{\pi_t})^k}
\end{aligned}
\tag{16}
$$

The $\partial \pi_{t-1}(\theta)$ is based on the initialization values and their partial derivatives given in Equation 9.

References

MA Al-Osh and AA Alzaid. Binomial autoregressive moving average models. *Stochastic Models*, 7(2):261–282, 1991.

Takeshi Amemiya. Qualitative response models: A survey. *Journal of Economic Literature*, 19 (4):1483–1536, 1981.

Robert M de Jong and Tiemen Woutersen. Dynamic time series binary choice. *Econometric Theory*, 27(04):673–702, 2011.

David John Finney. *Probit Analysis*. Cambridge University Press, 1971.

David Firth. Bias reduction of maximum likelihood estimates. *Biometrika*, 80(1):27–38, 1993.

Georg Heinze and Meinhard Ploner. *logistf: An R package for Firth-type bias-reduced logistic regression*, 2013. R package version 1.21.

Georg Heinze and Michael Schemper. A solution to the problem of separation in logistic regression. *Statistics in medicine*, 21(16):2409–2419, 2002.

Heikki Kauppi and Pentti Saikkonen. Predicting us recessions with dynamic binary response models. *The Review of Economics and Statistics*, 90(4):777–791, 2008.

Luke Keele and David K Park. Difficult choices: an evaluation of heterogeneous choice models. In *Paper for the 2004 Meeting of the American Political Science Association*, pages 2–5, 2006.

Henri Nyberg. Dynamic probit models and financial variables in recession forecasting. *Journal of Forecasting*, 29(1-2):215–230, 2010a.

Henri Nyberg. Testing an autoregressive structure in binary time series models. *Economics Bulletin*, 30(2):1460–1473, 2010b.

Henri Nyberg. Forecasting the direction of the us stock market with dynamic binary probit models. *International Journal of Forecasting*, 27(2):561–578, 2011.

Karl Pearson. Mathematical contributions to the theory of evolution. vii. on the correlation of characters not quantitatively measurable. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 195:1–405, 1900.

Tina Hviid Rydberg and Neil Shephard. Dynamics of trade-by-trade price movements: decomposition and models. *Journal of Financial Econometrics*, 1(1):2–25, 2003.

Richard Startz. Binomial autoregressive moving average models with an application to us recessions. *Journal of Business & Economic Statistics*, 26:1–8, 2008.

AN Tikhonov and V Ya Arsenin. Methods of solving incorrect problems. *Science, Moscow*, 1974.

A. Zeileis and T. Windberger. *glogis: Fitting and Testing Generalized Logistic Distributions*, 2011. R package version 0.1-0.

A. Zeileis, R. Koenker, and P. Doebler. *glmx: Generalized Linear Models Extended*, 2011. R package version 0.0-4.

Christopher Zorn. A solution to separation in binary response models. *Political Analysis*, 13(2): 157–170, 2005.