TRENDING IN PROBABILITY OF COLLISION MEASUREMENTS VIA A BAYESIAN ZERO-INFLATED BETA MIXED MODEL

Jonathon Vallejo[(1)], **Matt Hejduk**[(2)], **and James Stamey**[(3)]
[(1)]*a.i. solutions Inc.,10001 Derekwood Lane, Lanham, MD 20706*
[(2)]*Astrorum Consulting LLC, 10006 Willow Bend Drive, Woodway, TX 76712*
[(3)]*Department of Statistics, Baylor University, P.O. Box 97140, Waco, TX 76798*

Abstract: *We investigate the performance of a generalized linear mixed model in predicting the Probabilities of Collision (P_c) for conjunction events. Specifically, we apply this model to the* $\log_{10}$ *transformation of these probabilities and argue that this transformation yields values that can be considered bounded in practice. Additionally, this bounded random variable, after scaling, is zero-inflated. Consequently, we model these values using the zero-inflated Beta distribution, and utilize the Bayesian paradigm and the mixed model framework to borrow information from past and current events. This provides a natural way to model the data and provides a basis for answering questions of interest, such as what is the likelihood of observing a probability of collision equal to the effective value of zero on a subsequent observation.*

Keywords: *Conjunction Analysis, Probability of Collision, Trending, Bayesian, Prediction*

1. Introduction

The problem of deciding whether to maneuver a satellite that is in conjunction with another space object is often not straightforward, and a serious threat involves the deliberation and cooperation of various parties[1]. Quantifying the risk for any such conjunction is generally accomplished through the use of the predicted miss distance at time of closest approach (TCA) and the calculated probability of collision (P_c) at that same time. These measurements are generally taken beginning up to a week before TCA and continue until the opportunity for any active remediation of risk has passed. The calculated P_c values are affected by the uncertainty in the positions of the space objects, an uncertainty that generally decreases as one approaches TCA. This decrease in uncertainty eventually yields a decrease in P_c for most events, although the rate and manner of decrease is different for each.

In a recent paper, we investigated the use of a simple method for determining whether any given event's current P_c value is likely to be the peak P_c value for the entire event.[2] We found that this method was relatively successful and thus that some trend analysis in P_c value is possible. Though the model was competent at peak prediction, it did not have the sophistication to capture the overall behavior of the P_c values in all cases; and that one of the main causes of its difficulties was the large number of P_c values equal to essentially zero. In this paper, we propose a more sophisticated model designed to capture more such nuances of the data and which we ultimately test for prediction of P_c values. We also propose a simple method, which we name the "Look-Up Method," as a baseline against which to evaluate model performance. The Look-Up Method is intended to represent how an analyst might make intuitive judgments regarding the progression of P_c values over time.

2. Methods for Trending in Probability of Collision

2.1. Calculating the Probability of Collision

The basic procedure for calculating the probability of collision between two space objects is to obtain positions and position covariances propagated to TCA, and, using this information, to determine the probability of the two objects' passing within a stated small distance of each other. One generally makes assumptions of Gaussian uncertainty distributions, rectilinear motion in the neighborhood of the conjunction, and uncorrelated covariance matrices in order to reduce the problem to a more manageable two-dimensional formulation.[3] Alfano[4] and others have noted that this probability measure tends to follow a canonical progression of measured increase and then untimately rapid decrease in *Pc* as uncertainty in the satellite's positional data decreases, a trend that was examined in our previous paper.

We seek to model the path of the P_c values for any event as a set of observations for a subject in a mixed model. In contrast to the previous paper in which our main focus was peak prediction, here our ultimate goal is to predict the value of the next observed P_c measurement, although it is hoped that a robust peak prediction capability may flow from this. To verify the accuracy of our model, we test it against both a large archive of past conjunction information and the "Look Up" approach mentioned previously.

2.2. Data

When modeling the trend in P_c values, one is generally concerned with changes in order of magnitude; thus one generally models $\log_{10} P_c$ as opposed to the observed P_c values. This poses an interesting statistical question, namely the distribution of $\log_{10} P_c$ values. Distribution selection is more obvious for the P_c values, as they are bounded between 0 and 1; thus a statistical modeler generally would choose a beta distribution to model these values (although there are a few other less commonly used distributions, such as the simplex distribution, that could be deployed). Theoretically, there is no lower bound on $\log_{10} P_c$ values, as P_c values can be arbitrarily close to zero. Operationally, however, one often considers P_c values below 1E-10 to be effectively 0. To account for this effective terminus in P_c, in our previous efforts we "floored" the $\log_{10} P_c$ values at -10, so that the large number of small $\log_{10} P_c$ values did not overly influence the model. This simplification allows one to focus inference on the operationally relevant $\log_{10} P_c$ values, which tend to be around -5 and greater. We follow suit here, flooring all $\log_{10} P_c$ values at -10. Therefore, even in modeling the $\log_{10} P_c$ values, we have bounded data (between -10 and 0) that are "inflated," meaning that there are a notable number of points on the boundary of the defined interval–points that may be just beyond what is representable by a hypothesized distribution over that interval. In this case, the data are -10-inflated, but when the variable is rescaled to fit the Beta distribution, the data are zero-inflated. Such a situation can be accommodated by the zero-inflated beta distribution; the form of the equation and basic definitions are given below, with more elaborate discussion of the parameters and their meaning provided in subsequent sections:

$$f(y|\mu,\phi,p) = (1-p)\frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1-\mu)\phi)} y^{\mu\phi-1}(1-y)^{(1-\mu)\phi-1} I_{(0,1)}(y) + pI_{[0]}(y). \qquad (1)$$

Here, $I_A(\cdot)$ is the indicator function, so that the first term corresponds to values of y falling between 0 and 1 (or $\log P_c$ values falling between -10 and 0) and the second term corresponds to values equal to 0 (or $\log P_c$ values equal to -10). The parameter μ is the mean of the Beta distribution, which will be modeled in the Generalized Linear Model (GLM) framework, and the parameter ϕ is the corresponding dispersion parameter, which is a measure of variability. The parameter p can be interpreted to be the probability that one observes a 0 (that is, a $\log P_c$ of -10).

Additionally, as noted previously, the P_c values for each event tend ultimately to decrease with time but at a different rate in each conjunction. This suggests approaching the problem within a mixed model framework, allowing random terms for each conjunction. This is a natural approach to take, as the data are longitudinal in nature: one observes an overall trend in time; yet each subject (in this case, each conjunction) deviates somewhat from this trend, and observations within a subject are correlated with each other. In Figure 1, we visualize the longitudinal nature of the data. We plot the $\log_{10} P_c$ values of ten events over time, with each events' values connected by a line; and we also plot these values versus the ratio of combined covariance radius to miss distance. This plot exposes the canonical trend in Pc development: as the event moves closer to TCA, the covariance shrinks, bringing this ratio slowly to a peak and then a marked drop-off.

It is clear that the trend is more pronounced for the ratio of covariance radius to miss distance. Unfortunately, as was discussed in our previous paper, this value is not monotonically increasing or decreasing with time (due to unpredictable changes in the covariance size and the estimate of the mean miss distance between the two satellites as the event develops); so despite its closer linkage to the root phenomenology of the situation, it is actually a less desirable independent variable for performing trending and prediction. Therefore, as previously, model construction for P_c trending and prediction will use time to TCA as the independent variable.

We can visualize the trend of the $\log_{10} P_c$ values over time by considering a two-dimensional histogram, shown in Figure 2. Recall that we have replaced all $\log_{10} P_c$ values below -10 with -10. This figure indicates that the probability of observing a P_c value of 1E-10 or lower increases as one approaches TCA. In fact, at 2 days to TCA, about 40% of events observed have a P_c of 1E-10 or lower. At 7 days until TCA, the most observed value is about -5, which becomes less frequent over time, as more events observe a $\log_{10} P_c$ of -10. Interestingly, -5 seems to be the most likely value when one does *not* observe a -10, regardless of the time. We can use this information to construct prior information for the model in Equation (1), as the increase of observed -10 values gives us an idea of how p behaves over time, and the observed mode of -5 of the $\log_{10} P_c$ values above -10 gives us some information about the mean of the Beta distribution.

3. The Bayesian Beta Regression Model

To model a beta-distribution random variable with reference to a covariate (such as time), it is best to use a GLM. Although GLM's for all other members of the exponential family (Normal, Gamma, etc.) have been developed since 1972[5], the GLM for the beta distribution is relatively new, being introduced in 2001[6]. The reason for this late development is due to the fact that one must reparametrize the beta distribution in order to model the mean adequately, an expansion that was not explored until recently. We provide the derivation here for completeness. The probability density function (pdf) of a random variable X with a beta distribution is generally given as

$$f(x|\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1},$$

where $\Gamma(\cdot)$ is the gamma function. The mean of this distriubtion is $E(X) = \frac{\alpha}{\alpha+\beta}$. GLM's are generally specified by setting some function $g(\mu)$ of the mean equal to a linear combination of covariates. For instance, logistic regression uses the logit link $g(\mu) = \log(\frac{\mu}{1-\mu})$, which is then set equal to a linear combination of covariates, e.g. $\beta_0 + \beta_1 X$, where X is a covariate, such as time. However, as the beta distribution is specified, it is unclear how to model the mean. To facilitate direct modeling of the mean, let $\mu = \frac{\alpha}{\alpha+\beta}$ and $\phi = \alpha+\beta$. Then we can rewrite the beta pdf as

$$f(x|\mu,\phi) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1-\mu)\phi)} x^{\mu\phi-1}(1-x)^{(1-\mu)\phi-1}.$$

Now we may model the mean μ directly. For instance, we may choose the logit link and model

$$\log\left(\frac{\mu}{1-\mu}\right) = \beta_0 + \beta_1 x_{ij} + \ldots + \beta_p x_{ij}^p,$$

so that the log-odds of the mean has a linear relationship to X. Various link functions are possible, such as the probit link, the complementary log-log link, and the log link. Our simulations have shown that there is no significant advantage in choosing one over the other, so we proceed with the logit link, as it is comparatively easy to interpret.

Recall that for each conjunction, one observes a different progression of P_c values. Sometimes the P_c values drop off quickly well prior to TCA, other times they drop off much nearer TCA, and sometimes not at all. To model such a behavior, we include a random intercept for each event as follows. Let μ_{ij} be the mean of the j^{th} P_c value in the i^{th} event, scaled to be between 0 and 1. Since we have $\log_{10} P_c$ values bounded between -10 and 0, a suitable transformation is $\mu_{ij} = E(Y_{ij})/10+1$, where Y_{ij} is the $\log_{10} P_c$ value of the j^{th} P_c value in the i^{th} event. We may consider the model

$$\log\left(\frac{\mu_{ij}}{1-\mu_{ij}}\right) = \beta_0 + \beta_1 t_{ij} + \ldots + \beta_p t_{ij}^p + b_i,$$

where b_i is the random intercept for the i^{th} event, and t_{ij} is the time until TCA of the j^{th} P_c value within the same event. One may additionally consider a random slope or other random effects for higher order terms.

Recall that in Equation (1) we also introduced the parameter p. This parameter controls what percentage of the time we observe a zero. In our case, since about a third of our data are zeros, p might be close to 1/3. However, we also know that the closer an event approaches TCA, the more likely one is to observe a P_c value that is 0. As a result, we can also let p depend on our covariate. Since it is a probability, this parameter is also bounded between 0 and 1, so we can also use a logit link function here (or any of the other aforementioned link functions). Additionally, we may consider a random term for this model for each event, as the probability of observing a zero is higher for some events than others. Thus, we may consider a regression such as

$$\log\left(\frac{p}{1-p}\right) = \alpha_0 + \alpha_1 t_{ij} + ... + \alpha_p t_{ij}^p + a_i,$$

which is similar to the regression for μ above. Again, if we wish we can consider other random terms, *i.e.*, a random slope.

3.1. Model Selection

Given below are some selected results from an exploratory model selection. To evaluate the relative merits of different levels of model complexity, we use the penalized deviance construct[7], where lower values indicate a better fit. Specifically, $D(\theta)$ is defined as the "Bayesian Deviance," with form

$$D(\theta) = -2\log p(y|\theta) + 2\log f(y), \tag{2}$$

where $p(y|\theta)$ is the likelihood of y given θ and $f(y)$ is the saturated model, where $f(y) = p\{y|E(Y) = y\}$. We can rewrite $D(\theta)$ as

$$D(\theta) = -2\left(\log p(y|\theta) - \log f(y)\right), \tag{3}$$

which shows that $D(\theta)$ is -2 times the difference between the fitted model and the saturated model. Put simply, $D(\theta)$ measures how well a model has fit the data relative to a model that fits the data perfectly. We estimate $D(\theta)$ with $\overline{D(\theta)}$, which can be written as

$$\overline{D(\theta)} = D(\bar{\theta}) + p_D, \tag{4}$$

where $p_D = \overline{D(\theta)} - D(\bar{\theta})$. The estimate $\overline{D(\theta)}$ is known as the *penalized deviance*, as it is computed as the sum of $D(\bar{\theta})$, the mean deviance, and p_D, the penalty term. The term $D(\bar{\theta})$ measures how well a model fits the data, with lower values indicating better fit, and the term p_D penalizes this fit for more parameters, where higher values indicates a larger penalty. The penalty term p_D is also known as the *effective number of parameters*, so that one may interpret this term as an estimate of how many parameters the model is actually estimating in order to describe the data. This is to account for the expected outcome of models with more parameters fitting the data better and thus potentially over-fitting the data.

Given in Table 1 is the calculated mean deviance, penalty, and consequent penalized deviance for various models. This is provided in order to justify the selection of our final model, as we chose the model with the lowest penalized deviance. The variables Y_c and Y_d represent the continuous and discrete parts of the model given in Equation (1), respectively. That is, Y_c are the values produced by the beta distribution, and Y_d are the 0-1 variables that either indicate a zero (1) or a continuous variable (0). All added complexities are in addition to the baseline linear model specified in equations (5-12) below. Let Y_{ij} be the j^{th} scaled $\log_{10} P_c$ value of the i^{th} event. Also, let t_{ij} be the corresponding time until TCA (in days).

$$
\begin{align}
Y_{ij} &\sim f(y_{ij}|\mu_{ij}, \phi_{ij}, p) \tag{5}\\
\log\left(\frac{\mu_{ij}}{1-\mu_{ij}}\right) &= \beta_0 + \beta_1 t_{ij} + b_i \tag{6}\\
b_i &\sim N(0, \tau_b) \tag{7}\\
\tau_b &\sim Gamma(0.001, 0.001) \tag{8}\\
\log\left(\frac{p_{ij}}{1-p_{ij}}\right) &= \alpha_0 + \alpha_1 t_{ij} + a_i \tag{9}\\
a_i &\sim N(0, \tau_a) \tag{10}\\
\tau_a &\sim Gamma(0.001, 0.001), \tag{11}\\
\beta_k, \alpha_k &\sim Normal(0, 1), \quad k = 0, 1, 2. \tag{12}
\end{align}
$$

Table 1 shows that adding a random slope to either Y_c or Y_d did not produce a better fit, nor did specifying a correlation between the random effects:

Model	Mean Deviance	Penalty	Pen. Deviance
Linear	-17.23	74.93	57.7
Quad Term for Y_c	-23.02	77.32	54.31
Quad Term for Y_d	-26.78	76.45	49.67
Quad Term for Y_c and Y_d	-32.52	79.23	46.71
QuadTerm for both, RanSlope for Y_c	-31.12	81.07	49.95
QuadTerm for both, RanSlope for Y_d	-36.91	85.54	47.63
Cubic Term for Y_c	-31.89	81.1	49.21
Quadratic, linear for phi	-27.76	80.87	53.11

Table 1. Model Selection Output

Based on these results, we propose the following final form of the model:

$$
\begin{aligned}
Y_{ij} &\sim f(y_{ij}|\mu_{ij},\phi_{ij},p) && (13)\\
\log\left(\frac{\mu_{ij}}{1-\mu_{ij}}\right) &= \beta_0+\beta_1 t_{ij}+\beta_2 t_{ij}^2+b_i && (14)\\
b_i &\sim N(0,\tau_b) && (15)\\
\tau_b &\sim Gamma(0.001,0.001) && (16)\\
\log\left(\frac{p_{ij}}{1-p_{ij}}\right) &= \alpha_0+\alpha_1 t_{ij}+\alpha_2 t_{ij}^2+a_i && (17)\\
a_i &\sim N(0,\tau_a) && (18)\\
\tau_a &\sim Gamma(0.001,0.001), && (19)\\
\beta_k,\alpha_k &\sim Normal(0,1) \quad k=0,1,2. && (20)
\end{aligned}
$$

3.2. Inference

3.2.1. Inference for a new data point

Let $\theta=(\alpha_0,\alpha_1,\alpha_2,\beta_0,\beta_1,\beta_2,\phi,\tau_a,\tau_b)^{'}$. Suppose one has data $\mathbf{y}$ and one wishes to predict $y^*=\log P_c$ at a new data point at time t^*. One can make an inference on y^* by using the predictive distribution

$$g(y^*|x^*,\mathbf{y})=\int_{\Theta} g(y^*|\theta,x^*,\mathbf{y})\pi(\theta|\mathbf{y})d\theta,$$

which can be estimated by using the posterior samples of a Markov Chain Monte Carlo (MCMC) simulation. t^* can be a future time for which estimating the P_c would be desirable, or it can be set to the time of the next received P_c value in order to evaluate the model's predictive power through residual analysis. In conducting evaluations within the latter paradigm, we construct a 95% credible interval for y^* and check to see if the actual value of y is contained in the interval. The percentage of credible intervals which contain the true y value is known as coverage. If the coverage is close to the nominal value of 95%, we can assume that these predictions are reliable. These predictions are made starting with the second P_c observation for each event, as was done in our previously-referenced efforts.

3.2.2. Issues of Identifiability

The model proposed in equations (13)-(20) has a total of 7 parameters and 2 random effects, which suggests one must estimate a total of 9 quantities in order to make inferences and hence predictions. However, this issue can be ameliorated by using informative priors in a Bayesian framework. To acquire these informative priors, we run the proposed model on a training dataset of a large number of events, which are not used for model evaluation. We use the posterior distribution of the parameters as informative prior distributions by matching sample moments of the posterior samples with its prior distribution family. We do this for all of the population-level parameters, which are *phi*, α_k, and β_k for $k=0,1,2$. Then we are left with two parameters to estimate, the random intercepts a_i and b_i. Because we make predictions beginning with the second observation, these parameters are identifiable when making inference on a single event.

Motivated by the large number of events in our testing data set, we investigated whether prediction could be improved by making inferences on more than one event at once. In order to test this, we followed the mean squared prediction error (MSPE) when making predictions on one event, 5 events, 10 events, and 25 events. Including more events did not improve the MSPE, likely due to the fact that, in reference to a single event, other events contribute only to the population-level parameters, which are already well known due to the informative prior distributions. Thus, it is quite adequate to make predictions on a single event at a time.

4. The Look-Up Method

4.1. Intuition

As a basis of comparison to our Beta regression model, we propose a simple alternative model, which we call the "Look-Up Method." The Look-Up Method is based on the common expectation that, when one observes an event with relatively high P_c values, one can suppose this event to behave similarly to other events with other similarly high values. In order to formalize this intuitive approach into an explicit model, we need to establish how high "relatively high" is. A natural way to quantify this notion is in terms of quantiles. That is, we expect events with $\log P_c$ values in the qth quantile to behave similarly to other events with $\log P_c$ values in the qth quantile. The method we describe below is similar to methods involving "look-up tables," where one has quantiles for various scenarios and can look up the probability of an event within the table.

4.2. Method

Let x and y be the time and P_c value from the most recent observation. Furthermore, let x^{new} and y^{new} represent the time of prediction and the true P_c value at this time. The algorithm for the Look-Up Method is as follows

Algorithm

1. Choose an historical data set $\mathbf{Y}_h$ such that the events contained in $\mathbf{Y}$ are believed to behave similarly to the event of interest.
2. Choose a window w.
3. Calculate the empirical CDF $\hat{F}(y)$ of the $\log P_c$ values in the interval $(x-w, x+w)$.
4. Calculate the sample quantile $\hat{q}$ of y
5. Calculate the empirical CDF $\hat{F}(y)$ of the $\log P_c$ values in the interval $(x^{new}-w, x^{new}+w)$.
6. Predict y^{new} to be $\hat{q}(x^{new})$

We find that $w = 0.5$ days to be a reasonable window length, as it is a serviceable adjudication between compressing the time-span enough to contain data for only a single developmental stage yet be broad enough to allow a reasonable amount of sampling. This length depends on how much prior data is at hand, as w may need to be larger for datasets which are less dense at the time of interest. Note that this method only predicts an estimate of y^{new} and does *not* by default generate a prediction interval or any other confidence information.

The method above is simple: find the sample quantile of the observed P_c value at the given time, and assume that future P_c values will be at the exact same quantile. While simplicity has its virtues, in statistical analysis simple models often discard potentially useful information. For instance, the predictions are made based only on the sample quantile of the most recent observation and make no use of previous observations within the given event. However, one could argue that the most recent observation is the most (or only) meaningful observation, and thus one should make inferences based on this value rather than more immediate past values.

4.3. Prediction Intervals

As noted above, the Look-Up Method does not automatically generate prediction intervals; this is a consequence of the method making no distributional assumptions. However, one may still construct prediction intervals via bootstrapping or cross-validation[8]. These methods have been formally compared[9], and the results from this investigation indicate that estimators based on Repeated Cross Validation (RCV) tend to outperform other estimators (*e.g.*, bootstrap estimators). As a result, we implement RCV to generate prediction intervals. The method was initially proposed by Burman (1989)[10], which describes the algorithm in detail.

5. Measures of an Effective Model

In this section we discuss the manner in which we will compare our two models. We focus on model fit and decision-making performance.

5.1. Model Fit

The main concern in building predictive models is fitting the data well enough to predict new observations accurately. In order to quantify model fit performance, we check the bias, prediction errors, and upper bounds of the proposed models. Specifically, we would like our models to be unbiased, so that the prediction errors are centered at zero. Secondly, we check to see if the prediction intervals are bigger or smaller for different times, predicted values, and prediction intervals. Lastly, we check to see that our upper bounds have the correct coverage.

5.1.1. An Aside: Coverage from an Initial Simulation

In an initial exploratory simulation, we found that 97.5% prediction intervals constructed in the Beta model had 86% coverage. Though coverage with real data is often lower than the nominal rate, this coverage level is low enough to bring into question the model's operational utility. In investigating this phenomenon more deeply, we found that dividing the evaluation data into three parts, a high-, medium-, and low-risk group, ameliorated the issue of low coverage. Specifically, if an event had a high (above -4) $\log P_c$ value by 3 days' time to TCA , we called it high-risk. If an event had a medium (between -7 and -4) $\log P_c$ value by 3 days' time to TCA, we called it medium-risk. If an event had a low (below -7) $\log P_c$ value by 3 days' time to TCA, we called it low-risk. We shall refer to the high-, medium-, and low-risk groups as Red, Yellow, and Green hereafter.

Incidentally, the fact that our model performed well when the data were separated into different risk groups supports the notion that the $\log P_c$ value behaves differently depending on the quantile it inhabits. In terms of the Beta regression model, this successfuul stratification of data implies that the population-level trend is different for these different risk groups, which counsels that they ought to be modeled separately. In future work, we plan to explore more rigorously exactly how one should define these different risk groups. For the simulations presented in this paper, these definitions worked well and possess the additional advantage of aligning closely with thresholds presently in use operationally for categorizing conjunction event severity.

5.2. Decision-Making Efficacy

5.2.1. Framework

In order to assess our models within the framework of making decisions about whether to continue active monitoring of a conjunction event, we implement a simple decision-making paradigm and study its properties in both models. Because the typical period for conjunction assessment operational decision-making occurs 2-3 days; time to TCA, we focus on this region of the data. Specifically, we make predictions at 2 days' time to TCA and take a decision based on this prediction. Let $\hat{y}_2$ be an estimate of the $\log P_c$ predicted to occur at 2 days' time to TCA. We will make the decision that the $\log P_c$ values will remain above the threshold θ after 2 days; time to TCA if

$$\tilde{y}_2 > \theta \tag{21}$$

and will make the decision that the $\log P_c$ values will fall below the threshold θ otherwise. To couch this in the hypothesis testing framework, we write

$$H_0 : \tilde{y}_2 < \theta \quad vs. \quad H_1 : \tilde{y}_2 > \theta, \tag{22}$$

so that rejecting H_0 is synonymous with claiming the $\log P_c$ will remain high. In our simulations, we set $\theta = -5$ for the Red group and $\theta = -7$ for the Yellow group. Note that while -7 is the lower bound for being in the Yellow group at 3 days' time to TCA, -5 is below -4, the corresponding lower bound for the Red group. A lower threshold was chosen as these events are generally of much higher concern, thus one prefers an extra order magnitude of certainty before claiming the event is at a lower risk level. In order to explore this trade-off fully, we examined various quantiles of the distribution of $\tilde{y}_2$, which we describe below.

5.2.2. Type I and Type II Errors

As with most decision-making frameworks, our framework can admit Type I and Type II errors. The hypothesis in (22) is framed in terms of the event of a P_c value remaining high, as this is the event we are most concerned with. A Type I error in this case is the incorrect assertion of a high P_c value (*i.e.*, a false alarm), and a Type II error is a the more worrisome incorrect prediction of a low value (*i.e.*, a missed detection). Thus, while we may find it acceptable to trigger an alarm when the $\log P_c$ value has actually dropped off, it is almost *never* acceptable to miss detecting a high $\log P_c$ value. Of course, we can make our system as powerful against missed detections as we want, with the trade off of triggering more false alarms. It is worth noting that a false alarm for a high value is the same thing as missed detection for a value which has dropped off. Ideally, we would like to have an alarm that detects high values *and* low values with a high degree of accuracy. However, since we are more concerned with high values, we seek to quantify how often, if ever, can we detect these low values while still maintaining the high accuracy needed for detecting the high values.

6. Numerical Results

6.1. Data

The datasets previously mentioned for tuning (*i.e.*, setting the parameters for the informative prior distributions) and testing the model was taken from the NASA Conjunction Analysis and Risk Assessment historical Conjunction Message database. For the Yellow group, five hundred events' worth of data from calendar year 2013 was used for model tuning (the "training" dataset), and the tuned model was evaluated against approximately 2000 events from 2014 (the "validation" dataset); so there was no overlap in terms of time-period or actual data between the two datasets. For the Red group, 82 events were used for training and 70 were used for testing (this data set is far smaller, as these kinds of events are more rare). Data were taken from conjunctions against primaries in the orbital region defined by a perigee height between 500 and 750 km and an eccentricity less than 0.25. As mentioned perviously, data flooring at a $\log_{10} P_c$ value of -10 was performed on the dataset. To qualify for use in tuning or evaluation, an event must have had at least two CDMs with a $\log_{10} P_c$ greater than -10.

6.2. Simulation Setup

As described earlier, to train our Beta distribution model, we perform a Bayesian analysis on the training data using non-informative priors. We determine the distribution parameters for the informative priors used in the test data by matching the first and second moments of the observed distributions to the hypothesized prior distributions. These informative priors are then used in the predictive simulations with the validation dataset. All MCMC inference is conducted using the JAGS ("Just Another Gibbs Sampler")[11] software suite.

The simulation procedure for a given event is as follows. We attempt to make predictions for the peak y value only after the second received P_c calculation. We are interested in estimating the next $\log P_c$ value, which we predict by using the time at which it was observed. The predicted value is taken to be the mode of the posterior predictive distribution. In this context, it is important to use the posterior mode as opposed to the posterior mean, as the posterior predictive distribution is generally bimodal, with some mass at -10 and the remaining density between -10 and 0, inducing another peak. Thus, we choose the "most likely value" as opposed to the mean. The predictions using the Look-Up Method are performed in the straightforward manner described in section 4.2.

To further assess model fit, we also track a two-sided 95% credible interval for each prediction. We utilize the upper bound from the credible set for checking coverage. This is also done for the Look-Up Method, though here the interval is a confidence interval and is calculated using repeated cross-validation. In addition to coverage, we are also interested in how many of these upper bounds are low enough to be "useful". That is, we would like to know how many of these lower bounds are lower than the lower threshold of the Yellow and Red groups.

6.3. Results

Below we discuss model fit for a simulation run on the Red data set. The decision-making efficacy is discussed relative to the Yellow data set.

6.3.1. Coverage

The upper bounds we constructed for both the Beta and Look-Up models achieved 97.6% and 97.4% coverage, respectively. The fact that these both achieve the nominal coverage of 97.5% suggests that both models have been specified properly and are reliable in creating prediction intervals. Figure 3 shows the relationship of these upper bounds with time to TCA. Notice that because the upper bounds were found via cross-validation for the Look-Up Method, they do not always conform to the distribution of the data and can yield physically-impossible positive values. Also, the Look-Up Method method is better at detecting a $\log P_c$ of -10, and thus one observes many upper bounds at -7.2 (the 97.5% upper bound on errors was 2.8).

Figure 4 plots the probability densities of these upper bounds on top of each other. Again, we see the positive upper bounds from the Look-Up Method, as well as a large density of lower bounds between -8 and -6. These figures highlight deficiencies in both models: in the Look-Up Method some upper bounds are above 0, and in the Beta Regression method not enough upper bounds are in the -10 to -7 (Green) region.

6.3.2. Prediction Errors

To get a better understanding of the model fit, we can inspect the prediction errors. Figure 5 shows the prediction errors plotted against time. In both cases, they are unbiased (have essentially zero means) and may increase slightly in variability near TCA. Again, we can see that the Look-Up Method is competent at detecting $\log P_c$ values, as evidenced by the large number of residuals equal to 0. This is the case of a -10 being predicted by a -10.

We also examine predictive performance as a function of prediction interval. Figure 6 shows the residuals against the time until the predicted value. Most of the predictions are made within a day, suggesting that most event updates occur in this time frame. There seems to be no real trend, except possibly a slight decrease in variability as the time between the current observation and the prediction goes to 0, which is to be expected.

Next, we explore the relationship between the residuals and the actual $\log P_c$ value. It is apparent from Figure 7 that both models are somewhat biased. This bias is a result of the nature of the data and the models: it is impossible to observe a value below -10, and in both models is impossible to predict a value below -10. Thus, when one errs in these cases, one always errs high. The residuals from the Look-Up Method are somewhat better behaved in this setting, leveling out to a mean of zero around an actual $\log P_c$ value of about -5 or -4. The Beta regression residuals do not level off in this way until an actual $\log P_c$ value of about -3 or -2.

Figures 8 and 9 show a few more diagnostics for the residuals. It is worth noting that in Figure 9, there are a large number of prediction errors that are 0 for the Look-Up Method and that are large negative prediction errors in the Beta regression model. This alignment again has to do with the Look-Up Method correctly identifying the $\log P_c$ values of -10.

6.3.3. Making Decisions

Figure 10 is a Receiver Operating Characteristic (ROC)-like curve demonstrating the properties of the alarm system for detecting a high value in the Yellow group. Here, our threshold is $\theta = -7$. Notice that while the Beta model has a smaller distance between the correct detection of a high value and a false alarm, the Look-Up Method is less able to detect high events correctly. In practice, one would like to have an extremely high (say, 90-99%) rate of correctly detecting high events. Though the Look-Up Method is a potentially more useful alarm system at lower rates of correctly detecting a high value (it triggers fewer false alarms relative to the Beta model), its inability to correctly detect high values makes it virtually unusable in practice.

To further visualize this phenomenon, consider Figure 11. This is an ROC curve which plots the results of Figure 10 in a classic true positive vs. false alarm setting. Again, here a true positive is a correct detection of a high event. As we are interested in 90-99% correct detection of high events, we consider the upper right-hand portion of the graph. Notice that the Beta model has numerous points in this region (with false alarm rates ranging from about 70-100%), while the Look-Up Method only has one point: 100% true positives with 100% false alarms. That is, one triggers an alarm for every event. The Beta model has many points in this region, and they are all above $y = x$, indicating that the model correctly identifies a high value more often than triggering a false alarm.

It should be noted that in Figure 11 that the lower left-hand portion of the graph indicates that, because of the relative performance of the true positive and false alarm rate, the Look-Up Method is a better alarm system in this region. This improved performance corresponds to the 0th to 20th percentile region from Figure 10, which have a higher rate of correctly detecting a high value. Although we are not interested in this area of the graph as the probability of detection is too low to be useful, it is worth noting that this again shows that the Look-Up Method has some advantages relative to the Beta model, indicating that possibly an even better model could be constructed.

Lastly, one may wonder why the Look-Up Method is unable to correctly detect high values at a higher rate than shown. This is likely due to the fact that the method is ad-hoc, and thus there is no guarantee that the model will produce meaningful predictions or prediction intervals.

7. Conclusion and Future Work

In this paper, we have presented two models for predicting future $\log P_c$ values in conjunction assessments. The Look-Up Method was proposed as a reference method, and clearly has many desirable properties. The method is fast to compute an relatively easy to implement if one has sufficient data. One of the major drawbacks of this method is that it is ad-hoc, so it may be difficult to justify theoretically. Consequently, though this method had better prediction errors and other useful properties, the model was not able to be used in a decision-making context in a meaningful way. The Look-Up Method can be recommended as a quick way to make relatively accurate predictions, but one should be cautious in using it to make decisions without further development.

The Beta Regression method was proposed to handle the bounded nature of the data. Including a model for the excess -10 values proved useful, though the model often is conservative in predicting a -10 value. Thus, more exploration needs to be done to borrow strength across these two models, in the hopes of achieving tighter prediction intervals and ultimately better decision making. Though the model is inferior in some ways to the Look-Up Method, it ultimately performs better in a decision-making context and therefore can be operationally useful. Its theoretical underpinning allows one to make probabilistic statements about future events, which have been validated through simulation. Furthermore, the theoretical underpinning allows one to construct meaningful prediction intervals of any size, resulting in an ability to make decisions at various true positive and false alarm levels. This model would thus give operators the ability to forecast accurately and with confidence, especially knowing what the characteristics of the alarm system are (shown through simulation), although it would need to be determined whether a classified with such a large false-alarm rate would actually be operationally useful.

Our future work is focused on exploring the findings from these simulations more in order to construct a more powerful predictive model. This paper has shown that the quantile of the current observation is a useful piece of information, and we believe including it as a covariate or threshold mechanism could lead to better results. We also continue to explore methods for longitudinal data, thus deploying more powerful ways of borrowing information across time and events.

8. References

[1] Frigm, R. C., Levi, J. A., and Mantziaras, D. C. "Assessment, planning, and execution considerations for conjunction risk assessment and mitigation operations." "Proceedings of SpaceOps 2010 Conference: Delivering on the Dream, Huntsville, Alabama," pp. 25–30. 2010.

[2] Vallejo, J., Hejduk, M., and Stamey, J. "Trending in Probability of Collision Measurements." Some journal, 2015.

[3] Chan, F. K. Spacecraft collision probability. Aerospace Press El Segundo, CA, 2008.

[4] Alfano, S. "Relating Position Uncertainty to Maximum Conjunction Probability©." 2005.

[5] Nelder, J. A. and Baker, R. "Generalized linear models." Encyclopedia of Statistical Sciences, 1972.

[6] Paolino, P. "Maximum likelihood estimation of models with beta-distributed dependent variables." Political Analysis, Vol. 9, No. 4, pp. 325–346, 2001.

[7] Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van Der Linde, A. "Bayesian measures of model complexity and fit." Journal of the Royal Statistical Society: Series B (Statistical Methodology), Vol. 64, No. 4, pp. 583–639, 2002.

[8] Stine, R. A. "Bootstrap prediction intervals for regression." Journal of the American Statistical Association, Vol. 80, No. 392, pp. 1026–1031, 1985.

[9] Borra, S. and Di Ciaccio, A. “Measuring the prediction error. A comparison of cross-validation, bootstrap and covariance penalty methods.” Computational statistics & data analysis, Vol. 54, No. 12, pp. 2976–2989, 2010.

[10] Burman, P. “A comparative study of ordinary cross-validation, v-fold cross-validation and the repeated learning-testing methods.” Biometrika, Vol. 76, No. 3, pp. 503–514, 1989.

[11] Plummer, M. et al. “JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling.” “Proceedings of the 3rd international workshop on distributed statistical computing,” Vol. 124, p. 125. Vienna, 2003.

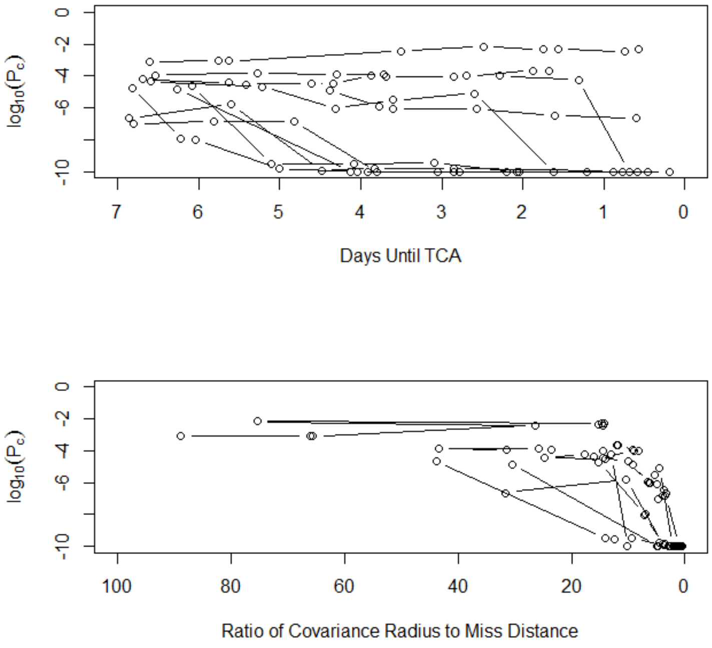

Figure 1.

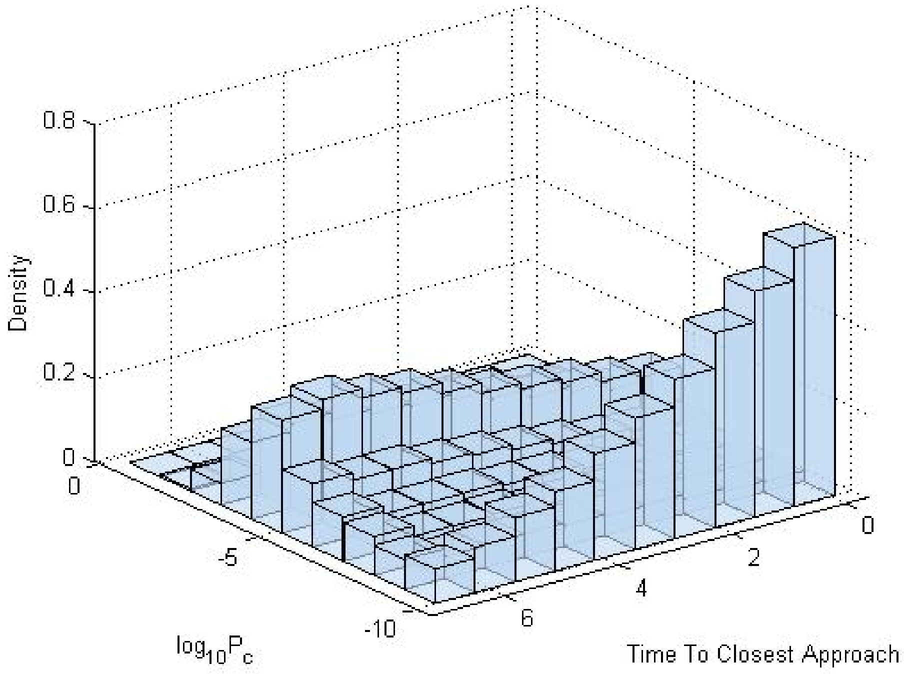

Figure 2.

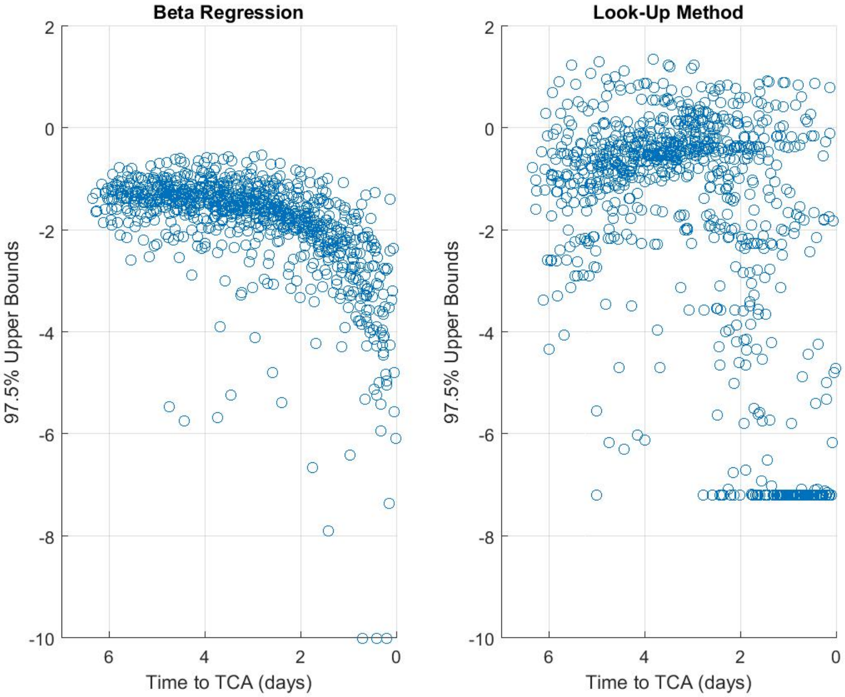

Figure 3.

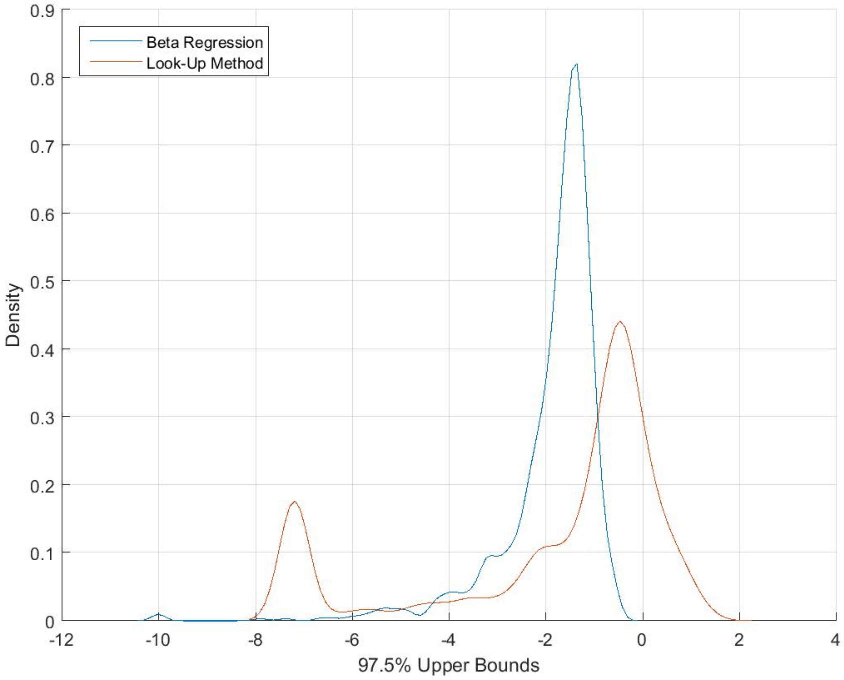

Figure 4.

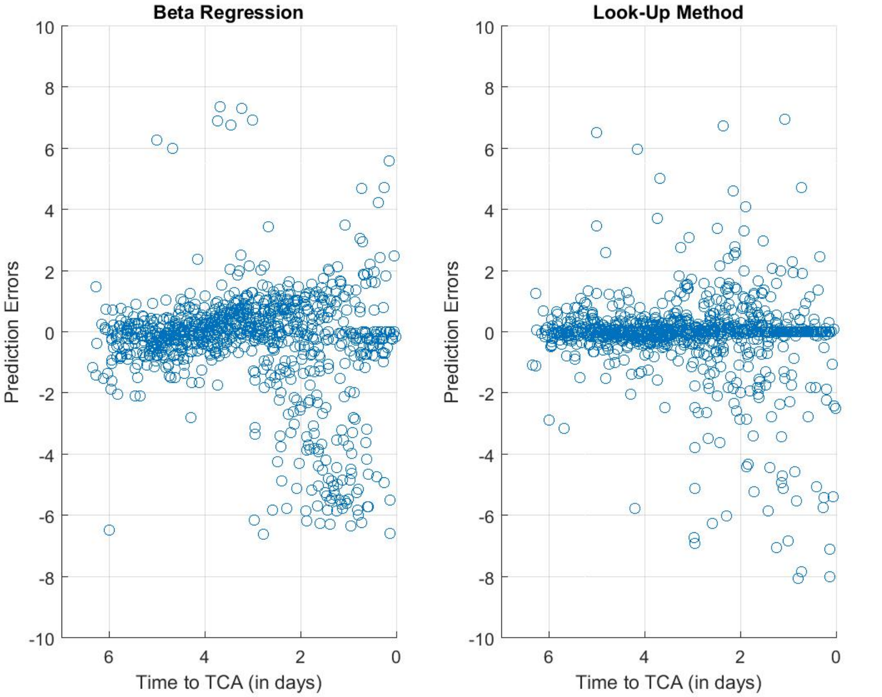

Figure 5.

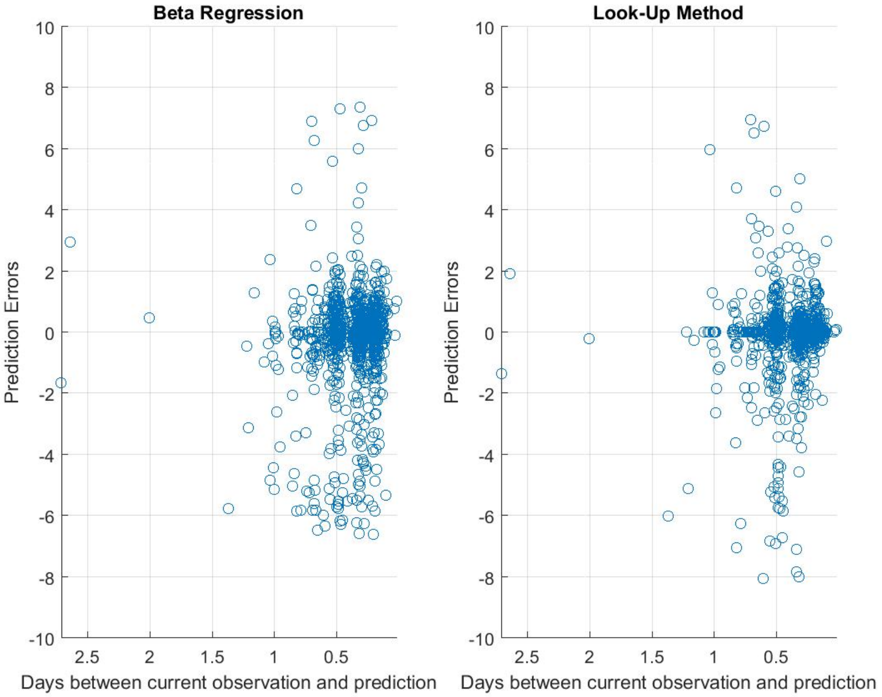

Figure 6.

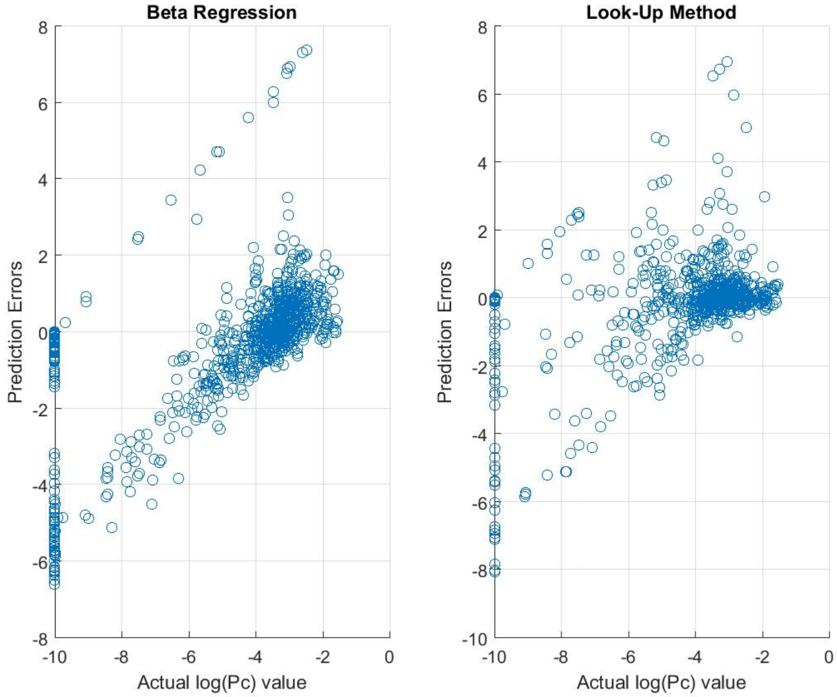

Figure 7.

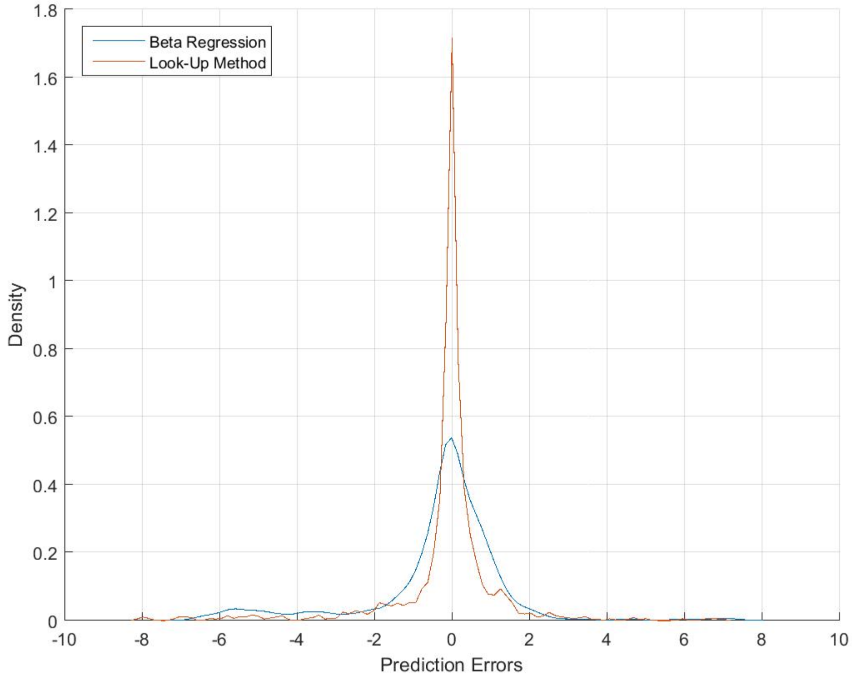

Figure 8.

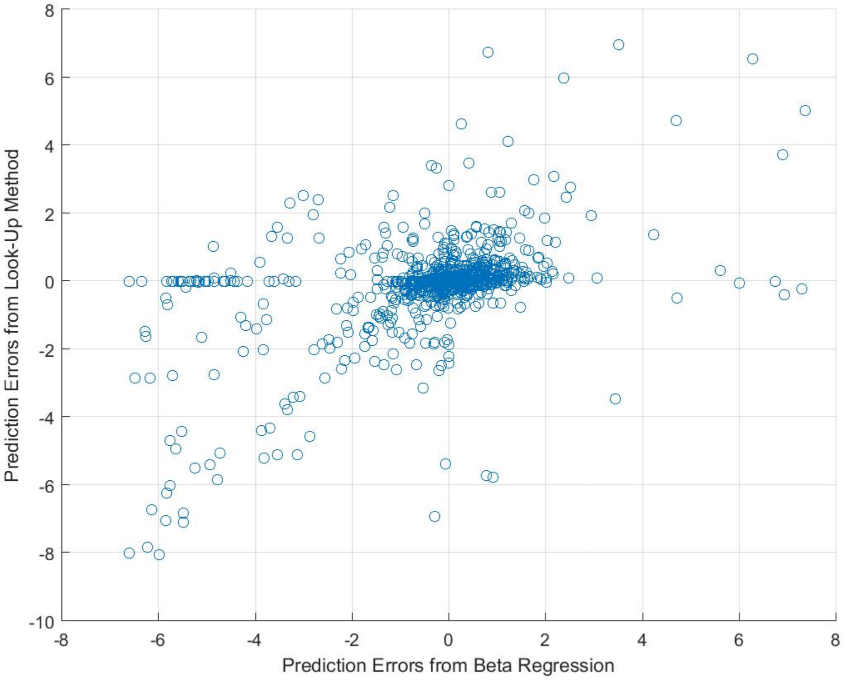

Figure 9.

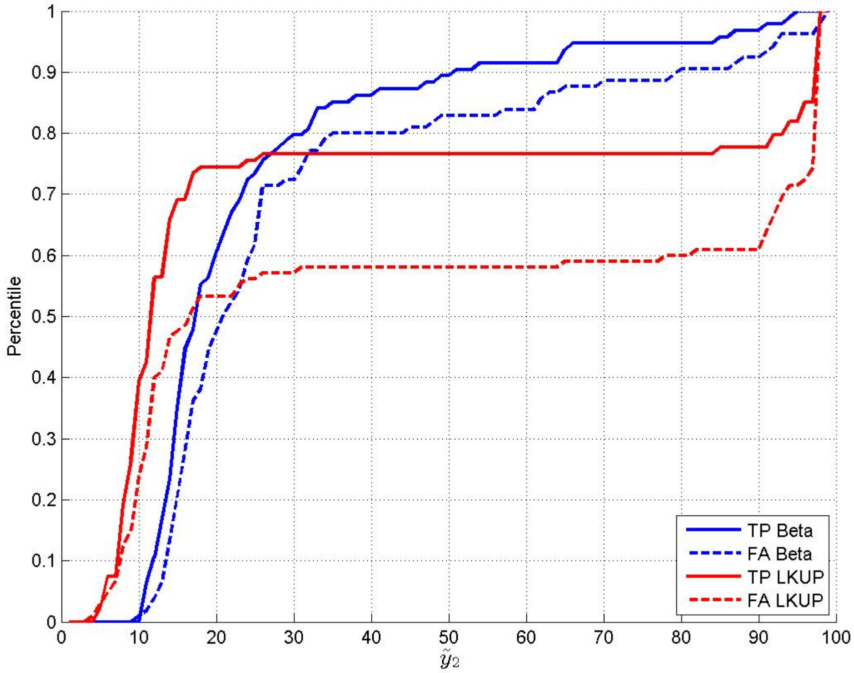

Figure 10.

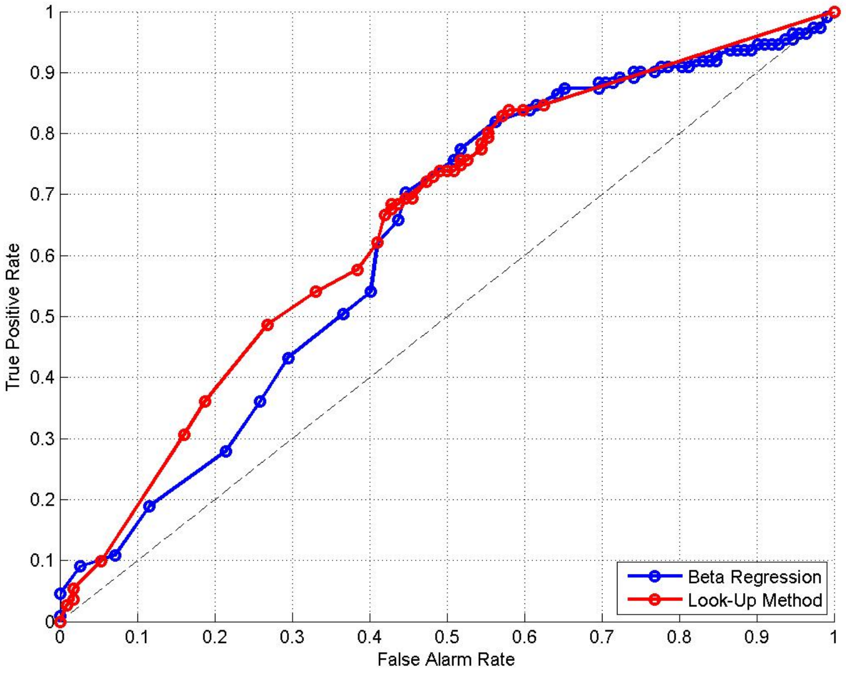
True Positive Rate
False Alarm Rate
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
Beta Regression
Look-Up Method

Figure 11.