Sheffield Economic Research Paper Series

SERP Number: 2012027

ISSN 1749-8368

Arnab Mukherji
Satrajit Roychowdhury
Pulak Ghosh
Sarah Brown

Estimating Healthcare Demand for an Aging Population: A Flexible and Robust Bayesian Joint Model

October 2012

Department of Economics
University of Sheffield
9 Mappin Street
Sheffield
S1 4DT
United Kingdom
www.shef.ac.uk/economics

Estimating Healthcare Demand for an Aging Population: A Flexible and Robust Bayesian Joint Model[*]

Arnab Mukherji[†]

Centre for Public Policy, Indian Institute of Management Bangalore

Satrajit Roychowdhury

Expert Statistical Methodologist, Novartis Pharmaceutical Company

Pulak Ghosh

Department of QM & IS, Indian Institute of Management Bangalore

Sarah Brown

Department of Economics, University of Sheffield

[*] We are grateful to Arne Risa Hole and Jennifer Roberts for excellent comments.

[†] Corresponding Author, NF-010, Center for Public Policy, Indian Institute of Management Bangalore, Bannerghatta Road, Bangalore, 560076, Karnataka, India. Email: arnab.mukherji@gmail.com. Tel: 91-80-2699 3750. Fax: 91-80-2658 4050)

Abstract

In this paper, we analyse two frequently used measures of the demand for health care, namely hospital visits and out-of-pocket health care expenditure, which have been analysed separately in the existing literature. Given that these two measures of healthcare demand are highly likely to be closely correlated, we propose a framework to jointly model hospital visits and out-of-pocket medical expenditure. Furthermore, the joint framework allows for the presence of non-linear effects of covariates using splines to capture the effects of aging on healthcare demand. Sample heterogeneity is modelled robustly with the random effects following Dirichlet process priors with explicit cross-part correlation. The findings of our empirical analysis of the U.S. Health and Retirement Survey indicate that the demand for healthcare varies with age and gender and exhibits significant cross-part correlation that provides a rich understanding of how aging affects health care demand, which is of particular policy relevance in the context of an aging population.

Keywords: Aging, Bayesian Methods, Healthcare Demand, Joint Model, Splines
JEL Code: C11, C14, I10

1 Introduction

The world population is aging: According to a joint report by the U.S. Department of State and the National Institute on Aging (NIA), almost 500 million people worldwide were 65 and older in 2006 (Dobriansky et al. 2007). This number is expected to increase to 1 billion, 1 in every 8 of the earths inhabitants, by the year of 2030. In the U.S., life expectancy has increased from 49 years for Americans born in 1900 to 78 years for those born in 2006 (Arias 2010). Rapid demographic change is expected to lead to an increase in healthcare spending by 25% by 2030 (Strunk et al. 2006; Dobriansky et al. 2007). While global aging represents a triumph of medical, social, and economic advances, it also poses tremendous challenges for healthcare systems. It is well understood that aging will change the mix of diseases in favour of chronic conditions for inpatient care and this alone is likely to increase the demand for healthcare (Strunk et al. 2006; Hartman et al. 2008). With limited long-term benefits under healthcare schemes such as Medicare in the U.S., such increases in demand will potentially lead to large out-of-pocket medical expenses for the elderly (Wei et al. 2004; Hartman et al. 2008). Thus, obtaining reliable estimates of the demand for healthcare has never been more important than now with aging becoming a worldwide challenge (Dobriansky et al. 2007).

Health economics has traditionally focused on healthcare demand and Duan et al.'s (1982) seminal work on healthcare demand explored different strategies to estimate medical expenditure to address data concerns specific to healthcare cost data. Another metric that is also frequently used to measure healthcare demand is the rate of hospital admissions (Atella and Deb 2008). Despite the likely relationship between hospital visits and medical expenditure, these two measures of healthcare demand have typically been modelled separately in the existing literature. Furthermore, the probability of needing healthcare increases with age, particularly with the onset of chronic conditions. Hence, it is important to understand the effects of aging when modelling the demand for healthcare. Thus, managing healthcare demand arguably requires an understanding of hospitalisations as well as medical expenditure in the context of an aging population. In this paper, we analyse the key factors affecting both hospital visits and medical expenditure by developing a novel joint modelling framework, which allows us to reliably study healthcare demand and the correlation between these two aspects of healthcare demand.

Modelling hospitalisations and medical expenditure requires consideration of a number of complications specific to the analysis of healthcare data. First, both hospitalisation and out-of-pocket expenditure at the individual level usually have a considerable amount of zero observations, which cannot be adequately described by a simple distribution such as a Poisson or lognormal distribution. For example, 90% of the sample have no hospital visits and 17% report zero out-of-pocket expenditure in wave 1 of the U.S. Health and Retirement Survey (HRS). Thus, with the number of observed zeros larger than the expected number of zeros from a Poisson distribution, we see overdispersion in the data. Recently, Naya et al. (2008) compared model fits of a Poisson model and a zero-inflated Poisson (ZIP) model to zero-inflated data and found that a ZIP model gave estimates closer to the true values. Thus, we need to modify parametric distributions to incorporate excess zeros in the distribution of the hospitalisations and out-of-pocket medical expenditure. Recent literature (such as Deb and Trivedi 1997, Winkelmann 2004 and Atella and Deb 2008) has developed zero-inflated distributions for modelling the count of hospital visits and medical expenditure; however, they are modelled independently. Second, hospital visits and medical expenditure are likely to be correlated with each other over time for the same individual. Accounting for this correlation may lead to a better understanding of healthcare demand. Third, some important individual characteristics, such as age, may have complex nonlinear effects. In addition, the potential nonlinear effects of age could vary with other demographic characteristics, such as gender, resulting in an interaction effect that influences healthcare demand in a nonlinear fashion. Fourth, both the count of hospital visits and medical expenditure are known to be skewed (Liu et al. 2010). Although, some authors have argued in favour of log transformations to deal with skewness, this can be problematic. Re-transformation presents no problems when errors accord with linearity, normality and homoscedasticity assumptions (Jones 2000). When any one of these does not hold, re-transformation bias arises on reverting back to the original scale. Since the log-transformed model results in geometric means rather than arithmetic means, log scale predictions will, in general, provide biased estimates of the impact of any explanatory variable on the arithmetic mean (Yu et al. 2011).

In this paper, we develop a joint framework for modelling counts of hospital visits and out-of-pocket medical expenditure in an integrated framework to accommodate the aforementioned complications as follows. We model the count of hospital visits made by an individual using a Poisson hurdle model (Mullahy 1986) and we model out-of-pocket medical expenditure using a semicontinuous model (Liu et al. 2010). The Poisson hurdle model (semicontinuous model) consists of two components: a Bernoulli component that models the probability of hospitalisation (any positive expense) and a truncated Poisson component (log-normally distributed component) that models the number of hospital visits (amount of money spent) among users. Together, these components accommodate both the high proportion of zeros and the right-skewness of the nonzero events. In addition, we explicitly account for interdependencies between these events by modelling the correlation between these two processes. While the literature on healthcare demand discusses "multi-part" models, such as in the original work of Duan et al (1982) or the more recent work of Liu et al. (2008), these differ from our model in a number of ways. These models focus on a single outcome and the multi-part model allows for flexibility in model parameters across sub-groups with different demands for health care. For example, Duan et al. (1982) focus on how the parameters vary by non-spenders, ambulatory spenders, and inpatient spenders; more recently, Liu et al. (2008) are interested in the differences between non-spenders, out-patient spenders and inpatient spenders. Our model provides a richer specification of healthcare demand that not only captures healthcare expenditure but also hospital visits within the same joint model with explicitly modelled random effects.

In addition, our sample is drawn from a predominantly aging population and the effects of age on hospital visits and medical expenditure are arguably poorly understood in the existing literature, yet, as argued above, are of utmost policy importance. We thus adopt a semi-parametric approach using spline models to flexibly capture the potentially nonlinear effects of age. This approach not only protects the model from the possible misspecifications of age effects but also allows us to explore if this nonlinear effect varies across gender. For the distribution of the latent random effects terms of the joint model, a standard assumption is to use a parametric distribution, such as the multivariate normal distribution. The importance of such a choice has received a lot of attention in the joint modelling literature. In particular, it has been shown that a restrictive parametric assumption for this distribution could influence the results (Tsonaka et al. 2009 and Naskar and Das 2006). Thus, in order to protect the derived inferences against potential misspecification effects, we opt for a semi-parametric approach based on a Dirichlet Process prior. A similar approach to modelling random effects, but with a single outcome and without splines, has been proposed in the existing literature (Jochmann and Leon-Gonzalez 2004).

The rest of the paper is organised as follows: Section 2 presents the four part model as well as details of the Bayesian inference. Section 3 discusses the HRS data and the results of our empirical analysis. Finally, Section 4 concludes.

2 A Four Part Robust Semi-parametric Joint Model

Our joint model consists of three components: a semiparametric Poisson hurdle mixed effects model for the number of hospitalisations, a semiparametric semicontinuous model for out-of-pocket medical expenses, and a Dirichlet process for the joint distribution of the latent random effects from the Poisson hurdle and the semi-continuous models.

2.1 The Poisson Hurdle Model for the Count of Hospital Visits

The Poisson hurdle model is a two-component mixture model consisting of a point mass at zero followed by a truncated Poisson for the nonzero observations (Mullahy 1986). For independent and identically distributed (i.i.d.) responses, the hurdle model is given by

$$\begin{aligned} \Pr(Y_i = 0) &= 1 - p, \quad 0 \leq p \leq 1 \\ \Pr(Y_i = k) &= p\frac{\mu^k e^{-\mu}}{k!(1 - \mathrm{e}^{-\mu})}, \quad k = 1, \ldots, \infty, : 0 < \mu < \infty, \end{aligned} \tag{1}$$

where Y_i denotes the response for individual $i = 1, \ldots, n$, and μ is the mean for an untruncated Poisson distribution. As the zeros and nonzero counts are modelled uniquely, the hurdle model accommodates both an excess number of zeros and a right-skewed distribution for the positive counts. By comparison, a standard Poisson regression would have to compromise between these two competing goals, since excess zeros would tend to lower the Poisson mean while large nonzero values would tend to increase it. The expected count under the Poisson hurdle model is given by $\mathrm{E}(Y) = p\mu / (1 - \mathrm{e}^{-\mu})$.

In health services research, p is known as the *usage probability*—i.e., the probability of using services at least once. When $(1 - p) > \mathrm{e}^{-\mu}$, the data are zero inflated relative to an ordinary Poisson; when $(1-p) < e^{-\mu}$ there is zero deflation (i.e., fewer than expected zeros). In the extremes, $p = 0$ or 1. When $p = 1$, there are no zero counts and the model reduces to a truncated Poisson, and when $p = 0$, there are no users (i.e., all counts equal zero), and the model is degenerate at zero. Typically, one assumes that p lies strictly between 0 and 1, so that all individuals have a nonzero probability of usage and are, therefore, considered "potential" users even if they do not actually use health services during the study period. A special case of (1) is the zero-inflated Poisson model (Lambert 1992), which consists of a degenerate distribution at zero mixed with an untruncated Poisson distribution:

$$P(Y_i = 0) = (1 - p) + p\mathrm{e}^{-\mu}, \quad 0 < p < 1 \tag{2}$$

$$P(Y_i = k) = p\frac{\mu^k \mathrm{e}^{-\mu}}{k!}, \quad k = 1, \ldots, \infty, : 0 < \mu < \infty. \tag{3}$$

Note that the zero-inflated Poisson model can be rewritten as a hurdle model with mixing probability $\theta = p(1 - \mathrm{e}^{-\mu})$. Unlike the hurdle model, which accommodates zero deflation as well as zero inflation, the ZIP allows only for zero inflation and, thus, allows for greater flexibility (Neelon et al. 2010). Let Y_{ij}^H be the count of the number of hospital stays reported by the ith individual in the jth wave, $i = 1, 2, \cdots, m$; $j = 1, 2, \cdots, n$, where m represents the number of individuals in the study, and n is the total number of waves over which the individual is surveyed. Depending on whether an individual is hospitalised or not, a large number of zeros is observed in Y_{ij}^H. Also, let X_{ijk} be the k^{th} covariate for individual i at time j; such covariates include baseline and time-varying variables.

Each individual's total count of hospital visits is determined simultaneously by needing some healthcare (p_{ij}) as well as the level of care needed given that the person needs care λ_{ij}. Given that these are jointly determined, and that the determinants of either may or may not be relevant for the other, we consider simultaneous modelling of both λ_{ij} and p_{ij}. The hurdle model can be extended to accommodate covariates and random effects as follows:

$$\begin{aligned} p(y_{ij}^H|\boldsymbol{\phi}_i) &= (1-p_{ij}^H)1_{(y_{ij}^H=0)} + p_{ij}^H \mathrm{Tpois}(y_{ij}^H;\mu_{ij}^H)1_{(y_{ij}^H>0)} \\ \mathrm{logit}(p_{ij}^H) &= \mathbf{X}_{ij1}^T\boldsymbol{\beta}_1^p + \mathbf{Z}_{ij1}^T\boldsymbol{b}_{i1} + f^p(W_{ij}) \\ \log(\mu_{ij}^H) &= \mathbf{X}_{ij2}^T\boldsymbol{\beta}_1^\lambda + \mathbf{Z}_{ij2}^T\boldsymbol{b}_{i2} + f^\lambda(W_{ij}) \end{aligned} \tag{4}$$

where, $\mathbf{X}_{ij1,}\mathbf{X}_{ij2}$ are the vectors of covariates corresponding to the fixed effects and $\mathbf{Z}_{ij1}, \mathbf{Z}_{ij2}$ are the vectors of covariates corresponding to the random effects. Note that the zero-state and the Poisson state do not need to have the same set of covariates. The b_{i1} and b_{i2} are the random individual effects on p_{ij} and λ_{ij}, respectively. We will discuss the distribution of the random individual effects later. In many situations, such as our application, the effect of some covariates, viz., W_{ij} on p_{ij}^H and μ_{ij}^H, may not be linear. Thus, the effects of those covariates can be modelled by unspecified nonparametric functions $f^p(W_{ij})$ and $f^\lambda(W_{ij})$. These unknown smoothing functions reflect the nonlinear effects of the covariate. However, these functions only represent the population averages for a single population.

We now consider a modified model for multiple factors/populations. Instead of fitting one nonparametric smoothing spline for a single population, we can include multiple nonparametric smoothing splines for multiple populations in one model. We consider:

$$\begin{aligned}
\text{logit}(p_{ij}^H) &= \mathbf{X}_{ij1}^T \boldsymbol{\beta}_1^p + \mathbf{Z}_{ij1}^T \boldsymbol{b}_{i1} \\
&+ f_1^p(W_{ij})d_{ij1}^p + f_2^p(W_{ij})d_{ij2}^p + \cdots + f_L^p(W_{ij})(1 - d_{ij1}^p - d_{ij2}^p - \cdots - d_{ij(L-1)}^p) \quad (5) \\
\log(\mu_{ij}^H) &= \mathbf{X}_{ij2}^T \boldsymbol{\beta}_1^\lambda + \mathbf{Z}_{ij2}^T \boldsymbol{b}_{i2} \\
&+ f_1^\lambda(W_{ij})d_{ij1}^\lambda + f_2^\lambda(W_{ij})d_{ij2}^\lambda + \cdots + f_L^\lambda(W_{ij})(1 - d_{ij1}^\lambda - d_{ij2}^\lambda - \cdots - d_{ij(L-1)}^\lambda) \quad (6)
\end{aligned}$$

where, d_{ijk}; $k = 1, 2, \cdots, L$ are indicator variables for multiple populations. With L populations, the first group is indicated by $(d_{ij1} = 1,\ d_{ij2} = 0, \cdots, d_{ij(L-1)} = 0)$, the second group is indicated by $(d_{ij1} = 0,\ d_{ij2} = 1, \cdots, d_{ij(L-1)} = 0)$ and the last group is indicated by $(d_{ij1} = 0,\ d_{ij2} = 0, \cdots, d_{ij(L-1)} = 0)$. The $f_1, f_2, \cdots, f_L$ are their respective nonparametric smoothing splines.

We approximate the spline function $f(W_{ij})$ (suppressing the subscripts) by a piecewise polynomial of degree τ. The knots $\tilde{w} = (\tilde{w_1}, \tilde{w_2}, \cdots, \tilde{w_m})$ are placed within the range of W_{ij}, such that $\min(W_{ij}) < \tilde{w_1} < \tilde{w_2} < \cdots < \tilde{w_m} < \max(W_{ij})$. Then $f(W_{ij})$ is approximated by

$$f(W_{ij}) = \nu_1 W_{ij} + \nu_2 W_{ij}^2 + \cdots + \nu_\tau W_{ij}^\tau + \sum_{c=1}^{C} u_c \gamma_c (W_{ij} - \tilde{w}_c)_+^\tau$$

where $X_+ = x$ if $x > 0$, and 0 otherwise, $\nu = (\nu_1, \cdots, \nu_\tau)$, $\tilde{w}$ are the vectors of regression coefficients in the polynomial regression spline. Note that there is no intercept in the polynomial regression to avoid identifiability. We assume $u_c \sim^{iid} N(0, \sigma_u^2); i = 1, \ldots, C$. In the above formulation, one of the important issues is the choice of the number of knot points and where to locate them. If there are too few knots or they are poorly located, estimates may be biased, while too many knots will inflate the local variance. Thus, following Smith and Kohn

1996), we incorporate selector indices, γ_c, that allow the spline coefficients to be included or excluded and that are defined for each knot. The γ_c are then drawn independently from a Bernoulli prior, viz., $\gamma_c \sim$ Bernoulli(0.5). By introducing this, we can select a subset of well supported knots from a larger space. For each knot point u_c, the γ_c will weight the importance of a particular knot point.

2.2 Semicontinuous Model for Out-of-Pocket Medical Expenditure

In this section, a semi-continuous model for longitudinal data on out-of-pocket medical expenditure is introduced. Since in some years the individual may not have incurred any medical expenditure, this type of data has a mix of zeros and positive continuous observations. To formulate the model, let y_{ij}^M be the medical expenditure of individual i at year j. Let R_{ij} be a random variable denoting annual medical expenditure where,

$$R_{ij} = \begin{cases} 0, & \text{if } y_{ij}^M = 0 \\ 1, & \text{if } y_{ij}^M > 0, \end{cases} \tag{7}$$

with conditional probabilities

$$\Pr(R_{ij} = r_{ij}) = \begin{cases} 1 - p_{ij}^M, & \text{if } r_{ij} = 0 \\ p_{ij}^M, & \text{if } r_{ij} = 1. \end{cases}$$

For this semicontinuous data, we introduce an analogous semi-continuous model consisting of a degenerate distribution at zero and a positive continuous distribution, such as a lognormal

(LN), for the nonzero values:

$$
\begin{aligned}
f(y_{ij}^{M}|\mathbf{p}_{i}^{M}) &= (1-p_{ij}^{M})^{1-r_{ij}}\left\{p_{ij}^{M}\times \mathrm{LN}(y_{ij}^{M};\mu_{ij}^{M},\sigma^{2})\right\}^{r_{ij}} \\
\mathrm{logit}(p_{ij}^{M}) &= \mathbf{X}_{ij}^{T}\boldsymbol{\beta}_{1}^{Mp}+\mathbf{Z}_{ij1}^{T}\boldsymbol{b}_{i3} \\
&+ h_{1}^{p}(W_{ij})e_{ij1}^{p}+h_{2}^{p}(W_{ij})e_{ij2}^{p}+\cdots \\
&+ h_{L}^{p}(W_{ij})(1-e_{ij1}^{p}-e_{ij2}^{p}-\cdots-e_{ij(L-1)}^{p}) \qquad (8)\\
\log(\mu_{ij}^{M}) &= \mathbf{X}_{ij}^{T}\boldsymbol{\beta}_{1}^{M\lambda}+\mathbf{Z}_{ij2}^{T}\boldsymbol{b}_{i4} \\
&+ h_{1}^{\lambda}(W_{ij})e_{ij1}^{\lambda}+h_{2}^{\lambda}(W_{ij})e_{ij2}^{\lambda}\cdots \\
&+ h_{L}^{\lambda}(W_{ij})(1-e_{ij1}^{\lambda}-e_{ij2}^{\lambda}-\cdots-e_{ij(L-1)}^{\lambda}) \qquad (9)
\end{aligned}
$$

where, r_{ij} is an indicator as defined above, and μ_{ij}^{M} and σ^{2} are the mean and variance of $\log(y_{ij}^{M})$, respectively. The interpretation of e_{ijk} is the same as d_{ijk} in the ZIP model and the nonparametric spline function $h(.)$ is also defined in a similar fashion. The model given by equations (9,10) is a semiparametric counterpart of the correlated two-part model proposed by Olsen and Schafer 2001); a gamma or log-skew-normal distribution may also be used to model the nonzero values.

2.3 The Latent Random Effects Distribution: Dirichlet Process Priors

Without loss of generality, we assume that all $\mathbf{b}_{ik}$ in equations (5, 6, 8, 9) are $r\times 1$ unobserved vectors. Let $\mathbf{b}=(\mathbf{b}_{1},\ldots,\mathbf{b}_{m})$ denote the random effects for all m individuals, where $\mathbf{b}_{i}=(\mathbf{b}_{i1}^{\top},\mathbf{b}_{i2}^{\top},\mathbf{b}_{i3}^{\top},\mathbf{b}_{i4}^{\top})^{\top}$, $i=1,\ldots,m$, is a $4r\times 1$ vector representing the random effects for the ith individual. To allow for the correlation structure between repeated observations for the same individual taken over different years and also to account for uncertainty in the probability distributions of the random effects, usually one assumes a multivariate normal distribution (Neelon et al., 2011).

However, in an aging population the subjects' responses may result in increased heterogeneity in the population. In addition, the endpoints are skewed and thus a parametric normal distribution may be restrictive for the latent random effects. Thus, instead of a normal distribution, we employ a Dirichlet process (DP) prior based on a stick-breaking scheme ((Ferguson 1973); (Sethuraman 1994)) that makes fewer assumptions about the distribution function.

To proceed, we assume latent variables $\mathbf{b}_i$ are drawn from an arbitrary distribution G, where G has a DP prior, denoted by $\mathbf{b}_i \sim \mathrm{DP}(\nu, G_0)$, $G_0 \sim \mathtt{N}_{4r}(\mathbf{0}, \mathbf{\Sigma})$ and ν is an unknown concentration parameter. Usually a uniform prior is assumed for ν. Thus, the DP prior is essentially a distribution defined on the space of distributions and parameterized by a known base distribution G_0 and by a positive concentration parameter ν that represents variability around G_0. The G_0 can be viewed as the "mean" distribution in the space of distributions covered and ν is a measure of the "variance" of realizations of G around G_0. For a comprehensive review of DP, see Hjort, Holmes, Walker and Muller (2010). Formally our model for $\mathbf{b}_i$ can be hierarchically expressed as;

$$\begin{aligned} \mathbf{b}_i|G &\overset{\text{iid}}{\sim} G,\ i = 1, \ldots, m, \\ G|a, G_0 &\sim \mathtt{DP}(aG_0), \qquad \text{with } G_0 = \mathtt{N}_{4r}(\mathbf{0}, \mathbf{\Sigma}), \end{aligned} \tag{10}$$

(Sethuraman 1994) provided an explicit characterization of G in terms of a stick-breaking construction where G is represented as an infinite mixture of discrete atoms m_h with probabilities w_h ($\sum_{h=1}^{\infty} w_h = 1$). In our context, the m_h are drawn i.i.d from $G_0 \sim \mathtt{N}_{4r}(\mathbf{0}, \mathbf{\Sigma})$. For w_h, imagine a probability stick of unit length and break off a portion $w_1 = \pi_1$, where π_1 is drawn from a beta distribution , $\mathrm{Beta}(1, \nu)$. The length of the remaining stick is $(1-\pi_1)$. Let π_2 be another independent draw from the same beta distribution, representing the portion of the remaining probability stick that is broken off. Thus, $w_2 = \pi_2(1 - \pi_1)$ denotes the probability associated with the second independent draw m_2 from G_0. Continuing this, we obtain:

$$
\begin{aligned}
G \quad &= \quad \sum_{h=1}^{\infty} w_h \delta_{m_h}; \quad \text{with} \quad w_h = \pi_h \prod_{l=1}^{h-1}(1-\pi_l), \text{ for } l = 1, 2, \cdots, \infty \\
&\text{where} \quad \pi_h|\nu \sim \text{Beta}(1,\nu), \quad \text{and} \quad m_h \overset{\text{iid}}{\sim} G_0
\end{aligned}
$$

Here, δ_{m_h} denotes a discrete distribution with all its probability mass at m_h. For all values of $\nu(\nu \approx 1)$ the first four or five m_h account for 99% of the distribution G while for a large value of $\nu(\nu \approx 10)$, 99% of the distribution of G is accounted for by the first 50 m_h's (Hjort et al, 2010). Due to this fact the G can be reduced to a truncated DP by truncating at a large number R.

2.4 The Bayesian Inference

Under the joint model described by equations (5,6,9,10), the likelihood of the observed data for the ith individual, denoted by $\mathbf{Y}_{i1}, \dots, \mathbf{Y}_{in}$, with $\mathbf{Y}_{ij} = (y^H_{ij}, y^M_{ij})^\top$ for $j = 1, \dots, n$, based on the parameter set Ω and the random effects $\mathbf{b}_i$ is proportional to

$$
\begin{aligned}
L_i(\mathbf{Y}_{i1}, \dots, \mathbf{Y}_{in}|\Omega, \mathbf{b}_i) \quad &= \quad \prod_{j=1}^{n} \left[(1-p^H_{ij})\right]^{I_{[y^H_{ij}=0]}} \times \left[\frac{p^H_{ij} {\mu^H_{ij}}^{y^H_{ij}} e^{-\mu^H_{ij}}}{y^H_{ij}!(1-e^{-\mu^H_{ij}})}\right]^{1-I_{[y^H_{ij}=0]}} \\
&\times \quad (1-p^M_{ij})^{1-r_{ij}} \left\{p^M_{ij} \times \text{LN}(y^M_{ij}; \mu^M_{ij}, \sigma^2)\right\}^{r_{ij}} \qquad (11)
\end{aligned}
$$

Assuming independence between observations from different individuals, the resulting likelihood for all the observations from the m individuals is the product of these individual likelihood values. Then, marginalising out all the random effects, which are modelled by a DP, as given in equation (10) with a fixed $a > 0$, leads to the likelihood of all the observed data being proportional to

$$
L(\Omega|\text{data}) = \int_{\mathbb{R}^r} \cdots \int_{\mathbb{R}^r} \prod_{i=1}^{m} L_i(\Omega, \mathbf{b}_i|\mathbf{Y}_{i1}, \dots, \mathbf{Y}_{in}) \mathbf{m}(\mathbf{b}) d\mathbf{b}_1 \cdots d\mathbf{b}_m,
$$

which is an m-folded integral. To complete the Bayesian specification of the model, we assign priors to the unknown parameters in the above likelihood function. Thus, the set of parameters from the model may be listed as:

$$\begin{aligned}\Omega &= \Big(\beta_{11}^{Hp}, \beta_{21}^{H_\lambda}, \beta_{31}^{Mp}, \beta_{41}^{M_\lambda}, \ldots, \beta_{19}^{Hp}, \beta_{29}^{H_\lambda}, \beta_{39}^{Mp}, \beta_{49}^{M_\lambda}, \nu_1^{Hp}, \ldots, \nu_\tau^{Hp}, \sigma^2_{H_p}, \\ &\quad \nu_1^{H_\lambda}, \ldots, \nu_\tau^{H_\lambda}, \sigma^2_{H_\lambda}, \nu_1^{Mp}, \ldots, \nu_\tau^{Mp}, \sigma^2_{M_p}, \nu_1^{M_\lambda}, \ldots, \nu_\tau^{M_\lambda}, \sigma^2_{M_\lambda}, \Sigma, a\Big) \qquad (12)\end{aligned}$$

For each parameter in Ω, we then specify a prior: for each model specific regression coefficient (β_{ij}^{θ}) and each spline specific regression coefficient (ν_i^{θ}), we assume a normal density prior; for each variance parameter (σ^2_θ), we assume an inverse-gamma (IG) prior and, finally, for the cross-part variance covariance matrix (Σ), we assume an inverse Wishart prior. Further, for the total mass, a, we assume a uniform distribution (Ohlssen et. al. 2007).

An IG prior with shape parameter c and scale parameter d is denoted by $x \sim IG(c, d)$ and its density is given by $f(x) \propto x^{-c}e^{(d/2x^2)}$. Additionally, we assume a Wishart distribution for the inverse of a variance covariance matrix, where $W_G(\rho, s)$ is a G-dimensional Wishart distribution, with ρ degrees of freedom and a mean of ρs^{-1}. Thus, we specify the following priors for the model parameters:

$$\begin{aligned}\pi(\underset{\sim}{\beta}) &= \Big(\beta_{11}^{Hp}, \ldots, \beta_{49}^{M_\lambda},\Big) \sim N(\underset{\sim}{\mu}_\beta, \Sigma_\beta) \\ \pi(\underset{\sim}{\nu}^{Hp}) &= \Big(\nu_1^{Hp}, \ldots, \nu_\tau^{Hp}\Big) \sim N(\mu_\nu^{Hp}, \Sigma_\nu^{Hp}) \\ \pi(\underset{\sim}{\nu}^{H_\lambda}) &= \Big(\nu_1^{H_\lambda}, \ldots, \nu_\tau^{H_\lambda}\Big) \sim N(\mu_\nu^{H_\lambda}, \Sigma_\nu^{H_\lambda}) \\ \pi(\underset{\sim}{\nu}^{Mp}) &= \Big(\nu_1^{Mp}, \ldots, \nu_\tau^{Mp}\Big) \sim N(\mu_\nu^{Mp}, \Sigma_\nu^{Hp}) \\ \pi(\underset{\sim}{\nu}^{M_\lambda}) &= \Big(\nu_1^{M_\lambda}, \ldots, \nu_\tau^{M_\lambda}\Big) \sim N(\mu_\nu^{M_\lambda}, \Sigma_\nu^{M_\lambda})\end{aligned}$$

For the remaining variance parameters, the variance covariance matrix and a, we assume:

$$\begin{aligned}
\pi(\sigma^2_{Hp}) &\sim IG(c_{Hp}, d_{Hp}) \\
\pi(\sigma^2_{H_\lambda}) &\sim IG(c_{H_\lambda}, d_{H_\lambda}) \\
\pi(\sigma^2_{Mp}) &\sim IG(c_{Mp}, d_{Mp}) \\
\pi(\sigma^2_{M_\lambda}) &\sim IG(c_{M_\lambda}, d_{M_\lambda}) \\
\pi(\Sigma^{-1}) &= Wishart(\rho, s) \\
\pi(a) &= Uniform(e, f)
\end{aligned}$$

The joint posterior distribution of the parameters of the models conditional on the data are obtained by combining the likelihood and the prior densities using Bayes Theorem:

$$\begin{aligned}
Post(\Omega, \mathbf{b}|\mathbf{Y}) \;\propto\; & \int_{\mathbb{R}^r} \cdots \int_{\mathbb{R}^r} \prod_{i=1}^{m} L_i(\mathbf{Y}_{i1}, \ldots, \mathbf{Y}_{in}|\Omega, \mathbf{b}_i)\mathbf{m}(\mathbf{b})\pi(\underset{\sim}{\beta})\pi(\underset{\sim}{\nu}^{Hp})\pi(\underset{\sim}{\nu}^{H_\lambda}) \\
& \pi(\underset{\sim}{\nu}^{Mp})\pi(\underset{\sim}{\nu}^{M_\lambda})\pi(\sigma^2_{Hp})\pi(\sigma^2_{H_\lambda})\pi(\sigma^2_{Mp})\pi(\Sigma^{-1})\pi(a)d\mathbf{b}_1 \cdots d\mathbf{b}_m \qquad (13)
\end{aligned}$$

The posterior distributions are analytically intractable. However, the models described above can be fitted using Markov Chain Monte Carlo (MCMC) methods such as the Gibbs sampler (Gelfand et al 1992). Since the full conditional distributions are not standard, a straightforward implementation of the Gibbs sampler using standard sampling techniques may not be possible. However, sampling methods can be performed using adaptive rejection sampling (ARS; Gilks and Wild 1992). In this paper, we follow their procedure, which first uses a data augmentation step to sample the values of the latent variables based on the current values of the parameters, and then samples the parameters using the ARS method given the latent variables. Samples were directly obtained from the joint posterior distribution of the parameters as well as the latent variables. The samples from the posterior distribution obtained from the MCMC allow us to achieve summary measures of the parameter estimates and to obtain credible intervals (CIs) of the parameters of interest. We present two simulation exercises in the appendix to justify the relative complexity of the proposed model, where the complexity of the proposed model arises from two aspects: (1) using a DP for the skewed distributed random effects $\boldsymbol{b_i}$ and (2) spline-based modelling of nonlinear time effects. The simulation exercises verify the performance of the model fitting procedure over more conventional models.

3 Empirical Analysis

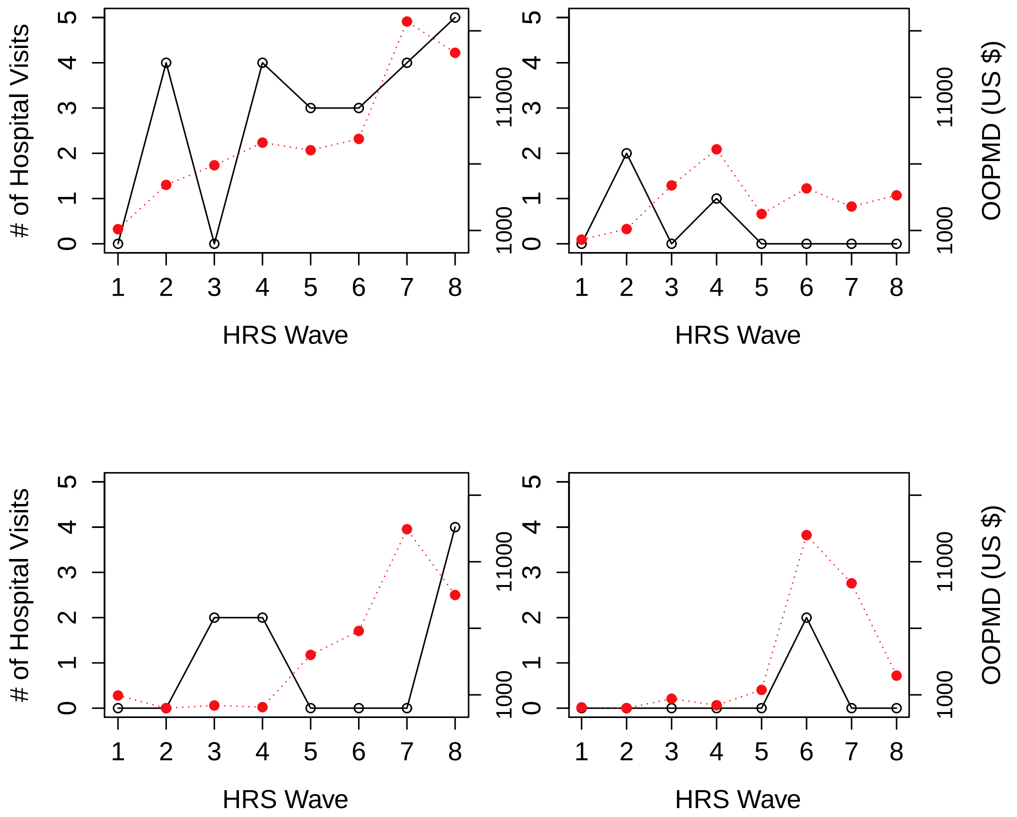
3.1 Data

In order to explore the relationship between out-of-pocket medical expenditure and hospitalisations, we use data from the University of Michigan's Health and Retirement Study (HRS). The HRS is a longitudinal survey of Americans over the age of 50, with a follow-up frequency of two years and is designed to provide multi-disciplinary data to understand the challenges of aging. In this paper, we use data from the 1931-41 cohort - the HRS cohort. Baseline observations for the HRS cohort begin in 1992 when individuals were between 52-62 years of age and were near retirement. The data we use is maintained by RAND's Center for the Study of Aging and has been comprehensively cleaned and documented. In practice, we also restrict the HRS cohort to include only those who did not drop-out of the study in the first 5 of the 8 waves of the study to allow for sufficient length in the panel.

For our outcome measures, we use the number of hospital visits made since the previous interview which is based on responses to the following question: How many different times were you a patient in a hospital overnight in the last 12 months? On the other hand, the total out-of-pocket medical expenses variable (OOPMD) covers total medical costs for all medical services since the previous interview and excludes all costs that were reimbursed or paid through insurance. It covers four groups of services, namely: hospital/nursing; doctor/outpatient/dental; prescription drugs; and home healthcare/special services.

Figure 1 presents plots of hospital visits and out-of-pocket medical expenditure (OOPMD) over time for four randomly chosen individuals. It is apparent that individuals vary widely in the number of hospital visits that they make. Not only do individuals have different

Figure 1: Co-movement of Hospital Visits and OOP-Medical Expenditure



Note: On each x-axis we plot the survey wave in which the case was observed. On the left y-axis we plot the count of hospital visits made and on the right y-axis we plot the amount of out-of-pocket medical expenditure (OOPMD) in US dollars. The dotted line captures the OOPMD incurred at each wave while the straight line captures the count of hospital visits.

intensities of hospital visits but the corresponding OOPMD varies. These graphs present preliminary evidence suggesting that the number of hospital visits and OOPMD are correlated - as the number of hospital visits increases (or decreases) so does the OOPMD. This co-movement in these outcomes endorses a joint modelling approach.

Table 1 presents summary statistics for both outcome measures for cases where we have non-missing observations for the first four waves. The count of hospital stays exhibits increasing frequency of missing observations in later waves; additionally, there is a high but declining fraction of the sample in each wave with zero hospital visits. This change in healthcare demand over time is also captured through narrower ranges of outcomes in earlier waves than in later waves, emphasising the important effects of aging. Thus, while in wave 1 over 90% of the sample did not visit a hospital, by the last wave this proportion has declined to 40%. This high frequency of zeros supports the use of a Poisson hurdle model for hospital visits. Similarly, OOPMD also shows significant zero-inflation suggesting that treating it as a continuous variable would be problematic. As people age, the frequency of hospital visits rises, and so does OOPMD. We see this in Table 1 as the frequency of zeros declines the average OOPMD rises from $1,108 to $2,516.

Table 1: Distribution of Outcomes

	Count of Hospital Visits				Out-of-Pocket Medical Expense			
HRS	%	Non-zeros			%	Non-zeros		
Wave	Zeros	Mean	Min	Max	Zeros	Mean	Min	Max
1	91.33	1.50	1	3	16.33	1,108	2	18,494
2	81.00	1.72	1	8	12.33	1,175	9	26,629
3	76.33	1.63	1	6	11.67	1,767	10	58,250
4	75.33	1.99	1	10	8.00	1,325	6	39,800
5	68.33	1.80	1	8	8.00	1,522	15	24,800
6	58.33	2.82	1	60	9.00	3,710	35	232,400
7	51.00	1.64	1	5	7.00	3,422	10	301,000
8	40.33	2.47	1	25	4.00	2,516	5	45,200

Table 2: Summary Statistics of Response and Predictors

Variables	**Mean**	**SD**	**Min**	**Max**
Time Invariant				
Is Female? ($female_i$)	54.30%	49.89%	0	1
Eduction: GED or Higher? ($gedplus_i$)	71.67%	45.14%	0	1
Time Varying				
Count of Hospital Visits	0.44	1.72	0	60
OOPMD	2563.26	9707.71	0	301000
Do Health problems limit work? ($rhlthlm_{ij}$)	0.26	0.44	0	1
Has no difficulty in dressing ($rnodiffdress_{ij}$)	0.95	0.22	0	1
Self-reported expectation of living 10+ years ($rlive_{ij}$)	61.31	29.60	0	100
Change of health at current wave (coh_{ij})	0.09	0.86	-2	2
Change of health at previous wave ($coh_{i,j-1}$)	-0.01	1.43	-4	4

Descriptive statistics for the baseline and time varying covariates are presented in Table 2. A key aspect of aging is a loss of functional abilities (muscular strength, ventilatory capacity, incontinence, or cardiovascular output); however the rate of this decay varies with lifestyle and environmental factors (Wei et al. 2004). Information on many of these factors, such as gender, education and functional independence is available in the HRS. Hence, we control for being female as well as for whether the individual's education is at the level of the General Education Diploma (GED) or higher. The variables used to capture functional independence are whether the individual reports that he/she experiences no difficulty in dressing and whether his/her health limits their ability to work. Additionally, data on each individual's self-reported health status in the current and past wave, distinguishing between excellent, very good, good, fair and poor health, is used as it is known to be predictive of health status (McGee et al 1999). Specifically, we include variables capturing the change in self-assessed health between the current and previous waves, where positive (negative) values indicate a deterioration (improvement) in self-reported health between waves. The HRS also includes information on each respondent's expectation of being alive for the next ten years or more on a 0 to 100 scale; this is known to predict mortality (Hurd and McGarry 2002). We include this variable to explore the influence of expectations on health seeking behaviour. We also use the age of the respondent; in almost any aging study, the age of the respondent is an important predictor of health outcomes (Strunk et al. 2006; and Wei et al. 2004), hence it is believed that the age of the respondent is predictive of his/her healthcare demand and out-of-pocket medical expenditure.

3.2 Model Specifications and Results

Before discussing our results, we firstly compare our model with some alternative models to test the quality of model fit that characterises our model. To compare alternative models, we compute $P(Y_i|Y_{-i})$, which is the posterior predictive distribution of Y_i conditional on the observed data with a single data point deleted. This value is known as the conditional predictive ordinate (CPO) and has been widely used for model diagnostics and assessment (Gelfand et al 1992). For the i^{th} individual, the CPO statistic under model $M_l : 1 \leq l \leq L$ is defined as:

$$CPO_i = P(Y_i|Y_{-i}) = E_{\underset{\sim}{\theta} l}\left[P(Y_i|\underset{\sim}{\theta} l)|Y_{-i}\right] \tag{14}$$

where $-i$ denotes the exclusion of individual i from the sample. The $\underset{\sim}{\theta_l}$ is the set of parameters of the M_l and $P(Y_i|\underset{\sim}{\theta_l})$ is the sampling density of the model evaluated at the i^{th} observation. The preceding expectation is taken with respect to the posterior distribution of the model parameter $\underset{\sim}{\theta_l}$ given the cross-validated data, Y_{-i}. For individual i, the CPO_i can be obtained from the MCMC samples by computing the following weighted average:

$$C\hat{P}O_i = \left(\frac{1}{M}\sum_{m=1}^{M}\frac{1}{f(Y_i|\theta_l^{(m)})}\right)^{-1} \tag{15}$$

where M is the number of simulations and $\theta_l^{(m)}$ denotes the parameter samples at the m^{th} iteration. A large CPO value indicates a better fit. A useful summary statistic of the CPO_i is the logarithm of the Psuedo-marginal Likelihood (LPML) defined as:

$$LPML = \sum_{i=1}^{n} log(C\hat{P}O_i) \tag{16}$$

Greater LPML values represent a better fit. The LPML is well defined under the posterior predictive density where it is computationally stable. We compare the following models using the LPML values: Model 1, the four part model proposed in this paper, the results from which are discussed below; Model 2, a four part model where each part is modelled independently without random effects; Model 3, a four part model with correlated random effects in a multivariate normal distribution; and Model 4, a four part model with robust random effects but no age splines or interactions. The LPML values for Models 1 to 4 are -5405.7, -7198.4, -6201.8 and -61332.4, respectively. Thus, Model 1 has the highest LPML value suggesting that it has the best fit amongst the alternative models. The large difference in the LPML values between our proposed model and the alternative models indicates the presence of a nonlinear age effect and the importance of DP for our analysis.

We formulate an empirical version of the four-part model discussed above to be applied to the HRS data as follows. Equations 17 and 18 present the zero-inflated semi-continuous component of the model that seeks to explain hospital stays. The same covariates are allowed to differentially impact on the propensity for visiting a hospital (in equation 17) and the count of such visits made (in equation 18).

$$\begin{aligned}
\text{logit}(p_{ij}^{H}) &= \beta_{11}^{p} + \beta_{12}^{p} t_{ij} + \beta_{13}^{p}\text{gedplus}_i + \beta_{14}^{p}\text{female}_i + \beta_{15}^{p}\text{rhlthlm}_{ij} + \beta_{16}^{p}\text{rnodiff}_{ij} \\
&+ \beta_{17}^{p}\text{rlive}_{ij} + \beta_{18}^{p}\text{coh}_{ij} + \beta_{19}^{p}\text{coh}_{i,j-1} + f_1^{p}(age_{ij})d_{ij1}^{p} + f_2^{p}(age_{ij})(1-d_{ij1}^{p}) + b_{i1} \quad (17) \\
\log(\mu_{ij}^{H}) &= \beta_{11}^{\lambda} + \beta_{12}^{\lambda} t_{ij} + \beta_{13}^{\lambda}\text{gedplus}_i + \beta_{14}^{\lambda}\text{female}_i + \beta_{15}^{\lambda}\text{rhlthlm}_{ij} + \beta_{16}^{\lambda}\text{rnodiff}_{ij} \\
&+ \beta_{17}^{\lambda}\text{rlive}_{ij} + \beta_{18}^{\lambda}\text{coh}_{ij} + \beta_{19}^{\lambda}\text{coh}_{i,j-1} + f_1^{\lambda}(age_{ij})d_{ij1}^{\lambda} + f_2^{\lambda}(age_{ij})(1-d_{ij1}^{\lambda}) + b_{i2} \quad (18)
\end{aligned}$$

Similarly, equations 19 and 20 are the two components of the semi-continuous hurdle model for out-of-pocket medical expenses incurred. For both, the Poisson hurdle model and the semicontinuous model, age is allowed to flexibly affect both the propensity and the level of healthcare demand through a smoothing spline that is allowed to vary by gender:

$$\begin{aligned}
\text{logit}(p_{ij}^{M}) &= \beta_{11}^{M_p} + \beta_{12}^{M_p} t_{ij} + \beta_{13}^{M_p}\text{gedplus}_i + \beta_{14}^{M_p}\text{female}_i + \beta_{15}^{M_p}\text{rhlthlm}_{ij} + \beta_{16}^{M_p}\text{rnodiff}_{ij} \\
&+ \beta_{17}^{M_p}\text{rlive}_{ij} + \beta_{18}^{M_p}\text{coh}_{ij} + \beta_{19}^{M_p}\text{coh}_{i,j-1} + h_1^{p}(age_{ij})e_{ij1}^{p} \\
&+ h_2^{p}(age_{ij})(1-e_{ij1}^{p}) + b_{i3} \quad (19) \\
\log(\mu_{ij}^{M}) &= \beta_{11}^{M_\lambda} + \beta_{12}^{M_\lambda} t_{ij} + \beta_{13}^{M_\lambda}\text{gedplus}_i + \beta_{14}^{M_\lambda}\text{female}_i + \beta_{15}^{M_\lambda}\text{rhlthlm}_{ij} + \beta_{16}^{M_\lambda}\text{rnodiff}_{ij} \\
&+ \beta_{17}^{M_\lambda}\text{rlive}_{ij} + \beta_{18}^{M_\lambda}\text{coh}_{ij} + \beta_{19}^{M_\lambda}\text{coh}_{i,j-1} + h_1^{M_\lambda}(age_{ij})e_{ij1}^{M_\lambda} \\
&+ h_2^{M_\lambda}(age_{ij})(1-e_{ij1}^{M_\lambda}) + b_{i4} \quad (20)
\end{aligned}$$

Finally, in equations 17, 18, 19 and 20, the random effects $\mathbf{b_i} = (b_{i1}, b_{i2}, b_{i3}, b_{i4})$ are jointly modelled as a DP $(aG_0 \equiv N_4(0, \Sigma)$ and $a \sim Uniform(0.4, 10)$. To fully specify the Bayesian model, we assign weakly informative conjugate priors for the parameters. For each aggregate-level coefficient, we assume a normal density prior of $\text{N}(0, 100)$. For the variance parameters, we assume inverse-Gamma (IG) priors of $\text{IG}(2.01, 1.01)$, giving rise to a prior mean of 1 and a prior variance of 100. Lastly, we take an inverse-Wishart prior for the variance-covariance matrix by assuming $\Sigma^{-1} \sim \text{Wishart}(4, 0.1\text{I}_4)$, where I_4 is the 4×4 identity matrix. Each component of this multi-part joint model with robust random effects captures important aspects of healthcare demand.

Table 3: Poisson Hurdle Model for Hospital Visits

	parameter	mean	95% Credible Interval
Logit: p^H			
Intercept	$\beta_{11}^{H_p}$	**6.32**	[0.46, 12.26]
Wave	$\beta_{12}^{H_p}$	**1.40**	[0.75, 2.35]
Education: GED or Higher?	$\beta_{13}^{H_p}$	0.81	[-0.99, 2.69]
Is Female?	$\beta_{14}^{H_p}$	**0.76**	[0.45, 2.24]
Does health limit work?	$\beta_{15}^{H_p}$	**0.58**	[0.07, 2.05]
Has no difficulty in dressing	$\beta_{16}^{H_p}$	**-1.09**	[-3.78,- 0.63]
Self-reported expectation of living 10+ years	$\beta_{17}^{H_p}$	**-0.17**	[-0.24,- 0.08]
Self-reported health: Δ in current wave	β_{18}	**0.57**	[0.21, 1.04]
Self-reported health: Δ in previous wave	β_{19}	**0.59**	[0.1, 1.41]
Log: μ^H			
Intercept	$\beta_{11}^{H_\lambda}$	**-3.32**	[-4.52,- 0.91]
Wave	$\beta_{12}^{H_\lambda}$	-0.32	[-1.04, 0.06]
Education: GED or Higher?	$\beta_{13}^{H_\lambda}$	-0.94	[-3.22, 0.37]
Is Female?	$\beta_{14}^{H_\lambda}$	-1.02	[-2.83, 0.03]
Does health limit work?	$\beta_{15}^{H_\lambda}$	**0.56**	[0.03, 0.85]
Has no difficulty in dressing	$\beta_{16}^{H_\lambda}$	**-0.11**	[-1.4,- 0.08]
Self-reported expectation of living 10+ years	$\beta_{17}^{H_\lambda}$	**0.08**	[0.04, 0.15]
Self-reported health: Δ in current wave	$\beta_{18}^{H_\lambda}$	-0.19	[-0.45, 0.001]
Self-reported health: Δ in previous wave	$\beta_{19}^{H_\lambda}$	-0.06	[-0.15, 0.002]

The estimates for the two part Poisson hurdle model given by equations 17 and 18 are reported in Table 3. The top panel reports the determinants of the propensity for hospital stays while the bottom panel presents the determinants of the count of hospital stays conditional on stays. It is apparent that flexibility to differentially affect the logit and log portions is important with almost every variable behaving differentially in the two components. Two exceptions to this are: if the respondent states that his/her health condition limits his/her ability to work and if he/she has any difficulty in dressing. If the health condition limits work and if there is difficulty in dressing both raise the propensity for hospital stays as well as the number of hospital stays conditional on there being any stays at all, hence indicating the importance of functional ability in determining the demand for this aspect of healthcare. From the top panel of Table 3, it is evident that, holding all other effects constant, over time (i.e. wave) the propensity for hospital stays increases, but the number of hospital stays does not appear to increase over time. Males are found to be less likely to visit a hospital, on average, than women, yet gender is not found to exert a statistically significant effect on the number of hospital stays. Respondents with higher self-reported expectations of being alive for the next ten years have a lower propensity for hospital stays. Interestingly, conditional on there being any hospital stays, this self-reported expectation is positively associated with the number of stays, which may reflect optimism regarding the effects of any hospital treatments received. The difference in the influence of this variable across the two components of the model once again highlights the importance of using a modelling framework which allows variables to flexibly differentially affect the various components of the model. Changes in self-reported health status, associated with worsening health, on the other hand, are also associated with increases in the propensity for hospital stays, yet do not influence the number of stays. Such findings highlight the important role played by individual's expectations and perceptions of their health status as predictors of the demand for healthcare.

Table 4 reports estimates from the semicontinuous model for out-of-pocket medical expenditure (OOPMD). A number of interesting differences with the Poisson hurdle model are noted. First, the propensity for any OOPMD is unaffected by education, gender, health conditions that may affect work, or changes in self-assessed health status. Holding other things constant over time, however, with no difficulty in dressing themselves, and with a higher self-reported expectation of being alive for the next ten years, respondents have a lower propensity for incurring OOPMD, which ties in with intuition and, once again, highlights the importance of functional ability and perceptions regarding health status as predictors of the demand for healthcare. Interestingly, in the second part of the model, the self-reported expectation also has an inverse association with the amount of medical expenditure, whereas having no difficulty in dressing is positively associated with the amount of medical expen-

Table 4: Two Part Model for Out-of-Pocket Medical Expenses

	parameter	mean	95% Credible Interval
Logit: p^M			
Intercept	$\beta_{11}^{M_p}$	-0.09	[-3.06,2.86]
Wave	$\beta_{12}^{M_p}$	**-1.00**	[-3.41,-0.35]
Education: GED or Higher?	$\beta_{13}^{M_p}$	0.02	[-2.74,3.08]
Is Female?	$\beta_{14}^{M_p}$	-0.01	[-3.04,2.84]
Does health limit work?	$\beta_{15}^{M_p}$	-0.07	[-1.08,1.85]
Has no difficulty in dressing	$\beta_{16}^{M_p}$	**-0.13**	[-2.32,-0.07]
Self-reported expectation of living 10+ years	$\beta_{17}^{M_p}$	**-1.65**	[-3.32,-0.46]
Self-reported health: Δ in current wave	$\beta_{18}^{M_p}$	0.05	[-1.43,1.7]
Self-reported health: Δ in previous wave	$\beta_{19}^{M_p}$	-0.04	[-1.15,0.8]
log: μ^M			
Intercept	$\beta_{11}^{M_\lambda}$	**-3.69**	[-4.56,-2.95]
time	$\beta_{12}^{M_\lambda}$	**0.31**	[0.08,0.42]
Education: GED or Higher?	$\beta_{13}^{M_\lambda}$	0.49	[-0.24,1.29]
Is Female?	$\beta_{14}^{M_\lambda}$	**0.45**	[0.14,1.11]
Does health limit work?	$\beta_{15}^{M_\lambda}$	-0.01	[-0.26,0.23]
Has no difficulty in dressing	$\beta_{16}^{M_\lambda}$	**0.11**	[0.02,1.5]
Self-reported expectation of living 10+ years	$\beta_{17}^{M_\lambda}$	**-0.23**	[-0.38,-0.01]
Self-reported health: Δ in current wave	$\beta_{18}^{M_\lambda}$	0.01	[-0.1,0.12]
Self-reported health: Δ in previous wave	$\beta_{19}^{M_\lambda}$	0.03	[-0.06,0.13]

diture. Once we condition on incurring OOPMD, we also find that, holding other things constant, subsequent waves are characterised by higher OOPMD and that women experience higher health care expenditure than men.

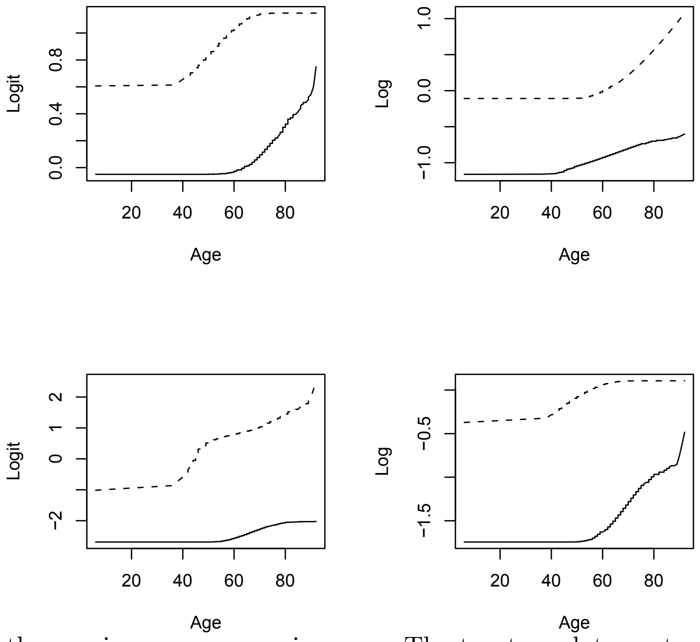
With both the self-reported expectation of being alive for 10+ years variable and the difficulty in dressing variable being statistically significant in each of the four components of the four part model, it seems natural to expect significant correlation across the random effects from each of the components. Table 5 presents estimates for the correlation coefficients across the four components of the model. The two correlation coefficients between the random effects that are non-zero are the correlation between the random effects of the logit and log components of the Poisson hurdle sub-model and that between the random effects of the log portion of the Poisson hurdle model and the log portion of the semi-continuous hurdle model. The first is negative and suggests that individuals with larger unobserved effects on the propensity of hospitalisation tend to have lower unobserved effects on the conditional count of hospital visits. While statistically significant, the correlation coefficient is much smaller (0.20) than the correlation between the random effects from the conditional count of hospital visits from the Poisson hurdle model and the random effects from the conditional OOPMD component of the semi-continuous model (0.66). The high correlation between the unobserved components of the conditional count of hospital visits and the conditional OOPMD is expected as unobserved factors that determine hospital visits are likely to be closely related to unobserved factors that explain OOPMD. Interestingly, there is no correlation between the random effects from the propensity to visit a hospital and the random effects from the conditional OOPMD model. This suggests that, while the conditional count of hospital visits and the conditional OOPMD are closely related to each other, the propensity to visit a hospital per se is determined differently.

Table 5: Correlation between Random Effects Across Models

	mean	95% Credible Interval
corr between logit and log of ZIP	**-0.20**	[-0.49,-0.07]
corr between logit of ZIP and logit of semi-continuous	0.01	[-0.89,0.78]
corr between log of ZIP and logit of semi-continuous	**-0.28**	[-1.91,-0.09]
corr between logit of ZIP and log of semi-continuous	0.02	[-0.09,0.16]
corr between log of ZIP and log of semi-continuous	**0.66**	[0.11,1.25]
corr between logit and log of semi-continuous	0.01	[-0.13,0.16]

Finally, we analyse the effect of age on healthcare demand and how it varies as people age and with gender. Figure 2 plots the effect of aging on each component of the four part model. Figure 2 shows that the demand for healthcare varies significantly across the life cycle and across gender; note this is not apparent from the baseline effects in the regression tables. The first quadrant of the figure shows that there is a large difference in the baseline levels of the demand for healthcare with women having a higher propensity for making any hospital visits. For women, the baseline demand for healthcare does not change until the age of 40 after which it rises linearly until the age of 60. After the age of 60, further aging

Figure 2: Non-linear Effects of Aging for each part of the 4PM



Note: On each plot the x-axis measures age in years. The top two plots capture the gender effects of the Poisson Hurdle model. The top left captures the difference in the propensity for any hospital visit and no hospital visit for women (dotted line) and men (straight line). The top right captures the conditional count of hospital visits. The bottom two captures gender effects in semicontinuous model with the bottom left capturing gender differences in propensity for any OOPMD while the bottom right captures the gender difference in conditional OOPMD.

appears to have almost no additional impact on the propensity to use hospital facilities. Men, on the other hand, have no change in the baseline propensity to visit a hospital until the age of almost 60. Thereafter, the propensity to visit a hospital at least once increases exponentially. The conditional demand for healthcare in terms of the count of visits behaves somewhat differentially - women visit more frequently over their entire life time, while men maintain their baseline rates of hospitalisation almost until the age of 60. Thereafter, men start visiting a hospital more frequently than they had in the past. However, the increase is slower than the increase in the conditional counts observed for women.

Similarly, with OOPMD, we find that women are more likely to incur expenditure and they also tend to incur higher expenditure than men at each stage of the life cycle. From the age of 40, the propensity to incur expenditure rises rapidly until the age of 60 and, thereafter, it increases at a much more modest rate for women. For men, there is no change in the baseline propensity of incurring OOPMD until the age of 60. Thereafter, there is a modest increase in the propensity for incurring any OOPMD. In terms of OOPMD expenditure conditional on there being some expenditure, it is clear that women incur substantially higher costs throughout their lifetime than men, with a modest increase after the age of 40. Consistent with the Poisson hurdle model, men have a much lower level of baseline conditional OOPMD expenditure until the age of 60. After the age of 60, conditional OOPMD expenditure increases very rapidly and the gap between male and female medical expenditure declines rapidly, but does not fully go away.

4 Conclusion

In this paper, we analyse healthcare demand for an aging population using a Bayesian semi-parametric joint modelling framework. We incorporate a number of interesting adaptations to this joint model to ensure that our model is appropriate for this application as well as being robust and allowing us to flexibly estimate a key covariate for an aging population, namely the effects of age itself. In the Bayesian framework, we allow for zero-inflation that is a key characteristic for both hospital stays and out-of-pocket medical expenditure (Duan et al. 1982; Olsen and Schafer 2001; Liu et al. 2010). Thus, our four-part model differentially captures the propensities for usage as well the levels of use across these two measures of healthcare demand. This enables us to uncover complex patterns of correlations across a range of covariates and at different portions of the distribution of each outcome. Using DP priors to specify random effects for each participant allows us to reliably estimate health care demand after accounting for unobserved heterogeneity. Finally, the correlation across the components allows us to borrow information across the two measures of healthcare demand to better understand the co-movement in our joint model in a way that has not been previously applied to healthcare demand.

The four part model allows us to capture a number of important aspects relating to how aging influences healthcare demand. Age splines and their interaction with gender allow us to ascertain that at younger ages healthcare demand is higher for women, whilst after the age of 60, healthcare demand for men increases very rapidly. This affects both hospital visits and out-of-pocket medical expenditure. These findings have different implications: For example, with increased aging, there is need for greater profiling of men as they near 60, which has implications for the health sector, while greater out-of-pocket medical expenses will have important implications for the financial planning of individuals and households as well as for the design of health insurance systems. We hope that our findings will stimulate further research into this area of economics, which is clearly set to increase in terms of its policy relevance in the future.

A Appendix: The Simulation Study

In this appendix, we present two simulation exercises, the purpose of which is to verify the performance of our proposed model in comparison to simpler and parsimonious, but parametric, models.

A.1 Using the DP model for skewed distributed random effects $\boldsymbol{b_i}$

This simulation exercise evaluates the performance of our model when the random effects are from a skewed distribution. For this simulation exercise, we consider the following models:

$$\begin{aligned}
\text{logit}(p_{ij}^H) &= \beta_{11} + \beta_{12}t_{ij} + \beta_{13}X_i + \beta_{14}Z_{ij} + b_{i1} \\
\log(\lambda_{ij}) &= \beta_{21} + \beta_{22}t_{ij} + \beta_{23}X_i + \beta_{24}Z_{ij} + b_{i2} \\
\text{logit}(p_{ij}^M) &= \beta_{31} + \beta_{32}t_{ij} + \beta_{33}X_i + \beta_{34}Z_{ij} + b_{i3} \\
\log(s_{ij}+1) &= \beta_{41} + \beta_{42}t_{ij} + \beta_{43}X_i + \beta_{44}Z_{ij} + b_{i4} + e_{ij} \qquad (21)
\end{aligned}$$

In this model, we consider an individual-specific baseline covariate X_i, random intercepts $b_i = (b_{i1}, b_{i2}, b_{i3}, b_{i4})'$, and a time-varying covariate Z_{ij}, where $i = 1, 2, \ldots, 100$ and $j =$

$1, 2, \ldots, 16$. Data are generated from equation (21) to mimic the HRS data analysed in Section 3. The data is generated using the following steps:

1. The X_i's are assumed to be continuous and generated from a univariate normal distribution with mean $\mu_X = 0$ and $\sigma_X = 0.5$.

2. Time dependent covariates (Z_{ij}) for 16 time-points are generated using $MVN(\boldsymbol{\mu_Z}$, $\boldsymbol{\Sigma_Z})$. In order to maintain a correlation between the Z values in adjacent time-points within one individual, $\boldsymbol{\Sigma_Z}$ was assumed to have an AR(1) variance-covariance structure.

3. Random intercepts b_{i1} are generated from a skewed bimodal distribution (a balanced mixture of the N(-1, 2.25) and log normal (2.30, 0.48) distributions). In order to create correlated random effects, b_{il} was generated as a linear combination of $b_{i1}, b_{i2}, \ldots b_{il-1}$ and a skewed bimodal distribution as described above ($l = 2, 3, 4$).

4. The e_{ij}'s of the two part model are generated from a normal distribution.

5. Finally, we generated Y_{i1} from a hurdle Poisson distribution ($\boldsymbol{p^H}$, $\boldsymbol{\lambda}$) and Y_{i2} from TP($\boldsymbol{p^M}$, $\boldsymbol{s}$). The parameter values used in the simulation exercise are chosen to produce data that are similar to the HRS data. In particular, we take $\beta_{11} = 6.23$, $\beta_{12} = 1.41$, $\beta_{13} = 0.81$, $\beta_{14} = -0.21$, $\beta_{21} = -3.32$, $\beta_{22} = -0.32$, $\beta_{23} = -0.94$, $\beta_{24} = 0.08$, $\beta_{31} = -0.10$, $\beta_{32} = -0.34$, $\beta_{33} = 0.02$, $\beta_{34} = -1.65$, $\beta_{41} = -3.70$, $\beta_{42} = 0.30$, $\beta_{43} = 0.49$ and $\beta_{44} = -0.23$.

6. One thousand simulated data sets are used in the simulation study.

Using the generated data described above, we fit our proposed model with normal random effects and DP random effects. Model performance is evaluated for both the normal and the DP model for random effects b_i. Our results are presented in Table 6 below. We have computed the bias, mean square error (MSE) and coverage probability (CP). The numbers in parentheses in column 1 of Table 6 are the true population values of the parameters. Our simulation results show that the DP model produces better estimates of the model parameters with minimal bias and better coverage probabilities as compared to the normal model.

Table 6: Results for Normal and DP models in the presence of skewed random effects

	Normal Model				DP Model			
Parameter	Mean	Bias	MSE	CP	Mean	Bias	MSE	CP
Logit-p^H								
$\beta_{11}^{H_p}$(6.23)	4.33	-1.91	2.444	0.85	7.02	0.79	1.087	0.93
$\beta_{12}^{H_p}$(1.41)	0.62	-0.79	0.457	0.82	1.77	0.15	0.202	0.95
$\beta_{13}^{H_p}$(0.81)	0.87	0.06	0.313	0.90	0.85	0.04	0.133	0.97
$\beta_{14}^{H_p}$(-0.21)	-0.13	0.08	0.271	0.89	-0.25	-0.03	0.274	0.90
Log-linear-μ^H								
$\beta_{11}^{H_\lambda}$(-3.32)	-5.25	-1.93	1.476	0.86	-3.93	-0.61	1.003	0.91
$\beta_{12}^{H_\lambda}$(-0.32)	-0.23	0.09	0.333	0.89	- 0.29	0.03	0.091	0.94
$\beta_{13}^{H_\lambda}$(-0.94)	-0.81	0.13	0.452	0.90	-0.91	-0.03	0.512	0.91
$\beta_{14}^{H_\lambda}$(0.08)	-0.01	-0.07	0.122	0.84	0.05	-0.03	0.006	0.95
Logit-p^M								
$\beta_{11}^{M_p}$(-0.10)	-0.08	0.02	0.022	0.93	-0.11	-0.01	0.062	0.94
$\beta_{12}^{M_p}$(-0.34)	0.12	0.46	0.013	0.76	-0.36	-0.02	0.014	0.96
$\beta_{13}^{M_p}$(0.02)	0.005	-0.01	0.070	0.83	0.02	0.00	0.086	0.98
$\beta_{14}^{M_p}$(-1.65)	-0.47	1.18	0.097	0.85	-0.99	0.66	0.035	0.91
Log-μ^M								
$\beta_{11}^{M_\lambda}$(-3.70)	-7.12	-3.42	1.303	0.79	-4.17	-0.47	1.406	0.94
$\beta_{12}^{M_\lambda}$(0.30)	0.09	-0.21	0.082	0.81	0.33	-0.03	0.047	0.92
$\beta_{13}^{M_\lambda}$(0.49)	0.48	-0.01	0.013	0.96	0.48	-0.01	0.029	0.96
$\beta_{14}^{M_\lambda}$(-0.23)	-0.11	0.12	0.107	0.90	-0.35	-0.12	0.103	0.95

Note: Number in parenthesis next to each parameter indicates its true population value.

A.2 Spline-based modelling of nonlinear time effects

This simulation exercise illustrates the performance of our proposed model under the complexity of nonlinear time effects. For this simulation exercise, we have considered the following model:

$$\begin{aligned}
\text{logit}(p_{ij}^H) &= \beta_{11} + \beta_{12}t_{ij} + \beta_{13}X_i + \beta_{14}Z_{ij} + b_{i1} + f^p(t_{ij}) \\
\log(\lambda_{ij}) &= \beta_{21} + \beta_{22}t_{ij} + \beta_{23}X_i + \beta_{24}Z_{ij} + b_{i2} + f^\lambda(t_{ij}) \\
\text{logit}(p_{ij}^M) &= \beta_{31} + \beta_{32}t_{ij} + \beta_{33}X_i + \beta_{34}Z_{ij} + b_{i3} + f^M(t_{ij}) \\
\log(s_{ij} + 1) &= \beta_{41} + \beta_{42}t_{ij} + \beta_{43}X_i + \beta_{44}Z_{ij} + b_{i4} + f^s(t_{ij}) + e_{ij} \qquad (22)
\end{aligned}$$

Table 7: Results for Parametric and Spline models in the presence of nonlinear time effects

	Linear Model				Model with Spline			
Parameter	Mean	Bias	MSE	CP	Mean	Bias	MSE	CP
Logit-p^H								
$\beta_{11}^{H_p}$(6.23)	5.33	0.90	1.044	0.90	6.02	0.21	1.087	0.93
$\beta_{12}^{H_p}$(1.41)	0.92	-0.59	0.457	0.89	1.57	0.16	0.202	0.95
$\beta_{13}^{H_p}$(0.81)	0.87	0.06	0.313	0.90	0.85	0.04	0.133	0.97
$\beta_{14}^{H_p}$(-0.21)	0.13	0.44	0.471	0.83	-0.25	-0.03	0.274	0.90
Log-linear-μ^H								
$\beta_{11}^{H_\lambda}$(-3.32)	-5.25	-1.93	1.476	0.90	-3.93	-0.61	1.003	0.92
$\beta_{12}^{H_\lambda}$(-0.32)	-0.23	0.09	0.333	0.89	- 0.29	0.03	0.091	0.93
$\beta_{13}^{H_\lambda}$(-0.94)	-0.81	0.13	0.452	0.90	-0.91	-0.03	0.512	0.91
$\beta_{14}^{H_\lambda}$(0.08)	-0.01	-0.07	0.122	0.84	0.05	-0.03	0.006	0.95
Logit-p^M								
$\beta_{11}^{M_p}$(-0.10)	-0.21	-0.11	0.070	0.90	-0.08	0.02	0.074	0.92
$\beta_{12}^{M_p}$(-0.34)	-0.89	-0.55	0.116	0.89	-0.16	0.18	0.131	0.90
$\beta_{13}^{M_p}$(0.02)	0.05	0.03	0.162	0.95	0.03	0.01	0.057	0.97
$\beta_{14}^{M_p}$(-1.65)	-2.78	-1.23	0.172	0.89	-1.41	0.24	0.078	0.95
Log-μ^H								
$\beta_{11}^{M_\lambda}$(-3.70)	-3.51	0.19	0.109	0.90	-3.56	0.14	0.112	0.93
$\beta_{12}^{M_\lambda}$(0.30)	0.66	0.36	0.082	0.88	0.54	0.24	0.068	0.90
$\beta_{13}^{M_\lambda}$(0.49)	0.82	0.39	0.207	0.90	0.33	-0.16	0.058	0.91
$\beta_{14}^{M_\lambda}$(-0.23)	-0.84	-0.61	0.066	0.87	-0.11	0.12	0.042	0.93

Note: Number in parenthesis next to each parameter indicates its true population value.

In this model, f^p, f^λ, f^M and f^s are the nonlinear time effects for 16 time-points, while the remaining variables have the same interpretation as in equation 21. Data for the simulation study is generated from equation 22 using the following steps:

1. X_i, Z_{ij} and e_{ij} are generated in the same way as described in steps 1, 2 and 4 in the first simulation exercise described above.

2. The random effects b_i's are generated from a multivariate normal distribution.

3. The nonlinear time effects are generated using the nonlinear functions $f^p(t) = 1/9 \cos^2((t+9)/17)$, $f^\lambda(t) = -0.9 + 0.005 \exp((12+t)/12)$, $f^M(t) = 1/2 \cos((t+12)/12) \sin(t/19)$ and $f^s(t) = -1.7 + 0.005 \exp(t/2) I_{\{t \geq 8\}}$.

4. Y_{i1} and Y_{i2} are generated from a hurdle model and semi-continuous distribution, respectively, as described in step 5 of the first simulation exercise using $\boldsymbol{p^H}$, $\boldsymbol{\lambda}$, $\boldsymbol{p^M}$ and $\boldsymbol{s}$ from equation 22.

5. One thousand data sets are generated for analysing model performance.

The spline model and the linear time effects model are fitted with normal random effects. The results of the simulation exercise are presented in Table 7. We have once again computed the bias, mean square error (MSE) and coverage probability (CP). In the presence of nonlinear time effects, the linear time effects model often produces higher bias and a substantially lower CP in the time-varying covariates, although the estimates of the other covariates appear comparable.

Based on the findings of both simulation exercises, we conclude that the model used in the analysis is characterised by good performance in modelling the data. Despite the increased complexity, the approach developed in this paper provides a safeguard against the potential effects of misspecification of the time effects, thus preventing the occurrence of large biases in the estimation of time-varying effects.

References

Atella, V. and P. Deb (2008). Are primary care physicians, public and private sector specialists substitutes or complements? Evidence from a simultaneous equations model for count data. *Journal of Health Economics 27*(3), 770–785.

Arias, E. (2010). United States Life Tables, 2006. *National Vital Statistics Reports 58*(21). http://www.cdc.gov/nchs/data/nvsr/nvsr58/nvsr58 21.pdf.

Blackwell, D., and MacQueen, J. B. (1973). Ferguson Distributions via Polya Urn Schemes, Annals of Statistics, 1, 353–355.

Deb, P. and Trivedi, P. K. (1997). Demand for medical care by the elderly: a finite mixture approach. *Journal of Applied Econometrics 12*(3), 313–336.

Dobriansky, P.J. and Suzman, R.M. and Hodes, R.J. (2007). Why population aging matters: A global perspective *National Institute on Aging, National Institutes of Health, US Department of Health and Human Services, US Department of State.*

Duan, N., J. P. Newhouse, C. N. Morris, and W. G. Manning (1982). A comparison of alternative models for the demand for medical care. Report R-2754-HHS, RAND, Santa Monica, CA.

Ferguson, T. S. (1973). A Bayesian Analysis of Some Nonparametric Problems. *Annals of Statistics 1*, 209—230.

Gelfand, A. E., Dey, D. K., and Chang, H. (1992). Model determination using predictive distributions with implementation via sampling-based methods. in J. M. Bernardo J. O. Berger, A. P. Dawid, and A. F. M. Smith (Ed.),*Bayesian Statistics, (Vol. 4)*, Oxford:Oxford University Press, Oxford,147—-159.

Gilks, W. R., and Wild, P. (1992) Adaptive Rejection Sampling for Gibbs Sampling *Applied Statistics 41*, 337—348.

Hartman, M., A. Catlin, D. Lassman, J. Cylus, and S. Heffler (2008). U.S. health spending by age, selected years through 2004. *Health Affairs 27*(1), w1–w12.

Hjort, N. L., C. Holmes, P. Muller and S. G. Walker (2010),*Bayesian Nonparametrics.* Cambridge University Press.

Hurd, M. D. and K. McGarry (2002). The predictive validity of subjective probabilities of survival. *The Economic Journal 112*(482), 966–985.

Jochmann, M. and R. Leon-Gonzalez (2004). Estimating the demand for health care with panel data: a semiparametric bayesian approach. *Health Economics 13*, 1003–1014.

Jones, A. M. (2000). *Handbook of Health Economics*, Volume 1, Chapter Health econometrics, pp. 265–344. Elsevier: Amsterdam.

Lambert, D. (1992) Zero-Inflated Poisson Regression, with an Application to Defects in Manufacturing. *Technometrics 34*(1), 1–14.

Liu, L., M. R. Conaway, W. A. Knaus, and J. D. Bergin (2008, May). A random effects four-part model, with application to correlated medical costs. *Computational Statistics & Data Analysis 52*(9), 4458–4473.

Liu, L., R. L. Strawderman, M. E. Cowen, and Y.-C. T. Shih (2010). A flexible two-part random effects model for correlated medical costs. *Journal of Health Economics 29*, 110–123.

McGee, D. L., Y. Liao, G. Cao, and R. S. Cooper, (1999). Self-reported Health Status and Mortality in a Multiethnic US Cohort *American Journal of Epidemology*, *149*(1), 41–46.

Mullahy, J. (1986). Specification and testing of some modified count data models. *Journal of Econometrics 33*(3), 341–365.

Naskar, M. and Das, K. (2006). Semiparametric analysis of two-level bivariate binary data. *Biometrics 62*, 1004—1013.

Naya, H., Urioste, J. I., Chang, Y.-M., Rodrigues-Motta, M., Kremer, R. and Gianola, D. (2008). A Comparison between Poisson and Zero-Inflated Poisson Regression

Models with an Application to Number of Black Spots in Corriedale Sheep. *Genetics Selection Evolution 40* (4), 379–394.

Neelon B. H., O'Malley A. J. and Normand S-L. T. (2010). A Bayesian model for repeated measures zeroinflated count data with application to outpatient psychiatric service use. *Statistical Modelling* , *10*, 421–439.

Neelon B. H., O'Malley A. J. and Normand S-L. T. (2011). A Bayesian two-part latent class model for longitudinal medical expenditure data: assessing the impact of mental health and substance abuse parity. *Biometrics 67*, 280—289

Olsen, M. K. and J. L. Schafer (2001). A two-part random-effects model for semicontinuous longitudinal data. *Journal of the American Statistical Association 96*(454), 730–745.

Ohlssen D., L. D. Sharples, and D. J. Spiegelhalter (2007),Flexible Random-effects Models using Bayesian Semi-parametric Models: Application to Institutional Comparisons, *Statistics in Medicine*, *26*, 2088–2112.

Sethuraman, J. (1994), A Constructive Definition of Dirichlet priors, *Statistica Sinica*, *4*, 639-650.

Smith, M. and R. Kohn (1996).Nonparametric regression using Bayesian variable selection. *Journal of Econometrics 75*, 317–343

St.Clair, P., D. Blake, D. Bugliari, S. Chien, O. Hayden, M. Hurd, S. Ilchuk, F.-Y. Kung, A. Miu, C. Panis, P. Pantoja, A. Rastegar, S. Rohwedder, E. Roth, J. Carroll, and J. Zissimopoulos (2009). *RAND HRS Data Documentation Version 1.* Santa Monica.

Strunk, B. C., P. B. Ginsburg, and M. I. Banker (2006). The effect of population aging on future hospital demand. *Health Affairs 25*(3), w141–w149.

Su L., Tom B. D. M. and Farewll V. T. (2009). Bias in 2-part mixed models for longitudinal semicontinuous data. *Biostatistics*,*10*, 374—389.

Tsonaka, R., Verbeke, G., and Lesaffre, E. (2009). A semi-parametric shared parameter model to handle nonmonotone nonignorable missingness. *Biometrics 65*, 81—87.

Wei, Y., A. Ravelo, T. H. Wagner, and P. G. Barnett (2004). The relationships among age, chronic conditions, and healthcare costs. *The American Journal of Managed Care 10*(12), 909–916.

Winkelmann, R. (2004). Health care reform and the number of doctor visits: An econometric analysis. *Journal of Applied Econometrics*,*19*(4), 455–472.

Yu, B., O'Malley, A. J., and Ghosh, P. (2011). Linear mixed models for multiple outcomes using extended multivariate skew-t distributions. *Journal of Statistical Planning and Inference forthcoming.*