Multivariate Anomaly Detection in Medicare Using Model Residuals and Probabilistic Programming

Richard A. Bauder, Taghi M. Khoshgoftaar

Florida Atlantic University

rbauder2014@fau.edu, khoshgof@fau.edu

Abstract

Anomalies in healthcare claims data can be indicative of possible fraudulent activities, contributing to a significant portion of overall healthcare costs. Medicare is a large government run healthcare program that serves the needs of the elderly in the United States. The increasing elderly population and their reliance on the Medicare program create an environment with rising costs and increased risk of fraud. The detection of these potentially fraudulent activities can recover costs and lessen the overall impact of fraud on the Medicare program. In this paper, we propose a new method to detect fraud by discovering outliers, or anomalies, in payments made to Medicare providers. We employ a multivariate outlier detection method split into two parts. In the first part, we create a multivariate regression model and generate corresponding residuals. In the second part, these residuals are used as inputs into a generalizable univariate probability model. We create this Bayesian probability model using probabilistic programming. Our results indicate our model is robust and less dependent on underlying data distributions, versus Mahalanobis distance. Moreover, we are able to demonstrate successful anomaly detection, within Medicare specialties, providing meaningful results for further investigation.

Introduction

Healthcare is a major industry in the U.S. with both private and government run programs. The costs of healthcare continue to rise, in part due to the increasing population of the elderly. U.S. healthcare spending from 2012 to 2014 increased by 6.7% to reach $3 trillion (National Center for Health Statistics 2015). One U.S. healthcare program catering primarily to individuals over the age of 65 is Medicare (Medicare.gov 2016). To further illustrate the significance of healthcare-related costs, as it pertains to the elderly population combined with the Medicare program, the number of elderly persons has increased 28% since 2004, whereas those under 65 have seen only a 6.5% increase in population (Administration on Aging 2015). This rising elderly population, combined with the increased costs of Medicare, need cost-cutting solutions, where the reduction

in fraud is one way to help recover costs and reduce overall payments. Medicare spending accounts for 20% of all healthcare spending in the U.S. at about $600 billion, with the recovery of 10% to 15% of these costs possible through fraud detection (Munro 2014).

In order to detect Medicare fraud and recover costs to reduce overall Medicare spending, we propose an innovative outlier detection method. Our method combines both multivariate regression and probability modeling. We are able to detect anomalous activities using a two-part approach, as follows: 1) creating a Multivariate Adaptive Regression Splines (MARS) model, and 2) using the regression residuals as inputs into a general, fully Bayesian univariate outlier detection model by leveraging the probabilistic programming paradigm. Our probability model is able to inherently represent uncertainty, or variability, in the underlying data to make probabilistic inferences. Moreover, our approach provides probability distributions of "outlierness" per point value, rather than simply an arbitrary score which can require further scrutiny in determining an actual outlying value. Our method not only effectively detects outliers, but also provides meaningful results markedly assisting with further investigations.

In our paper, we focus on detecting anomalous provider payments, in Medicare claims data, that could indicate possible fraud. Because objective evaluation of outlier detection is difficult, due to the lack of known outlying values, any detected outliers need to be investigated further to confirm a data point as being a true outlying value. The involvement of humans and/or other external processes, for further investigation and identification, is still necessary given the low number of actual documented Medicare fraud cases. Regardless of this extra involvement, our method is still able to prioritize and reduce the number of observations requiring further investigation. In this study, similar to Bauder and Khoshgoftaar (Bauder and Khoshgoftaar 2016a), we apply our method to two Medicare specialties, or provider types, exploring possible fraudulent observations. Furthermore, we show that our method does not exhibit the same concerns as seen with Mahalanobis distance (Miller, Vandome, and John 2010), a popular multivariate method to detect outliers. Note that due to space limitations, we do not provide additional comparative analysis with, or references for, other multivariate outlier detection techniques. Our study is

about Medicare fraud in general, thus we do not detail specific types of fraud, such as self-referrals or upcoding, but more information can be found in (Bauder, Khoshgoftaar, and Seliya 2017).

The rest of the paper is organized as follows. Related Works discusses works related to the current research. In the Methodology section, we detail the dataset used, probabilistic programming, the regression and probability models, and our experimental design. The Outlier Detection Method Comparison section provides a brief discussion on our method versus Mahalanobis distance. In the Results and Discussion section, we provide discussion on the application of our model using Medicare data. Finally, the Conclusion section summarizes our paper and future work.

Related Works

The majority of related research (Stevens 1984) is on the effects of outliers on regression models, specifically the use of outlier detection to improve linear regression models. Our research does not focus on how outliers affect the regression model, but rather the detection of outliers based on regression model outputs. We do, however, incorporate model residuals and Bayesian methods to detect outliers. With that, our review of related works focuses on these regression and Bayesian models, with emphasis on healthcare. To the best of our knowledge, no other study leverages both of these techniques together.

The use of studentized residuals, as well as several common techniques to detect outliers to improve regression models, is discussed in a paper by Mínguez et al. (Mínguez et al. 2012). The authors employ methods to automatically find outliers, such as hurricanes, that may be present in instrumental records (e.g. buoys), to protect the results of an analysis from these rare events. They remove or flag these events in order to create a clean, baseline dataset for further analysis. Chaloner et al. (Chaloner and Brant 1988) present a Bayesian approach using residuals in regression models to detect outliers. Using a simple linear model, and not probabilistic programming, the authors assess outliers via standardized residuals.

As in our work, several papers use distributions of expected versus actual, or observed, values to detect various types of possibly fraudulent behaviors. Thornton et al. (Thornton et al. 2014) explores several outlier-based detection frameworks using Medicaid claims data, specifically for dental providers. Their study involves multiple analysis techniques and outlier detection methods based on specific metrics, such as number of unique beneficiaries and claim payments. They employ three univariate methods and one multivariate method via clustering. Another study, by Hu et al. (Hu et al. 2012), involves the application of both patient utilization profiling and anomaly detection. The authors use patients' clinical characteristics to detect anomalies. They generate expected patient utilization levels from observations using three regression models (Regression Trees, Random Forest, and MARS), then use Grubb's test to find outliers in the expected versus actual values.

Both healthcare fraud and Bayesian methods were incorporated in a study by Ekina et al. (Ekina et al. 2013).

The authors apply Bayesian co-clustering to identify potentially fraudulent individuals using simulated data. Their Bayesian model assumes Dirichlet priors for the marginal membership probabilities, and independent Beta priors for the Bernoulli random variable parameters. Samples are drawn from the posterior probability distributions to infer co-clusters of providers and beneficiaries. These posterior distributions are used to detect fraud activities based on unusual cluster memberships.

Despite their being a large body of research on anomaly detection, outliers continue to be difficult to detect and thus research into their detection remains an important topic. We develop a new approach using a multivariate regression model in conjunction with a probability model to create a generalizable outlier detection method, applied to Medicare.

Methodology

In this section, we summarize the Medicare data used, discuss the two parts of our multivariate outlier detection method (regression and probability model), and outline the design of our experiment.

Medicare Data

The data that the Centers for Medicare and Medicaid Services (CMS 2016) has released, at the point of this publication, are for calendar years 2012, 2013, and 2014. We use the Physician and Other Supplier Data 2012 - 2014 dataset, from the Centers for Medicare and Medicaid Services, which contains payment and utilization claims data on services and procedures provided to Medicare beneficiaries. Due to the large data size, we decided to limit the data to office clinics in Florida (excluding larger facilities, such as hospitals). Furthermore, Florida is a good candidate for our research due to its high number of Medicare beneficiaries (The Henry J. Kaiser Family Foundation 2015). Table 1 summarizes the data used.

Table 1: 2012 - 2014 Medicare Physician and Other Supplier Data Summary

Dataset	Number of Instances	Number of Features	Unique Providers	Procedure Types
United States	27,757,455	26	1,049,362	6,741
Florida	1,197,238	26	48,230	2,922

Table 2: Provider Type Dataset Summary

Provider Type	Number of Instances	Number of Providers	Distinct Services	Average Payment
Thoracic Surgery	1,136	129	64	$110.79
Cardiology	76,105	1,685	559	$125.77

The Physician and Other Supplier PUF dataset is grouped by provider type/specialty, Healthcare Common Procedure Coding System (HCPCS) (CMS 2016) code, and National Provider Identifier (NPI) (Cognosante, LLC. 2004). For privacy reasons, the NPI numbers are purposefully obscured

in this paper. From the 26 total features or variables (detailed in (CMS 2016)), we selected 13 features. Out of these 13 features, 7 are numeric variables used for the regression models, which include the following: zip code, year, line service count (number of services provided/procedures performed per provider), count and sum of procedures performed across providers, beneficiary day service count (number of distinct beneficiary per day services), average allowed amount (allowed amount for the services with deductible and coinsurance amounts), and average payment amount (amount paid the provider for services performed). The other features, such as NPI, name and location, can be used for the identification of possibly fraudulent activities and corresponding provider(s). The remaining 13 unused features are primarily non-contiuous variables, such as free text physician notes, are not currently incorporated into our method and are left for future work. Additionally, for brevity, we limit the experiments to two of the 83 provider types (specialties) seen in Table 2. These specialties were chosen due to the large differences in the number of instances and distinct procedures performed.

Multivariate Outlier Detection Method

Regression Model MARS is a non-parametric regression model that accounts for the non-linearity between variables and their interactions (Friedman 1991). MARS utilizes a hinge function as piecewise linear functions (fitting the data) and performs automatic variable selection. It is suitable for large datasets, and is more flexible than traditional linear models. The advantages of MARS include the following: the ability of the hinge function to automatically partition the input data (which can contain some of the effects of outliers in the input), automatic feature selection (reducing possible masking when using the model outputs to detect outliers), and fast predictions. Comparisons with other multivariate and/or nonparametric models are an option for future work.

To detect possible anomalous activities, we use residuals, or model errors, from the MARS model. The absolute residuals are the difference of the actual (observed) and predicted values, or $\varepsilon_i = y_i - \hat{y}_i$. We calculate internally studentized, or standardized, residuals (Stevens 1984; Mínguez et al. 2012).

Probability Model In our study, we use probabilistic programming (Gelman et al. 2014; Davidson-Pilon 2015; Carpenter 2015) methods to detect outliers in Medicare claims data. Probabilistic programming employs a high-level language to create probability models and automatically solve them via statistical inference. Due to a 2013 initiative through the Defense Advanced Research Projects Agency (DARPA) (Jagannathan 2013), probabilistic programming continues to gain in popularity.

With the regression model residuals as the input, we use probabilistic programming to perform full Bayesian inference (Box and Tiao 2011). Bayesian inference provides a way of combining new evidence with prior beliefs, or assumptions, through the application of Bayes' rule: $P(A|X) = \frac{P(X|A)P(A)}{P(X)}$. In Bayes' rule, P(A) is the prior belief in event A (previous assumptions or beliefs based on

some prior knowledge), P(X) is the prior probability of the evidence, P(X|A) is the likelihood of evidence X given event A (a single value X for a hypothesis A), and P(A|X) is the posterior probability, which is the updated belief.

Bayesian methods also have more interpretable results returning distributions of probabilities per data point. Furthermore, Bayesian techniques provide credible intervals for the different parameters in the model. The credible intervals show that a value or parameter has an 80% or 95% probability of being within the actual interval bands. This is compared to traditional confidence intervals which indicate that, if an experiment is repeated many times, the values will be within this interval 80% or 95% of the time.

For our study, we use the probabilistic programming language known as Stan (Carpenter 2015). The posterior distributions are drawn from the full conditional of each unknown parameter. This is done using Hamiltonian Monte Carlo (HMC) and the No-U-Turn Sampler (NUTS), implemented in Stan. The model is fit by specifying the full likelihood function and the prior distributions of all unknown parameters. Below is our univariate outlier detection probability model (Bauder and Khoshgoftaar 2016b) showing inputs, unknown variables, distributions, and generated outputs. This is not a tutorial on Stan, but the interested reader can find detailed information in (Carpenter 2015).

```
data{
    int<lower=0> N;
    vector[N] value;        // input (from MARS model)
    vector[N] check_value;  // values to check
}
parameters{
    real mean_value;
    real stdev_value;
    real nu;                // degrees of freedom
}
model{
    mean_value ~ normal(100, 100);  // mean prior
    stdev_value ~ normal(100, 100); // stdev prior
    nu ~ cauchy(7, 5) T[0.0,]; // degrees of freedom
    for(i in 1:N) // Student's t-dist for outliers
      value[i] ~ student_t(nu, mean_value, stdev_value);
}
generated quantities{
    vector[N] cdf_prob;  // cdf of probabilities
    vector[N] prob; // final outlier probabilities
    for(i in 1:N){
      cdf_prob[i] = student_t_cdf(check_value[i],
                nu, mean_value, stdev_value);
      prob[i] = 2*(cdf_prob[i]*(1-cdf_prob[i]));
    }
}
```

Experimental Design

As mentioned, we incorporate two parts in order to detect outliers. The first part involves a nonparametric multivariate regression model. We use the R programming language (R Core Team 2016), with the earth package (Milborrow, Hastie, and Tibshirani 2016) implementation of MARS, to create and validate each model. The Classification And REgression Training (CARET) (Kuhn et al. 2016) package

is used to create the final MARS model, with 10-fold cross-validation to reduce overfitting of trained models. CARET is a set of functions to streamline the process for creating predictive models. For this study, we use the default model parameters, and do not consider parameter tuning.

The second part of our method detects outlying values from the studentized residuals. We used the rstan (Guo et al. 2016) implementation of the Stan probabilistic programming language. Each Stan probability model was run with 4,000 iterations and 2 Markov chains (Gelman et al. 2014) (for performance/run-time reasons). Additionally, the *value* and *check_value* variables (from the Stan model code) are identical vectors. This indicates that we are looking for outliers in the full population.

From the model, we compute the probability, at each data point, of observing a more extreme value. For example, we observe that a probability of some point is 50% which says that there is a 50% chance of seeing a value more extreme than this current value. This is similar to a value in the middle of a typical "bell curve", thus unlikely to be an outlier. In contrast to this example, a probability of 1% on a data point would indicate only a 1% chance of seeing a more extreme value, which could indicate an outlier since there are not many values more extreme. This is akin to being at the tails of the "bell curve" distribution.

Outlier Detection Method Comparison

For this comparison, due to space limitations, we only contrast our approach with Mahalanobis distance, a commonly used multivariate outlier detection method. We will consider other multivariate detection techniques for future work. Mahalanobis distance (Miller, Vandome, and John 2010) considers the scale of the data from many distributions (i.e. multivariate) expressing the probability of an observation. This method gives the distance from a case to the centroid of all cases for the predictor variables. A large distance indicates an observation that is an outlier in the space defined by the predictors (Stevens 1984). One weakness of using Mahalanobis distance is that it behaves best with data that are approximately multivariate normal (Cousineau and Chartier 2015). If the data are not multivariate normal, the means might not be a good representation of the center of the data and general trends in the data may not be identified correctly using variance as a measure of spread.

Figure 1 shows the detection of outliers for our method and Mahalanobis distance, by payment and count of services per day. Notice that the Mahalanobis distance method captured the high counts per day as outliers (red circles on the x-axis), but our method did not. It could be assumed that these are actual outliers because the number of procedures performed is higher than the rest, but these are more likely false alarms due to the multivariate normal requirement with Mahalanobis distance. Two tests were done in order to check for normality, the first was Anderson-Darling's Normality test (NIST 2013) on the individual predictors. The p-values were zero indicating each predictor's distribution as non-normal, at a 95% confidence. The second check involved Henze-Zirkler's and Mardia's Multivariate Normal-

ity tests (Korkmaz, Goksuluk, and Zararsiz 2014), both of which indicated this dataset was not multivariate normal.
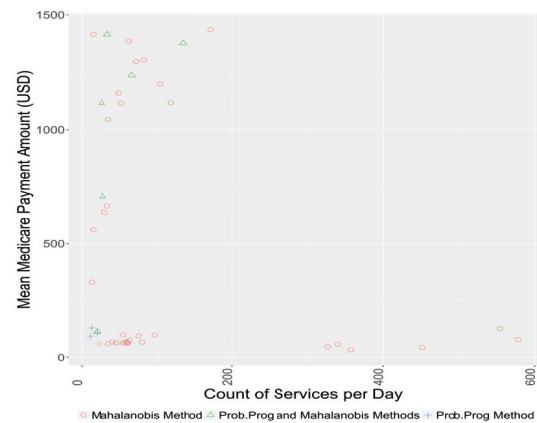

Figure 1: Thoracic Surgery Multivariate Outlier Comparison

The assumption of multivariate normal data can lead to incorrect outlier detection, especially in non-normally distributed datasets such as the Medicare claims data. Because our method is not dependent on the underlying data distribution, it can be used as is on Medicare claims data to detect outliers; whereas, a method like Mahalanobis distance may require data transformation and/or extra scrutiny when assessing outlying values.

Results and Discussion

The MARS model was created using 10-fold cross-validation (Witten and Frank 2005) for training. The performance measures include: Mean Absolute Error (MAE) - average of the absolute differences of predicted and observed values (lower is better) and Root Mean Squared Error (RMSE) - estimated standard deviation of unexplainable variations in the dependent variable (lower is better).

For our study, Thoracic Surgery had a MAE of 3.57 and RMSE of 5.81, and Cardiology had a MAE of 3.63 and RMSE of 11.65. The better model fit for Thoracic Surgery appears to indicate that this specialty has a more specific and homogeneous nature in the procedures types performed (i.e. different HCPCS codes). This implies that specialties with a greater heterogeneous mixture of procedure types will have possibly worse model fit results. Additionally, the absolute range of Medicare claims payment amounts vary greatly between and within each specialty, lending to further model performance fluctuations.

Next we run the regression residuals, per specialty, through our probability model resulting in possible fraudulent activities that could require further investigation. The flagged fraudulent values can be associated with various characteristics such as NPI and last name. These identification variables can be used to either bypass a flagged value (if it really isn't a fraud event) or help to narrow the investigation. In this study, the possible fraud labels are a combination of the first name and masked NPI for confidentiality.

Our outlier detection method provides a distribution of probabilities per event to better focus on the most plausible

fraudulent activities. Thus, the mean probability of being an outlier can be used to determine the appropriate probability threshold needed to indicate an outlier. Figure 2 shows possible fraudulent values, labeled at probabilities of 5% or less and payments greater than $10,000 by location for Cardiology. The use of these thresholds are arbitrarily chosen for graphical labeling clarity. As can be seen, possible fraudulent activities are labeled and show groupings of labels at locations more prone to these findings. The name circled in red indicates a Cardiologist under investigation for fraud. Cardiology appears to have more possible fraud events in Ocala and Bradenton, which could warrant investigation.
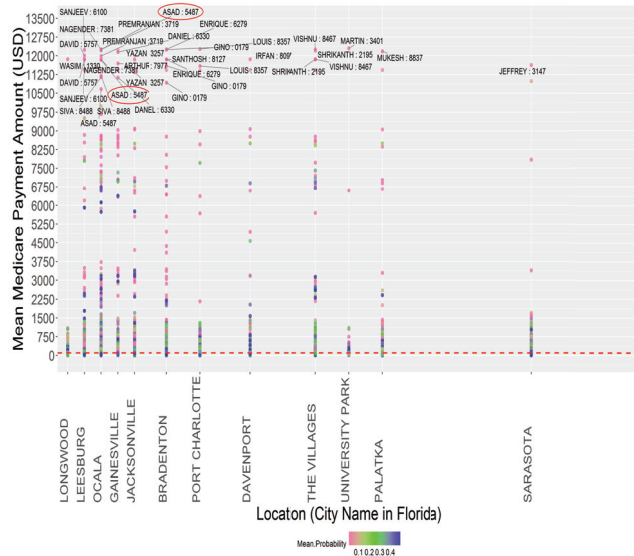

Figure 2: Cardiology Possible Fraud Events by Location

Beyond providing a distribution of mean probabilities, our model generates credible intervals for each value showing that a particular value has an 80% or 95% probability of being within its probability distribution. Figure 3 shows the credible intervals for Thoracic Surgery, for probabilities below 5%. The yellow dots indicate the mean probability, the red horizontal line is the credible level with 80% intervals, and the black horizontal line is the outer level with 95% intervals. The credible intervals help determine our confidence that a value is indeed an outlier. Figure 4 depicts the credible intervals for Cardiology. Notice that the intervals are wider for Thoracic Surgery indicating more variance in the dataset. This information captures inherent uncertainty and could be used to help create better indicators as to what constitutes an outlier given a more variable dataset or, conversely, assume tighter bounds when flagging outliers for less variable data.

To support our outlier detection method, we use the NPI numbers to search news reports for known fraud cases in Florida. When searching for providers via last name and NPI, we discovered a Cardiologist under investigation for fraud (HiersStaff 2016). It was reported that this provider billed Medicare for medically unnecessary peripheral artery interventions, which is flagged with a label of *Asad:5487* in the town of Ocala, Florida, circled in Figure 2. Even with
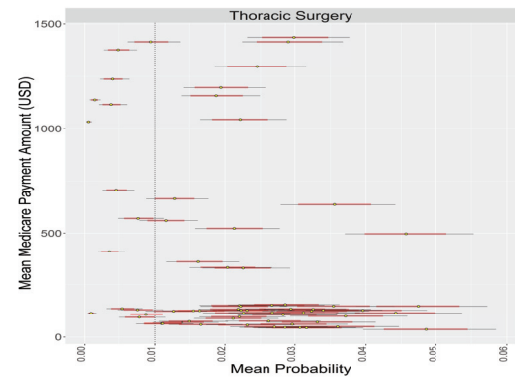

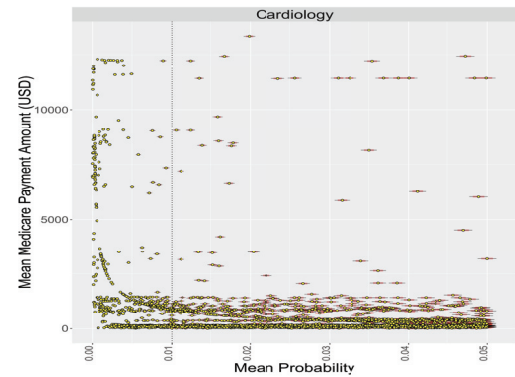Figure 3: Thoracic Surgery Credible Intervals

Figure 4: Cardiology Credible Intervals

this successful detection by our model, there is a gap in assessing fraud detection performance due to the limited number of real-world fraud cases in the 2012 - 2014 Medicare data, thus continued validation are left for future research. We have demonstrated the ability of our outlier detection method, which was shown to take in multiple input variables and produce meaningful outputs consisting of probability distributions, and credible intervals, per value.

Conclusion

The rising costs of Medicare in the U.S., and the corresponding increased fraud potential, require new and innovative solutions. One way to combat fraud is through the use of outliers, or anomaly, detection techniques. Anomaly detection gives investigators a way to discover activities or behaviors that could be conducive of fraud. In this paper, we outline a novel outlier detection approach applied to Medicare claims data. Our detection method is multivariate and does not rely heavily on underlying distribution assumptions, while providing meaningful probabilities per payment value in order to better determine whether a value is indeed an anomaly. Our model consists of a MARS model producing studentized residuals, which are inputs into our probability model created via the Stan probabilistic programming language.

We apply our method on two Medicare specialties using the claims dataset to detect possible fraud events. Even

though we focus on Medicare data, our method is general and can be applied other domains to detect anomalous values. In this paper, we discussed how our method differed from Mahalanobis distance by not requiring linear relationships between variables, or assuming multivariate normal data. We then conducted experimental investigations through the application of our method on two Medicare specialties. Cardiology had a documented Medicare fraud case under investigation, for which we were able to successfully show the detected anomalies. Our method's initial findings demonstrate the power and usefulness of our probability model in detecting outliers, and providing meaningful results via probability distributions. Future work involves expanding the number of specialties. We also intend to provide detailed comparisons of our model versus other multivariate outlier detection methods. Finally, additional validation with real-world cases should be done.

Acknowledgment

References

Administration on Aging. 2015. Profile of older Americans.

Bauder, R. A., and Khoshgoftaar, T. M. 2016a. A novel method for fraudulent medicare claims detection from expected payment deviations (application paper). In *2016 IEEE 17th International Conference on Information Reuse and Integration (IRI)*, 11–19.

Bauder, R. A., and Khoshgoftaar, T. M. 2016b. A probabilistic programming approach for outlier detection in healthcare claims. In *2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA)*, 347–354.

Bauder, R.; Khoshgoftaar, T. M.; and Seliya, N. 2017. A survey on the state of healthcare upcoding fraud analysis and detection. *Health Services and Outcomes Research Methodology* 17(1):31–55.

Box, G. E., and Tiao, G. C. 2011. *Bayesian inference in statistical analysis*, volume 40. John Wiley & Sons.

Carpenter, B. 2015. Stan: A probabilistic programming language. *Journal of Statistical Software*.

Chaloner, K., and Brant, R. 1988. A bayesian approach to outlier detection and residual analysis. *Biometrika* 75(4):651–659.

CMS. 2016. Centers for Medicare and Medicaid Services: Research, Statistics, Data, and Systems.

Cognosante, LLC. 2004. National Plan & Provider Enumeration System (NPPES): National Provider Identifier.

Cousineau, D., and Chartier, S. 2015. Outliers detection and treatment: a review. *International Journal of Psychological Research* 3(1):58–67.

Davidson-Pilon, C. 2015. *Bayesian Methods for Hackers: Probabilistic Programming and Bayesian Inference*. Massachusetts, USA: Addison-Wesley Professional.

Ekina, T.; Leva, F.; Ruggeri, F.; and Soyer, R. 2013. Application of bayesian methods in detection of healthcare fraud. *Chemical Engineering Transaction* 33.

Friedman, J. H. 1991. Multivariate adaptive regression splines. *The annals of statistics* 1–67.

Gelman, A.; Carlin, J. B.; Stern, H. S.; and Rubin, D. B. 2014. *Bayesian data analysis*, volume 2. Chapman & Hall/CRC Boca Raton, FL, USA.

Guo, J.; Gabry, J.; Goodrich, B.; Lee, D.; Sakrejda, K.; Sklyar, O.; Oehlschlaegel-Akiyoshi, J.; Wickham, H.; de Guzman, J.; Fletcher, J.; Heller, T.; Niebler, E.; and the R Core Team. 2016. rstan: R Interface to Stan.

HiersStaff, F. 2016. Cardiologist plagued by legal woes files for Chapter 11 bankruptcy protection.

Hu, J.; Wang, F.; Sun, J.; Sorrentino, R.; and Ebadollahi, S. 2012. A healthcare utilization analysis framework for hot spotting and contextual anomaly detection. In *AMIA*.

Jagannathan, S. 2013. DARPA probabilistic programming for advancing machine learning (PPAML).

Korkmaz, S.; Goksuluk, D.; and Zararsiz, G. 2014. Mvn: An r package for assessing multivariate normality. *The R Journal* 6(2):151–162.

Kuhn, M.; Wing, J.; Weston, S.; Williams, A.; Keefer, C.; Engelhardt, A.; Cooper, T.; Mayer, Z.; Kenkel, B.; the R Core Team; Benesty, M.; Lescarbeau, R.; Ziem, A.; Scrucca, L.; Tang, Y.; and Candan, C. 2016. *caret: Classification and Regression Training*. R package version 6.0-68.

Medicare.gov. 2016. US Medicare Program.

Milborrow, S.; Hastie, T.; and Tibshirani, R. 2016. earth: Multivariate Adaptive Regression Splines.

Miller, F.; Vandome, A.; and John, M. 2010. *Mahalanobis Distance*. VDM Publishing.

Mínguez, R.; Reguero, B. G.; Luceno, A.; and Méndez, F. J. 2012. Regression models for outlier identification (hurricanes and typhoons) in wave hindcast databases. *Journal of Atmospheric and Oceanic Technology* 29(2):267–285.

Munro, D. 2014. Annual U.S. healthcare spending hits $3.8 trillion.

National Center for Health Statistics. 2015. Total National Health Expenditures.

NIST. 2013. NIST/SEMATECH e-Handbook of Statistical Methods (Anderson-Darling).

R Core Team. 2016. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Stevens, J. P. 1984. Outliers and influential data points in regression analysis. *Psychological Bulletin* 95(2):334.

The Henry J. Kaiser Family Foundation. 2015. State Health Facts - Medicare.

Thornton, D.; Capelleveen, G.; Poel, M.; Hillegersberg, J.; and Müller, R. M. 2014. Outlier-based health insurance fraud detection for us medicaid data.

Witten, I. H., and Frank, E. 2005. *Data Mining: Practical Machine Learning Tools and Techniques, Second Edition*. Morgan Kaufmann.