Determinants of General Practitioner Utilization: Controlling for Unobserved Heterogeneity*

Logan McLeod†
McMaster University

May 16, 2007

Abstract

A flexible nonparametric kernel conditional density estimator is used to estimate a model of GP utilization. The main determinants of interest are demographic (gender and immigration status); socioeconomic (income and education); and health (self-reported health status) factors. Canadian longitudinal data permits the model to control for unobserved individual heterogeneity. The nonparametric estimator, controlling directly for unobserved individual heterogeneity, performs better than longitudinal parametric models on the dimensions of model performance statistics and predictive ability.

Socioeconomic factors (income and education) as well as immigration status appear to have no influence on a person's decision to contact a GP or on the number of visits to a GP. Gender and health status appear to be important determinants in both the probability of contacting a GP and the number of visits made to a GP.

Keywords: GP utilization; nonparametric kernel conditional density; latent class models; count data; longitudinal data; Canada

*I would like to thank Jerry Hurley, Jeff Racine and Kate Cuff for their invaluable advice and support. I am grateful to Lonnie Magee, Paul Contoyannis and McMaster Graduate Seminar participants for their helpful comments on earlier versions of this paper

†Department of Economics, McMaster University. Kenneth Taylor Hall, Room 330. 1280 Main Street West, Hamilton, Ontario. L8S 4M4. e-mail address: mcleodlc@mcmaster.ca

1 Introduction

Models of health care utilization are central to the development of health care policy. Estimates of model parameters are informative for policy makers. It is important to encourage the application of appropriate estimation methods; where appropriate refers to methods that impose assumptions closely reflecting the characteristics of the data. Inappropriate estimation methods applied to models of GP utilization produce results that can be misleading. Estimation methods imposing assumptions that closely match important features of the observed distribution of GP utilization, are generally more reliable for inference and policy development. The past decade has witnessed an explosion of intellectual effort directed at developing more flexible estimation methods and providing a better match to the observed distribution.

A nonparametric kernel conditional density estimator is applied to model determinants of the probability of visiting a GP. The number of GP visits is a count variable, since the number of visits are reported as a non-negative integer and a count variable takes on only non-negative integer values. Standard count data models make appropriate assumptions for the bounded nature of the dependent variable.

There are four main strands in the literature on modeling count data: (i) modeling endogeneity; (ii) developing more flexible parametric models; (iii) developing longitudinal count models; and (iv) comparing the performance of different models. I will focus on the application of a more flexible nonparametric model in comparison to parametric longitudinal models. The use of longitudinal data permit for control of unobserved individual heterogeneity. Parametric models often impose restrictive distributional assumptions, although less restrictive models have been proposed in recent years. While a number of people have striven to create more flexible parametric specifications, the changes to models tend to be incremental.

Therefore, I opt for a more flexible nonparametric method to estimate a model of general practitioner (GP) utilization. To my knowledge, no previous study has used a nonparametric kernel conditional density estimator. Kernel methods use the underlying data structure to construct functional forms and thus impose no restrictive assumptions prior to estimation.

This study makes three contributions to the literature: (i) the application of a flexible nonparametric kernel conditional density estimator; (ii) the use of Canadian longitudinal data; and (iii) a comparison between longitudinal parametric and nonparametric estimation methods.

Using Canadian longitudinal data, the nonparametric model performs better than existing parametric models in explaining the observed distribution for the number of GP visits. Socioeconomic factors of income and education do not appear to influence peoples decisions to make contact with a GP or the number of visits made to a GP. Self reported health status appears to influence both the probability of contacting a GP and the number of visits made to a GP.

2 Literature

There are relatively few Canadian studies looking at the number of visits to a General Practitioner (GP), and those that have primarily used a cross-section of survey data. In the past 15 years, studies have used the General Social Survey[1] (Birch et al. (1993), Eyles et al. (1995)); the National Population Health Survey (Dunlop et al. (2000), Deri (2005), Sarma and Simpson (2006)); and the Canadian Community Health Survey (Allin (2006), van Doorslaer et al. (2006)). These surveys are nationally representative samples and have a specific focus on the health of the population. The General Social Survey and the Canadian Community Health Survey are both cross sectional surveys. The first three cycles of the National Population Health Survey contain both cross-sectional and longitudinal components. To date, no study has used the full National Population Health Survey longitudinal component to look at GP visits and control for unobserved individual heterogeneity.

The evidence on the effect of income is mixed. A number of studies have found, after controlling for need, no relationship between income and GP utilization (Birch et al. (1993), Eyles et al. (1995), Dunlop et al. (2000) and Sarma and Simpson (2006)). Other studies have found people with higher income have a greater probability to visit their GP than lower income (Allin (2006) and van Doorslaer et al. (2006)). There is some evidence that conditional on making at least one visit, people with lower income visit the general practitioner more than higher income people (van Doorslaer et al. (2006)). Mixed evidence on the affect of income should not be surprising given visits to the GP are free for all residents of Canada.

It must be noted that access is not the same as utilization. We may expect to see people with low income have greater needs-adjusted GP utilization relative to high income people, if

[1] The two cycles of the General Social Survey referred to here are the General Social Survey Cycle 1 (1985) and Cycle 6 (1991). Both have a specific focus on health.

we think people with lower incomes incur a lower opportunity cost of time when making a visit. On the other hand, we may expect people with high income to have greater needs-adjusted GP utilization due to their consumption of preventative care in an attempt to avoid the higher opportunity cost from being ill.

There is also mixed evidence on the effect of education on needs-adjusted GP utilization. A number of studies have found no relationship (Birch et al. (1993), Eyles et al. (1995) and Sarma and Simpson (2006)); while one study found people with more formal education tended to visit their general practitioner more, relative to people with less formal education (Dunlop et al. (2000)). A person with more formal education and in better health appears to use more general practitioner services, and is postulated to do so for preventative care (Birch et al. (1993)). The mixed results on formal education could be expected, as one may expect a more educated person to visit their general practitioner more as they better understand the need for preventative care; or they are better able to recognize the need for treatment due to their current health state. On the other hand if a more formally educated person is a more efficient producer of health, then we would expect to see them consume less health care (Grossman (1972)).

Self-reported health status is found on all Canadian health surveys and is frequently used as a proxy measure of an individuals "actual" health status. One of the strongest, and most consistent, findings is the negative relationship between self-reported health status and GP utilization. People with lower levels of self-reported health visit their GP more than people with higher levels of self-reported health (Birch et al. (1993), Eyles et al. (1995), Dunlop et al. (2000), Allin (2006) and Sarma and Simpson (2006)). This is not surprising as we may expect, if not hope, people in poorer health would seek out treatment to improve their lot.

The number of GP visits are influenced by a number of other demographic and geographic factors: women visit more than men (Eyles et al. (1995), Sarma and Simpson (2006) and Dunlop et al. (2000)); immigrants visit as much or more than Canadian born (Sarma and Simpson (2006)); married individuals visit more than unmarried individuals (Sarma and Simpson (2006)); and there are differences across Provinces (Birch et al. (1993), Sarma and Simpson (2006) and Allin (2006)).

Health related behaviours have also been found to influence the number of GP visits. People who consume more alcohol visit more (Sarma and Simpson (2006)); as do people who smoke relative to non-smokers (Sarma and Simpson (2006)).

In nearly all studies, GP utilization is measured as the number of visits to a GP in the 12 months prior to the survey. The number of GP visits is a count variable. Count data takes on only non-negative integer values. The reported number of visits made to a general practitioner tends to have a large proportion of zeros, with the majority of responses less than five. The resulting empirical distribution is over-dispersed, as the mean is less than the variance. Some have argued that the presences of unobserved individual heterogeneity causes the over-dispersion (Cameron and Trivedi (1998)).

The features of the datum have important implications for appropriate estimation techniques. Canadian studies have, in general, not used estimation methods designed to accommodate count data. Estimation methods to look at socioeconomic factors affecting GP utilization used to date include ordinary least squares on the log-transformation of a count variable; two-part models; and cross-sectional count data models.

Taking a log-transformation of the number of GP visits lessening the effect of higher counts and makes the interpretation of coefficients from ordinary least squares easier. Zero reported visits are lost or modified, as the log-transformation can not account for them.

Deri (2005) employed this method in her estimation, and found the results to be consistent with those from logit, probit and instrumental variable estimation. Log-transformation ignores the restricted support of the dependent variable, which can lead to significant deficiencies unless the mean of the count is large. If the mean is large, normal approximations (such as ordinary least squares) may be appropriate (Cameron and Trivedi (1998)). The ordinary least squares on a application of a log-transformation to the number of GP visits is not appropriate since the observed distribution of GP visits is over-dispersed, has a high proportion of zeros, and a low mean.

Birch et al. (1993), Eyles et al. (1995) and Dunlop et al. (2000) use two-part models. The two-part model is thought to account for the two distinct processes governing the decision to use a GP: (i) making contact with a GP; and (ii) the number of visits made to a GP, conditional on making at least one visit. The two-part model does this by separating non-users from users and modeling separately the probability distributions of zero and positive outcomes. The two-part model fits well with the principal-agent framework thought to determine the use of health care services.

The first stage is modeled as a binary choice model, usually as a probit (Birch et al. (1993) and

Eyles et al. (1995)) or logit (Dunlop et al. (2000)). The second stage estimates the quantity of use, conditional on making contact. Birch et al. (1993) and Eyles et al. (1995) use a sample selection method. Dunlop et al. (2000) use a multivariate logistic model to estimate the probabilities of making: (i) 1-5 visits; or (ii) 6 or more visits.

Converting a count variable into a binary variable removes variation, by masking the underlying data structure, which could be informative. Binary models, such as logit and probit, make strong parametric assumptions about the underlying distribution of the latent variable. Choosing an inappropriate model will impact the estimates.

To my knowledge, there is only one Canadian study employing count data models. Sarma and Simpson (2006) compare the negative binomial, zero-inflated negative binomial, two-part and latent class models using the cross-section from the third cycle of the National Population Health Survey.

The negative binomial model accounts for over dispersion by relaxing the equi-dispersion assumption of the Poisson model. The zero-inflated negative binomial builds on the negative binomial model to account for the high proportion of zeros. The two-part model accounts for the two different decision processes determining the number of GP visits. The latent class model provides a more flexible framework compared to the two-part model. Classification of people into homogeneous groups, such as "ill" and "non-ill," is often estimated using observable characteristics. The latent class formulation is more flexible given it mixes the probability distributions for zero and positive outcomes.

All these models account for one or more of the characteristics of count data. They build on each other by imposing less restrictive assumptions. None of these can adequately control for unobserved individual heterogeneity given they use cross-sectional data.

While most Canadian studies have not employed count data models, there is an extensive literature over the past 25 years that has developed these methods. In the past 10 years there has been an explosion of intellectual effort to further this literature. This literature can be divided into four main strands: (i) modeling endogeneity; (ii) developing more flexible parametric models; (iii) developing longitudinal count models; and (iv) comparing the performance of different models.

First, in certain health care systems, individuals must purchase private insurance to guarantee access to services. An econometric problem arising from this institutional structure is endogeneity, as insurance status and number of visits to a general practitioner are determined endogenously. A number of alternative methods to deal with this have been developed (i.e. Mullahy (1997), Windmeijer and Santos Silva (1997) and others). Given universal insurance coverage for GP services in Canada, the endogeneity of insurance choice is not of concern.

Second, the development of more flexible models is built on the Poisson model. The motivation for more flexible models is driven by the restrictive assumptions of the Poisson and negative binomial and their inability to adequately describe the empirical distribution. Mullahy (1986) introduced the two-part model and Lambert (1992) introduced the zero-inflated model to account for the large proportion of zeros found in the data. Pohlmeier and Ulrich (1995)) introduced the two-part model to account for differences between users and non-users. Deb and Trivedi (1997) introduced the latent class model as a more flexible way to distinguish between different types of individuals. The two-part model has sharp distinction between groups of "users" and "non-users." The latent class model has a more fuzzy distinction between classes of "ill" and "non-ill" people, where class membership is frequently estimated based on observable characteristics.

A third strand of the literature has extended cross-sectional count models to panel data. The panel models explicitly, or implicitly, control for unobserved individual heterogeneity. Hausman et al. (1984) developed the Poisson and negative binomial fixed and random effect models, controlling for unobserved individual heterogeneity by conditioning the likelihood function on the sum of counts for an individual. Different specifications for two-part models have been explored by Van Ourti (2004); and Bago d'Uva (2005) and Bago d'Uva (2006) have extended the latent class models to a panel framework.

Finally, Sarma and Simpson (2006), Bago d'Uva (2005) and Bago d'Uva (2006) have looked at a series of models and used model performance statistics. Model performance statistics frequently used include the Akaike Information Criterion, Bayesian Information Criterion and log-likelihoods.

Proper parametric models still impose certain restrictions. For example, distributional assumptions on the dependent variable or the dispersion parameter of the Poisson process have a specific distribution (gamma, log-normal etc.). The literature discussed herein introduces more flexible parametric models by tweaking a component of previous models, such as how the latent class model blurs the sharp distinction between non-users and users in the two-part model. Instead, I introduce a new direction by employing a nonparametric estimator to the literature on estimating general practitioner utilization.

3 Methodology

3.1 Poisson Fixed Effects

Individual i is observed over a number of cycles t, where $t = 1, 2, \ldots, T_i$. Individual i reports the number of GP visits they have made in the previous 12 months (y_{it}) and series of other socioeconomic, demographic and health information ($\mathbf{x}_{it}$).

The starting point in panel count data modeling is the Poisson model.[2] The Poisson model assumes the dependent variable (y_{it}) follows a Poisson distribution, with Poisson parameter μ_{it}.[3] The Poisson parameter is initially parameterized to be a deterministic function of the independent variables ($\mathbf{x}_{it}$) with an exponential functional form: $\mu_{it} = e^{\mathbf{x}_{it}\beta}$.

The probability distribution of y_{it} is given by:

$$f(y_{it}|\mathbf{x}_{it}) = \frac{e^{-\mu_{it}}\mu_{it}^{y_{it}}}{y_{it}!},\ y_{it} = 0, 1, 2, \ldots \tag{1}$$

The central assumption of the Poisson model is the equi-dispersion property, which states the expected conditional mean is equal to the conditional variance:

$$E[y_{it}|\mathbf{x}_{it}] = \mu_{it} = Var[y_{it}|\mathbf{x}_{it}] \tag{2}$$

If the fundamental assumption of time independence is not violated then the Poisson model allows for aggregation of y_{it} over time. If μ_{it} is the Poisson parameter of y_{it} in period t, then the aggregated observations $y_i = \sum_t y_{it}$ will have a Poisson parameter $\mu_i = \sum_t \mu_{it}$.

To incorporate a fixed effect into the Poisson distribution, the unobserved individual heterogeneity that may be correlated with $\mathbf{x}_{it}$ is introduced through the Poisson parameter. Define the Poisson parameter to be a function of an individual specific effect, α_i, and the original Poisson

[2] See Hausman et al. (1984) for a more detailed presentation of this model.

[3] See Cameron and Trivedi (1998) and Cameron and Trivedi (2005).

parameter, μ_{it}, such that: $\tilde{\mu}_{it} = \alpha_i \mu_{it}$. Under conditioning, no distributional assumptions need to be made about $\tilde{\mu}_{it}$.

Re-writing (1) gives:

$$f\left(y_{it}|\mathbf{x}_{it}\right) = \frac{e^{-\tilde{\mu}_{it}}\tilde{\mu}_{it}^{y_{it}}}{y_{it}!} \tag{3}$$

The likelihood function is conditioned on the sum of the number of GP visits ($\sum_t y_{it}$). Since y_{it} has an assumed Poisson distribution, the Poisson parameter $\sum_t \tilde{\mu}_i = \alpha_i \sum_t \mu_{it}$. Thus, the distribution of y_{it} conditional on $\sum_t y_{it}$ is a multinomial distribution:

$$\tilde{f}\left(Y_{i1} = y_{i1}, \ldots, Y_{iT_i} = y_{iT_i}| \sum_t Y_{it}\right) = \frac{f\left(Y_{i1} = y_{i1}, \ldots, Y_{iT_i} = y_{iT_i}\right)}{f\left(\sum_t y_{it}\right)} \tag{4}$$

Where

$$\begin{aligned} f\left(Y_{i1} = y_{i1}, \ldots, Y_{iT_i} = y_{iT_i}\right) &= \prod_{t=1}^{T_i} \tilde{f}\left(y_{it}|\mathbf{x}_{it}\right) \\ &= \prod_{t=1}^{T_i} \frac{e^{-\alpha_i} e^{\mu_{it}}\left(\alpha_i \mu_{it}\right)^{y_{it}}}{y_{it}!} \\ &= \frac{e^{-\sum_t \tilde{\mu_{it}}} \prod_{t=1}^{T_i} (\tilde{\mu_{it}})^{y_{it}}}{\prod_{t=1}^{T_i} (y_{it}!)} \end{aligned} \tag{5}$$

and

$$f\left(\sum_t y_{it}\right) = \frac{e^{-\sum_t \tilde{\mu}_{it}} \sum_t \tilde{\mu}_{it}^{\sum_t y_{it}}}{\left(\sum_t y_{it}\right)!} \tag{6}$$

Using (5) and (6) in (4), we get

$$\begin{aligned}
\tilde{f}\left(Y_{i1}=y_{i1},\ldots,Y_{iT_i}=y_{iT_i}\Big|\sum_t Y_{it}\right) &= \left(\frac{e^{-\sum_t \tilde{\mu_{it}}}\prod_{t=1}^{T_i}(\tilde{\mu_{it}})^{y_{it}}}{\prod_{t=1}^{T_i}(y_{it}!)}\right) \Big/ \left(\frac{e^{-\sum_t \tilde{\mu}_{it}}\sum_t \tilde{\mu}_{it}^{\sum_t y_{it}}}{(\sum_t y_{it})!}\right) \\
&= \frac{(\sum_t y_{it})!}{\prod_{t=1}^{T_i}(y_{it}!)}\prod_{t=1}^{T_i}\left(\frac{\tilde{\mu}_{it}}{\sum_t \tilde{\mu}_{it}}\right)^{y_{it}} \\
&= \frac{(\sum_t y_{it})!}{\prod_{t=1}^{T_i}(y_{it}!)}\prod_{t=1}^{T_i}\left(\frac{\mu_{it}}{\sum_t \mu_{it}}\right)^{y_{it}} \qquad (7)
\end{aligned}$$

What can be seen from (7) is the probability density of y_{it} when conditioned on $\sum_t y_{it}$ is independent of the individual effect α_i.

The contribution of the i^{th} cross-sectional unit to the conditional log-likelihood function is derived by taking the natural logarithm of (7).

$$\begin{aligned}
\ln L_{P,i}(\beta) &= \ln \tilde{f}\left(Y_{i1}=y_{i1},\ldots,Y_{iT_i}=y_{iT_i}\Big|\sum_t Y_{it}\right) \\
&= \ln\left[\frac{(\sum_t y_{it})!}{\prod_{t=1}^{T_i}(y_{it}!)}\prod_{t=1}^{T_i}\left(\frac{\mu_{it}}{\sum_t \mu_{it}}\right)^{y_{it}}\right] \\
&= \ln\left[\frac{\Gamma(\sum_t y_{it}+1)}{\sum_t \Gamma(y_{it}+1)}\prod_{t=1}^{T_i}\left(\frac{\mu_{it}}{\sum_t \mu_{it}}\right)^{y_{it}}\right] \\
&= \ln\Gamma\left(\sum_t y_{it}+1\right)-\sum_t \ln\Gamma(y_{it}+1)+\sum_t y_{it}\ln\left(\frac{\mu_{it}}{\sum_t \mu_{it}}\right) \\
&= \ln\Gamma\left(\sum_t y_{it}+1\right)-\sum_t \ln\Gamma(y_{it}+1)+\sum_t y_{it}\ln\left(\frac{e^{\mathbf{x}_{it}\beta}}{\sum_t e^{\mathbf{x}_{it}\beta}}\right) \qquad (8)
\end{aligned}$$

The conditional log-likelihood function is the:

$$\begin{aligned}
\ln L_P(\beta) &= \sum_i \ln L_{P,i}(\beta) \\
&= \sum_i\left(\ln\Gamma\left(\sum_t y_{it}+1\right)-\sum_t \ln\Gamma(y_{it}+1)+\sum_t y_{it}\ln\left(\frac{e^{\mathbf{x}_{it}\beta}}{\sum_t e^{\mathbf{x}_{it}\beta}}\right)\right) \qquad (9)
\end{aligned}$$

3.2 Negative Binomial Fixed Effects

The negative binomial fixed effects model relaxes the equi-dispersion assumption made in the Poisson fixed effects model.[4]

In the Negative Binomial framework the dependent variable (y_{it}) has a Poisson distribution, with a Poisson parameter λ_{it}. As defined by (1), the probability distribution of y_{it} is given by:

$$f(y_{it}|\lambda_{it}) = \frac{e^{-\lambda_{it}}\lambda_{it}^{y_{it}}}{y_{it}!}, \; y_{it} = 0, 1, 2, \ldots \tag{10}$$

Using the properties of the Gamma function (see Appendix C), (10) can be rewritten as:

$$f(y_{it}|\lambda_{it}) = \frac{e^{-\lambda_{it}}\lambda_{it}^{y_{it}}}{\Gamma(y_{it}+1)} \tag{11}$$

For the negative binomial model, the Poisson parameter (λ_{it}) is defined to be a random variable from a gamma distribution with gamma parameters μ_{it} and δ. The individual and time specific effect is captured by μ_{it}, and the unobserved common effect across individuals and time is captured by δ. The individual and time specific effect is parameterized to be a deterministic function of the independent variables ($\mathbf{x}_{it}$) with an exponential functional form: $\mu_{it} = e^{\mathbf{x}_{it}\beta}$. If we define λ_{it} to have a Gamma (δ, μ_{it}) density:

$$g(\lambda_{it}|\delta, \mu_{it}) = \frac{\delta^{\mu_{it}}\lambda_{it}^{\mu_{it}-1}e^{-\delta\lambda_{it}}}{\Gamma(\mu_{it})} \tag{12}$$

Note that this gamma distribution has $E[\lambda_{it}] = \mu_{it}/\delta$ and $Var[\lambda_{it}] = \mu_{it}/\delta^2$. Since $E[\lambda_{it}] \neq Var[\lambda_{it}]$, the equidispersion property of the Poisson model is broken. The Poisson model is a special case of (12), when $\delta = 0$.

The marginal density of y_{it}, unconditional on λ_{it} but conditional on the deterministic parameters μ_{it} and δ, is given by integrating out λ_{it}.

$$h(y_{it}|\mu_{it}, \delta) = \int_0^\infty f(y_{it}|\lambda_{it})g(\lambda_{it}|\delta, \mu_{it})d\lambda_{it} \tag{13}$$

[4] See Hausman et al. (1984), Cameron and Trivedi (1986) and Cameron and Trivedi (1998) for detailed presentation of this model

Using (11) and (12) in (13) shows the negative binomial as a gamma mixture of Poisson random variables:

$$\begin{aligned} h(y_{it}|\mu_{it},\delta) &= \int_0^\infty \left(\frac{e^{-\lambda_{it}}\lambda_{it}^{y_{it}}}{\Gamma(y_{it}+1)}\right)\left(\frac{\delta^{\mu_{it}}\lambda_{it}^{\mu_{it}-1}e^{-\delta\lambda_{it}}}{\Gamma(\mu_{it})}\right)d\lambda_{it} \\ &= \left(\frac{\delta^{\mu_{it}}}{\Gamma(\mu_{it})\Gamma(y_{it}+1)}\right)\int_0^\infty e^{-(\lambda_{it}+\delta\lambda_{it})}\lambda_{it}^{y_{it}+\mu_{it}-1}d\lambda_{it} \\ &= \left(\frac{\delta^{\mu_{it}}}{\Gamma(\mu_{it})\Gamma(y_{it}+1)}\right)\left(\frac{\Gamma(y_{it}+\mu_{it})}{(1+\delta)^{\mu_{it}+y_{it}}}\right) \\ &= \frac{\Gamma(y_{it}+\mu_{it})}{\Gamma(\mu_{it})\Gamma(y_{it}+1)}\left(\frac{1}{1+\delta}\right)^{\mu_{it}}\left(\frac{\delta}{1+\delta}\right)^{y_{it}} \end{aligned} \tag{14}$$

As in the Poisson fixed effects model, the joint probability of the counts for each individual is conditioned on the sum of the counts for all individuals $\sum_t y_{it}$. The resulting distribution of y_{it} conditional on $\sum_t y_{it}$ is a multinomial distribution:

$$\tilde{h}\left(Y_{i1}=y_{i1},\ldots,Y_{iT_i}=y_{iT_i}|\sum_t Y_{it}\right) = \frac{h\left(Y_{i1}=y_{i1},\ldots,Y_{iT_i}=y_{iT_i}\right)}{h\left(\sum_t y_{it}\right)} \tag{15}$$

where:

$$\begin{aligned} h\left(Y_{i1}=y_{i1},\ldots,Y_{iT_i}=y_{iT_i}\right) &= \prod_{t=1}^{T_i} h(y_{it}|\mu_{it},\delta) \\ &= \prod_{t=1}^{T_i}\frac{\Gamma(y_{it}+\mu_{it})}{\Gamma(\mu_{it})\Gamma(y_{it}+1)}\left(\frac{1}{1+\delta}\right)^{\mu_{it}}\left(\frac{\delta}{1+\delta}\right)^{y_{it}} \\ &= \left(\frac{1}{1+\delta}\right)^{\sum_t \mu_{it}}\left(\frac{\delta}{1+\delta}\right)^{\sum_t y_{it}}\prod_{t=1}^{T_i}\frac{\Gamma(y_{it}+\mu_{it})}{\Gamma(\mu_{it})\Gamma(y_{it}+1)} \end{aligned} \tag{16}$$

and

$$h\left(\sum_t y_{it}\right) = \frac{\Gamma(\sum_t y_{it}+\sum_t \mu_{it})}{\Gamma(\sum_t \mu_{it})\Gamma(\sum_t y_{it}+1)}\left(\frac{1}{1+\delta}\right)^{\sum_t \mu_{it}}\left(\frac{\delta}{1+\delta}\right)^{\sum_t y_{it}} \tag{17}$$

Using (16) and (17) in (15), we get (18).

$$\begin{aligned}
\tilde{h}\left(Y_{i1}=y_{i1},\ldots,Y_{iT_i}=y_{iT_i}|\sum_t Y_{it}\right) &= \left(\left(\frac{1}{1+\delta}\right)^{\sum_t \mu_{it}}\left(\frac{\delta}{1+\delta}\right)^{\sum_t y_{it}}\prod_{t=1}^{T_i}\frac{\Gamma(y_{it}+\mu_{it})}{\Gamma(\mu_{it})\Gamma(y_{it}+1)}\right) \\
&\times\left(\frac{\Gamma(\sum_t y_{it}+\sum_t \mu_{it})}{\Gamma(\sum_t \mu_{it})\Gamma(\sum_t y_{it}+1)}\left(\frac{1}{1+\delta}\right)^{\sum_t \mu_{it}}\left(\frac{\delta}{1+\delta}\right)^{\sum_t y_{it}}\right)^{-1} \\
&= \frac{\Gamma(\sum_t \mu_{it})\Gamma(\sum_t y_{it}+1)}{\Gamma(\sum_t y_{it}+\sum_t \mu_{it})}\prod_{t=1}^{T_i}\frac{\Gamma(y_{it}+\mu_{it})}{\Gamma(\mu_{it})\Gamma(y_{it}+1)} \qquad (18)
\end{aligned}$$

The contribution of the i^{th} cross-sectional unit to the conditional log-likelihood function is derived by taking the natural logarithm of (18).

$$\begin{aligned}
L_{NB,i}(\beta) &= \ln\tilde{h}\left(Y_{i1}=y_{i1},\ldots,Y_{iT_i}=y_{iT_i}|\sum_t Y_{it}\right) \\
&= \ln\left[\frac{\Gamma(\sum_t \mu_{it})\Gamma(\sum_t y_{it}+1)}{\Gamma(\sum_t y_{it}+\sum_t \mu_{it})}\prod_{t=1}^{T_i}\frac{\Gamma(y_{it}+\mu_{it})}{\Gamma(\mu_{it})\Gamma(y_{it}+1)}\right] \\
&= \ln\Gamma\left(\sum_t \mu_{it}\right)+\ln\Gamma\left(\sum_t y_{it}+1\right)-\ln\Gamma\left(\sum_t \mu_{it}+\sum_t y_{it}\right) \\
&\quad +\sum_t\left(\ln\Gamma(\mu_{it}+y_{it})-\ln\Gamma(\mu_{it})-\ln\Gamma(y_{it}+1)\right) \qquad (19)
\end{aligned}$$

The conditional log-likelihood function is the:

$$\ln L_{NB}(\beta) = \sum_i \ln L_{NB,i}(\beta) \qquad (20)$$

3.3 Latent Class Negative Binomial Panel

The Latent Class Negative Binomial Panel Model (LCNB-Pan), presented by Bago d'Uva (2005), is based on the finite mixture models proposed by Aitkin and Rubin (1985) and more recently by Deb and Trivedi (1997). The Latent Class Negative Binomial methodology is argued to provide a more flexible parametric approach to modelling utilization (Deb and Trivedi (1997), Deb and Trivedi (2002) and Jiménez-Martín et al. (2002)).

The latent class approach assumes the sample of individuals are drawn from a population consisting of a finite number of latent classes. Each person in the sample is thought to have been drawn from one of the C latent classes, each class with a different underlying distribution.

Individual i is observed over a number of cycles t, where $t = 1, 2, \dots, T_i$ and belongs to latent class j, where $j = 1, \dots, C$. Conditional on a vector of covariates ($\mathbf{x}_{it}$), individuals are homogeneous within latent class j, but are heterogeneous across latent classes. The probability individual i belongs to latent class j is equal to π_{ij} such that $0 < \pi_{ij} < 1$ and $\sum \pi_{ij} = 1$. The parameter π_{ij} can be defined as a constant or further parameterized.

Let $y_i = [y_{i1}, \dots, y_{iT_i}]$ be the dependent variable for individual i over the panel. Conditional on being a member of latent class j, y_{it} has a probability density function $f_j(y_{it}|\mathbf{x}_{it}, \theta_j)$, where $\theta_j = (\alpha_j, \beta_j)$ are vectors of parameters. The joint density density of y_i is the product of the independent probability densities such that:

$$g(y_i|\mathbf{x}_i; \pi_{i1}, \dots, \pi_C; \theta_{i1}, \dots, \theta_C) = \sum_{j=1}^{C} \pi_{ij} \prod_{t=1}^{T_i} f_j(y_{it}|\mathbf{x}_{it}, \theta_j) \tag{21}$$

As in Bago d'Uva (2005), the conditional density of y_{it} is determined by a negative binomial model.

$$f_j(y_{it}|\mathbf{x}_{it}, \theta_j) = \frac{\Gamma(\psi_{j,it} + y_{it})}{\Gamma(\psi_{j,it})\Gamma(y_{it}+1)} \left(\frac{\psi_{j,it}}{\mu_{j,it} + \psi_{j,it}}\right)^{\psi_{j,it}} \left(\frac{\mu_{j,it}}{\mu_{j,it} + \psi_{j,it}}\right)^{y_{it}} \tag{22}$$

where $\Gamma(\cdot)$ is the gamma function, $\mu_{j,it} = e^{\mathbf{x}'_{it}\beta_j}$ and $\psi_{j,it} = (1/\alpha_j)\mu^k_{j,it}$. The parameters $\alpha_j > 0$ are overdispersion parameters, and k is an arbitrary constant.

Conditional on the latent class an individual belongs to, the joint distribution of y_i is the product of aggregated $f_j(y_{it}|\mathbf{x}_{it}, \theta_j)$ obtained by replacing $f_j(y_{it}|\mathbf{x}_{it}, \theta_j)$ in (21) with (22).

There are assumed to be only two latent classes ($j = 1, 2$). Two latent classes have been found sufficient to explain GP utilization (Deb and Trivedi (1997), Deb and Trivedi (2002) and Bago d'Uva (2006)). Class membership probabilities (π and $1 - \pi$) are taken as parameters to be estimated and are parameterized as a logit model:

$$\pi = \frac{e^{z_i \gamma_j}}{1 + e^{z_i \gamma_j}} \tag{23}$$

where z_i are time-invariant individual characteristics, and γ_j are the parameter estimates for class j.

As in Bago d'Uva (2005), the model is restricted so all slopes are equal across latent classes. Namely, the model places the restriction $\theta_i = \theta_j, \ \forall i \neq j$. This represents the case where there is unobserved individual heterogeneity, but not in the responses to the covariates. It is also assumed the arbitrary constant is equal to zero ($k = 0$). Applying these two assumptions to (22), the distribution function becomes:

$$\begin{aligned}
f_j(y_{it}|\mathbf{x}_{it}, \theta_j) &= \frac{\Gamma((1/\alpha_j)\mu_{j,it}^k + y_{it})}{\Gamma((1/\alpha_j)\mu_{j,it}^k)\Gamma(y_{it}+1)} \left(\frac{(1/\alpha_j)\mu_{j,it}^k}{\mu_{j,it} + (1/\alpha_j)\mu_{j,it}^k}\right)^{(1/\alpha_j)\mu_{j,it}^k} \left(\frac{\mu_{j,it}}{\mu_{j,it} + (1/\alpha_j)\mu_{j,it}^k}\right)^{y_{it}} \\
&= \frac{\Gamma((1/\alpha_j)(1) + y_{it})}{\Gamma((1/\alpha_j)(1))\Gamma(y_{it}+1)} \left(\frac{(1/\alpha_j)(1)}{\mu_{j,it} + (1/\alpha_j)(1)}\right)^{(1/\alpha_j)(1)} \left(\frac{\mu_{j,it}}{\mu_{j,it} + (1/\alpha_j)(1)}\right)^{y_{it}} \\
&= \frac{\Gamma((1/\alpha_j) + y_{it})}{\Gamma((1/\alpha_j))\Gamma(y_{it}+1)} \left(\frac{(1/\alpha_j)}{e^{\mathbf{x}'_{it}\beta_j} + (1/\alpha_j)}\right)^{(1/\alpha_j)} \left(\frac{e^{\mathbf{x}'_{it}\beta_j}}{e^{\mathbf{x}'_{it}\beta_j} + (1/\alpha_j)}\right)^{y_{it}} \\
&= \frac{\Gamma((1/\alpha_j) + y_{it})}{\Gamma((1/\alpha_j))\Gamma(y_{it}+1)} \left(1 + \alpha_j e^{\mathbf{x}'_{it}\beta_j}\right)^{-(1/\alpha_j)} \left(1 + \frac{1}{\alpha_j e^{\mathbf{x}'_{it}\beta_j}}\right)^{-y_{it}} \qquad (24)
\end{aligned}$$

This model is estimated using maximum-likelihood. The the log-likelihood function is constructed from the probability of belonging to latent class j (π) and (24).

$$L(\theta_j) = \ln\left[\pi f_1(y_{it}|\mathbf{x}_{it}, \theta_1) + (1-\pi) f_2(y_{it}|\mathbf{x}_{it}, \theta_2)\right] \qquad (25)$$

3.4 Nonparametric Estimator

The nonparametric estimator employed is a kernel conditional density estimator. For the following sections, let y denote the number of GP visits, and $\mathbf{x}$ denote the independent variables. Given that $\mathbf{x}$ is a mixture of continuous ($\mathbf{x}^c$) and discrete ($\mathbf{x}^d$) variables, denote $\mathbf{x} = (\mathbf{x}^c, \mathbf{x}^d)$.

The conditional density of y given $\mathbf{x}$ is $g(y|\mathbf{x})$. Note that $g(y|\mathbf{x})$ is the ratio of the joint density of $(\mathbf{x}, y)$ to the marginal density of $(\mathbf{x})$:

$$g(y|\mathbf{x}) = \frac{f(\mathbf{x}, y)}{\mu(\mathbf{x})} \qquad (26)$$

The actual functions $f(\mathbf{x}, y)$ and $\mu(\mathbf{x})$ are unknown, so kernel estimators of these functions are used. Let $\hat{f}(\mathbf{x}, y)$ and $\hat{\mu}(\mathbf{x})$ denote kernel estimators of $f(\mathbf{x}, y)$ and $\mu(\mathbf{x})$.

$$\hat{g}(y|\mathbf{x}) = \frac{\hat{f}(\mathbf{x}, y)}{\hat{\mu}(\mathbf{x})} \tag{27}$$

$$\hat{f}(\mathbf{x}, y) = \frac{1}{n}\sum_{i=1}^{n} K_\gamma(\mathbf{x}, X_i) k_{h_0}(y, Y_i) \tag{28}$$

$$\hat{\mu}(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^{n} K_\gamma(\mathbf{x}, X_i) \tag{29}$$

where h_0 is the smoothing parameter associated with y, γ are the smoothing parameters for x^c and x^d respectively, such that $\gamma = (h, \lambda)$ and $k_{h_0}(y, Y_i) = h_0^{-1} k\left((y - Y_i)/h_0\right)$. The kernels are defined as:

$$K_\gamma(\mathbf{x}, X_i) = W_h(\mathbf{x}^c, X_i^c) L(\mathbf{x}^d, X_i^d, \lambda) \tag{30}$$

$$W_h(\mathbf{x}^c, X_i^c) = \prod_{s=1}^{p} \frac{1}{h_s} w\left(\frac{x_s^c - X_{is}^c}{h_s}\right) \tag{31}$$

$$L(\mathbf{x}^d, X_i^d, \lambda) = \prod_{s=1}^{q} \left(\frac{\lambda_s}{(c_s - 1)}\right)^{[N_{is}(\mathbf{x})]} (1 - \lambda_s)^{[1 - N_{is}(\mathbf{x})]} \tag{32}$$

where $N_s(\mathbf{x})$ is an indicator function equaling one if $X_{is}^d \neq x_s^d$ (zero otherwise) and c_s is the number of possible outcomes for discrete variable s.[5] A second-order Gaussian kernel is used for the continuous variables; a Wang-Van Ryzin kernel is used for the ordered discrete variables; and a Li-Racine kernel is used for the unordered discrete variables.

The task of selecting appropriate smoothing parameters (bandwidths), especially in the context of multivariate mixed data, is not trivial. Employing a data driven method of bandwidth selection is most appropriate in this case.

The method used to select bandwidths is least squares cross validation which selects bandwidths (h, λ) by minimizing the weighted integrated square error (ISE).

$$ISE = \sum_{x^d} \int \{\hat{g}(y|\mathbf{x}) - g(y|\mathbf{x})\}^2 \mu(\mathbf{x}) W_h(\mathbf{x}^c) d\mathbf{x}^c dy \tag{33}$$

where $\hat{g}(y|(\mathbf{x})$, $\mu(\mathbf{x})$ and $W_h(\mathbf{x}^c)$ are as defined above.

[5] Refer to Li and Racine (2007) for a thorough presentation of this method.

One advantage of this method is that it automatically smooths out irrelevant variables by selecting bandwidths appropriately. One disadvantage is that this method is computationally intensive.[6]

All bandwidth selection and nonparametric model estimation was done on a cluster at the McMaster Research Data Centre. The use of a cluster reduces processing time due required for the computationally intensive methods. When least squares cross validation is run to select bandwidths, the bandwidths selected resulted in the model oversmoothing. Over smoothing was caused by the short panel and lack of insufficient variation to optimize bandwidth selection for the individual identifier variable (ID). The insufficient variation was due to the short panel (5 cycles). To prevent the model from over smoothing, the bandwidth for the ID variable was set manually to balance the in- and out-of-sample predictions.

4 Data

The first five cycles (1994/1995 - 2002/2003) of the Canadian National Population Health Survey are used. The National Population Health Survey collects information on the health and socio-demographic information of the Canadian population.

The National Population Health Survey started in 1994/1995, with subsequent cycles collected every two years. The first three cycles (1994/1995, 1996/1997 and 1998/1999) contained both a cross-sectional and longitudinal component. However, starting in cycle 4 (2000/2001) the National Population Health Survey became strictly longitudinal. At the present, there are 5 cycles available. The National Population Health Survey longitudinal sample includes 17,276 persons from all ages and is not renewed over time.[7]

The target population of the longitudinal National Population Health Survey is comprised of household residents in 1994/1995 age 12 and older in all provinces, excluding populations in the Territories, residents of health institutions, on Indian Reserves, Canadian Forces Bases and some remote areas in Quebec and Ontario.

The dependent variable of interest is the number of consultations with a family doctor or general practitioner (GP). The specific question asked of all respondents is:

[6]Refer to Hall et al. (2004) for a thorough presentation of the method of Least Squares Cross Validation.

[7]See Statistics Canada (2004).

> *(Not counting when you were an overnight patient) In the past 12 months, how many times have you seen or talked on the telephone with a family doctor or general practitioner about your physical, emotional or mental health?*

Responses take the form of non-negative integer values. This is important as it will inform our choice of estimation method.

The independent variables have been broken down to account for demand and supply side factors affecting utilization.[8] The main independent variables of interest are two measures of demographics (gender and immigration status), two measures of socioeconomic status (education and household income) and one measure of health (self-reported health status).

Immigration status defines three categories: (i) recent immigrant; (ii) long-term immigrant; and (iii) Canadian born. Recent immigrants are defined as people who have immigrated in the past 10 years. Long-term immigrants are defined as people who immigrated more than 10 years ago.

Household income is defined as a continuous income variable in real 1996 dollars. In order to account for inflation over the 10 years of the survey, a continuous income variable needed to be constructed. Construction of the income variable was necessary due to the absence of reported continuous income on the first two cycles. Cycle three through five give people the option of reporting continuous income. Continuous income is modeled on cycle three through five then fit to all cycles to predict a the continuous income variable. The predicted income is then converted to real 1996 dollars using the consumer price index reported on the survey.

The education variable is a derived variable for the highest level of education attained, classified into four categories: 'less than high school,' 'high school graduate,' 'some post secondary,' and 'post secondary graduation.'

Self-reported health status is asked of all respondents. They are asked to answer the question "*In general, would you say your health is:*" with a response of excellent, very good, good, fair or poor. Other measures of health status being controlled for include the number of chronic conditions an individual has and activity limitation.

The sample is restricted to those individuals with a complete response patterns across all five cycles of the survey; 18 years of age or older in cycle 1; without missing information on the

[8]See Appendix A Table 4 for definitions of how each independent variable.

number of GP visits; and who report making less than 30 GP visits. A few people move to a territory or to the United States, and hence are removed from the sample as there is limited information collected as these geographies are beyond the scope of the survey. These restrictions result in a final sample size of just over 8,500 people.

Descriptive statistics for the number of GP visits are reported in Table 1.

Table 1: Summary of GP Visits

	Mean	Median	0 visits (%)	> 5 visits (%)
All	3.0	2.0	20.9	15.5
Males	2.4	1.0	27.0	11.1
Females	3.6	2.0	15.3	19.5
Immigrants	3.3	2.0	18.6	17.0
Canadian Born	3.0	2.0	21.4	15.2

The overall mean number of GP visits in the sample is 3.0. The mean for males is slightly lower (2.4) and is slight higher for females (3.6). There is a smaller difference in means between immigrants (3.3) and Canadian born (3.0).

A more robust measure of central tendency is the median. The overall median for the sample is 2.0. All groups, except males, have a median of 2.0. Males have a lower median than other groups (1.0).

Overall, the proportion of individuals reporting zero GP visits in a year is roughly 21%. Nearly twice the proportion of males (27.0%) than females (15.3%) report making zero visits to their general practioner. This proportion is slightly lower for immigrants (18.6%) than for Canadian born (21.4%).

The proportion of people reporting 5 or more GP visits is (15.5%); and is greater for females (19.5%) than for males (11.1%). There is not as large a difference between immigrants (17.0%) and Canadian born (15.2%).

These results indicate the number of GP visits in this sample are over-dispersed, have a high proportion of zeros and a low mean.

Descriptive statistics for all independent variables are reported in Table 5 in Appendix B. A select list of independent variables for all respondents is reported in Table 2.

In the sample, just over half the respondents are female (52.7%). The average age of all respondents starts at 42.9 years of age and increases at in roughly two year intervals, as would be expected, over length of the panel (from 42.9 to 50.9 years of age). Nearly 1 in 5 respondents is an immigrant. The proportion of recent immigrants decreases over the panel, as no new respondents are captured by the survey and recent immigrants are no longer recent after 5 cycles of the survey.

The average income in all years for all respondents is just over $52,000, with the median at just over $45,000. The influence of inflation has been mitigated as income is reported in real terms. Changes in the mean between cycles may be a result of the improving economic climate in Canada during the 1990s, or the changing age structure of the sample.

Nearly two in five respondents report having attained a post secondary education. The proportion of individuals reporting some post-secondary education, high school or less than high school decrease over the panel, and is expected as education takes time and as the respondents age through the panel they are able to attain higher levels of education. By the end of the panel, there is still roughly one in five people who do not have a high school education.

The proportion of respondents in all health states remains stable over the panel, with only minor variation in levels from cycle to cycle. Roughly nine in ten respondents report being in excellent (23.3%), very good (39.5%) or good health (28.0%); leaving just under one in ten people reporting fair or poor health. Note that while the proportion of individuals in fair or poor health is low overall in terms of levels, there are large percentage changes over the panel (43% and 108%).

Table 2: Descriptive Statistics

	1994	1996	1998	2000	2002	Total
Basic Demographics						
Age (mean)	42.9	44.9	46.9	49.0	50.9	46.9
Male (%)	47.3	47.3	47.3	47.3	47.3	47.3
Female (%)	52.7	52.7	52.7	52.7	52.7	52.7
Recent Immigrant (%)	3.9	3.3	2.6	1.3	0.6	2.4

continued on next page

continued from last page

	1994	1996	1998	2000	2002	Total
Long-Term Immigrant (%)	13.7	14.3	15.1	16.3	17.0	15.3
Canadian Born (%)	82.3	82.3	82.3	82.3	82.3	82.3
Income (Real 1996 $)						
Mean	51,139	48,665	52,369	55,001	54,506	52,336
Median	42,779	41,135	47,747	47,143	54,897	45,022
5th Percentile	11,772	11,366	11,144	10,756	10,334	11,242
95th Percentile	109,198	105,083	102,849	98,534	93,986	104,598
Education						
Less than High School (%)	21.8	20.5	20.0	19.4	19.0	20.1
High School (%)	16.1	15.3	14.7	14.3	13.8	14.8
Some Post-Secondary (%)	26.4	27.6	26.9	25.8	25.7	26.5
Post-Secondary (%)	35.5	36.5	38.3	39.7	40.2	38.1
Health						
Excellent (%)	27.4	24.8	24.3	21.6	18.4	23.3
Very Good (%)	39.2	40.7	41.6	38.8	37.5	39.5
Good (%)	25.5	27.2	26.5	28.8	31.9	28
Fair (%)	6.7	6.3	6.5	8.8	9.6	7.6
Poor (%)	1.2	1.2	1.1	2.0	2.5	1.6

5 Results

5.1 Model Performance

Models are compared using standard measures of model performance (Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC) and the log-likelihood (*log* L)) and a measure of prediction accuracy (correct classification ratio (CCR)).[9] Note that n refers to the number of observations in the model and k is the number of parameters in the model.

[9]Discussion of these measures and their interpretations are found in Appendix D

Model performance statistics reported in Table 3 are consistent with the findings of Bago d'Uva (2005) and Bago d'Uva (2006). The latent class negative binomial model outperforms the Poisson fixed effects and negative binomial fixed effects in terms of log-likelihood, AIC and BIC.

Table 3: Model Performance Statistics

	n	k	$\log L$	AIC	BIC	CCR
Poisson	41965	43	-67,991.6	136,069.3	136,441.0	23.6%
Negative Binomial	41965	44	-60,238.9	120,563.8	120,935.6	15.1%
Latent Class Negative Binomial	4443	74	-46,480.3	92,812.6	93,582.1	9.6%
Nonparametric	28805	20	-30,990.3	62,020.6	62,186.0	79.6%

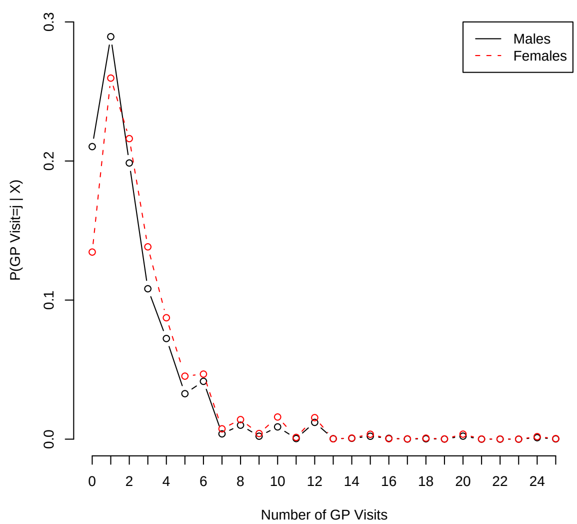
While the standard model specification criteria suggest the more flexible latent class parametric model performs better than the more restrictive Poisson and negative binomial model; the CCR indicates the more flexible parametric model is performing better at a cost of predictive ability. This is seen by the lower AIC, BIC and CCR value for the latent class negative binomial model compared to the Poisson fixed effects or negative binomial fixed effects models. The CCR value for the latent class negative binomial model that is less than half of the CCR for the Poisson fixed effects model.

The nonparametric conditional kernel density estimator outperforms all three parametric models along all statistics. The log-likelihood, AIC and BIC all indicate the nonparametric model is the model of choice among this set. The CCR also indicates that the nonparametric model is predicting greater than three times the actual outcomes than the nearest parametric model.

5.2 Model Results

Reported here are the kernel estimates of the conditional probability of visiting a GP zero, one and five times. The conditional density in each figure is generated by holding all other variables at their median. The interpretation of each conditional probability is different. The conditional density of making zero visits to a GP is interpreted as the probability of a person not making contact with a GP. The interpretation of the conditional probability of making one visit to a GP could be the probability of a person being a low user. The interpretation of the conditional probability of making five visits to a GP could be the probability of a person being a high user.

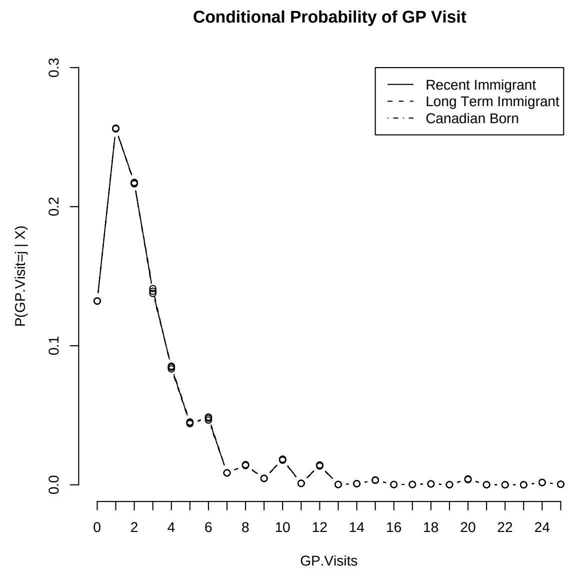
Figure 1: Conditional Probability of GP Visit: by Gender



The point estimates show females appear more likely to make contact with a GP than males, since the probability of having zero visits a general practitioner is lower for females than males (Figure 1). Females tend to be higher users than males, as show by the lower conditional probability of making one visit in conjunction with the higher conditional probability of making five visits. This is consistent with previous findings (Eyles et al. (1995), Sarma and Simpson (2006) and Dunlop et al. (2000)).

Immigration status appears not to be influential in the decision to make contact with a GP or on the decision of how often to visit. Point estimates indicate no difference in visitation patterns by recent or long-term immigrants relative to Canadian born (Figure 2). This contrasts with previous findings that immigrants tend to visit more than Canadian born (Sarma and Simpson (2006)).

Figure 2: Conditional Probability of GP Visit: by Immigration Status



Income does not appear to influence making contact with, or visiting, a GP (Figure 3). This is consistent with previous findings that there is no relationship between visiting a general practitioner and income (Birch et al. (1993), Eyles et al. (1995), Dunlop et al. (2000) and Sarma and Simpson (2006)). These findings contradict previous studies that claim people with higher income have a greater probability of contacting a GP than people with lower income (Allin (2006) and van Doorslaer et al. (2006)). There is no evidence that conditional on making one visit, people with lower income visit the general practitioner more than higher income people. This contradicts van Doorslaer et al. (2006).

Patterns of GP utilization do not appears to be affected by education (Figure 4), consistent with the findings of Birch et al. (1993), Eyles et al. (1995) and Sarma and Simpson (2006); and contradicts the findings of Dunlop et al. (2000).

Figure 3: Conditional Probability of GP Visit: by Income

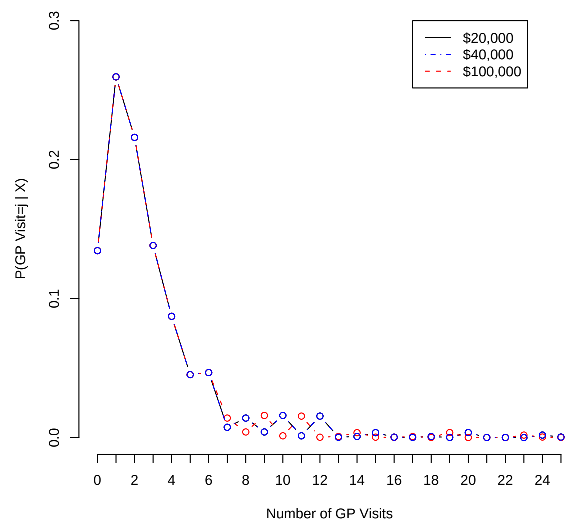
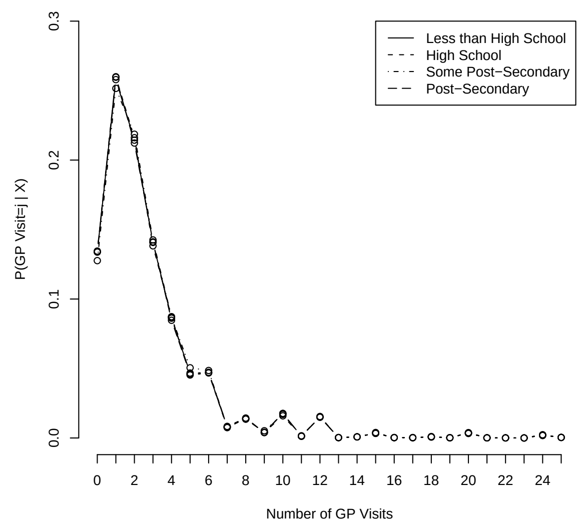
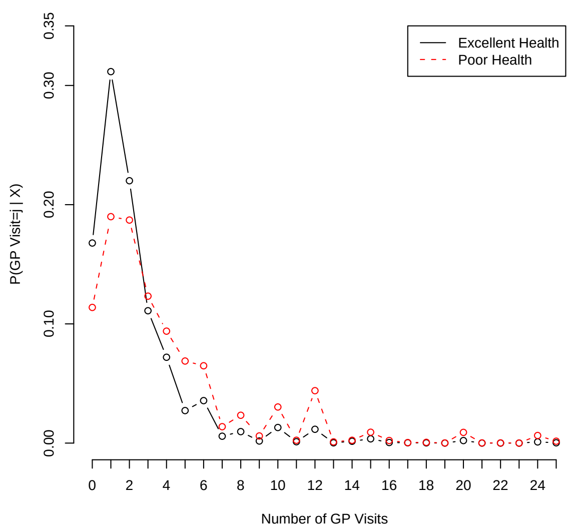


Figure 4: Conditional Probability of GP Visit: by Education



Health appears to be the greatest determinant of GP utilization (Figure 5). Point estimates show a clear gradient in health affecting the probability of visiting a general practitioner. People in better health: have a higher probability of making zero visits to their GP; have a higher probability of making one GP visit; and have a lower probability of making five visits to their GP. This is consistent with previous findings (Birch et al. (1993), Eyles et al. (1995), Dunlop et al. (2000), Allin (2006) and Sarma and Simpson (2006)).

Figure 5: Conditional Probability of GP Visit: by Health Status



6 Discussion

The more flexible nonparametric kernel conditional density estimator provides a better fit with the observed distribution of GP utilization and produces results derived directly from the data not constrained by model assumptions.

If the purpose of the Canadian health care system is to provide health care to those in need, then finding people in poorer health visiting their GP more could be perceived as a health care system achieving its intent. In combination with no apparent influence from income or formal education on needs adjusted utilization, we get the impression GP utilization is not determined by socioeconomic factors, but primarily determined by a person's need.

7 Conclusion

Standard practice in the literature of applied count data models is to use a more flexible parametric model and compare its performance to its predecessor. Here, a nonparametric framework is employed and is shown to out perform standard parametric models using model performance statistics and a measure of prediction accuracy.

The nonparametric kernel conditional density estimator indicate there is no effect from socioeconomic factors (income and education), but health factors (self-reported health) play a large role. The combination of the influence from self reported health and the lack of influence from income and education, give the impression that need is the greatest determinant of GP utilization and not socioeconomic factors.

This study has made three contributions: (i) applied a nonparametric kernel conditional density estimator; (ii) utilized Canadian longitudinal data; and (iii) compared between the longitudinal parametric and nonparametric estimation methods.

References

Aitkin, M. and D. B. Rubin (1985). Estimation and hypothesis testing in finite mixture models. *Journal of the Royal Statistical Society. Series B (Methodological) 47*(1), 67–75.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control 19*(6), 716–723.

Allin, S. (2006, October). Equity in the use of health services in Canada and its provinces. LSE Health, Working Paper Number: 3/2006.

Bago d'Uva, T. (2005). Latent class models for use of primary care: Evidence from a British panel. *Health Economics 14*, 873–892.

Bago d'Uva, T. (2006). Latent class models for utilisation of health care. *Health Economics 15*, 329–343.

Birch, S., J. Eyles, and K. B. Newbold (1993). Equitable access to health care: methodological extensions to the analysis of physician utilization in Canada. *Health Economics 2*(2), 87–101.

Cameron, A. and P. Trivedi (1986). Econometric models based on count data: Comparisons and applications of some estimators and tests. *Journal of Applied Econometrics 1*, 29–53.

Cameron, A. C. and P. Trivedi (1998). *Regression Analysis of Count Data*. Econometric Society Monographs.

Cameron, A. C. and P. Trivedi (2005). *Microeconometrics*. Cambridge University Press.

Deb, P. and P. K. Trivedi (1997). Demand for medical care by the elderly: A finite mixture approach. *Journal of Applied Econometrics 12*(3, 3, Special Issue: Econometric Models of Event Counts), 313–336.

Deb, P. and P. K. Trivedi (2002). The structure of demand for health care: latent class versus two-part models. *Journal of Health Economics 41*, 601625.

Deri, C. (2005). Social networks and health service utilization. *Journal of Health Economics 24*, 1076–1107.

Dunlop, S., P. Coyte, and W. McIsaac (2000). Socio-economic status and the utilization of physicians' services: results from the Canadian national population health survey. *Social Science & Medicine 5*, 123–133.

Eyles, J., S. Birch, and K. B. Newbold (1995, December). Delivering the goods? Access to family physician services in Canada: A comparison of 1985 and 1991. *Journal of Health and Social Behavior 36*(4), 322–332.

Greene, W. (2003). *Econometric Analysis: Fifth Edition*. Prentice Hall.

Grossman, M. (1972). On the concept of health capital and the demand for health. *Journal of Political Economy 80*(2), 223–55.

Hall, P., J. Racine, and Q. Li (2004, December). Cross-validation and the estimation of conditional probability densities. *Journal of the American Statistical Association 99*(468), 1015–1026.

Hausman, J., B. H. Hall, and Z. Griliches (1984, July). Econometric models for count data with an application to the patents - R&D relationship. *Econometrica 52*(4), 909–938.

Jiménez-Martín, S., J. M. Labeaga, and M. Martínez-Granado (2002). Latent class versus two-part models in the demand for physician services across the European Union. *Health Economics 11*, 301–321.

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

Li, Q. and J. S. Racine (2007). *Nonparametric Econometrics*. Princeton University Press.

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

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

Pohlmeier, W. and V. Ulrich (1995). An econometric model of the two-part decisionmaking process in the demand for health care. *The Journal of Human Resources 30*, 339–361.

Sarma, S. and W. Simpson (2006). A microeconometric analysis of Canadian health care utilization. *Health Economics 15*, 219–239.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics 6*(2), 461–464.

van Doorslaer, E., C. Masseria, and X. Koolman (2006, January 17). Inequalities in access to medical care by income in developed countries. *Canadian Medical Association Journal 174*(2), 177–183.

Van Ourti, T. (2004). Measuring horizontal inequity in Belgian health care using a gaussian random effects two part count data model. *Health Economics 13*, 705–724.

Windmeijer, F. and J. Santos Silva (1997). Endogeneity in count data models: An application to the demand for health care. *Journal of Applied Econometrics 12*, 281–294.

A Appendix: Variable Descriptions

Table 4: Variable Description

Demand Side Variables	
Demographic	
Age	Persons age in years
Children	Whether or not there is a child under the age of 12 in the household
Gender	Male (= 0) or Female (= 1)
Immigrant Status	Recent Immigrant (previous 10 years); long-term immigrant (10 or more years); Canadian born
Living Alone	Person lives alone (= 1)
Marital Status	Married or common law; single; and widowed or divorced
Socioeconomic	
Education	High school not completed; high school completed; some post secondary; or post secondary completed
Employment Status	Currently working, not currently working or did not work in the last year
Income	Predicted household income in real 1996 $'s (modeled household income on cycle 3-5, predicted on all cycles, then converted to real dollars)
Health	
Self Reported Health Status	Perceived health relative to other persons of comparable age: excellent; very good; good; fair; or poor
Number of Chronic Conditions	number of reported different chronic health

continued on next page

continued from last page

	problems with the opportunity to volunteer
	other conditions
Activity Limitation	Individuals reporting a health problem causing them to be limited in their activities / to have a long term disability or handicap
	Health Behaviour
Alcohol Consumption	the number of alcoholic drinks per week (0; 1-11; or 12 +)
Household Smoker	respondents were asked if anyone smokes regularly inside the house (= 1)
Body Mass Index	Low weight (BMI$<$ 18.5); Normal Weight (18.5 $\leq$BMI$<$ 25)); Over Weight (25 $\leq$BMI$<$ 30); Obese (30 $\leq$BMI)
Regular Medical Doctor	Does the person have a regular medical doctor? (= 1)
	Supply Side Variables
Number of GPs	the number of GPs practicing within a province in a given year
Province	Newfoundland, P.E.I., Nova Scotia, New Brunswick, Quebec, Ontario, Manitoba, Saskatchewan, Alberta or British Columbia
Urban	Respondent lives in an urban area (= 1)

B Appendix: Descriptive Statistics

Table 5: Descriptive Statistics

	1994	1996	1998	2000	2002	Total
Demographics						
Age (mean)	42.9	44.9	46.9	49.0	50.9	46.9
Male (%)	47.3	47.3	47.3	47.3	47.3	47.3
Female (%)	52.7	52.7	52.7	52.7	52.7	52.7
Recent Immigrant (%)	3.9	3.3	2.6	1.3	0.6	2.4
Long-Term Immigrant (%)	13.7	14.3	15.1	16.3	17.0	15.3
Canadian Born (%)	82.3	82.3	82.3	82.3	82.3	82.3
Single (%)	18.1	17.0	14.9	13.4	12.0	15.1
Married/Common Law (%)	70.1	69.8	70.1	70.1	69.6	69.9
Widowed/Divorced (%)	11.8	13.2	15.0	16.5	18.4	15.0
Child (%)	31.7	30.4	28.7	26.4	23.7	28.2
Lives Alone (%)	11.1	13.4	14.8	15.7	17.0	14.4
Currently Working (%)	65.7	66.5	66.3	66.7	64.7	66.0
Not Currently Working (%)	7.5	6.4	5.3	4.3	4.0	5.5
No Work in the last Year (%)	25.9	23.8	23.6	21.6	21.8	23.4
Working Not Stated (%)	0.9	3.3	4.9	7.4	9.5	5.2
Geography						
Newfoundland (%)	2.1	2.0	1.9	1.9	1.9	2.0
P.E.I. (%)	0.5	0.5	0.5	0.5	0.5	0.5
Nova Scotia (%)	3.2	3.2	3.2	3.2	3.3	3.2
New Brunswick (%)	2.7	2.7	2.7	2.7	2.7	2.7
Quebec (%)	25.7	25.6	25.6	25.6	25.6	25.6
Ontario (%)	36.6	36.5	36.5	36.5	36.5	36.5
Manitoba (%)	3.7	3.8	3.7	3.7	3.7	3.7

continued on next page

continued from last page

	1994	1996	1998	2000	2002	Total
Saskatchewan (%)	3.5	3.5	3.5	3.4	3.4	3.5
Alberta (%)	9.6	9.7	9.8	10.0	10.1	9.8
British Columbia (%)	12.3	12.6	12.6	12.4	12.4	12.5
Lives in Urban Area (%)	81.8	82.2	79.6	79.0	78.8	80.3
Income (Real 1996 $)						
Mean	51,139	48,665	52,369	55,001	54,506	52,336
Median	42,779	41,135	47,747	47,143	54,897	45,022
5^{th} Percentile	11,772	11,366	11,144	10,756	10,334	11,242
95^{th} Percentile	109,198	105,083	102,849	98,534	93,986	104,598
Education						
Less than High School (%)	21.8	20.5	20.0	19.4	19.0	20.1
High School (%)	16.1	15.3	14.7	14.3	13.8	14.8
Some Post-Secondary (%)	26.4	27.6	26.9	25.8	25.7	26.5
Post-Secondary (%)	35.5	36.5	38.3	39.7	40.2	38.1
Health Status						
Excellent (%)	27.4	24.8	24.3	21.6	18.4	23.3
Very Good (%)	39.2	40.7	41.6	38.8	37.5	39.5
Good (%)	25.5	27.2	26.5	28.8	31.9	28.0
Fair (%)	6.7	6.3	6.5	8.8	9.6	7.6
Poor (%)	1.2	1.2	1.1	2.0	2.5	1.6
Activity Limitation (%)	17.8	17.5	17.6	18.7	24.2	19.2
0 Chronic Conditions (%)	45.3	38.3	37.1	34.8	27.7	36.7
1-3 Chronic Conditions (%)	49.3	53.4	52.8	54.6	57.9	53.6
4-5 Chronic Conditions (%)	4.3	6.2	7.5	7.3	9.8	7.0

continued on next page

continued from last page

	1994	1996	1998	2000	2002	Total
6 or more Chronic Conditions (%)	1.1	2.0	2.5	3.2	4.6	2.7
Health Behaviour						
Regular Medical Doctor (%)	86.7	87.6	87.6	89.1	89.5	88.1
Household Smoker (%)	34.6	32.7	30.2	25.8	20.3	28.7
No Drinks (%)	33.0	32.7	31.0	32.7	30.4	32.0
Drinks 1-5 (%)	30.3	29.6	30.1	29.4	30.7	30.0
Drinks 6 or more (%)	18.9	18.9	19.9	18.9	19.2	19.2
Drinking Not Stated (%)	17.9	18.8	19.0	19.0	19.7	18.9
Low Weight (%)	1.8	1.3	1.2	1.2	1.1	1.3
Normal Weight (%)	42.3	40.2	37.6	34.2	30.7	37.0
Over Weight (%)	30.9	31.1	30.9	30.4	30.5	30.8
Obese (%)	11.5	11.5	13.3	14.5	15.6	13.3
BMI Not Stated (%)	13.6	15.9	17.0	19.6	22.1	17.6

C Appendix: Gamma Functions and Densities

As noted in Cameron and Trivedi (1998) page 374–75 and Greene (2003) page 927–28, the Gamma function is denoted by $\Gamma(x)$ where x is a positive constant.

The Gamma function $\Gamma(x)$ is defined by the following relationship:

$$\Gamma(x) = \int_0^\infty e^{-t} t^{x-1} dt, \; x > 0 \tag{34}$$

The Gamma function has the following property:

$$\Gamma(x) = (x-1)\Gamma(x-1) \tag{35}$$

This implies that the Gamma function can be rewritten as:

$$\Gamma(x) = (x-1)! \tag{36}$$

D Appendix: Model Comparison Statistics

D.1 Akaike Information Criterion

The Akaike Information Criterion (AIC), first proposed by Akaike (1974), is a measure of model performance that trades off goodness of fit with parsimony.

$$AIC = -2\log L + 2k \tag{37}$$

where $\log L$ is the value of the maximized log-likelihood function, k is the number of parameters in the model and n is the sample size. When comparing between models, those with smaller AIC values are preferred.

D.2 Bayesian Information Criterion

The Bayesian Information Criterion (BIC) is another measure of model performance. It was first proposed by Schwarz (1978), and is also known as the Schwarz Information Criterion (SIC).

$$BIC = -2\log L + k\log(n) \tag{38}$$

where $\log L$ is the value of the maximized log-likelihood function, k is the number of parameters in the model and n is the sample size. When comparing between models, those with smaller BIC values are preferred.

D.3 Correct Classification Ratio

The correct classification ratio (CCR) is defined to be the proportion of correctly predicted outcomes from a model. Mathematically, the correct classification ratio is defined as:

$$CCR = \frac{1}{n}\sum_{i=1}^{n} 1\left(\hat{y}_i = y_i\right) \tag{39}$$

where $1(\cdot)$ is an indicator function, $\hat{y}_i$ is the predicted value of the dependent variable y for individual i, y_i is the actual value of the dependent value for individual i. By construction, $0 \leq CCR \leq 1$. A model with stronger predictive power will have a CCR closer to 100%.