Predictive Modeling in Long-Term Care Insurance

Nathan R. Lally* and Brian M. Hartman†

May 3, 2015

Abstract

The accurate prediction of long-term care insurance (LTCI) mortality, lapse, and claim rates is essential when making informed pricing and risk management decisions. Unfortunately, academic literature on the subject is sparse and industry practice is limited by software and time constraints. In this paper we review current LTCI industry modeling methodology which is typically Poisson regression with covariate banding/modification and stepwise variable selection. We test the claim that covariate banding improves predictive accuracy, examine the potential downfalls of stepwise selection, and contend that the assumptions required for Poisson regression are not appropriate for LTCI data. We propose several alternative models specifically tailored towards count responses with an excess of zeros and overdispersion. Using data from a large LTCI provider, we evaluate the predictive capacity of random forests and generalized linear and additive models with zero-inflated Poisson, negative binomial, and Tweedie errors. These alternatives are compared to previously developed Poisson regression models.

Our study confirms variable modification is unnecessary at best and automatic stepwise model selection is dangerous. After demonstrating severe over-prediction of LTCI mortality and lapse rates under the Poisson assumption, we show that a Tweedie GLM enables much more accurate predictions. Our Tweedie regression models improve average predictive accuracy (measured by several prediction error statistics) over Poisson regression models by as much as four times for mortality rates and 17 times for lapse rates.

JEL classification: C53, G22.

Keywords: Tweedie, Random Forest, Overestimation, Overdispersion, Excess Zeros

*Department of Mathematics, University of Connecticut, nathan.lally@uconn.edu
†Department of Statistics, Brigham Young University, hartman@stat.byu.edu

1 Introduction and Literature Review

Private long-term care insurance (LTCI) policies in the United States are designed to pay a benefit when, due to disability, the policyholder can no longer perform routine activities necessary for their health and well-being without in-home assistance or institutional care (Brown and Finkelstein, 2009). These long-term contracts share features of both health and life insurance. They provide assistance with medical expenses while also being exposed to significant lapse and mortality risk which results in unique pricing assumptions.

Given the importance of the accurate prediction of mortality, lapse, and claim rates in pricing and risk management of LTCI, it is surprising that little research has been dedicated to the subject. Studies tend to discuss large scale macro-economic issues, behavioral economics, personal finance, and a variety of public health concerns (see examples by Brown and Finkelstein (2009), Smoluk (2009), and Peter Zweifel (1998)). Research modeling the lapse, claim, and death rates necessary for pricing long-term care contracts exists but is limited both data and scope. Our methods address these shortcomings by applying a variety of predictive modeling techniques to real US LTCI data provided by a major insurer.

We draw upon various sources to construct our predictive models. First, we review literature that attempts to model critical LTCI rates indirectly by estimating their values using related public health data. Next, we turn to more recent research with real LTCI data from private insurers or survey data from a population of LTCI policyholders. Finally, we conclude by reviewing industry practices as documented in an unpublished report by the University of Connecticut Goldenson Actuarial Research Center (2012) using 13 years of LTCI data.

Several use public health data to indirectly model LTCI rates. Pritchard (2006) constructs a continuous-time, multiple-state model using the U.S. National Long-Term Care Study from 1982, 1984, 1989, and 1994 to describe the disability process of aging and suggests adapting this model to price specific LTCI products. Meiners and Trapnell (1984) use traditional actuarial methods of the era (specified in an unpublished report) to provide premium estimates at the intersections of several important variables for prototype LTCI policies using data from the National Nursing Home Discharge Survey of 1976. These studies, while thorough and rigorous, do not directly model the critical rates necessary for fair and accurate pricing. Utilizing public health data of this nature predicts claims only by proxy, provides no information on lapse risk, and mortality and claim rates may not be accurate since the data does not represent the population of LTCI policyholders. Private long-term care insurance policyholders are likely to be wealthier than the median American and their risk profiles may differ as well (Brown and Finkelstein, 2009).

More recently, several papers focus on directly modeling lapsation and mortality rates for LTCI. Pinquet et al. (2011) builds a proportional hazards model for lapsation using data from a portfolio of individual Spanish LTCI contracts. The model includes the covariates gender, year of risk exposure and the logarithm of a health bonus-malus coefficient and demonstrates a link between age and lapse (lapse rate decreases as age increases). The authors conclude there is a connection between lack of knowledge of insurance products and higher lapse rates. Konetzka and Luo (2011) also presents models for lapsation, choosing logistic regression to describe the probability of lapsation for respondents who report having private LTCI in the 1996-2006 Health and Retirement Studies. The paper concludes that financial variables play more of a role in predicting lapsation than health status. Mortality rates as a function of age of claim occurrence and duration of care for heavy claimants in French LTCI portfolios are modeled in Tomas and Planchet (2013). The author utilizes a locally adaptive smoothing pointwise method (using the intersection of confidence intervals rule) and also a global method using local bandwidth correction factors.

We look to extend these papers in several important directions. We choose not to focus our attention on a specified subset of LTCI policyholders but rather a broad sample including all in-force policies for an insurance provider over more than a decade. In addition to our sample being more diverse, our dataset is much larger (over 69 million exposure units compared to thousands) than previous studies. This large sample size requires a unique approach to model selection. Residual plots are impractical and the calculation of information criteria strain computing resources for some likelihood functions (e.g. Tweedie). Instead we use several prediction error statistics (median absolute prediction error, median squared prediction error, and the equivalents for means) as measures for model comparison. We also compare the merits/disadvantages of a much broader range of models than other papers. Finally, since many publications on LTCI modeling

are European, it is worth noting that European LTCI policies often have smaller benefits and premiums than their U.S. counterparts and are frequently purchased only as a supplement to state provided care and insurance (Pinquet et al., 2011) making them difficult to compare directly.

Due to their specific nature and/or dissimilarity of data sources to our own, the current body of published research on LTCI predictive modeling cannot serve as the main reference for modeling techniques applied in our paper. Instead, we turn primarily to LTCI industry expertise. We are fortunate to have access to a currently unpublished 2012 report by the University of Connecticut Goldenson Actuarial Research Center using 13 years of aggregated data provided by a major LTCI provider. This report is the result of a joint effort between the University researchers, a major LTCI provider, and a major actuarial consultancy and is a fair representation of current predictive modeling practices in the LTCI industry. The Goldenson researchers and industry actuaries focus their efforts on developing Poisson regression models with an offset for exposure in months to predict mortality, lapse, and claim rates. Covariates are selected using automated stepwise routines and are often modified (factor grouping) in an effort to achieve improved model fit and computational performance. We contend that the assumptions required for Poisson regression are not appropriate in the case of LTCI data and disagree with the reliance on covariate modification as a means to achieve better fit.

Our study uses the Goldenson Center models as a baseline and compares their predictive qualities to several GLM, GAM, and non-parametric models better suited for the data. We focus on developing models for lapse and mortality rates. This is not strictly disadvantageous since lapse and mortality are both the most rich and sparse response variables in terms of frequency of occurrences respectively. Because of the polar nature of these two response variables, we are able to dedicate sufficient time to exploring the most extreme cases an actuary would encounter when building pricing models for LTCI. Furthermore, given the similarities (excess zeros and overdispersion) between the claims response and mortality and lapsation, we believe the approach outlined in this paper could be used to produce superior claims rate models. The majority of models developed in this paper are GLMs and GAMs known to perform well when applied to responses with an excess of zeros and significant overdispersion including zero-inflated Poisson, negative binomial, and Tweedie regression models. Finally we experiment with predictive models using random forests.

The remainder of this paper is organized as follows, Section 2 describes the data used in this study, Section 3 details and critiques the current predictive modeling practices in the LTCI industry today, Section 4 describes the modeling methods applied by our study, Section 5 outlines our results and compares our methods to contemporary industry practice, and section 6 provides our conclusions.

2 Data and Data Handling

A large dataset containing 13 years of aggregated LTCI policy information was provided by a major U.S. LTCI provider. The data is aggregated by rating factor and exposure base to make computations manageable. The data contains approximately 69 million policy-in-force months of mortality and lapse exposure and includes two response variables of interest to our predictive modeling effort; mortality count ($y_{\text{mort } i}$) and lapse count ($y_{\text{lapse } i}$).

These discrete response variables are easily converted to the desired continuous rate variables examined in this study through either the inclusion of an exposure offset ($t_{\text{mort } i}$ and $t_{\text{lapse } i}$) or by simply dividing the observed count by its associated exposure to create a rate variable ($r_{\text{mort } i}$ and $r_{\text{lapse } i}$).

Generally, the proportion of LTCI policyholders who die, lapse, or go on claim in any given time period is very low. As such the response variables tend to have an excess of zero observed counts. Our mortality and lapse count response variables are extremely sparse with 99.76% and 78.08% of exposures having zero counts respectively. This excess of zeros is problematic as it often invalidates common modeling assumptions such as heterogeneity. Section 3.4.1 of this paper explains this problem and its impact on our modeling decisions in more detail. Additionally, non-zero mortality and lapse rates appear to belong to a right skewed distribution (see Figure 1 section 4.3.2), albeit with apparent outliers in the tail.

In addition to the response variables, the dataset contains 22 predictor variables (covariates) and 2 offset variables (described in Table 3). Of the available covariates, 16 are categorical covariates describing either technical features of the policy or attributes of the policyholders. While there are a total of four

continuous covariates available, the variables duration, attained age, and issue age form a linear combination and therefore only two of the three can be included as fixed effects. For computational efficiency we grouped the exposures by these covariates creating approximately 17,000 or 600,000 groups with non zero exposures for mortality and lapse respectively.

Incidence Year*	Year of exposure, incidence year of decrement
Issue Year*	Year of issue
Stat Co	Statutory company
Gender	Gender
Ported	Indicates if the policyholder can continue coverage after leaving their employer
Duration*	Number of years since issue as of the exposure period, capped at 25
Attained Age*	Attained age as of the exposure period
Issue Age*	Age at issue
Coverage	Type of coverage (facility only, professional home care, total home care)
EP	Elimination period
Benefit Period	Benefit period
Benefit Amount	Monthly facility benefit amount
Funding	Funding type (employer funded, employer contributed, voluntary, etc...)
Relationship	Relationship to employee
Policy	Policy form group
State	Issue state
Inflation	Inflation protection type
Stop Bill	Indicates policy is in a status in which they are no longer billed
Home Pct	Percentage of the monthly facility benefit that is available for home care
Premiums*	Total premium of exposures
Premium Mode	The frequency at which premiums are required to be paid
In Mode Indicator	Indicates whether a premium is due in the next month or not
Lapse Exposure†	Reporting lag adjusted total insureds exposed to lapse in the exposure period
Mortality Exposure†	Reporting lag adjusted total insureds exposed to death in the exposure period

**indicates continuous or integer variable, all others are categorical*

†*indicates continuous exposure variable*

Table 1: Description of Covariates and Offset Variables

3 Current LTCI Industry Methodology

3.1 Poisson Regression with Offset

The authors of the Goldenson Center Report utilize generalized linear models with a Poisson error structure (Poisson regression) and a log link function to model critical rates used for pricing. Typically used to model count data, Poisson regression and other GLMs have been used in the actuarial profession since the early 1980s and have a wide variety of applications (Haberman and Renshaw, 1996). Since it is reasonable to assume the count of claims, lapses, and deaths are proportional to the time a policy is exposed to risk, the model includes an offset variable for exposure. In the case of this study, the offset variable $\ln(t)$ is defined as policy in force time in months and its inclusion allows us to make inferences about our response variables as rates ($\mathbb{E}[Y_i/t]$).

The validity of results derived from Poisson regression depends on three basic assumptions; perfect homogeneity throughout the sample (the rate parameter is the same for each unit of exposure with identical explanatory variables), each unit of exposure generates events (e.g. claims, lapses, or deaths) in accordance with a Poisson process, and the incidence of observing an event is mutually independent for all observations (Brockman and Wright, 1992). While some actuarial applications may adequately meet these assumptions,

modeling critical LTCI rates under these assumptions can be problematic as we will show in section 3.4.

3.2 Variable Selection and Models

To select covariates to include in their models, the Goldenson Center researchers used a combination of subject matter expertise and stepwise regression routines. P-values from Chi-Squared tests of the model parameters were used as variable selection criteria. Goodness-of-fit was evaluated by assessing the A/E ratio (actual mortality or lapse rate divided by their respective predicted/expected rates) at various values of the covariates. For example, the A/E ratio for mortality models was evaluated for every five-year grouping of attained age and duration, for males and females, and various intersections of attained age, duration, and gender. The A/E ratio was also calculated for the entire model. The Goldenson Researchers did not select training and test samples to evaluate out of sample predictions neither did they perform any cross validation, but rather fit and evaluated their models on the entire sample.

After fitting Poisson regression models for each critical rate, the Goldenson Center researchers employed a variable modification scheme at the suggestion of the LTCI insurance provider. The researchers suggest that some ranges of values for particular categorical covariates vary identically with respect to the response. These ranges of values were grouped by redefining them as a single factor level. The report states that this factor grouping improved statistical significance of associated parameters and enhanced computational performance. Continuous covariates were also modified to compensate for lack of exposure at "tail values". For example, there may be few observations for policyholders of attained age greater than 90 which has the potential to affect parameter significance and model convergence. Under the covariate modification scheme suggested by the Goldenson report all attained age values greater than 90 would be set to 90. Variable modification/banding is perhaps more common in insurance than other fields due to its usage in methodologies that predate sophisticated statistical techniques and also concerns about computing power in the early 90s (Brockman and Wright, 1992). The practice remains common due to popular industry software which bands variables automatically.

Table 2 presents the model formulas developed in the Goldenson report. A check mark indicates that the variable was included in the model while a check mark with an asterisk indicates the variable underwent some form of modification. Notice that many more variables are included in the lapse rate model and that few of the variables were banded.

Goldenson Center LTCI Poisson Regression Models		
Response & Covariates	Mortality Rate Model	Lapse Rate Model
Response	$y_{\text{mort } i}$	$y_{\text{lapse } i}$
Incidence Year		
Issue Year		
Stat Co		
Gender	✓	
Ported		✓
Duration	✓	✓
Attained Age	✓*	
Issue Age		✓
Coverage		✓
EP		
Benefit Period		✓*
Benefit Amount		✓*
Funding		✓
Relationship		✓
Policy		
State		
Inflation		✓*
Stop Bill		
Home Pct		✓
Premiums		
Premium Mode		✓
In Mode Indicator		✓
Exposure	$t_{\text{mort } i}$	$t_{\text{lapse } i}$

Table 2: Goldenson Center LTCI Poisson Regression Models

3.3 Critique of Poisson Regression and Common Variable Modification and Selection Techniques as Applied to LTCI Claims, Mortality, and Lapsation Data

We contend that for LTCI rate data the Poisson assumption (outcomes being generated by a Poisson process) is violated, Poisson regression models fail to address overdispersion and heterogeneity, and we highlight several properties of the Poisson distribution which make it less than ideal for these applications. Additionally, we provide a critique of variable modification schemes and stepwise regression, both common industry practices.

3.3.1 Violation of Assumptions required for Poisson Regression

For both response variables we observe an over-abundance of zero counts; more than would be expected for any rate parameter in a Poisson distribution that would also be likely to produce the observed positive counts. This suggests that our responses are not generated from a homogeneous Poisson process; thereby violating the first and second assumptions necessary for Poisson regression. Our response variables are indicative of unobserved population heterogeneity meaning it is likely that our data contains several sub-populations. Some of these sub-populations appear to have a very high probability of not experiencing a lapse or mortality while others may indeed behave in a Poisson-like manner. The heavy positive skewness of the distribution of our response variables is a major influence in the selection of error structures for models developed in section 4.

Overdispersion of count data is common in many sciences (Böhning et al., 1997) and before developing new models, we test the baseline Poisson models for its presence. The Lagrange Multiplier test can be used to test the Poisson dispersion against an alternative model (in our case negative binomial) (Greene, 1998). Its null hypothesis assumes equality between Poisson and alternative model dispersion and the test statistic is distributed χ^2_1. Table 3 contains the results of the Lagrange Multiplier test when applied to our baseline Poisson regression models for mortality and lapse rates. The models are fit to successively larger samples by incidence year per our methodology in section 4.3.1.

	Mortality Rate Model	Lapse Rate Model
Subset	p-value	p-value
2000-2002	**0.0008**	**0.0000**
2000-2003	**0.0364**	**0.0000**
2000-2004	0.0829	**0.0000**
2000-2005	0.1742	**0.0000**
2000-2006	0.1496	**0.0000**
2000-2007	**0.0446**	**0.0000**
2000-2008	0.1381	**0.0000**
2000-2009	0.0675	**0.0000**
2000-2010	0.1214	**0.0000**
2000-2011	0.0958	**0.0000**
2000-2012	0.1467	**0.0000**

Table 3: Lagrange Multiplier Test for Overdispersion in Baseline Poisson Models

With 95% confidence, 3 out of 11 mortality rate and all lapse rate models tested show evidence of overdispersion. The Lagrange Multiplier test reveals moderate to strong evidence for overdispersion in our baseline mortality rate models and extremely strong evidence for overdispersion in the baseline lapse rate models. Since overdispersion can be viewed as resulting from neglected unobserved heterogeneity, the homogeneity assumption necessary for Poisson regression is violated. Our proposed methods avoid the shortcomings of the single parameter Poisson error structure by utilizing more robust multi-parameter models designed to accommodate overdispersion and heterogeneity.

Another major drawback of using Poisson regression to make inferences about rates arises because there is a positive probability for every integer y_i. Since y_i can take any value from zero to infinity, when a finite and constant exposure variable t is included in a Poisson regression model there is some positive probability that the predicted outcome (in our case count of lapses or mortalities per observation) will exceed the number of exposures. Consequently, a Poisson regression model may predict rates in excess of 100%. This is clearly problematic for our application in LTCI as it implies it is possible to observe more claims, mortalities, or lapses than there are policyholders in a given exposure period. Indeed, inspection of predicted results generated from Poisson regression models on our dataset reveal several predicted rates in excess of 100% at times exceeding 3,000%. Ideally an error structure should be chosen that provides a more restrictive bound (0%, 100%) for the range of possible responses or at least minimizes the probability a predicted rate will exceed 100%.

3.3.2 Shortcomings of Variable Modification

In addition to the aforementioned problems arising from the application of Poisson regression to LTCI rate data we also warn against the common industry practices of variable modification and over reliance on stepwise regression routines.

Grouping factor (categorical) variables by combining several factor levels into fewer but larger groups can improve processing times for model estimation. However, provided researchers have access to sufficient computing resources, use appropriate sample sizes, and have efficient algorithms, concerns about processing time should not dictate model formulation. Instead of improving model fit, excessive removing or grouping of

factors serves to mask within-cell variation from individual exposures and variation between rating factors is lost completely (Murphy et al., 2000). Though goodness-of-fit statistics and diagnostic plots may appear to be better, this is not because the model more accurately describes the variance in the response. The improved fit is an illusion because some observed variance has been intentionally removed, effectively making the data a better fit for the model. Murphy et al. (2000) provides an extreme example of the data being reduced to one point, the mean, in which case we are left with a perfect fitting model. Grouping factor levels and modifying numeric variables also limits the ability of a model to make reliable inferences using out-of-sample data making it particularly poor practice for predictive modeling. In section 5 we compare Poisson regression models with substantial variable modification to those with the same covariates that have not undergone modification and confirm that the results from the modified models show no substantial improvement in predictive capacity.

3.3.3 Problems with Stepwise Regression

In the early 1990s, British actuaries began to investigate the application of GLMs in ratemaking (Brockman and Wright, 1992; Renshaw, 1994; Haberman and Renshaw, 1996). During this time, computing power had developed sufficiently to enable automated variable selection routines such as stepwise regression. Due to their ease of use, and the appeal of quickly finding a "good" subset of covariates these methods became popular in many sciences (Huberty, 1989). Since then, software packages such as SAS and SPSS have offered stepwise variable selection routines for linear and generalized linear models and their utilization remains commonplace. Despite their popularity, stepwise routines have undergone a great deal of scrutiny. Huberty (1989) outlined several major flaws of stepwise routines. More recently, simulation studies and other analysis have revealed further problems with stepwise algorithms including bias in parameter estimation, problems with consistency and interpretation, and the limitations of deciding on a single "best" model (Whittingham et al., 2006).

Whittingham et al. (2006) performs a simple simulation using stepwise selection to choose between a simple linear regression model with one covariate and an intercept only model. The study concludes that the individual estimates of the parameter associated with the covariate are biased, either towards an underestimate of zero when the parameter was not deemed significant or values in excess of the true parameter if tests were deemed significant.

Consistency and interpretation of models discovered by stepwise regression are also problematic. Both the number of parameters and the order of parameter entry into or removal from a model influence the model selected by a stepwise algorithm. These factors, though unrelated to the underlying process generating the response, can lead to significantly different models being chosen. Apparent quality of a given model selected by stepwise regression is often inflated since the final model is a result of many hypothesis tests thereby increasing the probability of Type I errors (Wilkinson, 1979). Sometimes referred to as Freedman's Paradox, this result indicates that stepwise algorithms tend to produce models with an inflated R^2 value and in which too many parameters appear significant (Lukacs et al., 2010). These factors culminate in misinterpretation of the resultant model and potential overfitting.

Additionally, many scientists take issue with the philosophy of stepwise selection which aims to select a single best model derived from a particular sample. This methodology can cause scientists to be overly confident in a model, not explore similarly fitting alternatives, and not adequately express uncertainty in experimental results (Whittingham et al., 2006).

3.3.4 Overdispersion, Variable Modification, & Stepwise Regression: A Problematic Combination

Additionally, we address a concern particular to stepwise selection coupled with Poisson overdispersion. When fitting a Poisson regression model in the presence of overdispersion, variance is inevitably underestimated. This in turn deflates standard errors (already potentially unreliable given a sparse response (Murphy et al., 2000)) and inflated test statistics for model parameters. Consequently, covariates with little to no effect on the response will appear significant. Coupled with Freedman's Paradox, this stepwise selection

and overdispersion could be disastrous; producing models impossible to interpret and increasing the risk of overfitting. The Poisson regression model for lapse rate created in the Goldenson report is a potential product of this phenomena. This model includes 12 covariates where as our best Tweedie GLM and GAM models only include 4 yet predict out of sample lapse rates with approximately 2.6 - 17.3 times the accuracy depending on the statistic of interest.

Finally, if variable modification is carried out before variable selection a similar problem could arise. As discussed in the preceding subsection, in an attempt to acquire a "better fitting" model, a researcher who performs factor grouping and banding can significantly reduce variation between rating factors. This in turn will drive down standard errors for model parameters and inflate test statistics. It is likely that if a stepwise variable selection routine is applied the bias in parameter estimation described by Whittingham et al. (2006) will be further exaggerated. Again, the end result is likely an over-specified and difficult to interpret model.

Considering the convincing body of work decidedly opposed to variable modification and stepwise routines, the effects of overdispersion, and our access to industry expertise, we forgo the method in this paper. We instead start with a subset of covariates informed by LTCI actuaries and our own experience and then add or remove variables based on exploratory analysis of the data. This methodology is more consistent with actuarial best practices for predictive modeling in LTCI and other areas.

4 Methods

4.1 Parametric Models with Error Structures Suited for Overdispersion Over-Abundance of Zeros

In this section we present several popular GLM error structures for modeling overdispersed count data with an over-abundance of zeros including the negative binomial, zero-inflated Poisson, and Tweedie models. We also present two alternative methods of predictive modeling we will apply to our dataset, generalized additive models (GAMs) and the non-parametric statistical learning technique of random forests.

Notably absent from the list of models, and perhaps the most common approach for handling overdispersed count data, is the quasi-likelihood approach under Poisson like assumptions ("quasi-Poisson") (Ver Hoef and Boveng, 2007). While the quasi-Poisson model is more robust than a traditional Poisson regression model in that it allows the independent estimation of a variance parameter, it is identical in respect to the estimated mean. Due to this similarity, the quasi-Poisson regression model will produce the same predicted values (with respect to the same set of covariates) as the traditional Poisson model, albeit with different standard errors for the intercept and regression coefficients. Since our goal is to find the most precise predicted response values, we exclude the quasi-Poisson regression model from our analysis due to redundancy.

4.1.1 Strictly Discrete Parametric Regression Models (Negative Binomial & Zero-Inflated Poisson)

In addition to Poisson regression models, we test two popular alternatives for overdispersed count data; negative binomial regression and zero-inflated Poisson regression.

4.1.2 Generalized Additive Models

The GAM is a further generalization of the GLM where the linear form of the model $g(\mu) = \beta_0 + x_1\beta_1 + x_2\beta_2 + ... + x_p\beta_p$ is replaced by the sum of smooth functions $g(\mu) = \beta_0 + s_1(x_1) + s_2(x_1) + ... + s_p(x_p)$ where each $s_i(\cdot)$ is an unspecified smoothing function and $g(\cdot)$ is the link function (Hastie et al., 1986). The modeler has the option of including non-smoothed strictly linear terms in the model as they would with a GLM (ex. $x_i\beta_i$) and in fact this is the only option for categorical covariates. Additive models explore non-linear covariate effects efficiently; eliminating the trial and error of manually selecting complex interactions and polynomial terms.

In this paper, we evaluate the predictive power of GAMs with both Poisson and Tweedie errors to compare our baseline (Poisson) and best performing (Tweedie) GLMs with the addition of non-linear covariate effects.

Though there are many options available for smoothing methods, we choose cubic regression splines for their relative computational efficiency given the size of our dataset.

4.1.3 Tweedie Regression

The Tweedie family of distributions are exponential dispersion models which include both discrete distributions such as the Poisson and continuous distributions such as the normal and gamma. The Tweedie family also includes a set of compound Poisson-gamma distributions described by Mildenhall (1999) as having a "Poisson frequency component and gamma severity component." A convenient, albeit simplified, parameterization of the Tweedie distribution is given by Shono (2008) where μ is the location parameter, σ^2 is the diffusion parameter, and p is the power parameter.

$$f(y|\mu, \sigma^2, p) = a(y|\sigma^2, p)e^{-\frac{1}{2\sigma}d(y|\mu,p)}$$

The relationship between the mean and variance of the Tweedie distribution is described by the following,

$$\mathbb{V}[Y] = \sigma^2\mathbb{E}[Y]^p = \sigma^2\mu^p$$

This study focuses on the case where $1 < p < 2$ which corresponds to a compound Poisson-gamma distribution (a member of the exponential family) which allows modeling within the GLM framework.

Tweedie regression is common in many sciences and in actuarial science is often used to develop pure premium models (see Jørgensen and Paes De Souza (1994), Gordon K. Smyth (2002)). The assumption in this case is that claims arrive according to a Poisson distribution while average severity is gamma distributed (Jørgensen and Paes De Souza, 1994). The fact that the compound Poisson-gamma distribution has a positive point mass at zero makes its application in pure premium modeling possible and also is essential for our application to LTCI rate data due to the abundance of rates equal to zero.

We employ the compound Poisson-gamma Tweedie models to LTCI rate data under the assumption that deaths and lapses arrive according to a Poisson process while their exposure periods in months follow a gamma distribution. In this context we model mortality and lapse rates directly ($r_{\text{mort } i}$ or $r_{\text{lapse } i}$ per our notation) rather than through the inclusion of an offset variable as in the other strictly discrete models discussed earlier in this section.

Prior to fitting a Tweedie regression model it is necessary to first estimate the power parameter p. We accomplish this estimation using the R package "tweedie" which contains a useful algorithm for calculating this parameter using the method of maximum likelihood estimation conditional on the same covariates, link function, and offset as the desired GLM model. For the LTCI dataset used in this study we found that the estimated value of p (denoted $\hat{p}$) consistently falls within the approximate ranges $1.30 < \hat{p}_{\text{mort}} < 1.40$ for mortality models, and $1.5 < \hat{p}_{\text{lapse}} < 1.60$ for lapse models. This result supports the compound Poisson-gamma distribution and overdispersion assumptions of the rate data. After the power parameter is acquired, a Tweedie regression model is fit in the same manner as a typical GLM using maximum likelihood estimation.

4.2 Predictive Modeling Via Statistical Learning: Random Forests

Random forests are used in many academic disciplines, most notably genetics and other bio-sciences, and are regarded for their accurate predictions in so called "small n large p" (high dimensional) problems (Strobl, 2009). In this application to LTCI rate data the problem is of the opposite nature ("large n small p") however, their statistical properties, particularly consistency, robustness with respect to noise variables, and adaptation to sparsity make their use appealing.

Though mathematically appealing, the relative computational intensity (due to bootstrap aggregation) of random forests when compared to GLMs, limited the sample sizes of our training sets. As previously mentioned, random forest models generally perform well with smaller sample sizes however, we noticed a fairly strong correlation between sample size and predictive accuracy throughout our model selection process.

Random forests offer the additional benefit of simplifying the modeling process by eliminating the need to select specific covariates to include in the model because random forests always converge (Breiman, 2001).

In fact, one of the strengths of random forest regression is that the procedure can be used to evaluate the strength and correlation of individual predictors. It is still advisable to remove very weak predictors from the model to improve computational efficiency.

4.3 Model Selection

We consider a variety of model structures to predict critical LTCI rates; some parametric (GLMs and GAMs) and others non-parametric (random forests). As such, traditional measures of fit such as Akaike information criterion (AIC) and Bayesian information criterion (BIC) are not sufficient to compare the relative strengths of our various models. Random forest models have no likelihood function to maximize and therefore there is no statistic like AIC or BIC. Also, AIC values for Tweedie models are not possible to compute because of to reliance on the quasi-likelihood and approximations may be unreliable (Shono, 2008). This problem is exacerbated with large datasets and AIC approximations require an excessive amount of computing time and resources. Fitting a model to a subset of our data and using it to predict later observations in the same dataset affords us the opportunity to compare these predicted values to actual observed values. This method provides a means to evaluate the predictive power of a model regardless of its mathematical underpinnings making it possible to compare our parametric and non-parametric models using the same measures. Therefore, we purpose using a set of intuitive weighted and un-weighted prediction error statistics as measures of the predictive potential of a given model. We evaluate the mean absolute prediction error (MeanAPE), mean squared prediction error (MeanSPE), median absolute prediction error (MedAPE), and the median squared prediction error (MedSPE). All prediction error statistics presented in this paper are in terms of the rate response variables r_{mort} and r_{lapse}.

In addition, we also consider the same prediction error statistics weighted by exposure ($t_{\text{mort }i}$, $t_{\text{lapse }i}$) which we denote in this paper with the prefix "Wgt" (ex. WgtMeanSPE). Figure 1 displays a density plot of non-zero mortality and lapse rates from our LTCI data. The majority of the observed rates are relatively small and fall between the range $0 - 0.5$. However, for both rates there is a significant cluster of observations at exactly 1. Exploratory analysis revealed that these observations tend to have very low exposure periods. Weighted prediction error statistics devalue observations with low exposure periods and describe the model's predictive performance for policies with longer histories.

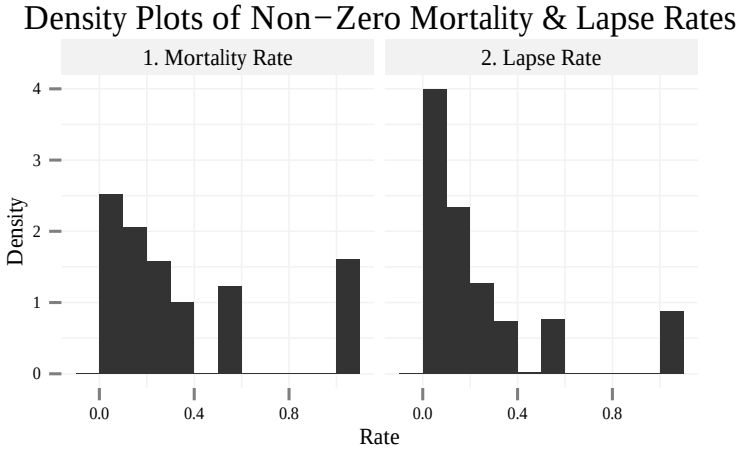

Figure 1: Density Plot of Non-Zero Mortality Rates

Both the un-weighted and weighted mean and median absolute prediction error statistics evaluate the average deviation of our predicted responses from the observed responses with the median being more robust

to extreme deviations. The squared prediction error statistics help to identify models with large deviations from the observed response.

4.3.1 Variable/Model Selection Process for GLM and GAM Models

Rather than employing an automated stepwise variable selection routine based on statistics such as p-values, AIC, BIC, or other common measures of goodness-of-fit, we choose covariates for GLM and GAM models based on guidance from industry actuaries, previous research, and our own experience modeling similar phenomena.

Since we are interested mainly in the predictive accuracy of our models rather than fit, we aim to minimize both the weighted and un-weighted MedAPE, MedSPE, MeanAPE, and MeanSPE. Our process for model selection is as follows. Initial candidate models were fit to a random training set from our data and used to predict observations in a test set. To avoid over-fitting and to evaluate their out-of-sample performance, a group of the best performing models from each family were then each fit to the first 3 years of our data and used to predict outcomes for the 4th year, then these models were fit to the first 4 years of data and used to predict the 5th. This process of fitting a model to the first n years of data and predicting the $(n+1)^{\text{st}}$ year was continued until the dataset was exhausted; constantly keeping track of prediction error statistics. The best model from each family was chosen based on considering the lowest average prediction errors across all years of data and also its ability to improve precision as it was fit to increasingly larger subsets of the data.

This method is similar in some sense to cross validation as it allows us to evaluate predictive accuracy of our models for out-of-sample data. However, this method allows us to evaluate a model's performance in a setting which is similar to the world of a practicing actuary. Typically, A ratemaking actuary updating a class plan would use all the historical data at their disposal to fit a model which predicts the future year(s) mortality or lapse rates. Our procedure emulates this process exactly and gives us insight into how well these models would perform in application rather than when used to make within-sample predictions which tend to be much more forgiving. Assuming no drastic changes in the underlying process generating moralities or lapses, we would prefer a model whose accuracy improves as more data is collected. Models with these characteristics should accurately describe the mortality and claims processes.

5 Results

Though the Poisson regression models presented in the Goldenson Center report include substantial variable modification, we selected models with equivalent covariates without modification as our baseline models for comparison to various proposed improved methodologies. In this manner we first test the popular hypothesis that variable modification improves predictive accuracy for LTCI rate models. We then compare the baseline Poisson GLM to a Poisson GAM, negative binomial GLM, zero-inflated Poisson GLM, Tweedie GLM and GAM, and finally random forest regression.

Though they may include slightly different sets of covariates, each model presented in this results section is the respective best predicting model given a particular error structure. The best predicting set of covariates for each error structure was tested for all other error structures. This helps to ensure predictive performance is a function of an improved mathematical relationship between the covariates and the response rather than variable selection technique.

5.1 Mortality Rate Results

5.1.1 Description of Mortality Model Structures

The baseline mortality model includes 3 covariates, attained age, gender, and duration. Though exact parametric values cannot be published due to proprietary concerns, results from this model are consistent with conventional logic. Mortality rate increases as attained age increases, increases as duration increases, and the model indicates that females have a lower overall mortality rate than males. Notably absent are the variables benefit amount and funding, which the LTCI insurer believes significantly affect mortality rates.

We include the same covariates in the Poisson GAM to see if dropping the linearity assumption improves model fit.

Our best fitting model for mortality rates was the Tweedie GLM model. It includes all the covariates recommended by the provider with the exception of funding which did not improve predictive accuracy. Interestingly both the interaction between issue age and duration, and the variable ported improved model performance. Again we include the same covariates in the Tweedie GAM as the GLM. Table 4 includes parameter estimates for the best predicting Tweedie GLM fit to mortality data from years 2000-2007.

Mortality Tweedie Model Parameter Estimates				
	Estimate	Std. Error	t-value	Pr(>|t|)
Intercept	-12.8890	0.1613	-79.9140	<2E-16
Gender (Male)	0.4066	0.0643	6.3280	2.48E-10
Issue Age	0.0745	0.0032	23.5200	<2E-16
Ported (Yes)	0.6674	0.0904	7.3790	1.60E-13
Funding (Level 1)	0.0101	0.1010	0.1000	0.9202
Funding (Level 2)	-0.2127	0.0941	-2.2610	0.0238
Funding (Level 3)	0.2857	0.0858	3.3290	0.0009
Issue Age : Duration	0.0021	0.0002	10.2870	<2E-16

Table 4: Mortality Tweedie Model Parameter Estimates

Based on our provider's experience, we expect the mortality rate to increase as issue age, and duration increase and also expect males to have a higher risk of mortality. The fact that funding level 2 is associated with a decreased mortality rate was expected as well. Our model confirms these initial assumptions and also reveals a strong association between the ability to port a policy and an increased mortality rate.

Both the zero-inflated Poisson and the negative binomial models proved difficult to fit. The zero-inflated models favored a simple covariate structure or they would not converge. This resulted in a count model with fewer covariates than the Poisson. The zero-inflated portion of the model also includes fewer covariates than we initially desired based on experimentation with fitting logistic regression models to a binary response (mortality occurred or not). The negative binomial model that produced the most reasonable mortality rate predictions in this study contains the same covariates as the best performing Tweedie model.

The mathematical properties of the random forest model discussed in section 4.2 allow us to include many potentially predictive variables in the training data. The variables identified as important for predicting LTCI mortality rates by the random forest algorithm generally agree with the suggestion of the provider. Issue age, funding, duration, and benefit amount, appear to reduce MSE and increase node purity as would be expected. However, for random forests, premiums and issue year are the first and fourth most predictive variables respectively for both measures. Given sound underwriting practice we would expect premium levels to be predictive of a policyholder's mortality rate. Interestingly, including premiums in the parametric regression models often caused a failure to converge possibly because of strong correlation with other covariates. Issue year is also a logical variable for mortality rates since it has been observed that the average age of mortality has been increasing steadily in the 20th and 21st century in the United States. The random forest model did not however rank the variable gender as being an important predictor. This seems to defy the conventional understanding that male mortality rate would be significantly higher than female mortality rate in a given population and warrants further investigation.

For a complete description of mortality rate models see table 5.

Mortality Rate Models						
Response & Covariates	Poisson GLM/GAM	Negative Binomial GLM	ZIP GLM Count	ZIP GLM Zero Inflation	Tweedie GLM/GAM	Random Forest
Response	$y_{\text{mort } i}$	$y_{\text{mort } i}$	$y_{\text{mort } i}$	$y_{\text{mort } i}$	$r_{\text{mort } i}$	$r_{\text{mort } i}$
Incidence Year						
Issue Year						✓
Stat Co						✓
Gender	✓	✓	✓	✓	✓	✓
Ported		✓			✓	✓
Duration	✓	✓			✓	✓
Attained Age	✓					
Issue Age		✓	✓	✓	✓	✓
Issue Age*Duration		✓		✓	✓	
Coverage						✓
EP						✓
Benefit Period						✓
Benefit Amount						✓
Funding		✓			✓	✓
Relationship						✓
Policy						✓
State						✓
Inflation						✓
Stop Bill						✓
Home Pct						✓
Premiums						✓
Premium Mode						✓
In Mode Indicator						✓
Mortality Exposure: $t_{\text{mort } i}$	✓	✓	✓			

Table 5: Mortality Rate Models

5.1.2 Mortality Rate Model Performance Comparison

Table 6 presents average prediction error statistics for all mortality rate models and average the improvements over the baseline Poisson model. The table also contains the average baseline Poisson mortality model prediction error statistics (multiplied by 1,000) for reference. Averages were calculated from results from the series of mortality models fit per the methodology explained in section 4.3.1. Appendix 7.1 contains a table with a complete summary of the series of fit models.

Summary of Mortality Rate Prediction Error Statistics							
	Negative Binomial	Poisson w/ Mod	Poisson GAM	Zero-Inf Poisson	Random Forest	Tweedie GLM	Tweedie GAM
MedAPE Improvement	0.506	1.022	1.010	1.053	1.791	2.121	2.188
MedSPE Improvement	0.250	1.083	1.000	1.083	3.250	4.333	4.333
MeanAPE Improvement	0.544	1.002	1.007	1.040	1.598	1.910	1.942
MeanSPE Improvement	0.905	0.999	1.001	1.002	1.029	1.031	1.031
WgtMedAPE Improvement	0.541	0.940	0.994	1.023	7.915	7.179	7.396
WgtMedSPE Improvement	0.289	0.882	0.990	1.048	58.096	52.788	55.678
WgtMeanAPE Improvement	0.614	1.009	1.014	1.030	10.735	12.112	12.355
WgtMeanSPE Improvement	0.453	1.163	1.039	1.057	5.392	5.409	5.412
Average Baseline Poisson Model Error Statistics							
1000 × MedAPE	0.3526						
1000 × MedSPE	0.0001						
1000 × MeanAPE	1.2094						
1000 × MeanSPE	0.2002						
1000 × WgtMedAPE	0.8386						
1000 × WgtMedSPE	0.0007						
1000 × WgtMeanAPE	4.0585						
1000 × WgtMeanSPE	0.1915						

Table 6: Summary of Mortality Rate Prediction Error Statistics

Tweedie GLMs and GAMs produced the best mortality models in this study. Un-weighted metrics show a large improvement and weighted prediction error statistics were even more dramatic. This drastic change between weighted and un-weighted prediction error statistics suggests that the Tweedie mortality model more accurately predicts rates for observations with large exposure values. Predictive error statistics for the Tweedie GAM are similar to the Tweedie GLM. However, these gains in predictive accuracy come at the cost of model interpretability. We discuss this problem in more detail in section 5.3.7.

Random forest regression models also yielded better predictions than the baseline Poisson models. Like the Tweedie model the largest gains were in MedAPE and MedSPE. MeanAPE and MeanSPE, like the Tweedie model, experienced more modest gains. Weighted median prediction error statistics (WgtMedAPE and WgtMed) for random forests outperformed all other models for mortality rates implying that these models generally make more accurate predictions given longer exposure periods. However, the improvement in weighted mean prediction error statistics was slightly lower than the improvement realized by the Tweedie models. This may be affected by the presence of more outlying prediction errors.

Poisson GLMs with variable modification, Poisson GAMs, zero-inflated Poisson GLMs, and negative binomial models did not improve predictive accuracy substantially when compared to the baseline Poisson model. Surprisingly, the negative binomial regression model was the poorest performer in this study.

5.1.3 Evidence of Mortality Rate Overestimation

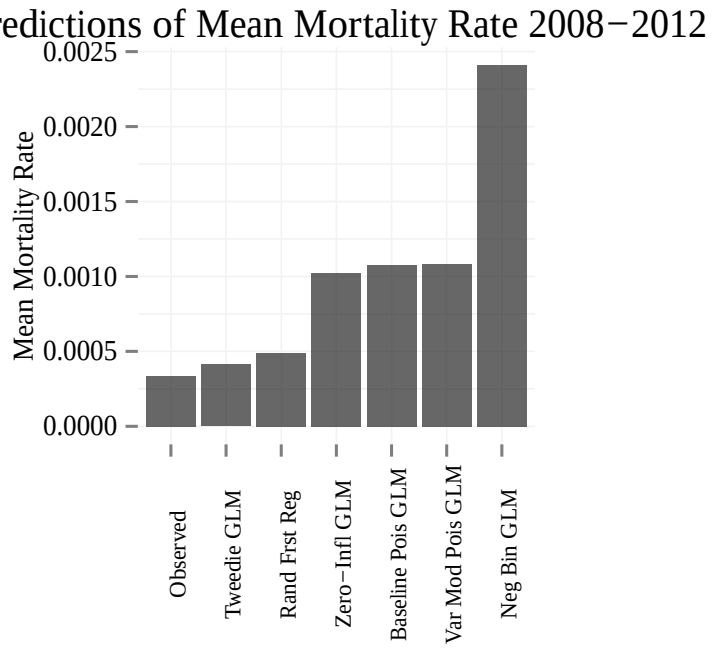

Figure 2: Predictions of Mean Mortality Rate 2008-2012

All of the models considered in this study tend to overestimate mean mortality rates. This phenomena is illustrated in figure 2 where models were fit to the data from the years 2000-2007 and then used to predict mortality rates in the years 2008-2012. Average predictions for these years was then compared to the observed mean during the same time period. Only the random forest and Tweedie models reduce the mean over-prediction significantly when compared to the baseline Poisson model which tends to overestimate by approximately a factor of 3. The random forest model over-predicts by a factor of about 1.45 and the Tweedie by a factor of 1.22. The worst performer was the negative binomial model which grossly overstated the mean mortality rate by a factor greater than 7.

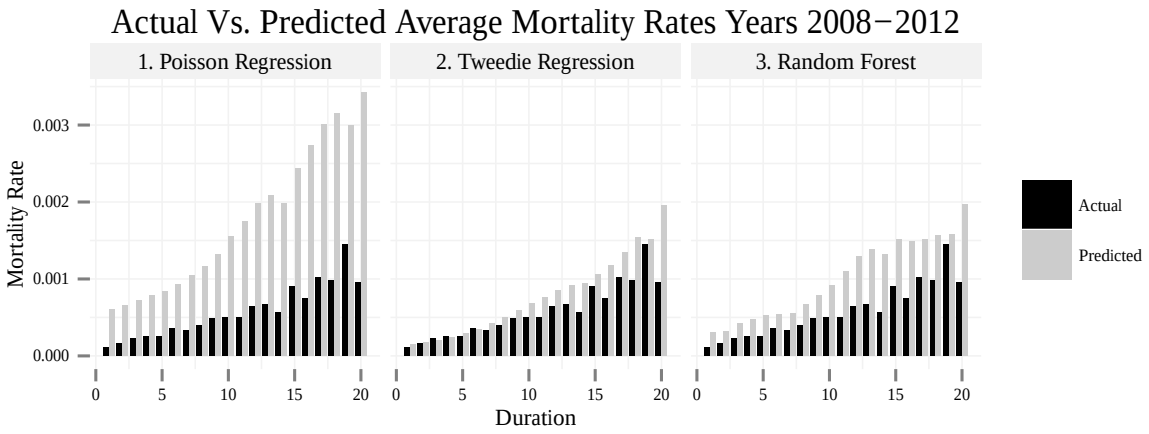

Figure 3: Poisson, Tweedie, and Random Forest Mortality Rate Prediction Comparison

In addition to overestimating overall mortality rates, some models in this study exaggerate observed trends between covariates and the mortality response. Figure 3 compares how well our two best performing (Tweedie and random forest) mortality rate models and the baseline Poisson model estimate a well-known trend in LTCI data. We expect mortality rate to increase exponentially as the variable duration increases. The baseline Poisson model tends to sharply overestimate this trend while the Tweedie model very closely mirrors the observed trend. Random forest regression captures the correlation between duration and the mortality rate response but with a more irregular trend which does not appear exponential as we would expect. This overestimation indicates that parameter estimates in the Poisson regression model are more extreme than the data would suggest.

5.2 Lapse Rate Results

5.2.1 Assessment of Final Lapse Model's Covariate Structure

The baseline lapse model includes 12 covariates. Documentation from the LTCI provider suggested that only 9 variables including, benefit period, benefit amount, home pct, inflation, attained/issue age, relationship, funding, duration, and coverage would be potentially predictive. In addition to the suggested variables, the Goldenson Center researchers also included the covariates ported, premium mode, and in mode indicator. Since the latter 3 variables are not known to have any significant effect on lapse rates, we postulate their inclusion was supported by a stepwise selection algorithm and this model may be overfit. The best performing zero-inflated Poisson, and negative binomial models had similar parameterizations to the baseline Poisson model.

Opposed to the baseline Poisson and other models which included many covariates, the Tweedie lapse models only included the four covariates (duration, issue age, benefit amount and funding). Though the model was simpler, it improved predictive accuracy in excess of a factor of two for all metrics, thereby helping to validate our hypothesis that the baseline Poisson model was over specified. Table 7 provides parameter estimates for a Tweedie GLM fit to lapse data from the years 2000-2007.

Lapse Tweedie Model Parameter Estimates				
	Estimate	Std. Error	t value	Pr(>|t|)
Intercept	-0.9331	0.2764	-3.3760	0.0007
Funding (Level 1)	-0.4121	0.0097	-42.5260	<2E-16
Funding (Level 2)	-0.8205	0.0089	-91.7870	<2E-16
Funding (Level 3)	-0.3132	0.0085	-36.7010	<2E-16
Issue Age	-0.0251	0.0003	-95.1040	<2E-16
Duration	-0.1274	0.0012	-106.4210	<2E-16
Coverage (Level 1)	-0.3650	0.0097	-37.7050	<2E-16
Coverage (Level 2)	-0.2606	0.0106	-24.5290	<2E-16
Benefit Amount (1,000-1,999)	-0.7471	0.2762	-2.7050	0.0068
Benefit Amount (2,000-2,999)	-0.8289	0.2762	-3.0010	0.0027
Benefit Amount (3,000-3,999)	-0.9963	0.2763	-3.6060	0.0003
Benefit Amount (4,000-4,999)	-1.0600	0.2763	-3.8350	0.0001
Benefit Amount (5,000-5,999)	-1.3610	0.2770	-4.9140	8.92E-07
Benefit Amount (6,000-6,999)	-1.3300	0.2767	-4.8060	1.54E-06
Benefit Amount (7,000-7,999)	-2.3720	0.2908	-8.1540	3.51E-16
Benefit Amount (8,000-8,999)	-2.1320	0.2820	-7.5590	4.07E-14
Benefit Amount (9,000-9,999)	-2.5270	0.3300	-7.6570	1.90E-14
Benefit Amount (10,000+)	-5.1930	0.4529	-11.4670	<2E-16

Table 7: Lapse Tweedie Model Parameter Estimates

The model predicts that lapse rates decrease as issue age, duration, and benefit amount increase and also vary according to funding and coverage type. The fact that all the model parameters are negative is generally consistent with the expectation of the LTCI provider. policyholders are less likely to lapse with a more valuable product (larger benefit amount) and are likely to value their policies more as they age or have paid into them for a long period of time. However, contrary to intuition, we observe that coverage level 2 (total home care), applies less downward pressure on lapse rates than other coverage levels. Previous efforts suggested this factor level may be associated with lower lapse rates due to policyholders viewing this coverage as more valuable.

As in the random forest mortality model, premium and issue age are important predictors of lapse rates. Additionally, benefit amount is a highly predictive covariate in the lapse rate random forest model. Intuitively, one would expect an LTCI policyholder to consider both the premium they pay and the potential benefit they may receive when making a lapsation decision. If the penalty from a high premium seems to outweigh a potential future benefit the policyholder may be more likely to lapse. In an analysis of LTCI lapse behavior Pinquet et al. (2011) draw a connection between a benefit-premium ratio and the age of the policyholder at inception of the contract. The variable significance of our random forest regression models appear to be consistent our intuition and other contemporary research.

For a complete description of mortality rate models see table 8.

Lapse Rate Models						
Response & Covariates	Poisson GLM/GAM	Negative Binomial GLM	ZIP GLM Count	ZIP GLM Zero Inflation	Tweedie GLM/GAM	Random Forest
Response	$y_{\text{lapse } i}$	$y_{\text{lapse } i}$	$y_{\text{lapse } i}$	$y_{\text{lapse } i}$	$r_{\text{lapse } i}$	$r_{\text{lapse } i}$
Incidence Year						
Issue Year						✓
Stat Co						✓
Gender						✓
Ported	✓	✓	✓			✓
Duration	✓	✓	✓	✓	✓	✓
Attained Age						✓
Issue Age	✓	✓	✓	✓	✓	✓
Issue Age*Duration						
Coverage	✓	✓	✓			✓
EP						✓
Benefit Period	✓	✓	✓			✓
Benefit Amount	✓	✓	✓		✓	✓
Funding	✓	✓	✓		✓	✓
Relationship	✓	✓	✓			✓
Policy						✓
State						✓
Inflation	✓	✓	✓			✓
Stop Bill						✓
Home Pct	✓	✓	✓			✓
Premiums						✓
Premium Mode	✓	✓	✓			✓
In Mode Indicator	✓	✓	✓			✓
Lapse Exposure: $t_{\text{lapse } i}$	✓	✓	✓			

Table 8: Lapse Rate Models

5.2.2 Lapse Rate Model Performance

Table 9 presents average prediction error statistics for all lapse rate models and average the improvements over the baseline Poisson model. The table also contains the average baseline Poisson lapse model prediction error statistics for reference. Averages were calculated from results from the series of lapse models fit per the methodology explained in section 4.3.1. Appendix 7.2 contains a table with a complete summary of the series of fit models.

Summary of Lapse Rate Prediction Error Statistics							
	Negative Binomial	Poisson w/ Mod	Poisson GAM	Zero-Inf Poisson	Random Forest	Tweedie GLM	Tweedie GAM
MedAPE Improvement	0.923	0.975	0.964	1.002	2.601	2.648	2.460
MedSPE Improvement	0.849	0.946	0.929	1.000	6.870	7.182	6.077
MeanAPE Improvement	0.918	0.996	0.994	1.000	3.597	3.597	3.214
MeanSPE Improvement	0.759	1.009	1.019	1.001	17.258	17.258	18.093
WgtMedAPE Improvement	0.923	0.989	0.989	1.000	7.909	8.114	10.252
WgtMedSPE Improvement	0.849	0.977	0.980	0.999	65.312	68.770	109.885
WgtMeanAPE Improvement	0.847	0.999	1.020	1.000	37.092	37.188	42.284
WgtMeanSPE Improvement	0.694	1.009	1.046	1.000	3080.628	3069.661	3348.091
MedAPE	0.0391						
MedSPE	0.0016						
MeanAPE	0.1151						
MeanSPE	0.1570						
WgtMedAPE	0.1432						
WgtMedSPE	0.0217						
WgtMeanAPE	0.9479						
WgtMeanSPE	7.2554						

Table 9: Summary of Lapse Rate Prediction Error Statistics

Tweedie lapse rate models provided an even more impressive improvement over the baseline Poisson models than Tweedie mortality models. The large gains in MeanSPE imply that the Tweedie model does not predict as many or as large deviations from the true mean as does the baseline Poisson model. Additionally, like mortality models, we see massive increases in weighted prediction error statistics with the same implications. The Tweedie GAM again was comparable to the Tweedie GLM with small improvements in most statistics except MeanSPE and WgtMeanSPE.

Random forest regression models for lapse rates were nearly as successful in making accurate predictions as were the Tweedie models when considering Un-weighted prediction error statistics. In fact, in this case the random forest regression was the most accurate model in terms of MeanSPE. These models were the best performers for all of the weighted prediction error statistics.

Again the Poisson GLM with variable modification, the Poisson GAM, and the zero-inflated Poisson GLM were not significantly better or worse than the baseline Poisson GLM in terms of predictive accuracy. However, as in the mortality rate models, the negative binomial model was the worst performer. Both un-weighted and weighted prediction error statistics suggest this model is from approximately 8%-30% less accurate than the baseline Poisson model.

5.2.3 Evidence of Lapse Rate Overestimation

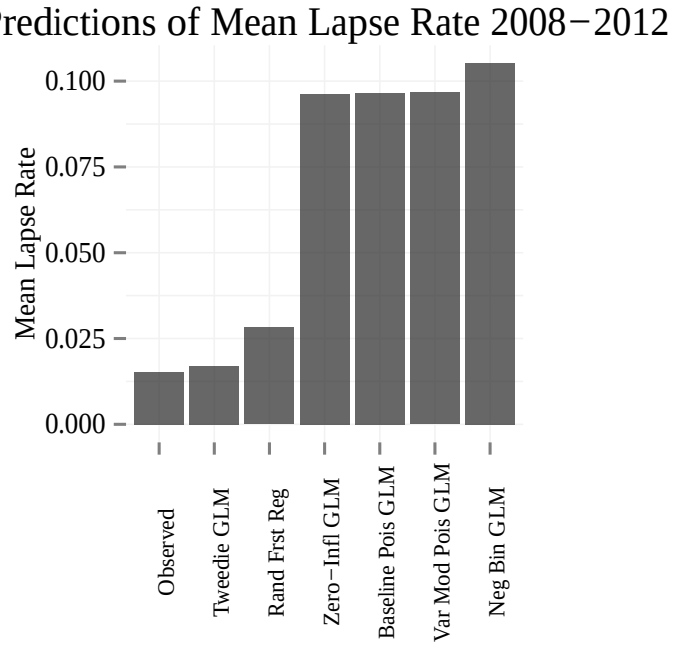

Figure 4: Predictions of Mean Lapse Rate 2008-2012

Baseline Poisson lapse rate models, possibly because of the influence of greater overdispersion, tend to overestimate to an even larger extent than mortality models. As seen in figure 4 The baseline lapse model overestimates the mean for the last 5 years of observed data by approximately 6.3 times. The zero-inflated and variable modified Poisson regression model overestimate by roughly the same amount.

The negative binomial model overstated the mean lapse rate more than all other models. Predictions on the last 5 years of data show the negative binomial model has overestimated the mean lapse rate by a factor slightly less than 7. As opposed to mortality rates, the increase in lapse rate over-prediction for negative binomial models was only marginally worse than the baseline Poisson models.

The Tweedie regression model only overestimates mean lapse rates by a factor of 1.12 and is again the model which over-predicts by the smallest margin.

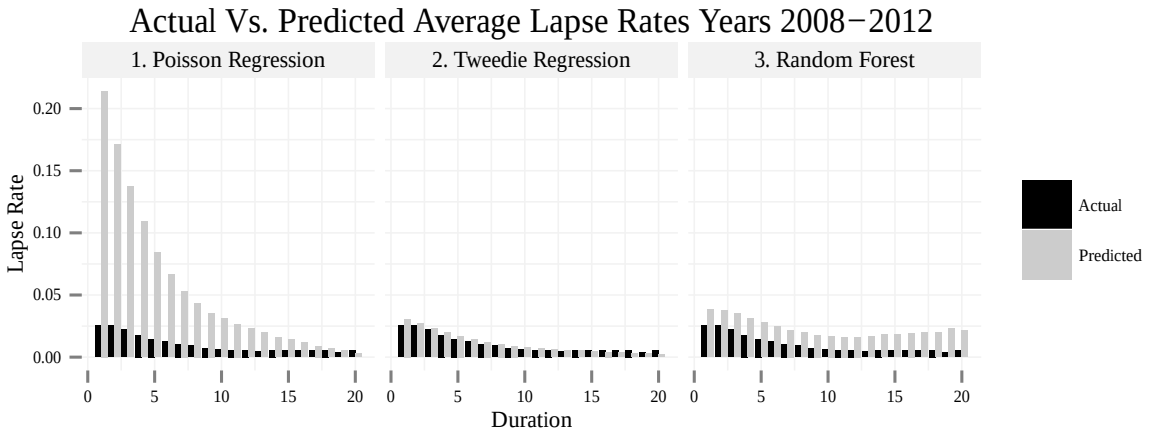

Figure 5: Poisson Vs. Tweedie Lapse Rate Prediction Comparison

Figure 5 examines how well the models reproduce the trend which relates duration to lapse rates to. We expect lapse rates to decrease as duration increases. The Poisson regression captures the correct correlation but produces a much exaggerated trend; overestimating lapse rate for most values of duration. The random forest regression model also tends to overestimate, albeit much less, and produces a more reasonably shaped trend. The Tweedie regression model captures the relationship between the response and duration precisely and is hardly discernible from the observed trend.

5.3 Additional Remarks and Observations Regarding Results

5.3.1 Success with Tweedie GLMs

Tweedie GLM and also GAM models were the best performing models in this study. The Tweedie GLM is the only parametric model in this study which has an error structure that directly accommodates positive semi-continuous data with a point mass at zero. In addition to being the best predicting models in this study, Tweedie models have appealing properties which lend themselves to the LTCI death and lapse processes.

The Tweedie models perform superiorly due to their ability to model Poisson overdispersion through the variance function with power parameter p, the positive point mass at zero, and because the compound Poisson-gamma distribution (as parameterized by the Tweedie distribution) accurately describes process generating the response variables of interest.

Because of their predictive accuracy, flexibility, mathematical appeal, and ease of interpretation, we believe the Tweedie compound Poisson-gamma GLM framework outperforms traditional Poisson GLM models enough to warrant their adoption by practicing actuaries.

5.3.2 Random Forest Regression: Accurate but Limited

Though random forest regression can produce accurate predictions, its non-parametric nature prevents the user from exploiting a priori assumptions such as the exponential growth relationship between mortality rate and duration. The model output gives no understanding of how the covariates affect the mean response nor any measures of uncertainty around estimated parameters (indeed there are no estimated parameters). For these reasons we hesitate to recommend the exclusive use models for LTCI rates. However, variable importance statistics and plots provide a meaningful and intuitive illustration of potentially predictive covariates.

This output from a random forest model could be used to inform the selection of covariates for a parametric model, especially if little is known about the process being analyzed.

5.3.3 Zero-Inflated Models: No Major Improvement

Given data with excess zeros, the extent to which a ZIP GLM outperforms a simple Poisson GLM in terms of fit and predictive accuracy is determined by how well the model identifies the zero generating process. Though we had selected covariates for the zero-inflated portion of the model (through logistic regression) we had difficulty incorporating these assumptions in our final ZIP models. Again convergence proved challenging and we were forced to dramatically simplify our model specifications to get reasonable parameter estimates. For this reason, our ZIP models did not differ significantly from our baseline Poisson models.

There are at least three possible reasons for this failure. The first two are fairly obvious, zeros and events such as deaths or lapses may not be generated by a binomial process, or zeros and events *are* generated by a binomial process but this LTCI dataset does not contain strongly predictive variables for this process. The third reason arises because of the properties of logistic regression models. Mathematically, successes, in a binary response model such as logistic regression, are more informative than failures. Subsequently, given a sparse response, these models tend to dramatically underestimate the probabilities of rare events, in our case mortalities or lapses (King and Zeng, 2001). It is likely a combination of the latter two problems affected predictions produced by our ZIP models and for these reasons we do not recommend these models for LTCI data.

5.3.4 An Overestimation Problem For Models With Poisson and Negative Binomial Error Structures

As shown throughout section 5.2, Poisson regression models tend to overestimate mean mortality and lapse rates. The overestimation seems to be driven by two factors. Primarily, since LTCI data has an excess of zero counts, we can expect that a model under the simple Poisson assumption will underestimate the number of zero observations (Lambert, 1992). Also, as discussed in section 3.3.1, there is no upper bound on the mortality or lapse rate a Poisson model can predict. As evidence, the baseline Poisson model predicted lapse rates as high as 3,200% when fit to the last 5 years of data. The combination of these two factors undoubtedly lead to over-prediction of our two response variables.

Prior to beginning our modeling efforts we expected negative binomial regression to be a natural superior alternative to Poisson regression due to the popular belief that the more sophisticated parameterization better accommodates overdispersion. It has even been shown that the probability of zero in a mixed Poisson distribution (such as the negative binomial) is greater than that for a simple Poisson distribution with the same mean (Mullahy, 1997). For our modeling efforts, negative binomial regression failed to improve upon the Poisson and in fact, performed significantly worse. Parameter estimates, especially estimates of the heterogeneity parameter, often failed to converge using the popular iteratively weighted least squares algorithm. Even with a very large number of iterations convergence was elusive for what appeared to be reasonable models under other error assumptions. For this reason the variable selection process for the negative binomial models was as much a function of potential for convergence as it was predictive accuracy.

Negative binomial regression models produced the most extreme overestimation of mean mortality and lapse rates when compared to our other models. Though somewhat counter-intuitive this problem has been observed in other studies. In an application to shark catch data with excess zeros, Minami et al. (2007) show that negative binomial regression tends to overestimate the trend over time in mean shark catch count. The negative binomial model accommodates the excess zeros by increasing its variance through the heterogeneity parameter α. The heterogeneity parameter would does affect parameter estimates if the negative binomial model is appropriate for the data and correctly specified (Minami et al., 2007). However, due mainly to an extreme excess of zeros, this LTCI dataset is not well described by the negative binomial assumption and therefore it appears the extreme heterogeneity is grossly exaggerating the trends in our data. For example, the LTCI provider's opinion and our data suggest there is a positive correlation between duration and mortality rate and a negative correlation between duration and lapse rate. The negative binomial model captures these

trends but appears to inflate the associated model parameters. Figure 6 illustrates the exaggerated trend estimations produced by the negative binomial model.

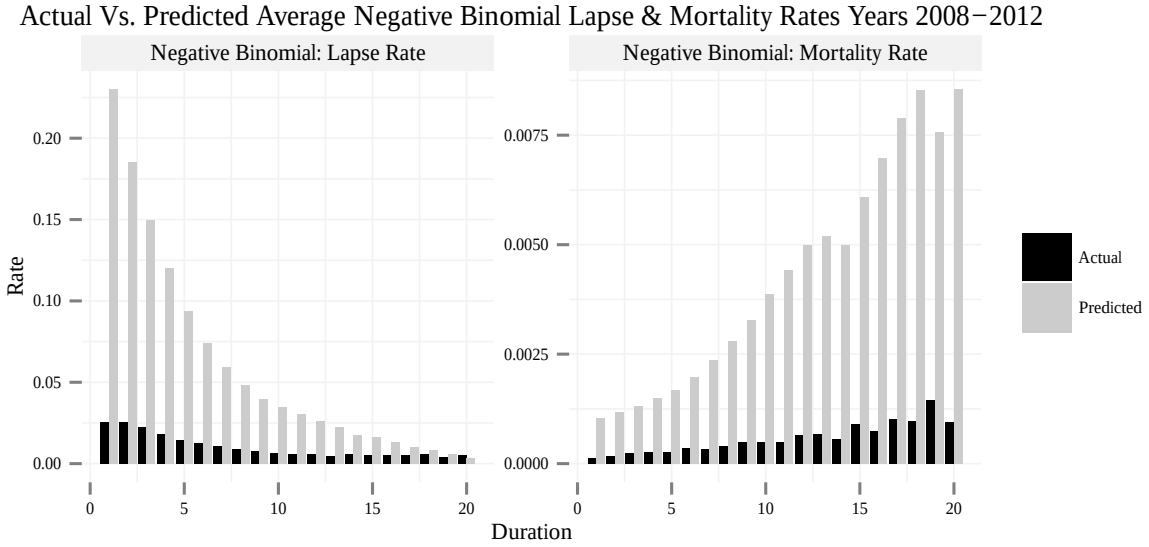

Figure 6: Actual Vs. Predicted Average Negative Binomial Lapse & Mortality Rates Years 2008-2012

Negative binomial GLMs overestimated our mean response variables to a greater degree than the Poisson regression models. We observed several outlying rate predictions in excess of 4,000% when the model was used to predict the last 5 years of data. Negative binomial models were by far our worst performing. We recommend caution before their application if the data exhibit extreme excess zeros.

5.3.5 Potential Problems Caused by Stepwise Regression

Due to the baseline Poisson lapse rate model's large number of covariates, conflict with the information provided by the LTCI insurer, and simpler models that vastly outperform it, we believe the combination of overdispersion and bias in parameter estimation (from stepwise selection) led to a uninterpretable, over fit, and poor predicting model. It is worth noting that this problem is far more pronounced for lapse rate models which show massive overdispersion than for mortality rate models where the Poisson assumption may be more reasonable.

5.3.6 Variable Modification: No Improvement

To assess industry practice which suggests variable modification can improve model fit and predictive accuracy, we fit Poisson mortality and lapse rate models with variable modification per the Goldenson Report. Our initial assumptions that variable modification would not improve predictive accuracy of a Poisson LTCI rate model were confirmed.

Importantly, given the outcomes in this study, variable modification did not alleviate any of the problems associated with Poisson over-prediction. Both the variable modified mortality and lapse rate models overestimated the mean to a slightly greater degree than the baseline models for the last 5 years of observed data.

5.3.7 Additive Models: Not Recommended for LTCI Mortality and Lapse Rates

When applied to our LTCI mortality and lapse data, smoothing of continuous variables offers little to no benefit. Additive models with either Poisson or Tweedie errors fit to the data produced either inconsequential gains in predictive accuracy or slight decreases. Poisson GAMs generally under-performed compared to their GLM counterparts, and at most we observed a 4.5% improvement in lapse MedSPE for the Tweedie GAM versus Tweedie GLM.

Even given the potential of a small performance boost for lapse rate models, we do not recommend additive models for LTCI rate data. Model interpretability and computational efficiency are sacrificed. Additionally, edf statistics for our smoothed predictors were very close to one for both mortality and lapse models. This implies that the log-linear relationship between continuous covariates and the mean response assumed in the baseline GLM models appears to be adequate.

6 Conclusion

Typical industry practice for modeling LTCI mortality and lapse data is fundamentally flawed. Assumptions required for Poisson regression are not supported by the data. Variable banding and stepwise selection algorithms can further exacerbate problems and provide false confidence in parameter estimates. Importantly, we showed that the Poisson GLMs and popular alternatives, zero-inflated Poisson, and negative binomial GLMs systematically overestimate mean LTCI mortality and lapse rates. This overestimation could have serious repercussions and will certainly increase the risk of making poor pricing decisions.

Tweedie GLMs produce vastly improved predictions of mean mortality and lapse rates for out of sample data. In our study, Tweedie regression models improved average predictive accuracy (measured by several prediction error statistics) over Poisson regression models by as much as four times for mortality rates and 17 times for lapse rates.

These performance gains are significant and the benefit of applying these models to LTCI data with excess zeros is clear. Though the Tweedie likelihood is more complex than the Poisson, the additional complexity should not be a barrier to implementation of Tweedie GLMs. Major statistical software packages such as SAS and R and some actuarial software (Emblem) include the capability to perform Tweedie regression so the method is accessible to majority of practicing actuaries. Interpretation of model outputs is similar to more popular generalized linear models and any learning curve is certainly justified by the results.

7 Appendix

7.1 Mortality Rate Prediction Error Statistics and Summary for All Methods

	Year	Poisson w/out Modification	Poisson w/ Modification	Poisson GAM CR Splines	Zero-Inflated Poisson	Negative Binomial	Tweedie GLM	Tweedie GAM CR Splines	Random Forest
MedAPE	Year 3	0.00024657	0.00024982	0.00024939	0.00024978	0.00038355	0.00010012	0.00009105	0.00008214
MedSPE	Year 3	0.00000006	0.00000006	0.00000006	0.00000006	0.00000015	0.00000001	0.00000001	0.00000001
MeanAPE	Year 3	0.00088495	0.00087865	0.00088191	0.00089254	0.00126704	0.00041660	0.00042003	0.00051640
MeanSPE	Year 3	0.00015798	0.00015815	0.00015804	0.00015812	0.00016292	0.00015524	0.00015524	0.00015547
MedAPE	Year 4	0.00028199	0.00028441	0.00028243	0.00027981	0.00055153	0.00013575	0.00012713	0.00010629
MedSPE	Year 4	0.00000008	0.00000008	0.00000008	0.00000008	0.00000030	0.00000002	0.00000002	0.00000001
MeanAPE	Year 4	0.00098153	0.00097149	0.00097953	0.00098097	0.00177184	0.00050071	0.00049711	0.00059004
MeanSPE	Year 4	0.00015731	0.00015736	0.00015731	0.00015740	0.00017270	0.00015386	0.00015384	0.00015408
MedAPE	Year 5	0.00031861	0.00031896	0.00031571	0.00031375	0.00063376	0.00015108	0.00014235	0.00013459
MedSPE	Year 5	0.00000010	0.00000010	0.00000010	0.00000010	0.00000040	0.00000002	0.00000002	0.00000002
MeanAPE	Year 5	0.00110578	0.00109537	0.00110118	0.00109790	0.00197740	0.00054534	0.00053775	0.00061208
MeanSPE	Year 5	0.00018552	0.00018573	0.00018553	0.00018560	0.00020472	0.00018051	0.00018050	0.00018069
MedAPE	Year 6	0.00032220	0.00031970	0.00031972	0.00030103	0.00069167	0.00016676	0.00015977	0.00016389
MedSPE	Year 6	0.00000010	0.00000010	0.00000010	0.00000009	0.00000048	0.00000003	0.00000003	0.00000003
MeanAPE	Year 6	0.00115867	0.00115023	0.00115517	0.00110640	0.00220467	0.00063079	0.00062944	0.00072707
MeanSPE	Year 6	0.00020121	0.00020146	0.00020119	0.00020082	0.00022328	0.00019561	0.00019555	0.00019585
MedAPE	Year 7	0.00034530	0.00034214	0.00034327	0.00033941	0.00073055	0.00016893	0.00016376	0.00020060
MedSPE	Year 7	0.00000012	0.00000012	0.00000012	0.00000012	0.00000053	0.00000003	0.00000003	0.00000004
MeanAPE	Year 7	0.00122445	0.00121877	0.00121676	0.00121492	0.00234255	0.00064469	0.00063628	0.00076551
MeanSPE	Year 7	0.00019694	0.00019728	0.00019678	0.00019703	0.00022026	0.00019090	0.00019090	0.00019103
MedAPE	Year 8	0.00037247	0.00036457	0.00036542	0.00036567	0.00075958	0.00017066	0.00016752	0.00017555
MedSPE	Year 8	0.00000014	0.00000013	0.00000013	0.00000013	0.00000058	0.00000003	0.00000003	0.00000003
MeanAPE	Year 8	0.00124801	0.00124557	0.00123968	0.00124177	0.00245406	0.00062096	0.00061044	0.00069950
MeanSPE	Year 8	0.00016015	0.00016054	0.00015993	0.00016020	0.00018663	0.00015355	0.00015351	0.00015373
MedAPE	Year 9	0.00039380	0.00038714	0.00038793	0.00038730	0.00078967	0.00017258	0.00016867	0.00019687
MedSPE	Year 9	0.00000016	0.00000015	0.00000015	0.00000015	0.00000062	0.00000003	0.00000003	0.00000004
MeanAPE	Year 9	0.00137290	0.00137284	0.00136020	0.00135615	0.00254526	0.00069122	0.00067921	0.00079451
MeanSPE	Year 9	0.00023370	0.00023398	0.00023337	0.00023348	0.00025945	0.00022618	0.00022618	0.00022628
MedAPE	Year 10	0.00042605	0.00041486	0.00042033	0.00041783	0.00085177	0.00018354	0.00018287	0.00023611
MedSPE	Year 10	0.00000018	0.00000017	0.00000018	0.00000017	0.00000073	0.00000003	0.00000003	0.00000006
MeanAPE	Year 10	0.00143085	0.00143230	0.00141679	0.00141226	0.00273659	0.00071090	0.00069464	0.00082252
MeanSPE	Year 10	0.00021853	0.00021891	0.00021818	0.00021823	0.00024893	0.00021027	0.00021023	0.00021047
MedAPE	Year 11	0.00045828	0.00043877	0.00045177	0.00044507	0.00090266	0.00019082	0.00018874	0.00024722
MedSPE	Year 11	0.00000021	0.00000019	0.00000020	0.00000020	0.00000081	0.00000004	0.00000004	0.00000006
MeanAPE	Year 11	0.00152666	0.00152942	0.00151164	0.00150152	0.00285612	0.00075109	0.00073261	0.00087620
MeanSPE	Year 11	0.00023766	0.00023806	0.00023727	0.00023728	0.00026715	0.00022862	0.00022861	0.00022876
MedAPE	Year 12	0.00048173	0.00046030	0.00047835	0.00038845	0.00093720	0.00019848	0.00019311	0.00037095
MedSPE	Year 12	0.00000023	0.00000021	0.00000023	0.00000015	0.00000088	0.00000004	0.00000004	0.00000014
MeanAPE	Year 12	0.00160479	0.00160937	0.00158569	0.00132930	0.00297557	0.00078656	0.00076572	0.00102028
MeanSPE	Year 12	0.00025185	0.00025242	0.00025139	0.00024822	0.00028294	0.00024201	0.00024201	0.00020070
MedAPE	Year 13	0.00023199	0.00021420	0.00022743	0.00019566	0.00043523	0.00019050	0.00018807	0.00025185
MedSPE	Year 13	0.00000005	0.00000005	0.00000005	0.00000004	0.00000019	0.00000004	0.00000004	0.00000006
MeanAPE	Year 13	0.00076531	0.00076737	0.00075754	0.00065788	0.00132329	0.00066480	0.00064841	0.00090071
MeanSPE	Year 13	0.00020111	0.00020116	0.00020105	0.00020073	0.00020498	0.00020000	0.00019997	0.00024226

Table 10: Mortality Rate Un-Weighted Prediction Error Statistics

	Year	Poisson w/out Modification	Poisson w/ Modification	Poisson GAM CR Splines	Zero-Inflated Poisson	Negative Binomial	Tweedie GLM	Tweedie GAM CR Splines	Random Forest
WgtMedAPE	Year 3	0.00065275	0.00069301	0.00066480	0.00066821	0.00091827	0.00006672	0.00006171	0.00004701
WgtMedSPE	Year 3	0.00000043	0.00000048	0.00000044	0.00000045	0.00000084	0.00000000	0.00000000	0.00000000
WgtMeanAPE	Year 3	0.00300735	0.00315711	0.00302741	0.00309691	0.00373197	0.00020528	0.00020505	0.00025243
WgtMeanSPE	Year 3	0.00008873	0.00008620	0.00008633	0.00009285	0.00011271	0.00002533	0.00002533	0.00002546
WgtMedAPE	Year 4	0.00071925	0.00075166	0.00072850	0.00071845	0.00130797	0.00009896	0.00009161	0.00006250
WgtMedSPE	Year 4	0.00000052	0.00000056	0.00000053	0.00000052	0.00000171	0.00000001	0.00000001	0.00000000
WgtMeanAPE	Year 4	0.00339420	0.00343950	0.00336710	0.00341707	0.00542290	0.00026850	0.00026525	0.00029076
WgtMeanSPE	Year 4	0.00011742	0.00010768	0.00011446	0.00011864	0.00021810	0.00002764	0.00002763	0.00002773
WgtMedAPE	Year 5	0.00079534	0.00083092	0.00081111	0.00079497	0.00148312	0.00010895	0.00010389	0.00007687
WgtMedSPE	Year 5	0.00000063	0.00000069	0.00000066	0.00000063	0.00000220	0.00000001	0.00000001	0.00000001
WgtMeanAPE	Year 5	0.00400023	0.00394265	0.00394569	0.00398223	0.00633981	0.00028619	0.00028132	0.00029823
WgtMeanSPE	Year 5	0.00016526	0.00014295	0.00015820	0.00016187	0.00030452	0.00002936	0.00002935	0.00002942
WgtMedAPE	Year 6	0.00083389	0.00088080	0.00084820	0.00080172	0.00164478	0.00011388	0.00010787	0.00008333
WgtMedSPE	Year 6	0.00000070	0.00000078	0.00000072	0.00000064	0.00000271	0.00000001	0.00000001	0.00000001
WgtMeanAPE	Year 6	0.00433911	0.00426195	0.00424385	0.00417502	0.00713053	0.00032244	0.00031935	0.00033500
WgtMeanSPE	Year 6	0.00019067	0.00016214	0.00018016	0.00017545	0.00038326	0.00003267	0.00003265	0.00003274
WgtMedAPE	Year 7	0.00084382	0.00090290	0.00085459	0.00083964	0.00167663	0.00011768	0.00011421	0.00010417
WgtMedSPE	Year 7	0.00000071	0.00000082	0.00000073	0.00000070	0.00000281	0.00000001	0.00000001	0.00000001
WgtMeanAPE	Year 7	0.00415148	0.00409062	0.00408729	0.00411317	0.00702445	0.00033876	0.00033259	0.00037751
WgtMeanSPE	Year 7	0.00018544	0.00015563	0.00017882	0.00018223	0.00039673	0.00003284	0.00003282	0.00003291
WgtMedAPE	Year 8	0.00089242	0.00096327	0.00090280	0.00089657	0.00171762	0.00011872	0.00011732	0.00009722
WgtMedSPE	Year 8	0.00000080	0.00000093	0.00000082	0.00000080	0.00000295	0.00000001	0.00000001	0.00000001
WgtMeanAPE	Year 8	0.00434785	0.00437123	0.00427375	0.00433822	0.00726435	0.00034285	0.00033610	0.00035822
WgtMeanSPE	Year 8	0.00018855	0.00016174	0.00018188	0.00018912	0.00040413	0.00002807	0.00002805	0.00002812
WgtMedAPE	Year 9	0.00095534	0.00103526	0.00095463	0.00095307	0.00180676	0.00012182	0.00011897	0.00010591
WgtMedSPE	Year 9	0.00000091	0.00000107	0.00000091	0.00000091	0.00000326	0.00000001	0.00000001	0.00000001
WgtMeanAPE	Year 9	0.00490743	0.00483337	0.00479410	0.00484418	0.00848582	0.00034919	0.00034126	0.00037641
WgtMeanSPE	Year 9	0.00026360	0.00021795	0.00024710	0.00025694	0.00073674	0.00003553	0.00003552	0.00003559
WgtMedAPE	Year 10	0.00098918	0.00106020	0.00098087	0.00098171	0.00185041	0.00012906	0.00012868	0.00012639
WgtMedSPE	Year 10	0.00000098	0.00000112	0.00000096	0.00000096	0.00000342	0.00000002	0.00000002	0.00000002
WgtMeanAPE	Year 10	0.00475785	0.00461748	0.00470320	0.00468671	0.00807648	0.00037363	0.00036483	0.00040716
WgtMeanSPE	Year 10	0.00025594	0.00021158	0.00024814	0.00025137	0.00065209	0.00003458	0.00003455	0.00003465
WgtMedAPE	Year 11	0.00105432	0.00111993	0.00104926	0.00104371	0.00194249	0.00013358	0.00013291	0.00013073
WgtMedSPE	Year 11	0.00000111	0.00000125	0.00000110	0.00000109	0.00000377	0.00000002	0.00000002	0.00000002
WgtMeanAPE	Year 11	0.00490214	0.00478200	0.00484409	0.00480451	0.00812499	0.00038970	0.00038032	0.00042361
WgtMeanSPE	Year 11	0.00027410	0.00023094	0.00026622	0.00026750	0.00065878	0.00003739	0.00003737	0.00003745
WgtMedAPE	Year 12	0.00110230	0.00116786	0.00110165	0.00096817	0.00200543	0.00013678	0.00013447	0.00019547
WgtMedSPE	Year 12	0.00000122	0.00000136	0.00000121	0.00000094	0.00000402	0.00000002	0.00000002	0.00000004
WgtMeanAPE	Year 12	0.00499408	0.00490305	0.00492521	0.00429266	0.00817796	0.00040129	0.00038931	0.00061461
WgtMeanSPE	Year 12	0.00027956	0.00024122	0.00027083	0.00021298	0.00065412	0.00003921	0.00003919	0.00006717
WgtMedAPE	Year 13	0.00038637	0.00041067	0.00038664	0.00035534	0.00069909	0.00013876	0.00013572	0.00013590
WgtMedSPE	Year 13	0.00000015	0.00000017	0.00000015	0.00000013	0.00000049	0.00000002	0.00000002	0.00000002
WgtMeanAPE	Year 13	0.00184191	0.00183952	0.00181069	0.00160808	0.00291011	0.00040813	0.00039806	0.00042476
WgtMeanSPE	Year 13	0.00009672	0.00009345	0.00009471	0.00008388	0.00012916	0.00006671	0.00006669	0.00003929

Table 11: Mortality Rate Weighted Prediction Error Statistics

7.2 Lapse Rate Prediction Error Statistics and Summary for All Methods

	Year	Poisson w/out Modification	Poisson w/ Modification	Poisson GAM CR Splines	Zero-Inflated Poisson	Negative Binomial	Tweedie GLM	Tweedie GAM CR Splines	Random Forest
MedAPE	Year 3	0.04658	0.04658	0.04721	0.04808	0.05274	0.01902	0.01847	0.01552
MedSPE	Year 3	0.00217	0.00217	0.00223	0.00231	0.00278	0.00036	0.00034	0.00024
MeanAPE	Year 3	0.14743	0.14743	0.14685	0.14775	0.16277	0.03787	0.03782	0.03666
MeanSPE	Year 3	0.28991	0.28993	0.27714	0.27422	0.38705	0.01017	0.01018	0.00950
MedAPE	Year 4	0.04580	0.04580	0.04661	0.04715	0.05036	0.01756	0.01708	0.01510
MedSPE	Year 4	0.00210	0.00210	0.00217	0.00222	0.00254	0.00031	0.00029	0.00023
MeanAPE	Year 4	0.13384	0.13384	0.13394	0.13410	0.14535	0.03626	0.03617	0.03629
MeanSPE	Year 4	0.19017	0.19017	0.18778	0.18645	0.24994	0.01040	0.01041	0.00969
MedAPE	Year 5	0.04342	0.04342	0.04485	0.04481	0.04772	0.01673	0.01629	0.01492
MedSPE	Year 5	0.00189	0.00189	0.00201	0.00201	0.00228	0.00028	0.00027	0.00022
MeanAPE	Year 5	0.12450	0.12450	0.12502	0.12502	0.13540	0.03555	0.03549	0.03671
MeanSPE	Year 5	0.15985	0.15986	0.15524	0.15753	0.21832	0.01038	0.01038	0.00971
MedAPE	Year 6	0.04079	0.03969	0.04249	0.04206	0.04444	0.01582	0.01546	0.01519
MedSPE	Year 6	0.00166	0.00158	0.00181	0.00177	0.00198	0.00025	0.00024	0.00023
MeanAPE	Year 6	0.11951	0.11986	0.12038	0.12103	0.13285	0.03374	0.03369	0.03633
MeanSPE	Year 6	0.16963	0.16742	0.16353	0.17552	0.24913	0.00987	0.00987	0.00928
MedAPE	Year 7	0.04084	0.04088	0.04219	0.04188	0.04380	0.01509	0.01474	0.01541
MedSPE	Year 7	0.00167	0.00167	0.00178	0.00175	0.00192	0.00023	0.00022	0.00024
MeanAPE	Year 7	0.11894	0.11896	0.11983	0.11949	0.13024	0.03312	0.03310	0.03672
MeanSPE	Year 7	0.15870	0.15870	0.15814	0.15996	0.21549	0.00958	0.00958	0.00907
MedAPE	Year 8	0.04094	0.04096	0.04253	0.04165	0.04357	0.01466	0.01445	0.01566
MedSPE	Year 8	0.00168	0.00168	0.00181	0.00173	0.00190	0.00021	0.00021	0.00025
MeanAPE	Year 8	0.11990	0.11991	0.12102	0.12004	0.13095	0.03215	0.03217	0.03612
MeanSPE	Year 8	0.18447	0.18447	0.18450	0.18644	0.24677	0.00920	0.00920	0.00872
MedAPE	Year 9	0.04051	0.04054	0.04248	0.04141	0.04322	0.01424	0.01407	0.01564
MedSPE	Year 9	0.00164	0.00164	0.00180	0.00171	0.00187	0.00020	0.00020	0.00024
MeanAPE	Year 9	0.12203	0.12208	0.12340	0.12211	0.13218	0.03145	0.03152	0.03548
MeanSPE	Year 9	0.19814	0.19815	0.19676	0.19402	0.24082	0.00901	0.00901	0.00853
MedAPE	Year 10	0.04004	0.04008	0.04214	0.04079	0.04253	0.01369	0.01360	0.01521
MedSPE	Year 10	0.00160	0.00161	0.00178	0.00166	0.00181	0.00019	0.00019	0.00023
MeanAPE	Year 10	0.11694	0.11697	0.11839	0.11688	0.12591	0.03133	0.03142	0.03547
MeanSPE	Year 10	0.13220	0.13219	0.13146	0.13013	0.16119	0.00962	0.00962	0.00909
MedAPE	Year 11	0.03781	0.03783	0.03981	0.03866	0.04042	0.01323	0.01318	0.01531
MedSPE	Year 11	0.00143	0.00143	0.00159	0.00149	0.00163	0.00017	0.00017	0.00023
MeanAPE	Year 11	0.10947	0.10949	0.11059	0.10995	0.11844	0.02806	0.02812	0.03296
MeanSPE	Year 11	0.11378	0.11377	0.11226	0.11432	0.14276	0.00753	0.00753	0.00727
MedAPE	Year 12	0.03564	0.03566	0.03754	0.03649	0.03807	0.01256	0.01244	0.01455
MedSPE	Year 12	0.00127	0.00127	0.00141	0.00133	0.00145	0.00016	0.00015	0.00021
MeanAPE	Year 12	0.10617	0.10616	0.10711	0.10684	0.11459	0.02706	0.02710	0.03178
MeanSPE	Year 12	0.10893	0.10891	0.10720	0.11043	0.13820	0.00729	0.00728	0.00701
MedAPE	Year 13	0.01733	0.01733	0.01777	0.01781	0.01849	0.01266	0.01246	0.02213
MedSPE	Year 13	0.00030	0.00030	0.00032	0.00032	0.00034	0.00016	0.00016	0.00049
MeanAPE	Year 13	0.04747	0.04747	0.04787	0.04789	0.05081	0.02541	0.02543	0.03935
MeanSPE	Year 13	0.02172	0.02172	0.02162	0.02249	0.02683	0.00708	0.00708	0.00765

Table 12: Lapse Rate Un-Weighted Prediction Error Statistics

	Year	Poisson w/out Modification	Poisson w/ Modification	Poisson GAM CR Splines	Zero-Inflated Poisson	Negative Binomial	Tweedie GLM	Tweedie GAM CR Splines	Random Forest
WgtMedAPE	Year 3	0.18550	0.18551	0.18461	0.18909	0.20751	0.02258	0.02191	0.01463
WgtMedSPE	Year 3	0.03441	0.03441	0.03408	0.03575	0.04306	0.00051	0.00048	0.00021
WgtMeanAPE	Year 3	1.40972	1.40975	1.32788	1.35914	1.67610	0.03115	0.03100	0.02350
WgtMeanSPE	Year 3	13.48813	13.48859	11.83220	12.03620	20.68636	0.00278	0.00279	0.00234
WgtMedAPE	Year 4	0.16878	0.16877	0.16920	0.17070	0.18290	0.02040	0.01980	0.01387
WgtMedSPE	Year 4	0.02849	0.02848	0.02863	0.02914	0.03345	0.00042	0.00039	0.00019
WgtMeanAPE	Year 4	1.08564	1.08565	1.06084	1.07811	1.29416	0.02800	0.02783	0.02210
WgtMeanSPE	Year 4	7.68571	7.68578	7.45913	7.70461	11.97500	0.00252	0.00253	0.00217
WgtMedAPE	Year 5	0.15558	0.15558	0.15652	0.15763	0.16838	0.01957	0.01911	0.01362
WgtMedSPE	Year 5	0.02420	0.02421	0.02450	0.02485	0.02835	0.00038	0.00037	0.00019
WgtMeanAPE	Year 5	1.00865	1.00868	0.98768	1.00396	1.21872	0.02727	0.02713	0.02237
WgtMeanSPE	Year 5	6.94005	6.94042	6.49356	6.87881	11.06026	0.00256	0.00257	0.00228
WgtMedAPE	Year 6	0.15369	0.15368	0.15518	0.15657	0.16842	0.01915	0.01862	0.01355
WgtMedSPE	Year 6	0.02362	0.02362	0.02408	0.02452	0.02837	0.00037	0.00035	0.00018
WgtMeanAPE	Year 6	1.10875	1.10879	1.07324	1.13138	1.37428	0.02650	0.02640	0.02235
WgtMeanSPE	Year 6	8.68262	8.68317	7.93864	9.23676	14.16662	0.00243	0.00244	0.00219
WgtMedAPE	Year 7	0.14848	0.14852	0.14971	0.15023	0.15998	0.01819	0.01774	0.01325
WgtMedSPE	Year 7	0.02204	0.02206	0.02241	0.02257	0.02559	0.00033	0.00031	0.00018
WgtMeanAPE	Year 7	0.98262	0.98266	0.96874	0.99167	1.17858	0.02589	0.02585	0.02242
WgtMeanSPE	Year 7	7.06375	7.06380	6.91797	7.21246	10.31060	0.00236	0.00237	0.00215
WgtMedAPE	Year 8	0.15042	0.15045	0.15242	0.15090	0.16148	0.01771	0.01730	0.01344
WgtMedSPE	Year 8	0.02263	0.02263	0.02323	0.02277	0.02608	0.00031	0.00030	0.00018
WgtMeanAPE	Year 8	1.08298	1.08300	1.07413	1.09681	1.28520	0.02505	0.02501	0.02178
WgtMeanSPE	Year 8	10.45092	10.45059	10.33251	10.71077	14.42853	0.00222	0.00223	0.00202
WgtMedAPE	Year 9	0.15419	0.15430	0.15704	0.15512	0.16606	0.01709	0.01675	0.01344
WgtMedSPE	Year 9	0.02377	0.02381	0.02466	0.02406	0.02758	0.00029	0.00028	0.00018
WgtMeanAPE	Year 9	1.13697	1.13700	1.13320	1.13367	1.28270	0.02459	0.02461	0.02160
WgtMeanSPE	Year 9	11.34643	11.34499	11.30350	10.87933	13.27683	0.00215	0.00216	0.00196
WgtMedAPE	Year 10	0.14496	0.14507	0.14797	0.14564	0.15513	0.01657	0.01624	0.01331
WgtMedSPE	Year 10	0.02101	0.02105	0.02190	0.02121	0.02407	0.00027	0.00026	0.00018
WgtMeanAPE	Year 10	0.83720	0.83721	0.83584	0.83743	0.94891	0.02377	0.02382	0.02091
WgtMeanSPE	Year 10	5.02064	5.01946	5.02054	4.86949	6.11325	0.00220	0.00222	0.00200
WgtMedAPE	Year 11	0.13736	0.13746	0.14004	0.13864	0.14794	0.01601	0.01565	0.01324
WgtMedSPE	Year 11	0.01887	0.01889	0.01961	0.01922	0.02189	0.00026	0.00024	0.00018
WgtMeanAPE	Year 11	0.76675	0.76674	0.76229	0.77587	0.88094	0.02255	0.02255	0.02006
WgtMeanSPE	Year 11	4.28795	4.28721	4.23261	4.39925	5.82944	0.00181	0.00182	0.00166
WgtMedAPE	Year 12	0.13216	0.13218	0.13473	0.13374	0.14230	0.01553	0.01517	0.01276
WgtMedSPE	Year 12	0.01747	0.01747	0.01815	0.01789	0.02025	0.00024	0.00023	0.00016
WgtMeanAPE	Year 12	0.73792	0.73789	0.73307	0.75075	0.85265	0.02204	0.02199	0.01958
WgtMeanSPE	Year 12	4.15067	4.15066	4.09956	4.40651	6.01267	0.00179	0.00180	0.00164
WgtMedAPE	Year 13	0.04375	0.04376	0.04451	0.04429	0.04697	0.01630	0.01580	0.01849
WgtMedSPE	Year 13	0.00191	0.00191	0.00198	0.00196	0.00221	0.00027	0.00025	0.00034
WgtMeanAPE	Year 13	0.26977	0.26978	0.26901	0.28023	0.31891	0.02430	0.02421	0.02993
WgtMeanSPE	Year 13	0.69218	0.69227	0.69788	0.80137	1.11338	0.00307	0.00307	0.00342

Table 13: Lapse Rate Weighted Prediction Error Statistics

References

Böhning, D., E. Dietz, and P. Schlattmann (1997). Zero-inflated count models and their applications in public health and social science. *Applications of Latent Trait and Latent Class Models in the Social Sciences. Waxman Publishing Co*.

Breiman, L. (2001). Random forests. *Machine Learning 45*, 5–32.

Brockman, M. J. and T. Wright (1992). Statistical motor rating: making effective use of your data. *Journal of the Institute of Actuaries 119*(03), 457–543.

Brown, J. R. and A. Finkelstein (2009). The private market for long-term care insurance in the united states: A review of the evidence. *Journal of Risk and Insurance 76*(1), 5–29.

Gordon K. Smyth, B. J. (2002). Fitting tweedie's compoung poisson model to insurance claims data: Dispersion modelling. *Astin Bulletin 32*, 143–157.

Greene, W. H. (1998). *Econometric Analysis*. Prentice Hall.

Haberman, S. and A. E. Renshaw (1996). Generalized linear models and actuarial science. *The Statistician 45*, 407–436.

Hastie, T., R. Tibshirani, et al. (1986). Generalized additive models. *Statistical science 1*(3), 297–310.

Huberty, C. J. (1989). Problems with stepwise methods - better alternatives. *Advances in Social Science Methodology 1*, 43–70.

Jørgensen, B. and M. C. Paes De Souza (1994). Fitting tweedie's compound poisson model to insurance claims data. *Scandinavian Actuarial Journal 1994*(1), 69–93.

King, G. and L. Zeng (2001). Logistic regression in rare events data. *Political analysis 9*(2), 137–163.

Konetzka, R. T. and Y. Luo (2011). Explaining lapse in long-term care insurance markets. *Health Economics 20*(10), 1169–1183.

Lambert, D. (1992). Zero-inflated poisson regression, with an application to defects in manufacturing. *Technometrics 34*, 1–14.

Lukacs, P. M., K. P. Burnham, and D. R. Anderson (2010). Model selection bias and freedman?s paradox. *Annals of the Institute of Statistical Mathematics 62*(1), 117–125.

Meiners, M. R. and G. R. Trapnell (1984). Long-term care insurance: Premium estimates for prototype policies. *Medical care 22*, 901–911.

Mildenhall, S. J. (1999). A systematic relationship between minimum bias and generalized linear models. In *Proceedings of the Casualty Actuarial Society*, Volume 86, pp. 393–487.

Minami, M., C. E. Lennert-Cody, W. Gao, and M. Roman-Verdesoto (2007). Modeling shark bycatch: the zero-inflated negative binomial regression model with smoothing. *Fisheries Research 84*(2), 210–221.

Mullahy, J. (1997). Heterogeneity, excess zeros, and the structure of count data models. *Journal of Applied Econometrics 12*(3), 337–350.

Murphy, K. P., M. J. Brockman, and P. K W Lee (2000). Using generalized linear models to build dynamic pricing systems for personal lines insurance. *Casualty Actuarial Society Winter Forum Winter 2000*, 107–140.

of Connecticut Goldenson Actuarial Research Center, U. and T. Watson (2012). Long term care predictive modeling. Technical report, University of Connecticut Goldenson Actuarial Research Center and Towers Watson.

Peter Zweifel, W. S. (1998). Long-term care insurance in a two-generation model. *The Journal of Risk and Insurance 65*, 13–32.

Pinquet, J., M. Guillén, and M. Ayuso (2011). Commitment and lapse behavior in long-term insurance: A case study. *Journal of Risk and Insurance 78*(4), 983–1002.

Pritchard, D. J. (2006). Modeling disability in long-term care insurance. *North American Actuarial Journal 10:4*, 48–75.

Renshaw, A. E. (1994). Modelling the claims process in the presence of covariates. *Astin Bulletin 24*(2), 265–285.

Shono, H. (2008). Application of the tweedie distribution to zero-catch data in cpue analysis. *Fisheries Research 93*(1), 154–162.

Smoluk, H. J. (2009). Long-term disability claims rates and the consumption-to-wealth ratio. *The Journal of Risk and Insurance 76*, 109–131.

Strobl, Carolin; Malley, J. T. G. (2009). An introduction to recursive partitioning: Rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychological Methods 14(4)*, 323–348.

Tomas, J. and F. Planchet (2013). Multidimensional smoothing by adaptive local kernel-weighted log-likelihood: Application to long-term care insurance. *Insurance: Mathematics and Economics 52*, 573–589.

Ver Hoef, J. M. and P. L. Boveng (2007). Quasi-poisson vs. negative binomial regression: how should we model overdispersed count data? *Ecology 88*(11), 2766–2772.

Whittingham, M. J., P. A. Stephens, R. B. Bradbury, and R. P. Freckleton (2006). Why do we still use stepwise modelling in ecology and behaviour? *Journal of Animal Ecology 75*(5), 1182–1189.

Wilkinson, L. (1979). Tests of significance in stepwise regression. *Psychological Bulletin 86*(1), 168.