

Medical Overpayment Estimation: A

Bayesian Approach

Rasim M Musal¹ and Tahir Ekin¹

¹ Department of Computer Information Systems and Quantitative Methods, McCoy College of Business, Texas State University, San Marcos, Texas USA

Address for correspondence: Rasim M Musal, McCoy College of Business, Texas State University, San Marcos, Texas USA.

E-mail: rm84@txstate.edu.

Phone: (+1) 512 245 3197.

Fax: (+1) 512 245 1452.

Abstract: Overpayment estimation using a sample of audited medical claims is an oft used method to determine recoupment amounts. The current practice based on Central Limit Theorem may not be efficient for certain kinds of claims data, including skewed payment populations with partial overpayments. As an alternative, we propose a novel Bayesian inflated mixture model. We provide an analysis of the validity and efficiency of the model estimates for a number of payment populations and overpayment scenarios. In addition, learning about the parameters of the overpayment distribution with increasing sample size may provide insights for the medical investigators. We present a discussion of model selection and potential modeling extensions.

Key words: Bayesian Hierarchical Modeling ; Gamma Mixture Models ; Inflation Models ; Medical Overpayment Estimation; Medical Audits

1 Introduction

Medical expenditures are a significant part of the governmental budgets. For instance, U.S. health care spending grew 3.6 percent in 2013, reaching \$2.9 trillion or \$9,255 per person, which accounts for 17.4 percent of the nation's Gross Domestic Product (CMS, 2014a). It is reported by U.S. governmental agencies that each year three to ten percent of the overall health care spending is lost to fraud, waste and abuse (Shin et al., 2012). Medical fraud is defined as an intentional deception or misrepresentation made by a person or an entity, with the knowledge that the deception could result in some kinds of unauthorized benefits (NHC, 2012) whereas waste and abuse differ by the level of intention and knowledge. We will use the term overpayment in reference to fraud, waste and abuse. These overpayments have direct cost implications to the government and to the taxpayers. In addition, overpayments diminish the ability of the medical systems to provide quality care to the deserving patients (Anderson and Hussey, 2001). The size and complex nature of the medical data make the use of sampling and estimation methods necessary for extrapolation of the overpayments. This paper proposes an overpayment model that is shown to be valid with respect to the governmental guidelines, and discusses a number of cases where it can be an efficient alternative in recovering overpayment.

In the U.S., governmental medical services are mainly provided through the federal and state programs of *Medicare* and *Medicaid* which are administered by *The Centers for Medicare & Medicaid Services* (CMS). There are a number of initiatives to oversee the health care spending. The 2013 annual report (OIG, 2013a) prepared by the joint efforts of *Department of Health and Human Services* and *Department of Justice* gives a broad overview of the current governmental efforts against overpayments. In

its five year strategic plan of 2013, *Office of Inspector General* (OIG) categorizes such efforts as identification and investigation of fraudulent activities that lead to overpayment, obtaining fair recovery amounts from wrongdoers, and fraud prevention (OIG, 2013b). The following subsection discusses the use of statistical methods in medical fraud assessment, particularly overpayment estimation.

1.1 Statistical Methods in Medical Fraud Assessment

Identification of the overpayments are ideally done by domain experts via audits of the medical claims. However, comprehensive auditing is only feasible for a small number of claims and for cases where overpayment can easily be identified due to irrefutable evidence. Such an example can be providers who bill for dead beneficiaries. In many cases, it is impractical to identify each overpaid claim because of the complex nature and the size of the medical data. This makes the use of statistical approaches a necessity in medical fraud assessment. Statistical approaches mainly include the use of data mining, sampling and estimation methods. Various data mining approaches are proposed to reveal existing patterns and flag potentially fraudulent claims, see (Li et al., 2008) for a comprehensive review. This paper focuses on sampling and estimation methods.

An important consideration in *Medicare* audits is the fair estimation of the recovery amounts. Sampling design is an important choice in this general framework. In the U.S., use of probability sampling methods for medical investigations has been accepted to be part of the legal framework since 1986. (Yancey, 2012) provides a comprehensive list about these legal sampling procedures and the parties involved in U.S. governmental medical insurance programs. Section 8.4 of CMS guidelines (CMS, 2011) lists a number of steps for the construction of valid sampling and estimation

methods;

1. Selection of the Provider
2. Selection of the Period
3. Definition of the Universe, Sampling Unit and Sampling Frame
4. Design of the Sampling Plan, Selection of the Sample
5. Review of Each Sampling Unit
6. Estimation of Overpayment

A requirement of a valid sampling design, that is listed in the 4th step, is that, each unit in the sample must have a known probability of selection that is greater than 0. Simple random sampling, systematic sampling, stratified sampling, and cluster sampling or a combination of these are listed as the most common acceptable sample designs in *Medicare* audits, see the relevant discussions in (CMS, 2014b), (OIG, 2013c) and (OIG, 2014a). (Daniel, 2011) presents the advantages and disadvantages of the various sampling designs, see also (Cochran, 2007) for an overview. To keep the paper parsimonious and focused on the overpayment estimation that is listed as the 6th step, we use simple random sampling.

The population of interest in these procedures is usually the payment amounts to a provider some of which result with overpayments. A payment amount associated with a claim can result in one of three outcomes when audited. A claim can be classified as completely legitimate, completely illegitimate or partially overpaid. Claims data where each claim is either a legitimate payment or a completely illegitimate payment is referred to as “all or nothing”. According to the current governmental sampling guidelines (CMS, 2001), in most situations the lower limit of a one sided 90

percent confidence interval for the total overpayments should be used as the recovery (recoupment) amount from the provider under investigation. Using the lower bound allows for a conservative recovery without requiring the tight precision to support the point estimate, sample mean. The sole application of Central Limit Theorem (CLT) is based on the assumption that overpayment population either follows the Normal distribution or that the sample size is reasonably large. However, it is very common that medical claims data exhibit skewness and non-normal behavior requiring large sample sizes for the valid application of CLT.

([Edwards et al., 2003](#)) shows that methods based on the CLT may not perform well for certain kinds of overpayment populations with small sample sizes. As an alternative, they propose the “minimum sum method”, a non-parametric inferential method which makes use of the hyper-geometric distribution and compute the respective lower bound estimates. If negative overpayments are non-existent or of negligible frequency, as is often the case, minimum sum method is shown to be mathematically valid such that it provides a lower confidence bound for total overpayment with confidence level greater than or equal to the nominal level of 90%. A number of extensions are proposed for the minimum sum method. ([Ignatova and Edwards, 2008](#)) utilize it within a two stage sampling procedure in that they use first stage (probe) samples to decide if the cost of additional sampling is justified. ([Gilliland and Feng, 2010](#)) provide an adaptation in order to address cases of varying payments. ([Gilliland and Edwards, 2010](#)) improve its efficiency via randomized lower bounds in which payment amounts are audited in equal sized packets. ([Edwards et al., 2015](#)) extend their packet sampling idea by using penny samples. The minimum-sum estimates are shown to be efficient in the recovery of overpayment from the claims that are essentially “all or nothing”. ([Edwards et al., 2003](#)) discuss a simple extension,

so called q-adaptation Minimum sum method, which is based on re-definition of illegitimate payments; so that the payments are defined as illegitimate if q percent of the payment is in error. In addition to these, standard stratified expansion ([Buddhakulsomsiri and Parthanadee, 2008](#)) and combined ratio estimators of the total are also proposed. ([Mohr, 2005](#)) also present a normality based overpayment model to capture “all or nothing” billing pattern.

These methods are shown to be practically robust and provide good coverage of lower bound estimates. However, ([Mohr, 2005](#)) acknowledges they are not able to capture the heterogenous nature of claims data, therefore may not be efficient for such cases. ([Edwards, 2011](#)) discusses the fact that there are also other certain kinds of illegitimate provider behaviors such as overcharging that correspond to partial overpayments. There are rare attempts to address this issue. ([King and Madansky, 2013](#)) explicitly model the overpayment percentage using a two-valued step function and a continuous exponential function. They assume overpayment to follow Gamma distribution, and propose a proportional stratification sampling method. ([Edwards et al., 2015](#)) consider one partial overpayment pattern in assessment of penny sample adaptation of the minimum sum method. The beta mixture model of ([Ekin et al., 2015](#)) address the mixed distribution characteristics of claims data. To the best of our knowledge, there is not an estimation model that is shown to be valid for a number of payment and overpayment scenarios, and that can potentially be an efficient alternative in the cases of multi-modal overpayment patterns.

1.2 Motivation

In some situations, an investigator may need to consider completely legitimate, completely and partially overpaid claims simultaneously to learn about the overpayment

population. Most of the existing models lose efficiency in the presence of partial overpayments in addition to “all or nothing” claims, since they do not explicitly model the characteristics of partial overpayment patterns.

Our interest is on such cases with heterogeneous medical data with multiple patterns and with a spike of zero values. These cases are not rare in medical claims data. For instance, an OIG investigation reports that “Skilled Nursing Facilities” (SNFs) billed one quarter of all claims in error in 2009, resulting in \$1.5 billion in inappropriate *Medicare* payments (OIG, 2012a). The majority of the claims in error were overcharged, meaning that they were billed at higher levels than warranted. In another OIG investigation, it is reported that *Medicare* inappropriately paid \$6.7 billion in 2010 for claims of “evaluation and management” (E/M) services (OIG, 2014b) which also included cases of overcharging. Recommendations of OIG to CMS include encouraging contractors to review E/M services billed by overcharging physicians, and follow up on claims for E/M services that were paid for in error. CMS agrees to reassess the effectiveness of reviewing claims billed by overcharging physicians. These are some of our motivating examples in which the claims data include partial overpayments in addition to completely legitimate claims. Our model can also be utilized when the population of interest is all claims paid to a given health care provider in a specified time frame, or individual claim lines for a particular procedure code with various modifier codes.

The existing practice includes the use of point estimates of overpayments and does not reveal the actual overpayment pattern. However, capturing and learning about the inflated and mixture characteristics of the overpayment distribution is crucial. This has recently gained more attention because of the shown persistence of the

overpayments despite the increased amounts of the recovery. As pointed out by (Musal, 2010), the federal budgetary report of “U.S. Office of Management and Budget” argues that the existing measures have not resulted with a trend of decrease in the overall level of health care overpayments. It should be noted that one of the ultimate goals of medical fraud assessment is to inhibit the sustained culture of medical overpayments. In a related area, there is an increasing awareness about the low level partial overpayments. For instance, a press release by CMS urges senior citizens to join fight against fraud despite the fact that the dollar amounts involved with these claims are relatively smaller (CMS, 2013a).

The proposed Bayesian inflated mixture based model links the known payment population and the information gathered from a sample of audited investigations. The initial objective of the model is to provide valid estimates that comply with the governmental guidelines. All overpayment methods are expected to provide a lower confidence bound with at least 90% confidence. Secondly, we investigate the efficiency of the model with respect to the recovery amount, with a focus on partial overpayments. Thirdly, our model can also help reveal the overpayment pattern and quantify learning. It allows the decision makers to conduct simultaneous estimation of both probability of each overpayment pattern and the percentage of partial overpayments, and can result in further insights. This can potentially have a positive impact on the culture of medical overpayments.

In section 2, we provide a brief review of the inflated zero and one mixture models, motivate the use of the Bayesian approach and describe how our model fits within the literature. Section 3 explains our modeling framework. Section 4 provides an analysis with respect to validity and efficiency of the model, and comparisons with

the current practice. Section 5 discuss model selection whereas Section 6 presents modeling extensions. The paper concludes in Section 7 with a discussion of findings and future work.

2 Literature Review

The review provided in this section is by no means exhaustive and mainly aims to show the proposed model's fit within the literature. We provide a discussion of the existing zero-inflated models, Gamma mixture models and show the relevance of Bayesian approaches for audits.

The existence of abundant number of a particular value in a data set has been a long known and well studied phenomena. Such models can be dated back to ([Aitchison, 1955](#)) that considers an inference problem for such a mixture population. The seminal work by ([Lambert, 1992](#)) discusses models for a manufacturing data set with abundant zeroes for the number of defective items. ([Min and Agresti, 2005](#)) studies biomedical and sociological applications in which there is a spike of zeroes, referring to this as having zero inflation. The authors propose Poisson and Negative Binomial distribution based models with random effects for pharmaceutical and occupational injury prevention. Another such application is the zero inflated Poisson model of ([Cruyff et al., 2008](#)) which uses data from a Social Security survey to model the respondent behavior with self-protection bias. Aside from Poisson and Negative Binomial distributions, the general class of zero or one inflated beta regression models are discussed by ([Ospina and Ferrari, 2012](#)). The zero inflated Gamma distribution is not as commonly applied. ([Feuerverger, 1979](#)) is one of the rare applications and attempts to model the rainfall data.

In terms of the use of Bayesian inference, (Ghosh et al., 2006) provide a description of the properties of zero-inflated regression models. (Neelon et al., 2010) present alternative Zero Inflated models using Poisson and Negative Binomial distributions. They discuss how Bayesian models allow incorporating expert information and result with improved parameter estimation. (Muralidharan, 2010) is an example that fits empirical Bayesian mixture models via the expectation maximization algorithm. On the other hand, (Erosheva et al., 2007) utilize individual level mixture models using variational approximation methods. Gamma mixture models with known and unknown number of distributions within the Bayesian framework is elaborated in (Wiper et al., 2001) which involves the estimation of the properties of a queue for email data. (Webb, 2000) present a Bayesian Gamma mixture model for target recognition using data from radar. In health care context, (Venturini et al., 2006) propose Gamma shape mixtures for estimation of the proportion of medical expenditures that exceed a given threshold.

The use of Bayesian models are not common in medical audits although they are shown to provide better estimates in the domain of tax auditing. (Guthrie, 1989) present a report on the use of non-standard distributions in auditing; discussing the need for mixtures of standard and degenerate distributions, which have masses at particular measurements such as zero. They compare different estimation methods used within audits, mainly focusing on the use of dollar unit sampling; which models the total population error as the product of the known payment amount and the mean tainting per dollar unit. They numerically illustrate that Bayesian methods such as (Cox and Snell, 1979) and (Tsui et al., 1985) provide better lower bounds for the adjustment population compared to the frequentist estimators such as the difference estimator, separate and combined ratio estimators. (Matsumura et al.,

1991) propose a multinomial Dirichlet Bayesian approach with comparisons. In a related work, (Ghosh and Meeden, 1997) present an empirical Bayesian estimator of the finite population mean.

3 Modeling Framework

This paper presents a Hierarchical Bayesian model for overpayment data that can belong to one of three populations such that an overpayment observation is either zero, is the payment value or is a value between zero and the respective payment. These respectively correspond to completely legitimate, completely illegitimate and partially overpaid claims. When the overpayment is partial, it belongs to one of the mixture sub-components with each having different mean values. We assume $\mathbf{X} = \{X_1, \dots, X_N\}$ to be the vector of known payment amounts for a population of N claims and let $\mathbf{Y} = \{Y_1, \dots, Y_N\}$ represent the unknown overpayment population. Samples from these populations are denoted as \mathbf{x} and \mathbf{y} . As underpayments are rare in medical audits, we assume that the overpayment distribution is in the non-negative region.

We introduce a latent variable vector, \mathbf{z} to indicate the membership of the main overpayment components for all N claims. The indicator vector, \mathbf{z} , is a categorically distributed random variable where each i^{th} claim belongs to one of the three main components; $z_i \in \{1, 2, 3\}$. For the i^{th} claim, z_i being equal to 1 implies that $Y_i = 0$, and when z_i is equal to 2, $Y_i = X_i$. When z_i is equal to 3, the claim has the case of a partial overpayment, overpayment taking values between zero and the respective payment; $0 < Y_i < X_i$. The probability vector, $\boldsymbol{\pi}$, has three elements π_1 , π_2 and $1 - \pi_1 - \pi_2$ that allow us to describe our uncertainty over the distribution of \mathbf{z} , such

that

$$z_i \sim \text{Categorical}(\boldsymbol{\pi}) \quad (3.1)$$

π_1 and π_2 are probabilities for the inflated portions of the overpayment distribution whereas $1 - \pi_1 - \pi_2$ corresponds to the probability of a partial overpayment. The $\boldsymbol{\pi}$ vector has a Dirichlet prior distribution.

We assume the number of mixture sub-components, K , to be unknown and utilize Dirichlet Process (Ferguson, 1973) within a semi-parametric framework via the reversible jump Markov chain Monte Carlo algorithm (Green, 1995) in OpenBugs (Thomas et al., 2006). This algorithm allows us to consider U different models with differing K values and obtain the uncertainty distribution over K . In doing so, a binary vector \mathbf{S} with size U is introduced so that the algorithm can switch between models with different values of K . The latent variable that indicates the mixture sub-component membership of i^{th} claim is denoted as m_i , and has a categorical distribution with the probability vector $\boldsymbol{\eta}$ of size K . The $\boldsymbol{\eta}$ has a Dirichlet prior distribution which takes the form of a uniform distribution for the special case of $K = 2$. In summary, the distribution of m_i in the case of K sub-components is

$$m_i \sim \text{Categorical}(\boldsymbol{\eta}), \text{ where } \boldsymbol{\eta} = \left\{ \eta_3, \dots, \eta_{K+1}, \eta_{K+2} = \left(1 - \sum_{k=3}^{k=K+1} \eta_k \right) \right\}. \quad (3.2)$$

Furthermore, the joint probability vector for all components of the model with K mixture sub-components can be written as

$$\{\pi_1, \pi_2, (1 - \pi_1 - \pi_2)\eta_3, (1 - \pi_1 - \pi_2)\eta_4, \dots, (1 - \pi_1 - \pi_2)\eta_{K+2}\}.$$

In addition, we explicitly model the mean and standard deviation of partial over-

payments. Partially overpaid claims are assumed to follow one of the K Gamma distributions with mean μ_i and standard deviation σ_k . A linear equation helps model μ_i as :

$$\mu_i = \prod_{k=3}^{k=K+2} (\rho_k + \epsilon_{k,i})^{\mathbb{I}(m_i=k)}, \quad (3.3)$$

where k indicates the mixture sub-component membership of a partial overpayment. The function \mathbb{I} is an indicator function that evaluates to 1 if $m_i = k$ and 0 otherwise. ρ_k is assumed to follow a Truncated Normal distribution with parameters μ_{ρ_k} and $\sigma_{\rho_k}^2$. $\epsilon_{k,i}$ is the individual level random effect for i^{th} claim that is a member of k^{th} mixture sub-component, and it is mainly used to achieve convergence.

$$\rho_k = T.Normal(\mu_{\rho_k}, \sigma_{\rho_k}^2); \text{ and } \epsilon_{k,i} = T.Normal(\mu_{\epsilon_k}, \sigma_{\epsilon_k}^2); \quad (3.4)$$

We specify σ_k as the standard deviation of a the k^{th} mixture sub-component

$$\sigma_k \sim Uniform(0, \sigma_{max}) \quad (3.5)$$

For the standard deviation, the Uniform distribution is suggested as a viable alternative by (Gelman, 2006). The upper bound of this Uniform distribution, σ_{max} is computed via Popoviciu's inequality (Sharma et al., 2010) which requires the standard deviation of a random variable to be smaller than a function of the range of its values.

$$\sigma_k^2 \leq \frac{1}{4} (Range(\mathbf{X}))^2.$$

Since the population of payment values are known, we use its range to provide the most conservative estimate for the upper bound of σ_k .

The choice of the Gamma distribution is appropriate due to the distribution's existence on the positive scale and its ability to represent skewness (Wiper et al., 2001). Alternative distributions that exist on the positive scale include the log-Normal and the exponential distributions. The log-Normal distribution's mean and variance may not be as easily separable and harder to interpret than the Gamma distribution. On the other hand, the exponential distribution's variance is simply the square of the distribution's mean and has a rather strict relationship. Gamma distribution is utilized since it can be easily re-parametrized to accommodate a flexible distribution with separate mean and variance terms.

Having defined μ_i and specified σ as $\sigma_k^{\mathbb{I}(m_i=k)}$ allow us to re-parametrize the Gamma distribution with α_i and β_i , shape and rate parameters.

$$Y_i \sim \text{Gamma}(\alpha_i, \beta_i) \tag{3.6}$$

where $\left(\alpha_i = \frac{(\mu_i)^2}{\sigma^2}, \beta_i = \frac{\mu_i}{\sigma^2} \right)$

Given the samples \mathbf{x} , \mathbf{y} , and $\mathbf{\Omega}$ denoting all of the parameters of interest; the likelihood of the overall model with a mixture size of K can be written as

$$L(\mathbf{\Omega}; \mathbf{x}, \mathbf{y}) = \prod_{i=1}^n \left[(\pi_1 \cdot 0)^{\mathbb{I}(z_i=1)} \cdot (\pi_2 \cdot x_i)^{\mathbb{I}(z_i=2)} \right. \\ \left. \left[(1 - \pi_1 - \pi_2) \prod_{k=3}^{K+2} \left(\eta_k \cdot \text{Gamma} \left(\frac{(\rho_k + \epsilon_{k,i})^2}{\sigma_k^2}, \frac{(\rho_k + \epsilon_{k,i})}{\sigma_k^2} \right) \right)^{\mathbb{I}(m_i=k)} \right]^{\mathbb{I}(z_i=3)} \right], \tag{3.7}$$

4 Application

This section presents the application of the proposed model with a number of real world payment populations and various overpayment scenarios. Overpayment scenarios are constructed using a variety of probability vectors in order to assess the model versatility. First, we describe the payment populations and explain the construction of the overpayment scenarios. Then we discuss the model specifications with a focus on prior selection. This is followed by the analysis with respect to validity and efficiency of the model as well as the evaluation of learning aspect of the method.

4.1 Data

We use four payment populations that represent characteristics of real world claims data. Table 1 lists the descriptive statistics and Figure 1 displays the box plots for these four payment populations.

Pop	N	Mean	Median	Std. dev.	IQR	Range
1	292	4019	4042	166	0	2655
2	244	4350	4721	455	861	1178
3	500	2000	2000	0	0	0
4	2976	3098	3000	674	600	3900

Table 1: Descriptive statistics of the payment populations

First two populations are replicated using the motorized wheelchairs claims data of (Edwards et al., 2003). First population has left skewness with low standard deviation and an inter-quartile range (IQR) value of zero. Whereas the second population

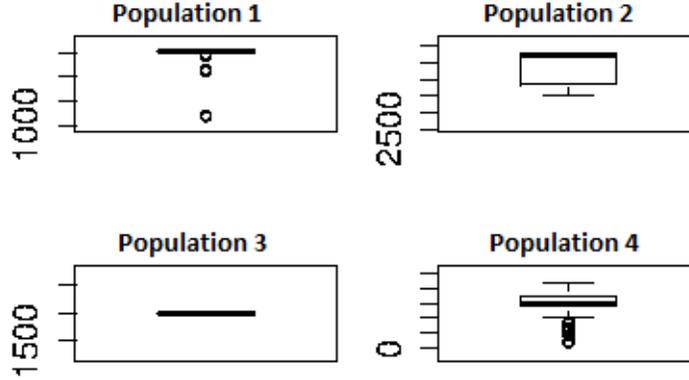


Figure 1: Box-plots of the payment populations

has a higher standard deviation and range, with strong separation from 0. Third population corresponds to the case where all payment values are same. The fourth represents a population with mixed characteristics, and it has a higher range and standard deviation. Payment data for the fourth population are retrieved from the servers of CMS (CMS, 2013b). They correspond to claims from the 2008 Outpatient Procedures file that are billed for the procedure code “J9041” (Injection of Bortezomib 0.1 mg). This procedure is selected because it was identified to have frequent overpaid billings in the recent investigations ((OIG, 2012b), (Noridian Healthcare, 2015), (Youngstrom, 2015)).

The density plot in Figure 2 shows the skewness of the payment distribution with multiple local maximums.

We consider a total of 13 overpayment scenarios to evaluate a variety of patterns. Table 2 reports the values for the joint probability vectors, $\boldsymbol{\pi} = \{\pi_1, \pi_2, (1 - \pi_1 - \pi_2)\eta_3, (1 - \pi_1 - \pi_2)(1 - \eta_3)\}$ for all these scenarios.

Scenarios 1-10 represent populations in which each claim is either completely legit-

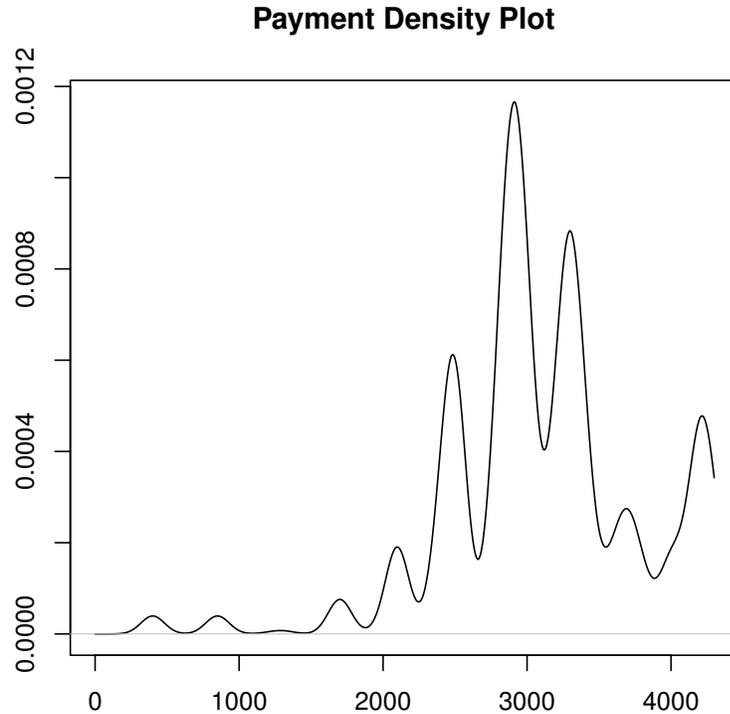


Figure 2: Density plot of the fourth payment population

imate and completely illegitimate, so called “all or nothing” case. The analysis in literature have focused on these cases, for example (Edwards et al., 2003) and (Mohr, 2005). On the other hand, Scenarios 11-13 include partial overpayments in addition to completely legitimate and completely illegitimate claims. It is assumed that there are two partial overpayment sub-populations in addition to the spikes that represent completely legitimate and completely illegitimate claims. These two partial overpayment sub-populations are simulated randomly using two Beta distributions, $Beta(0.15, 0.85)$ and $Beta(0.85, 0.15)$, which result in mean overpayment percentages of 0.15 and 0.85 respectively. Specifically, Scenario 11 represents the pattern in which the first partial overpayment pattern has the higher probability compared

Overp. Scenario	π_1	π_2	$(1 - \pi_1 - \pi_2)\eta_3$	$(1 - \pi_1 - \pi_2)(1 - \eta_3)$
Scenario 1	1.00	0.00	0.00	0.00
Scenario 2	0.95	0.05	0.00	0.00
Scenario 3	0.90	0.10	0.00	0.00
Scenario 4	0.70	0.30	0.00	0.00
Scenario 5	0.50	0.50	0.00	0.00
Scenario 6	0.30	0.70	0.00	0.00
Scenario 7	0.25	0.75	0.00	0.00
Scenario 8	0.10	0.90	0.00	0.00
Scenario 9	0.05	0.95	0.00	0.00
Scenario 10	0.00	1.00	0.00	0.00
Scenario 11	0.25	0.50	0.15	0.10
Scenario 12	0.50	0.25	0.10	0.15
Scenario 13	0.75	0.10	0.05	0.10

Table 2: Joint probability vectors for each overpayment scenario

to the second. Whereas in Scenario 12, second partial overpayment pattern has a higher probability. Scenario 11 also represents a population with higher probability of overpayment compared to no overpayments. In Scenario 12, these probabilities are equal to each other. Scenario 13 represents a case in which the majority of the population consists of completely legitimate claims.

4.2 Model Specification

This subsection describes the details of the model specification with a focus on prior selection. The prior selection in this paper reflects a relative lack of knowledge on

the part of the modeler about overpayments. The choice of weakly informative prior values let the data to drive the learning process about the posterior distribution of parameters.

It is assumed that the modeler lacks knowledge of the frequency of main components and mixture sub-components, and the respective probability vectors of $\boldsymbol{\pi}$ and $\boldsymbol{\eta}$. The probability vector, $\boldsymbol{\pi} = \{\pi_1, \pi_2, 1 - \pi_1 - \pi_2\}$ has a Dirichlet prior with the values (0.01, 0.01, 0.01) which makes the effect of prior negligible. Similarly, $\boldsymbol{\eta}$ has a Dirichlet prior with hyper-parameter values of 0.01. For the case where $K = 2$, the prior distribution of η_3 is assumed to be uniform with parameters 0 and 1, that corresponds to equally likely outcomes.

The prior for μ_{ρ_k} is chosen to follow a left truncated Normal distribution at 0, with a mean of 0 and a variance of 100. This allows ρ_k to vary while restricting the mean values to the non-negative region. The range of the standard deviation is determined by the choice of bounds of the Uniform distribution. In our model specification, the standard deviation, σ_{ρ_k} is assumed to follow a Uniform distribution with parameters 0 and 2150. We use the payment range to provide a conservative estimate for the upper bound of σ_k .

We set the hyper-parameters for $\epsilon_{k,i}$ as zero mean and a standard deviation of 0.1. The main purpose of using $\epsilon_{k,i}$ is to achieve convergence, and having $E[\epsilon_{k,i}] = 0$ results with a negligible impact on the μ_i .

4.3 Analysis

This subsection presents a discussion of the validity of the model estimates with respect to the current governmental guidelines. Then it proceeds to provide an efficiency comparison with the basic use of CLT. Lastly, an evaluation of learning aspect of the proposed method is presented. Proposed model is referred to as M.1 and the approach of CLT using the sample to retrieve the relevant estimates is called as M.2.

The proposed model is run for a sample size of 50 for each payment population and overpayment scenario listed in Table 2. Monte Carlo Markov Chain (MCMC) simulation is conducted using the algorithm of (Thomas et al., 2006) within OpenBugs software and the estimates are utilized using the R software (R Core Team, 2014). The MCMC simulations are run for 3 independent chains where 20,000 samples from these chains are analyzed after discarding the first 200,000 samples as burn-in. The Brooks-Gelman-Rubin (BGR) statistic (Brooks and Gelman, 1998) via the *R coda* package of (Plummer et al., 2006) is utilized to assess convergence. The chains are judged to have practically converged when BGR statistics are less than 1.05 for all parameters.

The posterior overpayment mean, \hat{Y} , is estimated using posterior parameter estimates. Whereas the estimate of the overpayment variance, $\hat{\sigma}_Y^2$ is retrieved from the posterior predictive distribution of overpayment. The lower bound of the one sided 90 % confidence interval for a given sample size n adjusted for the finite population size N is computed as the difference of mean estimate and the margin of error.

$$\hat{Y}_{lower,M.1} = \hat{Y} - t_{0.9}^{n-1} \sqrt{\frac{N-n}{N-1}} \frac{\hat{\sigma}_Y}{\sqrt{n}}$$

where $t_{0.9}^{n-1}$ is the 90th percentile of Student's t-distribution with $(n - 1)$ degrees of freedom. This lower bound estimate of the proposed model, $\hat{Y}_{lower,M.1}$ can be used as the recoupment amount according to the governmental guidelines.

Next, we present the computation of the lower bound of a one-sided 90 % confidence interval of the overpayment, $\hat{Y}_{lower,M.2}$,

$$\hat{Y}_{lower,M.2} = \bar{Y} - t_{0.9}^{n-1} \sqrt{\frac{N-n}{N-1} \frac{s_Y}{\sqrt{n}}}$$

where \bar{Y} and s_Y are respectively the sample mean and standard deviation. This is used as the estimate of the recoupment amount with respect to M.2.

4.3.1 Validity Assessment

In order to assess the validity of the methods, average coverage probabilities are computed for both models. For a given number of simulation replications, the frequency of the times that the sample mean overpayment value is greater than the lower 90 % bound provides the estimated average coverage probability.

Figure 3 provides the average coverage probabilities for the proposed model compared to the nominal 90 % confidence level which is represented as a dashed horizontal line. For payment populations 1 and 2, the proposed model is found to be valid for all cases other than Scenario 2. In fact, our lower bound estimates for most scenarios are conservative than necessary. For those two cases, model results in average coverage probabilities of 82% and 86%. This can potentially be explained by the relatively high standard deviation due to low overpayment rate and lack of learning. CLT based methods are shown to have coverage levels that are lower than the nominal 90% level for some populations including the ones with high overpayment rate (Edwards et al.,

2003). For such cases, the proposed model is found to be even conservative for many populations of interest.

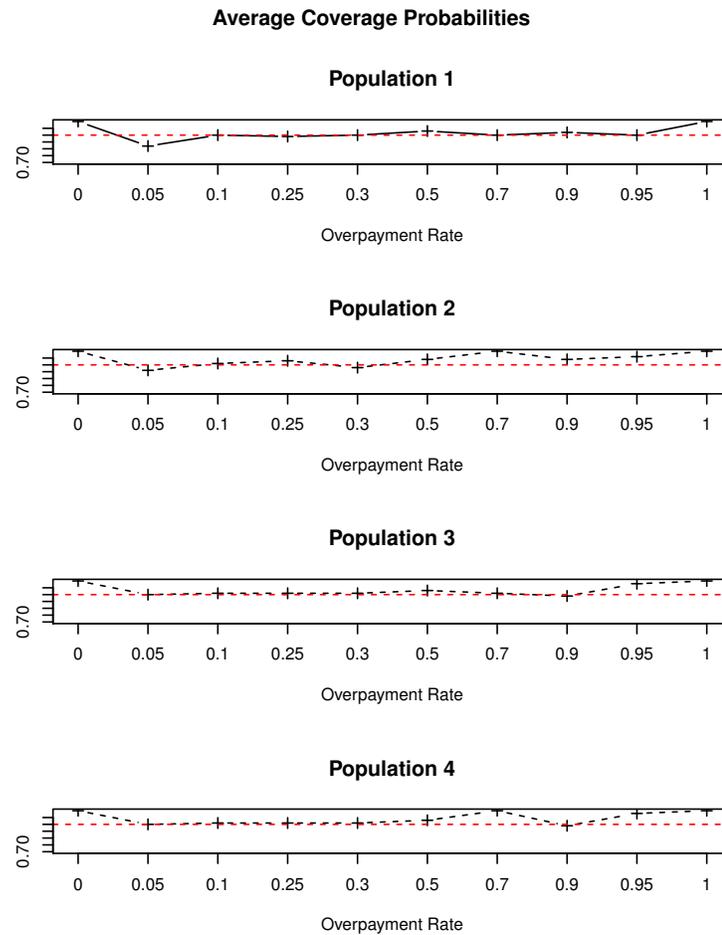


Figure 3: Average Coverage Probabilities for Overpayment Scenarios 1-10

Next, we analyze the validity with respect to the overpayment scenarios 11-13 that have partial overpayments. Table 3 lists the average coverage probabilities which show evidence for validity for the model.

In the case of Scenarios 12 and 13, for all payment populations the average coverage probabilities are found to be higher than or very close to the nominal level. This provides evidence for the validity of the model. However, for Scenario 11 the model

Payment Population	Scenario 11	Scenario 12	Scenario 13
1	80 %	90 %	95 %
2	93 %	94 %	91 %
3	81 %	91 %	94 %
4	89 %	93%	91 %

Table 3: Average Coverage Probabilities for Overpayment Scenarios 11-13

resulted in average coverage probabilities of 80% and 81% for payment populations 1 and 3, respectively. This can be potentially explained by the existence of 75% overpaid claims, a third of which have partial overpayments.

4.3.2 Efficiency Studies

Next, we explore the efficiency of the proposed model compared to the CLT based method. Particularly, we compare the overpayment recovery estimates of the proposed models $\hat{Y}_{lower,M.1}^m$ and $\hat{Y}_{lower,M.2}^m$ with the mean overpayment of the population, $E(Y)$. In so doing, the equations of Mean Absolute Percentage Error (MAPE) below are used for each model with $T = 100$ replications:

$$MAPE_{M.1}^m = \sum_{t=1}^{t=T} \left| \frac{E(Y)^m - \hat{Y}_{lower,M.1,t}^m}{E(Y)^m} \right| \text{ and } MAPE_{M.2}^m = \sum_{t=1}^{t=T} \left| \frac{E(Y)^m - \hat{Y}_{lower,M.2,t}^m}{E(Y)^m} \right| \quad (4.1)$$

where $E(Y)^m$ is the mean of the overpayment population for Scenario m and $\hat{Y}_{lower,M.1,t}^m$ and $\hat{Y}_{lower,M.2,t}^m$ are the respective model's overpayment estimates for the t^{th} replication in Scenario m .

Figure 4 presents the MAPE values for Scenarios 1-10 and each of the four payment populations for the proposed model. For Model 1, overall average efficiency is found

to increase (MAPE values decrease) for the scenarios that have high overpayment rates.

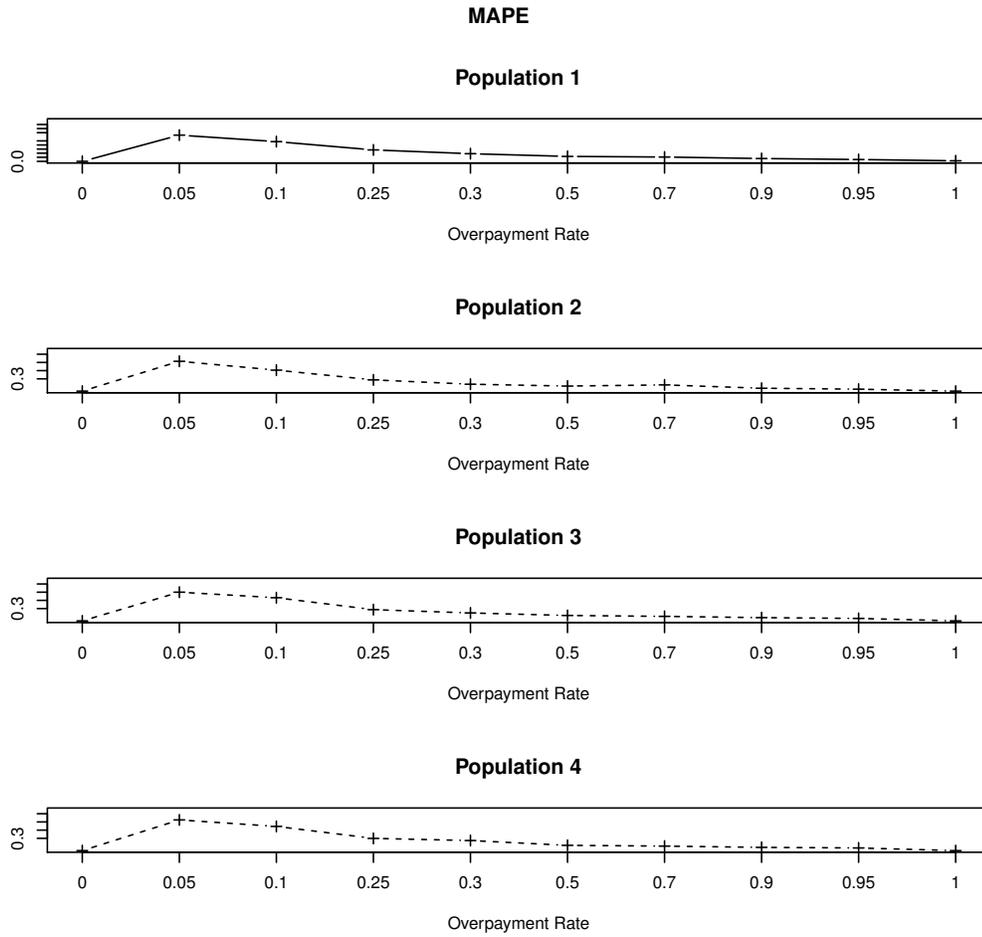


Figure 4: Mean Absolute Percentage Error for Overpayment Scenarios 1-10

Table 4 lists the MAPE values of both methods for Scenarios 11-13. In general, this implies that the proposed model outperforms the CLT based model when there exists partial overpayments.

In the cases of Scenarios 11 and 12, M.1 provides better estimation compared to M.2 for all payment populations. However, the superiority of M.1 is small for Populations 1 and 2, while they are same for the third population. For the fourth payment pop-

Payment Population.	Model	Scenario 11	Scenario 12	Scenario 13
1	M.1.	9 %	20 %	32 %
1	M.2.	15 %	24 %	35 %
2	M.1.	9 %	21 %	34 %
2	M.2.	13 %	23 %	34 %
3	M.1.	10 %	19 %	50 %
3	M.2.	15 %	24 %	50 %
4	M.1.	16 %	19 %	6 %
4	M.2.	26 %	41 %	55 %

Table 4: Mean Absolute Percentage Error for Overpayment Scenarios 11-13

ulation with mixed distribution characteristics, the proposed model provide highly superior efficiency compared to the M.2.

Despite the increase in efficiency, the errors are still relatively significant from the modeling perspective. Therefore, we also provide a discussion of a couple modeling extensions in Section 6.

When we repeat the analysis with a number of sample sizes, ($n=25$, $n=50$, $n=75$, $n=100$), there are not any significant changes in the results. As expected, the efficiency of the both methods improve due to better estimation. However, the patterns stay similar.

4.3.3 Evaluation of Learning

This paper focuses on the validity and efficiency of the Bayesian posterior estimates to be used as the recovery amount in medical audits. In addition, Bayesian estimation

	\bar{Y}	\bar{X}	$\bar{\pi}_1$	$\bar{\pi}_2$	$1 - \bar{\pi}_1 - \bar{\pi}_2$
n= 25	872.74	1606.80	0.32	0.24	0.44
n=100	606.21	1276.80	0.36	0.26	0.38

Table 5: Descriptive Statistics of the sample with sample sizes 25 and 100

and inference also provide probability interpretations on quantities of interest such as hypotheses, intervals of parameters, membership of a subject and model selection (Jackman, 2009). Proposed model can help the modeler to quantify the learning about a certain overpayment characteristic. For instance, the changes in the posterior distributions of $\boldsymbol{\rho}$ provide an understanding of the existence of partial overpayment patterns.

In order to illustrate this trait of the model, a new overpayment scenario is constructed with $\{(\pi_1, \pi_2, (1 - \pi_1 - \pi_2)\eta_3, (1 - \pi_1 - \pi_2)(1 - \eta_3)) = (0.25, 0.15, 0.2, 0.4)\}$. The descriptive statistics for these samples are summarized in Table 5. We run 3 independent MCMC chains for samples with sizes $n = 25$ and $n = 100$. When $n = 25$, after discarding 300,000 iterations on each chain, the convergence is attained and 100,000 samples are used with a thinning size of 20. For the case of $n = 100$, we achieved convergence after running the chains for 350,000 iterations which are discarded as burn-in and 126,000 iterations are used as posterior samples with a thinning size of 10.

Table 6 reports the posterior descriptive statistics of the parameters for samples with size of 25 and 100 respectively. In addition to the changes in the posterior means, the standard deviation values of the parameters decrease with an increase in sample size, providing evidence of learning about the parameters with additional

	Mean (SD), n=25	Mean (SD), n=100
π_1	0.32 (0.09)	0.27 (0.04)
π_2	0.20 (0.08)	0.49 (0.05)
$1 - \pi_1 - \pi_2$	0.48 (0.10)	0.24 (0.04)
η_3	0.46 (0.33)	0.32 (0.07)
σ	886.60 (814.44)	289.58 (39.92)
ρ_3	0.00 (0.02)	0.08 (0.01)
ρ_4	0.67 (0.45)	0.95 (0.04)

Table 6: Parameter estimates of M.1 for Scenario 5 with Sample Sizes 25 and 100

data. This is further illustrated via Figure 5 that show the posterior distributions of the parameters. The differences between the prior and posterior distributions indicate the extent of learning about the uncertainty of these parameters. Especially in the cases of heterogeneous and large size of claims data, investigators may want to use probe samples to have an initial validation on their hypotheses. (Ignatova and Edwards, 2008) presents such a sampling plan called 30-6-3, where thirty payments are randomly sampled, and the first six examined. If at least three of them are found completely illegitimate, the rest of the sample is examined.

The evidence from learning can also be used for model validation. The change between the prior and posterior distributions, learning, can help the modeler to validate prior choice, the mixture size and sample size. Figure 5 indicates there is an evidence of learning for $\boldsymbol{\pi}$ and ρ_3 . Whereas the negligible change in distributions of $\boldsymbol{\rho}$ signals a lack of learning regarding the uncertainty.

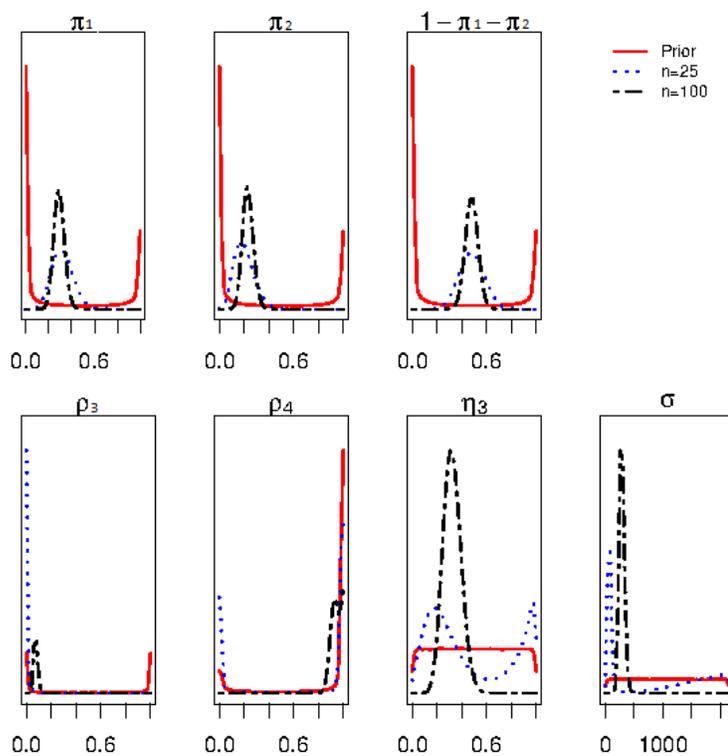


Figure 5: Density Plots of Posterior Distributions

5 Model Selection

([Edwards et al., 2015](#)) suggest that model based approaches may be vulnerable to abuse and they are open to arguments in a legal proceeding. In addition, they point out the unethical practice of some government contractors in order to increase recoupment in their extrapolations. This section provides suggestions and remarks about the use of our model with a focus on model selection.

Model uncertainty and selection are important issues in statistics ([Hoeting et al., 1999](#)). For instance, one can argue that medical overpayments are complex as it stands, and simple and parsimonious models might be preferred. However, in some cases the heterogeneous nature of the claims data results in inefficiencies which moti-

vates the need for more complex models. The proposed method is such a model that considers zero and one inflation as well as mixtures. It is shown to be an efficient alternative in the case of partial overpayments. However, our model may not be preferred in a number of cases because of its complexity. For instance, in cases without any observations of partial overpayments in the sample, our model may be more complicated than necessary. Instead, we recommend the auditors to use other approaches based on CLT or Hypergeometric likelihood for such “all or nothing” cases. Development of a more comprehensive method that is based on model averaging (Hoeting et al., 1999) is beyond the scope of this paper.

Another important matter in a Bayesian model is the choice of prior distributions. As Lindley notes in section 8 of his seminal paper (Lindley, 2000), the prior distribution is the quantification of a researcher’s uncertainty over model parameters which may have explicit physical interpretations. This can let the auditors to utilize the available knowledge via elicitation of priors from experts. In a related example, (Matsumura et al., 1991) use an informative prior that has more weight on zero and fully overpayments. However, this should be used with caution since it can certainly be argued that an analyst may unethically select priors that lead to his/her desired outcome. As a method of check and balance, the prior and posterior distributions can be compared to assess the impact of data and if learning has occurred. Nevertheless, in this paper, we recommend the use of weakly informative prior distributions. For instance, the prior over π quantifies our relative lack of knowledge on the probability of overpayment of a given claim. As suggested by Gelman (Gelman, 2009), weakly informative prior values are used to allow the data to drive the learning process about the posterior distribution of parameters.

([Edwards et al., 2015](#)) also point out that Bayesian approaches may be abused when the provider might be asked for recoupment although the sample shows no evidence of impropriety. Our computational results show the proposed model with the use of weakly informative priors is valid for such cases. In computing the 90% lower confidence bound, the small but positive posterior mean (0.01) is offset by the consideration of standard deviation and the use of lower bound of 90% confidence interval. Furthermore, the probability of no impropriety existing in a randomly drawn claim is computed to be higher than 99%. This points out a conservative result, which shows evidence for the fair treatment of the provider. Since we assume non-negative overpayments, potentially negative lower bound values are truncated to 0 as the recovery amount. Therefore, a case with a sample of all legitimate claims should be dismissed since the provider can be assumed to be innocent beyond a reasonable doubt. To give a concrete example, suppose we are provided a sample with 25 claims of which all are found to be legitimate. We assume that the prior for π is weakly informative, following a Dirichlet distribution with $\{0.01, 0.01, 0.01\}$. This prior distribution represents our relative lack of knowledge about zero, fully and partial overpayments; and it implies these events are equally likely. Due to the conjugacy of Dirichlet-Categorical distributions, the posterior probabilities are $(\{0.992, 0.004, 0.004\})$ for the fully legitimate, fully overpaid and partially overpaid claims respectively. The probability 0.004 represents our uncertainty of observing fully illegitimate and partially overpaid claims in the remainder of the population with $N - 25$ claims yet to be inspected. There is no way to know the overpayment values of all claims unless they are all audited. We can use such comparisons between prior and posterior distributions to quantify learning from data. Our models are similarly found to be valid for the case where all of the payments are illegitimate

as shown in Figure 3. The average coverage probabilities are found to be well above 90%, suggesting overly conservative estimates.

6 Modeling Extensions

This section presents two extensions of the proposed model, and provide brief, but not comprehensive, assessments of validity and efficiency.

6.1 Zero-one Inflated Finite Mixture Model

The proposed model with unknown number of mixtures may be unnecessarily complex if the auditor has enough evidence about the number of partial overpayment patterns. We introduce a more simple zero-one inflated model with known number of mixture sub-components and refer to it as *M.3*. For the sake of brevity, we do not present a comprehensive validity and efficiency analysis for these extensions. The results for the payment population 4 is presented. We have chosen Scenario 4 as a representative of the overpayment scenarios without any partial overpayments, and consider overpayment scenarios 4, 11, 12 and 13.

The average coverage probabilities with respect to the Scenarios 4, 11, 12 and 13 for a sample size of 50 are computed as 90%, 95%, 96%, 95%, respectively. Table 7 lists the mean absolute percentage error values of recovery estimates for sample sizes of 25, 50, 75, 100. As expected, the MAPE values for both models improve and become smaller for increasing sample sizes. *M.3* outperforms the CLT based model (*M.2*) for the first two scenarios based on the smaller MAPE values. The MAPE results are comparable for the third and fourth scenarios.

	Scenario 4	Scenario 11	Scenario 12	Scenario 13
M.3 n=25	16%	31%	52%	64%
M.2 n=25	40%	45%	52 %	77 %
M.3 n=50	12%	23%	42%	58%
M.2 n=50	26%	26%	41%	55%
M.3 n=75	10%	18%	35%	54%
M.2 n=75	22%	23%	30%	49 %
M.3 n=100	9%	18%	35%	50%
M.2 n=100	22%	24%	27%	43 %

Table 7: MAPE of the Overpayment Estimates

In addition to the magnitude, an analyst may also be interested in the direction of the estimation errors. This can be measured by the Mean Percentage Error (MPE) values. We compute MPE with $T = 100$ replications using the equations;

$$MPE_{m,M.3} = \sum_{t=1}^{t=T} \frac{E(Y)^m - \hat{Y}_{lower,M.3,t}^m}{E(Y)^m} \quad (6.1)$$

and

$$MPE_{m,M.2} = \sum_{t=1}^{t=T} \frac{E(Y)^m - \hat{Y}_{lower,M.2,t}^m}{E(Y)^m} \quad (6.2)$$

Positive MPE values indicate that the overpayment is underestimated whereas negative MPE values demonstrate the cases where the provider may unfairly be asked to pay back more than the actual overpayment amount. In order to have conservative recoupment demands from providers, we would prefer the MPE values to be positive rather than negative if there is same amount of the error. This is similar to the reasoning that leads to the recommended use of conservative estimates such as lower

bound of the 90 % confidence interval. We note that ideally the estimates would be close to zero as possible which would provide evidence for the unbiasedness of the estimates.

Figure 6 presents the box-plots of MPE values for both models when the sample size is 50 for all 4 overpayment scenarios. The long horizontal lines that accompany each box-plot in Figure 6 correspond to 0. For Scenario 4, the proposed model has lower positive median values and lower inter-quartile range, therefore provides a better performance. For scenarios 11-13, the median values are comparable albeit the variance and inter-quartile range of the proposed model are lower compared to M.2. In general, it can be argued that the smaller number of negative MPE values for M.3 indicates the conservative estimation by the proposed model compared to M.2. Overall, it can be suggested that even the simpler model M.3 provides at least as efficient estimates as M.2, if not better.

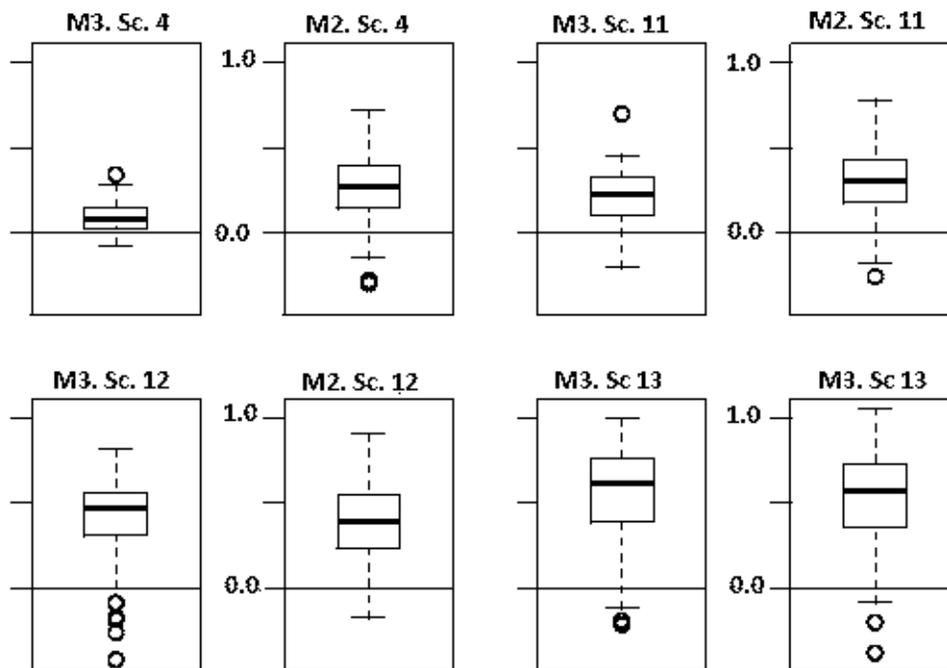


Figure 6: Mean Percentage Errors (MPE) of M.1 and M.2 for all 4 scenarios

6.2 Model with Covariates

Next, we explore the feasibility of a modeling extension of the proposed model that utilizes co-variate information. The main difference is the explicit modeling of the inverse mean overpayment ratio, $\frac{1}{\rho}$. This parameter is in positive region which motivated us to use the log transformation and fit a linear equation.

$$\log\left(\frac{1}{\rho}\right) = \phi_0 + \phi_1 X_p \quad (6.3)$$

We have selected the data of payment population 4 and Scenario 13 for demonstration. The main characteristic of Scenario 13 is the limited number of claims with partial overpayments in the sample. For instance in a sample with size 50, the num-

	n	Mean	St. Dev.
<i>Completely Legitimate Claims</i>	37	1281.08 (0)	1478.43 (0)
<i>Partially OverPaid Claims</i>	8	2362.50 (1218.75)	1424.22 (1102.78)
<i>Fully OverPaid Claims</i>	5	1092.00 (1092.00)	1426.65 (1426.65)
X_p	8	0.38	0.52

Table 8: Descriptive Statistics of the Payment(Overpayment)

ber of expected partial overpayments is 7.5. This results in poor estimation and lack of learning about model parameters. As a potential remedy, we utilize a binary co-variate such as provider type, X_p , while modeling partial overpayments. Table 8 presents the descriptive statistics of payment and overpayment values of different components and the co-variate for the sample with size 50.

Table 9 displays the descriptive statistics of the posterior parameters. The maximum number of mixture sub-components, U is assumed to be 8, since there are 8 partially overpaid claims in this particular sample. As can be seen in the table, K , the number of mixture sub-components, has a median of 2, which corresponds to two partial overpayment patterns. Despite the fact that $\boldsymbol{\eta}$ is a vector of length 8, we only report the top 3 η 's descriptive statistics since they cumulatively correspond to a total of 91% of the mixture sub-component probabilities. The standard deviation of the mixture sub-components are also reported where σ_k is the k 'th mixture sub-component's standard deviation. The $\boldsymbol{\sigma}$ does not exhibit much difference between its elements however considering the posterior value of K , and the median differences between σ_3 and σ_4 , we can infer that there are two sub mixture populations.

We obtain the estimates of ρ as 0.53 and 0.16 for the cases of $X_p = 0$ and $X_p = 1$,

	Mean	S.D.	2.5%	Median	97.5%
K	2.11	0.94	1.0	2.0	4.0
ϕ_0	0.64	0.27	0.14	0.63	1.23
ϕ_1	1.19	0.30	0.53	1.21	1.73
π_1	0.74	0.06	0.61	0.74	0.85
π_2	0.10	0.04	0.03	0.10	0.20
$(1 - \pi_1 - \pi_2)$	0.16	0.05	0.07	0.16	0.27
η_3	0.61	0.32	0.02	0.68	1.00
η_4	0.21	0.23	0.00	0.11	0.82
η_5	0.09	0.13	0.00	0.03	0.50
σ_3	1.67	1.02	1.01	1.34	4.70
σ_4	1.80	1.26	1.01	1.37	5.99
σ_5	1.62	0.92	1.01	1.33	4.22

Table 9: Descriptive Statistics of Posterior Parameters

respectively. A comprehensive analysis is required about coverage probabilities to assess whether this would be a valid alternative. This would require a large number of repetitions at the expense of a computational cost. Efficient computation methods for such models is left for future research.

7 Conclusion

This paper proposes a Bayesian zero-one inflated mixture based overpayment estimation model. First, we show the validity of the model with respect to the governmental guidelines. Then, the model has been shown to be efficient for some overpayment scenarios including the ones with partial overpayments. In addition, it can describe the uncertainty about the overpayment population. The learning aspect with the increasing sample size can be used by investigators to improve their understanding of a given claims data set. For instance, the proposed model can be valuable for OIG when asked to investigate the potential fraudulent activities in an hospital which has submitted claims within many different domains. Methodologically, this is one of the rare applications of Gamma Inflated mixture models in literature. We also present a discussion of model selection and discuss potential extensions.

We should emphasize that our approach requires explicit modeling of the overpayment probabilities and percentages, and it comes at an additional computational cost and some statistical knowledge. Therefore, we recommend the proposed model to be used for claims data with multi-modal overpayment patterns, for which the existing models may not be as efficient. For instance, one barrier to application can be model specification due to the relatively unknown aspects of Bayesian modeling

by medical auditors. This paper makes the prior selection with an assumption of relative lack of knowledge on the part of the modeler about overpayments. Although the use of Bayesian approaches is not common in medical audits, we believe this is a modest but crucial step to be able to represent overpayment better. Achieving MCMC convergence can also be an issue especially for cases with a high number of partial overpayment patterns. It has been recognized that for scenarios with many legitimate claims, the model may not be expected to perform well. The main reason is that the amount of learning does not occur due to the lack of non-zero overpayments. We have presented modeling extensions as potential remedies. The model fit can potentially be improved by considering other co-variates such as the monetary amount of the billing.

It would be of interest to incorporate the model within a dynamic sampling decision making procedure by utilizing posterior predictive distributions. As future research, a number of modeling extensions can be proposed. For instance, use of ordered Dirichlet parameters can improve the estimation of partial overpayment mean values and can resolve potential identifiability issues associated with the use of latent variables. Clustering algorithms can be considered to segment types of claims as part of the modeling framework. Lastly, the use of the Bayesian inflated mixture models in the domain of tax auditing may also be an interesting research topic.

References

- (2012). The problem of health care fraud. National Health Care Anti Fraud Association. <http://www.nhcaa.org/resources/>

[health-care-anti-fraud-resources/the-problem-of-health-care-fraud.aspx](#). Accessed: 1/18/2014.

Aitchison, J. (1955). On the distribution of a positive random variable having a discrete probability mass at the origin. *Journal of the american statistical association*, **50**(271), 901–908.

Anderson, G. and Hussey, P. (2001). Comparing health system performance in oecd countries. *Health Affairs*, **20**(3), 219–232.

Brooks, S. P. and Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, **7**(4), 434–455.

Buddhakulsomsiri, J. and Parthanadee, P. (2008). Stratified random sampling for estimating billing accuracy in health care systems. *Health Care Management Science*, **11**(1), 41–54.

CMS (2001). Program Memorandum Carriers Transmittal B-01-01. Accessed: 06/03/2015.

CMS (2011). Medicare Program Integrity Manual Chapter 8 Administrative actions and statistical sampling for overpayment estimates. <https://www.cms.gov/Regulations-and-Guidance/Guidance/Manuals/downloads/pim83c08.pdf>. Accessed: 03/14/2015.

CMS (2013a). Medicare urges seniors to join fight against fraud. URL <http://www.cms.gov/Newsroom/MediaReleaseDatabase/Press-Releases/2013-Press-Releases-Items/2013-06-06.html>. Accessed: 09/20/2016.

CMS (2013b). Basic stand alone (bsa) Medicare claims public use files (pufs). Accessed: 07/19/2015.

CMS (2014a). National health expenditure data. Accessed: 06/06/2015.

CMS (2014b). Medicare Program Integrity Manual Chapter 3 - verifying potential errors and taking corrective actions. Accessed: 10/19/2014.

Cochran, W. G. (2007). *Sampling techniques*. John Wiley & Sons.

Cox, D. and Snell, E. (1979). On sampling and the estimation of rare errors. *Biometrika*, **66**(1), 125–132.

Cruyff, M. J., Böckenholt, U., van den Hout, A., and van der Heijden, P. G. (2008). Accounting for self-protective responses in randomized response data from a social security survey using the zero-inflated Poisson model. *The Annals of Applied Statistics*, **2**(1), 316–331.

Daniel, J. (2011). *Sampling essentials: Practical guidelines for making sampling choices*. Sage Publications.

Edwards, D. (2011). On stratified sampling and ratio estimation in Medicare and Medicaid benefit integrity investigations. *Health Services and Outcomes Research Methodology*, **11**(1-2), 79–94.

Edwards, D., Ward-Besser, G., Lasecki, J., Parker, B., Wieduwilt, K., Wu, F., and Moorhead, P. (2003). The minimum sum method: a distribution-free sampling procedure for Medicare fraud investigations. *Health Services and Outcomes Research Methodology*, **4**(4), 241–263.

- Edwards, D., Gilliland, D., Ward-Besser, G., and Lasecki, J. (2015). Conservative penny sampling. *Journal of Survey Statistics and Methodology*, **3**(4), 504–523.
- Ekin, T., Musal, R. M., and Fulton, L. V. (2015). Overpayment models for medical audits: multiple scenarios. *Journal of Applied Statistics*, **42**(11), 2391–2405.
- Erosheva, E. A., Fienberg, S. E., and Joutard, C. (2007). Describing disability through individual-level mixture models for multivariate binary data. *The Annals of Applied Statistics*, **1**(2), 346–384.
- Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems. *The Annals of Statistics*, **1**(2), 209–230.
- Feuerverger, A. (1979). On some methods of analysis for weather experiments. *Biometrika*, **66**(3), 655–658.
- Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Analysis*, **1**(3), 515–534.
- Gelman, A. (2009). Bayes, jeffreys, prior distributions and the philosophy of statistics. *Statistical Science*, **24**(2), 176–178.
- Ghosh, M. and Meeden, G. (1997). *Bayesian methods for finite population sampling*, volume 79. CRC Press.
- Ghosh, S. K., Mukhopadhyay, P., and Lu, J.-C. J. (2006). Bayesian analysis of zero-inflated regression models. *Journal of Statistical Planning and Inference*, **136**(4), 1360–1375.

- Gilliland, D. and Edwards, D. (2010). Using randomized confidence limits to balance risk: An application to Medicare fraud investigations. *Statistics and Probability Research Memorandum RM-685, Michigan State University*.
- Gilliland, D. and Feng, W. (2010). An adaptation of the minimum sum method. *Health Services and Outcomes Research Methodology*, **10**(3-4), 154–164.
- Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, **82**(4), 711–732.
- Guthrie, D. (1989). Statistical models and analysis in auditing: Panel on nonstandard mixtures of distributions. *Statistical Science*, **4**(1), 2–33.
- Hoeting, J. A., Madigan, D., Raftery, A. E., and Volinsky, C. T. (1999). Bayesian Model Averaging: a tutorial (with comments by M. Clyde, David Draper and E. I. George, and a rejoinder by the authors). *Statistical Science*, **14**(4), 382–417.
- Ignatova, I. and Edwards, D. (2008). Probe samples and the minimum sum method for Medicare fraud investigations. *Health Services and Outcomes Research Methodology*, **8**(4), 209–221.
- Jackman, S. (2009). *Bayesian Analysis for the Social Sciences*. Wiley, New York.
- King, B. and Madansky, A. (2013). On sampling design issues when dealing with zeros. *Journal of Survey Statistics and Methodology*, **1**(2), 144–170.
- Lambert, D. (1992). Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics*, **34**(1), 1–14.
- Li, J., Huang, K.-Y., Jin, J., and Shi, J. (2008). A survey on statistical methods for health care fraud detection. *Health Care Management Science*, **11**(3), 275–287.

- Lindley, D. V. (2000). The philosophy of statistics. *Journal of the Royal Statistical Society: Series D (The Statistician)*, **49**(3), 293–337.
- Matsumura, E. M., Plante, R., Tsui, K.-W., and Kannan, P. (1991). Comparative performance of two multinomial-based methods for obtaining lower bounds on the total overstatement error in accounting populations. *Journal of Business & Economic Statistics*, **9**(4), 423–429.
- Min, Y. and Agresti, A. (2005). Random effect models for repeated measures of zero-inflated count data. *Statistical Modelling*, **5**(1), 1–19.
- Mohr, D. L. (2005). Confidence limits for estimates of totals from stratified samples, with application to Medicare Part B overpayment audits. *Journal of Applied Statistics*, **32**(7), 757–769.
- Muralidharan, O. (2010). An empirical bayes mixture method for effect size and false discovery rate estimation. *The Annals of Applied Statistics*, **4**(1), 422–438.
- Musal, R. (2010). Two models to investigate Medicare fraud within unsupervised databases. *Expert Systems with Applications*, **37**(12), 8628–8633.
- Neelon, B. H., OMalley, A. J., and Normand, S.-L. T. (2010). A bayesian model for repeated measures zero-inflated count data with application to outpatient psychiatric service use. *Statistical Modelling*, **10**(4), 421–439.
- Noridian Healthcare, S. (2015). Outpatient drug J9041 - CA service specific probe review notification. <https://med.noridianmedicare.com/web/jea/cert-reviews/mr/notifications-findings/outpatient-drug-j9041-ca-service-specific-probe-review-notification>. Accessed: 11/07/2016.

OIG (2012a). Inappropriate payments to skilled nursing facilities cost Medicare more than a billion in 2009. Accessed: 06/09/2015.

OIG (2012b). Review of Medicare outpatient billing for selected drugs at Essentia Health Duluth. Accessed:06/06/2015.

OIG (2013a). Annual report for fiscal year 2013. Accessed: 10/23/2014.

OIG (2013b). Oig strategical plan 2014-2018. URL <https://oig.hhs.gov/reports-and-publications/strategic-plan/files/OIG-Strategic-Plan-2014-2018.pdf>. Accessed: 03/03/2015.

OIG (2013c). Medicare Compliance review of University of Miami Hospital. Accessed: 10/14/2014.

OIG (2014a). Cigna Healthcare of Arizona, Inc. (Contract H0354), submitted many diagnoses to The Centers for Medicare and Medicaid Services that did not comply with federal requirements for calendar year 2007. <https://oig.hhs.gov/oas/reports/region7/71001082.pdf>. Accessed: 10/14/2014.

OIG (2014b). Improper payments for evaluation and management services cost Medicare billions in 2010. Accessed: 06/09/2015.

Ospina, R. and Ferrari, S. L. (2012). A general class of zero-or-one inflated Beta regression models. *Computational Statistics and Data Analysis*, **56**(6), 1609–1623.

Plummer, M., Best, N., Cowles, K., and Vines, K. (2006). Coda: Convergence diagnosis and output analysis for MCMC. *R News*, **6**(1), 7–11.

R Core Team (2014). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

- Sharma, R., Gupta, M., and Kapoor, G. (2010). Some better bounds on the variance with applications. *Journal of Mathematical Inequalities*, **4**(3), 355–363.
- Shin, H., Park, H., Lee, J., and Jhee, W. (2012). A scoring model to detect abusive billing patterns in health insurance claims. *Expert Systems with Applications*, **39**(8), 7441–7450.
- Thomas, A., O’Hara, B., Ligges, U., and Sturtz, S. (2006). Making BUGS open. *R News*, **6**(1), 12–17.
- Tsui, K.-W., Matsumura, E. M., and Tsui, K.-L. (1985). Multinomial-dirichlet bounds for dollar-unit sampling in auditing. *Accounting Review*, **60**(1), 76–96.
- Venturini, S., Dominici, F., and Parmigiani, G. (2006). Gamma shape mixtures for heavy-tailed distributions. *The Annals of Applied Statistics*, **2**(2), 756–776.
- Webb, A. R. (2000). Gamma mixture models for target recognition. *Pattern Recognition*, **33**(12), 2045 – 2054.
- Wiper, M., Insua, D. R., and Ruggeri, F. (2001). Mixtures of Gamma distributions with applications. *Journal of Computational and Graphical Statistics*, **10**(3), 440–454.
- Yancey, W. (2012). Sampling for Medicare and other claims. Accessed: 9/20/2016.
- Youngstrom, N. (2015). Medical-necessity audits gain steam, hit on chemo, cardiac; watch for new lcds. <http://www.racmonitor.com/rac-enews/1833-medical-necessity-audits-gain-steam-hit-on-chemo-cardiac-watch-for-new-lcds.html>. Accessed: 11/07/2016.