

Joint Modeling of Mixed Responses with Bayesian Modeling and Neural Networks: Performance Comparison with Application to Poultry Data

¹JC Hapugoda and ²MR Sooriyarachchi

¹Department of Management Studies, The Open University of Sri Lanka, Nawala, Sri Lanka

²Department of Statistics, University of Colombo, Colombo, Sri Lanka
jchap@ou.ac.lk and roshini@stat.cmb.ac.lk

Received 7th May 2019/ Revised: 20th December 2019 / Accepted: 5th January 2020
©IAppStat-SL2018

ABSTRACT

Joint modeling of mixed responses has become a popular research area due to its applicability in many disciplines. The interest of this study is joint modeling of survival and count data. Survival data is continuous in nature with censoring information combined to it, while count is a discrete variable. Due to this fact, joint modeling of these two variables will be a challenging task, but it will provide interesting and improved results than modeling these two variables separately. In this study, the concept of joint modeling of survival and count data has been carried out using two approaches: Bayesian modeling and Neural Networks, in order to compare their performances. The results of an application to the poultry data revealed that the Neural Network has a better fit in general.

Key words: Joint modeling, Bayesian modeling, Neural Network, Performance comparison, Poultry

1. Introduction

Joint modeling of data with mixed responses has become a popular research area in recent times. These models can be applied in different disciplines. Many researchers have used joint modeling techniques in various disciplines (Rizopoulos, 2012; Gardiner 2013; Elashoff, Li and Li 2016; Hapugoda et al., 2017). When there is an association between the two responses, a joint model will provide interesting and improved results than modeling the responses separately. According to the literature, many approaches of joint modeling could be identified.

In statistical modeling three types of estimation methods are of use: parametric, semi-parametric and nonparametric. (Darmon, 2013, Pace, 1995). Many

researchers have considered parametric estimation methods, which deal with many distributional assumptions (Rizopolous, 2012, Gardiner 2013, Elashoff et al. 2016). Recent literature reveals the popularity of semi-parametric approaches, such as Bayesian estimation methods, which deals with lesser distributional assumptions (Dunson and Herring, 2005, Bello et al., 2009, Chen et al., 2016). Though, it is not common among conventional statisticians to use nonparametric approaches like neural networks for model fitting, those can also be accommodated for such joint modeling scenarios (Raman and Sunilkumar 1995, Pradhan and Lee, 2010).

Many researchers have carried out comparative studies to identify the performances of different models using different techniques (Comrie 1997, Yay and Akinci 2009, Hapugoda and Sooriyarachchi 2017). It will be beneficial to be aware of the performances of these approaches in joint modeling, while concluding the better fit of the models to the data.

Survival time data (i.e. the duration of time between the initial point up to the occurrence of a particular event) and count data (i.e. the number of observations) are some of the commonly recorded data in many fields of studies. In many cases, though, these two variables can be correlated (Hapugoda and Sooriyarachchi 2017). In addition to these, the data being of hierarchical nature (i.e. in form of clusters) could also be seen in many fields of studies. If a particular set of data consists with all of the above mentioned forms, joint modeling of these two variables can be carried out considering the clustering effect as well. In that case, it is also better to identify which approach can be utilized to obtain more accurate results.

Thus, the objective of this study is to compare the performances of the models obtained by using the Bayesian modeling approach and artificial neural network approach for such data through an application to Poultry industry data from Sri Lanka.

The overview of the paper is as follows: the literature review is presented in the next section, the methodologies of the two approaches used in this study were explained in ‘Methodology’ section, the application of the concept to the poultry industry was described with a discussion in section ‘Analysis of Poultry Data’ and finally the conclusion of the study were presented in the ‘Conclusion’ section.

2. Literature Review

In the recent past, Bayesian methods have become popular among Statisticians as this method has some advantages over other classical statistical methods (Jayawardena and Sooriyarachchi 2014). With the availability of powerful,

high speed computers, Bayesian methods have become computationally feasible with the development of Markov Chain Monte Carlo (MCMC) methods, especially Gibbs sampling (Zeger and Karim 1991).

Bayesian models can be fitted using PROC MCMC procedure of SAS software (Chen 2009) and moreover, the mixed models developed using conventional parametric statistical models can be converted to Bayesian model by using the PROC MCMC procedure (Chen et al. 2016).

ANN is a computer program or hardwired machine that is designed to learn in a manner similar to the human brain. It's a data analyzing technique that was introduced as a result of the information technology advancements and was first designed by McCulloch and Pitts in 1943 and further developed by many other inventors all over the world. It is an information processing paradigm that is established based on the workings of the biological nervous systems. The key element of this paradigm is the novel structure of the information processing system. It is composed of a large number of highly interconnected processing elements (neurons) working in union to solve specific problems. ANNs are also known as the most powerful computational tools to model complex non-linear systems (Koutsoyiannis 2007).

However, ANN approach has advantages and disadvantages compared to the conventional statistical models. The advantages of ANN include, the requirement of less formal statistical training to develop, complex nonlinear relationships between independent and dependent variables can be implicitly detected, the ability to detect all possible interactions between predictor variables and development can be done using multiple different training algorithms. Disadvantages include, a "black box" nature, limited ability to explicitly identify possible causal relationships, requirement of greater computational resources, prone to overfitting, development is empirical, and many methodological issues remain to be resolved (Tu 1996).

Among the different types of ANN approaches present, it is vital to identify which technique/approach is the most appropriate to model the data under consideration. As the response variables are continuous and discrete, the ANN technique that can address continuous variables should be chosen.

Feed forward neural network (FFNN) and generalized regression neural network (GRNN) which belong to a class of neural networks widely used for mapping discrete/ continuous functions (Timonin and Savelieva 2005). Researchers have compared the performances of FFNN and GRNN to model various scenarios and identified that the GRNN performs well in the case of dis-

crete/ continuous variables (Comrie 1997, Düzgün 2010). Thus, the GRNN approach was adopted to use in this study as a better approach to model a discrete/ continuous function which has been first introduced by Specht in 1990.

Measuring performance of the developed models is a crucial factor to conclude the best model and conclusions. Many proposed criteria have a component that quantifies the goodness of model fit, along with a component that penalizes model complexity; namely, the Akaike information criterion (AIC), the Bayesian information criteria (BIC), and the deviance information criterion (DIC). In the context of a Bayesian hierarchical model, the number of independent parameters included in the model is difficult to determine, which makes the use of AIC or BIC problematic. DIC has been proposed for model comparison in this context (Jayawardena and Sooriyarachchi 2014). DIC is a generalization of AIC to be used for Bayesian modeling (Spiegelhalter et al. 2002). Literature suggests that conditional AIC (cAIC) is a comparable form to DIC (Vaida and Blanchard 2005). Thus, DIC of Bayesian model and cAIC of GRNN model was used to compare the model performance in this study.

3. Methodology

3.1 Data Processing

The authors of this study have previously proposed a joint model based on Generalized Linear Mixed Modeling (GLMM) approach (Hapugoda et al. 2017). Due to the complexity of combining survival and count data due to their different nature, the survival data was converted into a discrete format in a way that it is accommodated by Discrete Time Hazard Model (DTHM). This process has been explained in Hapugoda et al. 2017. At the end of the process, two types of variables are created: binary (survived/ not survived) and survival time indicators. In modeling, binary variable is used as the response variable, while survival time indicators are added as explanatory variables. This was carried out in SAS software using author written codes.

3.2 Bayesian Methods

The Bayesian approach is a sequential learning approach where the prior beliefs/ ideas are combined with the data collected to produce posterior beliefs/ideas to a problem. Then the previous posterior ideas act as prior knowledge and combined with data simulated from the joint posterior distribution using a sampler such as Gibbs sampler or Metropolis-Hastings approach. For example, for an unknown parameter θ when the prior beliefs are condensed to a prior distribution $p(\theta)$ and when the collected data y (with the distributional assumption) produces a likelihood function $L(y|\theta)$, which is

used as the function that maximum likelihood methods maximize. Then the prior distribution and the likelihood function are combined to produce the posterior distribution for θ , $p(\theta|y) \propto p(\theta)L(y|\theta)$ where this equation can be used to reach inferences about θ .

However, calculating the proportionality constant is an issue in this process. MCMC methods evade this problem as it does not calculate the exact form of the posterior distribution but instead produce simulated draws from it. MCMC methods are more general in that they can be used to fit many statistical models. It is a simulation based procedure so that rather than producing point estimates, the methods are run for many iterations and at each iteration an estimate for each unknown parameter is produced. These estimates will not be independent, as in each iteration the estimates from the last iteration are used to produce new estimates. The process is repeated many times using the previously generated set of parameter values to generate the next set. The chain of values generated by this sampling procedure is known as a Markov chain, as every new value generated for a parameter only depends on its previous values through the last value generated. The field of MCMC convergence diagnostics is concerned with calculating when a chain has converged to its equilibrium distribution (Jayawardena and Sooriyarachchi 2014). SAS software version 9.4 was used to model the data using PROC MCMC procedure which uses Metropolis-Hastings approach.

3.3 Generalized Regression Neural Networks (GRNN)

GRNN was proposed by Donald F. Specht in 1990 and falls into the category of probabilistic neural networks (Specht 1990). This type of neural networks needs only a fraction of training samples a back propogational neural network would need in order to train the network (Specht 1991). Therefore, the use of a probabilistic neural network is especially advantageous due to its ability to converge to the underlying function of the data with only few training samples available. The additional knowledge needed to get the fit in a satisfying way is relatively small and could be done without additional input by the user. This makes GRNN a very useful tool to perform predictions and comparisons of system performance in practice.

GRNN can be fitted to data using MATLAB software. The probability density function used in GRNN is the Normal Distribution. For each training sample, X_i is used as the mean of a Normal Distribution for n number of samples. The distance, D_i , between the training sample and the point of prediction, is used

as a measure of how well the each training sample can represent the position of prediction, X .

$$Y(X) = \frac{\sum_{i=1}^n X_i \exp(-D_i^2/2\sigma^2)}{\sum_{i=1}^n \exp(-D_i^2/2\sigma^2)} \quad (1)$$

$$D_i^2 = (X - X_i)^T \cdot (X - X_i) \quad (2)$$

The smoothness parameter (σ) is the only network parameter of this procedure (i.e. the parameter of the Gaussian curves). The search for the value of the smoothness parameter has to take several aspects into account depending on the application the predicted output is used for. It was suggested to use the holdout method to select a good value of σ (Specht 1990). In the holdout method, one sample of the entire set is removed and for a fixed σ GRNN is used again to predict this sample with the reduced set of training samples. The squared difference between the predicted value of the removed training sample and the training sample itself is then calculated and stored. The removing of samples and prediction of them again for this chosen σ is repeated for each sample-vector. After ending this process the mean of the squared differences is calculated for each run. Then the process of reducing the set of training samples and predicting the value for these samples is repeated for different values of σ . The σ for which the sum of the mean squared difference is a minimum is the σ that should be used for the predictions by using this set of training samples.

The method to obtain the best neural network is based on selecting the best parameters and best input combination. It was identified that the ten-fold stratified cross validation method is a better approach to select the best model in terms of the input combination (Kohavi 1995). Hence, the same method has been used in this study.

3.4 Performance Comparison

As stated in the literature review, DIC of Bayesian model and cAIC of GRNN model are used to compare the performances of the models.

3.5 Deviance Information Criterion (DIC)

The DIC is a generalization of AIC to a Bayesian setting (Spiegelhalter et al., 2002), where s is replaced by the Bayesian equivalent, namely PD, and the goodness of fit in the first term is replaced by a Bayesian estimate (e.g., posterior mean). Letting Θ be the parameters of the model, the DIC is defined as:

$$\text{DIC} = D(\bar{\Theta}) + 2P_D \quad (3)$$

where $D(\bar{\Theta}) = -2\log L(y/\bar{\Theta})$ is the deviance of the model evaluated at the means of the posterior distributions of the parameters, and $P_D = \overline{D(\bar{\Theta})} - D(\bar{\Theta})$ is the effective number of parameters. SAS version 9.4 PROC MCMC calculates DIC taking $\overline{D(\bar{\Theta})}$ to be the posterior mean of $-2\log L(y/\bar{\Theta})$, and evaluating $D(\bar{\Theta})$ as -2 times the log likelihood at the posterior mean of the stochastic nodes (Vallejo et al. 2014).

3.5 Conditional Akaike Information Criterion (cAIC)

The conditional AIC is similar in form to the marginal AIC; however, these focuses have different likelihood functions. and a different number of parameters. (Vallejo et al. 2014). The c-AIC is defined as :

$$cAIC = d + 2S_c \quad (4)$$

where the deviance (d) is minus 2 times the conditional log-likelihood function at convergence, and S_c is the effective number of parameters of the candidate model defined in Vaida and Blanchard (2005).

4. Analysis of Poultry Data

4.1 Background

Poultry industry is one of the most research intensive agricultural industries due to its technical advancement and high degree of quality concentration. The poultry industry in Sri Lanka has shown a significant growth and emerged as a dynamic industry within a short span of time. Locally, this industry operates at levels of breeder farms consisting of grandparent and parent flocks and commercial farms (often a profitable venture). Many researches have been carried out in various countries to identify factors affecting the quality of chicks in every level of the industry hierarchy.

The economic success of a breeder operation depends on prediction of ‘the number of chicks that can be obtained by hatching a known number of eggs’ (number of hatching eggs). This is dependent on breeder factors such as breeder age and fertility rate (Reis et al. 1997, Lapao et al. 1999, Tona et al. 2002, Petek and S. Dikmen 2006, Hapugoda et al. 2013). Though the storage of hatching eggs is a necessary part of the commercial incubation, dramatic decline in the number of hatching eggs is observed in broiler breeder eggs stored for extended periods (Lapao et al. 1999) and there is an interaction between length of storage and breeder age (Reis et al. 1997). Moreover, in practice, when the eggs are hatched based on the demand (selling of one-day-old

chicks is the objective of breeder operation), eggs from older breeders with less fertility are given the priority, irrespective of the length of storage (in days). Thus, a joint model to estimate the number of hatching eggs with their length of storage considering breeder factors such as age and fertility rate will be advantageous for the success of the operation of poultry industry.

As the breeders are grouped in different batches based on their age (i.e. 35 breeders from same age are considered as one batch), the number of eggs obtained from each batch in different time points are considered for this study. As the data were clustered within batches and consisted with data collected from same batch of breeders repeatedly over time, the data are correlated and clustered.

The data are processed according to the method mentioned in the methodology section. The variables used for this study are shown in table 1.

Table 1: Description of data

Variable	Category	Code
Length of storage (survival time indicators)	<14 days (t1)	1
	14 – 20 days (t2)	2
	>20 days (t3)	3
Hatching status	Hatched	1
	Not hatched	0
Breeder age	Continuous	
Fertility rate	Continuous	
Egg Count	The number of eggs from each batch in each day	

The responses of analysis are: Y_{gij1} (hatching status – binary variable) and Y_{gij2} (log of egg count - number of eggs – normally distributed variable). Here, the event (hatching) is happening in time interval g for the i th egg from the j th batch given that event does not occur in $(g-1)$ and x_{ij} are covariates at the batch j and egg i . Since the egg count is was found to be not distributed as

Poisson, the log of egg count was considered as a normally distributed variable (Changyong et al. 2014). Variables that impact $Y = (Y_1, Y_2)$ are the explanatory variables (breeder factors) denoted by X_{ij} .

4.2 Bayesian Model

To formulate a joint model, Generalized Linear Model (GLM) can be used to form marginal models for each response by considering mean $E(Y_k|X_i)$ and variance $\text{Var}(Y_k|X_i)$ where $k=1,2$. The responses are linked by structuring a covariance matrix $\text{Var}(Y_k|X_i)$ to include potential correlations.

In GLM,

$$l_k(E(Y_{ik} | X_i)) = X_{ik}'\beta_k \quad (5)$$

where $k=1,2$ and i denotes each record from each cluster and l_k is the link function. Here, $l_1(u)$ is the logit link and $l_2(u)$ is the identity link.

A structural formulation of the model is given as:

$$l_1(Y_{gij1}') = \text{Logit}(\pi_{gij}) = \beta_0 + u_{0j} + \sum_{g=1}^n \alpha_g T_{gij} + \beta X_{ij} \quad (6)$$

where $u_{0j} \sim N(0, \sigma_u^2)$

and

$$l_2(Y_{gij2}') = \text{Log}(\mu_{gij2}) = \beta_0 + u_{0j} + \beta X_{ij} \quad (7)$$

where $u_{0j} \sim N(0, \sigma_u^2)$

For simplicity, we assume that both random effects are the same (u_{0j}) and have the same variance (σ_u^2). The joint model variance matrix is the $[(\sigma_u^2, \rho_{12}\sigma_u^2), (\rho_{12}\sigma_u^2, \sigma_u^2)]$. The variance matrix of $Y = (Y_1, Y_2)$ is structured as $\text{Var}(Y_i|X_i) = A_i^{1/2}R_i A_i^{1/2}$ where R_i is a user specified 2×2 covariance structure and A_i is the diagonal matrix of the variances of (Y_1, Y_2) .

This model was fitted as a Bayesian model using PROC MCMC procedure of SAS software. Here, each of the two model statements is based on a different regression model, but they share the same random-effect (Chen et al. 2016).

4.3 GRNN Model

The best model could be obtained when the best σ value is considered, in which the sum of square error is minimized. In this analysis, the best σ value was recorded as 0.5.

For the analysis, the dataset was split to two categories, as training set (90% of data) and test set (10% of data). The ten-fold stratified cross validation method was used to identify the best division of data and this has been carried out by using different columns of the input matrix to include in the training set and test set. For example, the first combination is 1 to 9 columns of input matrix as the training set while 10th column taken as the test set. Likewise, 10 different models have been considered in order to choose the best division of the data. Thus, for different sets of training data and test data divisions used in this analysis, the plots of σ values with their corresponding sum of square error (SSE) values were obtained. The best model is selected by considering the best σ values identified, which generates the smallest sum of squares value. Here, the best division of data was identified as; taking the 1st column of input matrix as test set, while the rest of the data were considered as the training set. The recorded SSE value for this combination was 5.9645.

The obtained neural network consists of 6 input nodes, 2 output nodes and 2 hidden layers, where the two hidden layers consists of 988 and 2 nodes respectively.

4.4 Performance Comparison

In order to compare the performance DIC value of Bayesian model and cAIC value of GRNN method was obtained. The two values are 28,211.76 and 12,168 respectively, as shown in table 2. Since the value of Neural Network was the lowest, it could be mentioned that the GRNN has a better fit than Bayesian model.

Table 2: Performance Comparison

Measure	Bayesian	GRNN
cAIC	-	28,211.76
DIC	12,168	-

Conclusion

In this study, Bayesian approach and neural network approach for statistical modeling were compared using a real life application. According to the results, it can be concluded that GRNN has a better fit than the Bayesian model. However, in order to generalize the findings, a simulation study should be carried out.

Acknowledgments

Authors profoundly thank University of Colombo Research Grant 2014 for funding the research and Dr. RS Kalupahana & Dr. DA Satharasinghe of University of Peradeniya for providing the data.

References

1. Changyong FENG · Hongyue WANG · Naitii LI · Tian CHEN · Hua HF and Ying LI 2014 Log-transformation and its implications for data analysis. *Shanghai archives of psychiatry* 26(): 105.
2. Chen F 2009 Bayesian modeling using the MCMC procedure. In *Proceedings of the SAS Global Forum 2008 Conference, Cary NC: SAS Institute Inc.*
3. Chen F · Brown G and Stokes M 2016 Fitting Your Favorite Mixed Models with PROC MCMC. In *Proceedings of the SAS Global Forum 2016 Conference Cary NC: SAS Institute Inc* <https://support.sas.com/resources/papers/proceedings16/SAS5601-2016.pdf>.
4. Comrie, A.C. 1997. Comparing neural networks and regression models for ozone forecasting. *Journal of the Air & Waste Management Association* 47(6): 653-663.
5. Darmon, D. 2013. Parametric, Semiparametric and Nonparametric Models. Retrieved from <http://thirdorderscientist.org/homoclinic-orbit/2013/10/6/parametric-semiparametric-and-nonparametric-models>
6. Duzgun, R. 2010. Generalized regression neural networks for inflation forecasting. *International Research Journal Of Finance and Economic* 59-70.
7. Flashoff R and Li N. 2016. *Joint modeling of longitudinal and time-to-event data*. CRC Press.
8. Gardiner IC 2013 Joint modeling of mixed outcomes in health services research. In *SAS Global Forum Paper* (pp. 435-2013).

9. Hapugoda J C · Kalunahana R S · Sooriyarachchi M R and Satharasinghe D A 2017 Joint modeling of mixed responses: An application to Poultry Industry. In *International Conference on Computational Mathematics, Computational Geometry & Statistics (CMCGS). Proceedings* (p. 182). Global Science and Technology Forum.
10. Hapugoda J C · Perera D D S and Deshani K A D 2013 A Comparative Study of Generalized Estimating Equations (GEE) & General Regression Neural Networks (GRNN) to Assess the Production of Poultry Breeders. In *Proceedings of the Sri Lanka Association for The Advancement of Science (SLAAS) Annual Conference, Sri Lanka*.
11. Hapugoda J C and Sooriyarachchi M R 2017 A Comparative Study of Generalized Linear Mixed Modeling and Artificial Neural Network approach for the joint modeling of survival and incidence of Dengue patients. In *Journal of Physics Conference Series*890(1): 012135(DOI: 10.1088/1742-6596/890/1/012135), 2017.
12. Jayawardena N I and Sooriyarachchi M R 2014 A multilevel Bayesian analysis of university entrance eligibility for selected districts in Sri Lanka: methods and application to educational data. *Journal of the National Science Foundation of Sri Lanka*42(1).
13. Kohavi, R. 1995. A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Ijcai* 14 (2): 1137-1145.
14. Koutsoyiannis, D. 2007. Discussion of Generalized regression neural networks for evapotranspiration modelling. *Hydrological Sciences Journal*52(4):832-839.
15. Lanao C · Gama I T and Soares M C 1999 Effects of broiler breeder age and length of egg storage on albumen characteristics and hatchability. *Poultry Science* 78(5): 640-645.
16. McCulloch, W.S. and Pitts, W. 1943. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics* 5(4): 115-133.
17. Pace R K 1995 Parametric, semiparametric and nonparametric estimation of characteristic values within mass assessment and hedonic pricing models. *The Journal of Real Estate Finance and Economics*11(3):195-217.
18. Petek, M. and Dikmen, S. 2006. The effects of prestorage incubation and length of storage of broiler-breeder eggs on hatchability and subsequent growth performance of progeny. *Czech Journal of Animal Science* 51(2): 73.
19. Reis I H · Gama I T and Soares M C 1997 Effects of short storage conditions and broiler breeder age on hatchability, hatching time, and chick weights. *Poultry Science* 76(11):1459-1466.
20. Rizonoulos D 2012 *Joint models for longitudinal and time-to-event data: With applications in R*. Chapman and Hall/CRC.
21. Specht, D.F. 1990. Probabilistic neural networks. *Neural networks*3(1): 109-118.

22. Specht, D.F. 1991. A general regression neural network. *IEEE transactions on neural networks*2(6): 568-576.
23. Spiegelhalter, D.J.; Best, N.G.; Carlin, B.P. and Van Der Linde, A. 2002. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 64(4): 583-639.
24. Timonin, V. and Savelieva, E. 2005. Spatial prediction of radioactivity using general regression neural network. *Applied GIS*1(2): 19-01.
25. Tona K · Ramelis F · De Ketelaere K R · Bruogeman V and Decuyvere F. 2002. Effect of induced molting on albumen quality hatchability and chick body weight from broiler breeders. *Poultry science*81(3):327-332.
26. Tu, J.V. 1996. Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. *Journal of clinical epidemiology* 49(11):1225-1231.
27. Vaida, F. and Blanchard, S. 2005. Conditional Akaike information for mixed-effects models. *Biometrika* 92(2):351-370.
28. Vallejo Seco, G.; Tuero Herrero, E.; Núñez Pérez, J.C. and Rosário, P.J.S.L.D.F. 2014. Performance evaluation of recent information criteria for selecting multilevel models in behavioral and social sciences. *International Journal of Clinical and Health Psychology*.
29. Yav M and Akinci F.D. 2009. Application of ordinal logistic regression and artificial neural networks in a study of student satisfaction. *Cypriot Journal of Educational Sciences* 4(1): 58-69.
30. Zeger S I and Karim M R. 1991. Generalized linear models with random effects: a Gibbs sampling approach. *Journal of the American statistical association* 86(413):79-86.