To the Graduate Council:
I am submitting herewith a dissertation written by Panagiotis Kasteridis entitled "Three essays on count data estimation methods with applications to health economics." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Economics.

Robert A. Bohm, Major Professor

We have read this dissertation
and recommend its acceptance:

Murat K. Munkin

Matthew N. Murray

George Siopsis

Steven Yen

Accepted for the Council:

Carolyn R. Hodges

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

THREE ESSAYS ON COUNT DATA ESTIMATION METHODS WITH APPLICATIONS TO HEALTH ECONOMICS

A Dissertation
Presented for the
Doctor of Philosophy
Degree
The University of Tennessee, Knoxville

Panagiotis Kasteridis
May 2009

Copyright © 2009 by P. K. Kasteridis.
All rights reserved.

Dedication

This dissertation is dedicated to Maria — we did it! Also, to my parents Pavlos and Zoe for paying the ultimate price for this work in feeling my absence in their lives.

Acknowledgments

I would like to express my sincere gratitude to my advisor, Murat Munkin, for guiding this research and serving as a mentor throughout my graduate studies. Dr. Munkin introduced me to Bayesian econometrics and computational methods and motivated me to undertake research in health economics. I learned a lot from his own research which generously shared with me. His experience and scholarly advice were essential during the writing of the dissertation and made its completion possible. Most of all, he helped me to channel my enthusiasm toward the production of a well-defined scholarly product. Not only I deeply appreciate him as a scholar but I also consider him an earnest friend of mine.

I am indebted to Dr. Steven Yen who along with Dr. Munkin played a critical role in mentoring and guiding me through the third chapter of my dissertation.

I would like to express my appreciation to my other committee members, Dr. Robert Bohm, Dr. Matthew Murray, and Dr. George Siopsis for their willingness to serve and their constructive comments. Most especially my thanks go to to Dr. Bohm and Dr. Murray who provided me with continuous support and encouragement throughout my graduate study at The University of Tennessee and advised me on many professional matters.

I also wish to thank Donna Kemper and Susan McGee for providing secretarial support in a most cheerful manner. Finally, thanks are in order for the Economics Department at The University of Tennessee for providing me with the necessary financial resources to complete my Doctorate program.

Abstract

The purpose of this dissertation is to propose and estimate count data microeconometric models that take into account special features of health care data. Three problems in health economics are investigated. The first essay analyzes the effect of managed care insurance plans on the demand for outpatient physician visits. It provides a Bayesian method to empirically separate selection effects due to individual choice of coverage and incentive effects known as moral hazard. The proposed Endogenous Hurdle Poisson log-Normal model addresses two important econometric issues: the large proportion of zero outpatient visits and the endogeneity of managed care insurance status to utilization of outpatient services. The data are obtained from the Medical Expenditure Panel Survey, (MEPS), and the sample consists of privately insured individuals, aged 21-64, all of whom are employed but not self-employed. The analysis indicates both favorable selection bias and important moral hazard effects in the decision whether to utilize an outpatient visit or not. The second essay investigates the impact of health insurance coverage on utilization of doctor visits. A Bayesian panel count data model with correlated random effects and endogenous treatment of the insurance variable is applied to a subset of the Health and Retirement Study (HRS), a national longitudinal survey of individuals over 50 years old and their spouses. The age limit for Medicare eligibility at age 65 serves as an exclusion restriction. After controlling for selection effects insured individuals present higher utilization of doctor visits. The third essay analyzes the demand for cigarettes using data from the 1994-1996 Continuing Survey of Food Intakes by Individuals (CSFII). Since individuals tend to round their consumption of cigarettes in packs smoked per day the dependent variable - number of cigarettes - exhibits pile-ups of counts. The mixed binary-ordered probit approach accommodates this feature of the data and models the starting smoking, the quitting smoking and the how much to smoke decisions. The analysis is performed separately for men and women and provides strong evidence of gender differences in cigarette consumption.

Preface

This dissertation consists of three self-contained essays concerned with microeconometric models with a dependent variable in the form of a count, understood as the number of times an event occurs. Count data models have become particularly popular in health economics applications because of their ability to model health care utilization (such as the number of physician visits or the number of prescriptions dispensed over a given year) as well as other discrete aspects of health behavior such as cigarettes and alcohol consumption over a time interval.

The literature on count data models is extensive and develops from the Poisson Regression Model (PRM) — the starting point for count data analysis — to more sophisticated models that address the limitations of the PRM and take into account special features of the data. The present dissertation contributes to this literature by investigating three problems in health economics which involve non-standard count data modeling and estimation techniques.

The empirical methodology applied in the first two chapters yields a number of insights into the way utilization of health care services is modeled. Both essays focus on the impact of health insurance status on the count of medical services. The insurance status is a binary choice between managed care plans and fee for service plans in the first essay and between insurance coverage and no coverage in the second essay. In studies with observational data observed differences in utilization rates between the two insurance states may be partly due to selection bias. Insured individuals may have different characteristics than the uninsured and similarly those in managed care plans may differ from those covered by fee-for-service plans. While selection on observed characteristics is controlled for by including rich sets of explanatory variables in the model specification there may still be selection on unmeasured or unmeasurable factors. The model accounts for selection on unobservables by endogenizing the binary insurance variable such that unobserved factors that

affect the insurance decision may also affect utilization of medical services. It is shown that the addition of this feature in the models has important policy implications since it allows separation of the two forces that are responsible for observing different utilization rates between insurance states: the incentive effects, known as moral hazard, and the selection effects.

The first essay analyzes the effect of managed care insurance plans on the demand for outpatient care. The dependent variable, outpatient physician visits, is a count variable that displays a large proportion of zeros. Two part models for count data, known as hurdle models, accommodate such a high percentage of zeros by assuming two different data generating processes for zeroes and positive counts. The proposed model extends the hurdle model to account for endogeneity of the insurance status to outpatient care utilization. The model is applied to a sample of privately insured individuals, 21-64 years old, obtained from the Medical Expenditure Panel Survey (MEPS).

The second essay is methodologically and thematically related to the first essay. The setting is similar in that the dependent variable is a count and the explanatory variable of interest is an endogenously determined binary choice variable. The distinctive feature is that endogeneity is modeled in a panel framework. The proposed panel count data model with correlated random effects and explicit treatment of endogeneity is applied to investigate the impact of health insurance coverage on utilization of doctor visits. The data are compiled from the Health and Retirement Survey (HRS), a national longitudinal survey of individuals aged over 50 and their spouses. Estimation of the model indicates favorable selection bias on unobservables and positive incentive effects: insured individuals utilize more doctor visits than the uninsured. If selectivity were not explicitly modeled the estimate of insurance would compound selection and incentive effects leading to misleading estimates. This is particularly relevant to the current debate about the benefits and costs of nearly universal health care coverage. Policymakers are interested in predicting the amount of overcrowding in primary care physician offices resulting from offering insurance coverage to the currently uninsured. The model can also serve as an empirical tool to predict the impact of universal coverage on other measures of medical and Rx utilization.

Both essays adopt Bayesian and parametric estimation methods. Bayesian econometrics focus on the posterior density which combines prior beliefs about the parameters and information from the data embodied in the likelihood. Numerical summaries such as the posterior mean and variance

are calculated in order to present the information contained in the posterior distribution. The intractable analytic expressions of the posterior distributions motivate the use of Markov Chain Monte Carlo (MCMC) methods.

The third essay analyzes the demand for cigarettes fitting observed zero outcomes with a trivariate model consisting of an equation for the starting smoking decision, an equation for the quitting decision, and an equation that models the level of cigarettes consumed. As opposed to the first two essays the model is estimated by Maximum Likelihood. Data are compiled from the 1994-1996 Continuing Survey of Food Intakes by Individuals (CSFII) and exhibit an interesting pattern: the histogram of the number of cigarettes smoked per day has spikes at multiples of 5 cigarettes possibly because individuals round their consumption of cigarettes to half pack, one pack, 2 packs and so forth. This feature of the data set motivates the use of an ordered probit model for the level of consumption which provides the best fit among five competing specifications.

Contents

1 Bayesian Estimation of an Endogenous Hurdle Model: an Application to the Demand for Outpatient Care **1**

1.1 Introduction 1

1.2 Background on managed care 3

1.3 Count Data - Hurdle models 7

1.4 Model Specification 14

1.5 Estimation 16

1.6 Hypothesis testing 19

1.7 Treatment effects 20

1.8 Numerical Example 21

1.9 Application 23

1.10 Results 25

1.11 Conclusions 26

2 Bayesian Estimation of an Endogenous Panel Count Data Model with Correlated Random Effects: an Application to Universal Health Insurance Coverage **28**

2.1 Introduction 28

2.2 Model Specification 33

2.3 Estimation 35

2.4 Bayes factor and Treatment effects 38

2.5 Numerical example 40

2.6 Application 42
2.7 Conclusions 46

3 Demand for Cigarettes: A Mixed Binary-Ordered Probit Aproach 48

3.1 Introduction 48
3.2 Econometric Specification 50
3.3 Data and Variables 55
3.4 Application and Results 57
3.5 Conclusions 62

Bibliography 63

Appendices 70

A Tables — Chapter 1 71

B Tables — Chapter 2 75

C Tables — Chapter 3 81

D The trivariate model 91

E Figures — Chapter 2 93

Vita 96

List of Tables

A.1 Numerical Example 72
A.2 Utilization by Insurance Status 72
A.3 Summary of the Data 73
A.4 MCMC Estimates 74
A.5 Average Treatment Effects for the Treated 74

B.1 Numerical example 76
B.2 Source of data for entry cohorts in RAND HRS data file by wave. 76
B.3 Summary statistics. 77
B.4 Categorical income and wealth. 77
B.5 Utilization of Dr visits by insurance. 77
B.6 Differences in utilization and insurance by group. 78
B.7 MCMC estimates. 79
B.8 Average treatment effects for the treated. 80
B.9 Average treatment effects. 80
B.10 Health variables by insurance status. 80

C.1 Summary statistics 82
C.2 ML estimates for binary part: pooled sample 83
C.3 ML estimates for conditional part: pooled sample 84
C.4 Mean marginal effects and average discrete changes: pooled sample 86
C.5 Likelihood-ratio tests for gender differences and AIC for male and female subsamples 87

C.6 ML estimation of the ordered probit parts for male and female subsamples 88
C.7 Mean marginal effects and average discrete changes 90

List of Figures

E.1 Medicare coverage by age 94
E.2 Insurance coverage by age 94
E.3 Histogram of cigarettes smoked 95
E.4 Histogram of cigarette categories 95

Chapter 1

Bayesian Estimation of an Endogenous Hurdle Model: an Application to the Demand for Outpatient Care

1.1 Introduction

This chapter develops a model to analyze the effect of an endogenously determined binary choice variable on a count variable. Specifically, the model is applied to analyze the effect of managed care relative to the standard indemnity insurance on outpatient physician visits for the privately insured US population.

Endogeneity of insurance is a major concern for researchers that analyze the effects of managed care on utilization. When data are non-experimental, individuals are not randomly assigned to different insurance plans. Instead they self-select themselves into one of the plans based on many factors, some of which can be controlled for by employing a rich set of explanatory variables. For example, several studies have documented that health maintenance organization (HMO) populations are younger. Other studies have found that HMOs are likely to attract a larger proportion of new migrants, families who expect to utilize maternity care, and individuals who expect to use a high level of outpatient services. However, some of the factors that influence the choice of insurance, such as risk preferences, are unobservable. Since the same risk preferences are likely to affect utilization, endogeneity is generated by the omission of the relevant unobservables and statistical

control for it is necessary.

The dependent variable of interest, outpatient physician visits, is a count variable that displays a large proportion of zeros (more than 80 percent). This motivates the use of a two-part model for count data, often referred to as a hurdle model. Hurdle models have become popular in the literature on the demand for health care because they serve two purposes. First, they can fit patterns with high incidences of zero medical care usage, which are often observed in real data. Second, they can be given a structural interpretation that is consistent with a two part decision making process. At the first stage, the patient decides independently whether to visit a doctor at all. At the second stage, the physician decides the intensity of treatment. This stage is sometimes referred to as the frequency decision. An appropriate modeling would require the contact decision and the frequency decision treated as different stochastic processes. In the hurdle model this is achieved by assuming two different data generating processes for the zero and positive counts. Various specifications for the two parts have been proposed (Mullahy, 1986; Pohlmeier and Ulrich, 1994; Gurmu, 1998; Winkelmann, 2004). A semiparametric estimation method of the hurdle model has also been suggested by Gurmu (1997).

The essay contributes to the literature on hurdle models by proposing a Bayesian estimation procedure based on Markov Chain Monte Carlo (MCMC) methods and extending the hurdle model to account for endogeneity of a binary treatment variable. The MCMC algorithm builds on the recent advances made by Tanner and Wong (1987), Albert and Chib (1993), Chib et al. (1998), Li (1998), Chib and Hamilton (2000), Munkin and Trivedi (2003), and Deb et al. (2006b). The model is applied to a sample obtained from the Medical Expenditure Panel Survey (MEPS) and covers the period 1997-2002. We combine all managed care plans into one managed care category. The remaining privately insured individuals have a standard indemnity plan coverage. We perform hypothesis testing of endogeneity calculating Bayes factors using the Savage-Dickey density approach. We also calculate various average treatment effects based on the estimated posterior distributions of the parameters.

The rest of the chapter is organized as follows: Section 1.2 provides a brief background on managed care and reviews the literature of selection models on health care utilization. Section 1.3 discusses hurdle models and their contribution to health literature. Section 1.4 specifies the

model and prior distributions. Section 1.5 presents the MCMC algorithm. Section 1.6 discusses the computation of Bayes factors used for hypothesis testing. Section 1.7 calculates treatment effects. An example with artificially generated data is provided in section 1.8. Section 1.9 describes the empirical application and the data. Finally, section 1.10 discusses the results and section 1.11 concludes.

1.2 Background on managed care

Managed care organizations emerged as a result of the failure of traditional fee-for-service (FFS) plans to contain costs of health care through consumer cost sharing. Unlike indemnity plans, the early managed care plans, known as "staff" and "group" HMOs, used exclusively supply side methods to reduce costs (see Glied, 2000). Staff and group HMOs were fully vertically integrated organizations. In staff HMOs the provider of health care employed a limited number of physicians and owned the hospitals while in group HMOs a group of physicians and hospitals contracted exclusively with an insurance organization. Over-utilization of services was prevented by paying physicians a salary in the case of staff HMOs or a fixed payment for each patient (capitation payment) in the case of group HMOs. With a salary payment doctors have no incentive to provide more services while with capitation payment they bear the full cost of treatment for each patient and as a result they have incentives to avoid costly procedures and limit the provision of services. Treatment intensity was also monitored implicitly through the use of "gatekeepers" or explicitly through utilization review. In gatekeeper arrangements patients can consult a specialist only if they obtain a referral from a primary care physician (gatekeeper). On his part the primary care physician is restricted to limit the number of referrals he makes. Therefore, in cases where treatment can be offered by either the primary care physician or the specialist, the use of gatekeepers reduces significantly the cost of treatment. On the other hand, these early HMO plans provided more generous benefits especially in the form of preventive and maternity care in order to attract healthier population and prevent serious diseases and more expensive procedures.

Eventually, the market share of staff and group HMOs diminished and new less integrated types of HMOs emerged. The Independent Practice Association (IPA) type includes insurance

plans that contract with physicians who work independently in their own offices and retain the option to contract with other managed care organizations and treat patients enrolled in indemnity plans. Payments to participating physicians are made either on a capitation or a discounted fee-for-service basis. Depending on the market power of the HMO relative to the market power of physicians and hospitals in the region, an IPA-type HMO can discount fees paid to physicians and hospitals in exchange for including them in its network. Reimbursing on a discount fee-for-service basis is cost saving relative to paying the full price for the service (as in indemnity plans) but it does not remove the incentive for excessive service provision by physicians. Given that consumer cost-sharing is also limited in IPA-HMOs, the intensity of service provided to patients is controlled using gatekeeper arrangements, "second opinion" processes and utilization review. Utilization review is common policy especially for expensive procedures (surgery, hospitalization, major diagnostic tests). Enrollees or their physicians are required to receive prior authorization from the insurance plans to proceed with the proposed treatment. In second opinion arrangements, a physician paid by the insurer provides a second opinion usually regarding a surgery. However, he is not the provider of the treatment and therefore has no financial incentives to offer a bias recommendation.

Preferred Provider Organizations (PPO) which are nowadays the dominant form of employment based group insurance differ from IPAs in that they do not require enrollees to obtain referrals from a primary care physician before consulting a specialist. Subscribers have also the option to consult physicians or visit hospitals from outside the network. In this sense PPO is a less integrated form of managed care. However, the use of services provided by the plan's network of providers is strongly encouraged by applying higher co-payments as well as deductibles for out-of-network services. Discounted fee-for-service is the most usual payment method.

Finally, Point-of-Service (POS) plans share common features with HMOs, and PPOs. Like PPOs, point-of-service plans allow their policy holders to receive medical care outside the network, though use of facilities and physicians within the network is encouraged. However, subscribers are required to choose a primary care physician (who acts as a gatekeeper) from the provider's network, a feature of HMO plans.

Utilization of medical services across enrollees in different insurance plans can vary significantly

as the result of two underlying factors: the incentive effect (or moral hazard) and the selection effect. Moral hazard can arise if physicians provide excessive treatment to those individuals with generous insurance. On the other hand, the selection effect is present if individuals enrolled in an insurance plan have different characteristics than the average individual. For example, higher utilization rates for FFS enrollees relative to HMO enrollees may reflect in part that those with a FFS plan are less healthy. Indeed, several studies have documented that managed care and traditional plan enrollees may differ in factors such as socioeconomic and health status, expected future use of health care and risk preferences.

Traditional indemnity FFS plans attack the moral hazard problem exclusively through consumer (demand) cost-sharing. Co-payments and deductibles in FFS plans are a percentage of the total cost of treatment so that consumers bear a share of their consumption of health care. Managed care organizations emerged as a result of the failure of traditional FFS plans to contain costs of health care through consumer cost sharing. As explained above pure HMOs relied heavily on supply cost sharing (capitation payment), gatekeeper arrangements, utilization review, administrative regulations and other supply side mechanisms to control moral hazard while demand cost-sharing was limited. The evolution of managed care organizations from staff and group HMOs to IPA-type HMOs, PPOs, and POS plans is an effort to solve some of the problems of the original HMO plans. Higher co-payments (even though they are still a fixed fee and lower than in indemnity plans) enhance the role of consumer's cost-sharing in reducing the degree of moral hazard. On the other hand, these more recent forms of managed care are more flexible in the sense that they do not restrict consumers to in-network providers.

It is apparent from the above discussion that moral hazard and self selection are two different problems for the policymakers. The empirical distinction between incentive and selection effects would help the insurer design the insurance plan efficiently. If increased utilization is the result of adverse selection but the incentive effect is weak, the policy maker could reverse the situation by introducing strategies of favorable selection (e.g. target communities with younger and healthier populations, provide generous maternity care etc). On the other hand, if moral hazard is the primary effect, implementation of higher deductibles and co-payments as well as supply side strategies can be used.

Empirical research on managed care has become increasingly complicated over time due to the proliferation of different types of managed care plans, the adoption of managed care features by conventional insurance plans, and the recognition of self-selection issues. As the new forms of managed care organizations moved towards less integrated schemes, the distinction between different types of insurance plans has become a much more difficult task. Moreover, even plans with the same characteristics may actually differ due to unobserved (to the researchers) factors such as the stringency of each plan's utilization review or details regarding administrative regulations that aim to control the use of services. The picture is actually even fuzzier because the traditional fee-for-service plans have incorporated managed care features such as utilization review.

Nonetheless, the most serious challenge for researchers is the problem of selection bias. If the population enrolled in managed care plans has different characteristics from those subscribed in indemnity plans then observed differences in utilization of services may reflect in part adverse selection. Indeed, several studies have documented that HMO populations are younger and healthier. Other studies have found that HMOs are likely to attract a larger proportion of new migrants, families who expect to need maternity care, and individuals who expect to use a high level of outpatient services. In general, managed care and traditional plan enrollees may differ in factors such as socioeconomic and health status, expected future use of health care and risk preferences. Studies that do not control for these differences may suffer from selection bias and be unreliable. Furthermore, there might be unobserved factors such as unobserved severity of illness that result in self-selection but fewer studies focus on this.

Beside the problems, empirical research has attempted to identify the effects of managed care on various measures of health care utilization (hospital admissions, hospital stays, outpatient services, non-physician visits etc). Early studies using observational data have been criticized on the grounds of inadequate control of health status and other observed characteristics. (Luft (1981) provides a review of studies dating from 1950 to 1978).

In the context of the RAND Health Insurance Experiment, Manning et al. (1984), assigned people randomly into an HMO and several different FFS plans with varying levels of cost-sharing and benefits. Enrollees who were randomly assigned to the HMO plan had 40% lower inpatient

admission rates and significantly shorter hospital stays than those in a FFS plan with no cost-sharing. On the other hand there was no evidence that the number of outpatient visits was higher for HMO subscribers. In addition, patients randomly assigned to the HMO plan were compared with patients who self-selected themselves in the HMO. The comparison revealed only minor disparities between the two groups indicating that studies using non-experimental data may not suffer from selection effects. Although selection bias is not an issue with experimental data, the RAND study was criticized by Bruce et al. (1987) on the basis of a high percentage (29%) of refusals to participate in the experiment.

Miller and Luft (1994, 1997, 2002) updated Luft's review to include more recent studies while limiting their focus on those papers that controlled at least for observed characteristics. Reschovsky (2000); Reschovsky and Kemper (2000) argued that the threat of selection and the risk of unmeasured (though observed) differences reduces by employing an especially rich set of control variables. Comparing HMO effects on those with and without a ready choice between HMO and non-HMO plans eliminated the possibility of remaining selection bias.

1.3 Count Data - Hurdle models

The simplest model that deals with count outcomes is the Poisson regression model (PRM). In the PRM the number of events y has a Poisson distribution:

$$P(y_i = k \mid x_i) = \frac{\exp(-\lambda_i)\lambda_i^k}{k!},\ k = 0, 1, 2, ...$$

and the conditional mean $\lambda_i = E(y_i \mid x_i)$ is expressed as a function of observable individual attributes. To ensure nonegativity, λ_i is specified as $\lambda_i = \exp(x_i\beta)$ where β is a vector of parameters. Different values of x_i (i.e. individuals with different characteristics) result in different values for the conditional mean. However, individuals with exactly the same attributes have the same conditional mean. This feature of the PRM is known as observed heterogeneity.

Besides its simplicity, in practical applications the PRM performs poorly in fitting the observed data. A restrictive feature of the PRM is that it assumes equidispersion while in most count data applications the conditional variance exceeds the conditional mean (overdispersion). A possible

explanation for the overdispersed data is the presence of some form of unobserved heterogeneity. In this case, even individuals with same characteristics can have different conditional means. It is important that although unobserved heterogeneity results in overdispersion, the opposite does not always hold true. Hence, a formal test for overdispersion cannot be considered as a test for unobserved heterogeneity.

An additional problem is that many count data sets present a high proportion of zero counts relative to the frequency of zeros predicted by the PRM. Overdispersion and excess zeros have been investigated in the literature as different problems. However, Mullahy (1998) pointed out that they are both consequences of unobserved heterogeneity.

Mixture models attempt to formalize the concept of unobserved heterogeneity by introducing an additional error term in the conditional mean of the dependent count variable. The most widely used mixture is the Negative Binomial regression model (NB), which is obtained as a Poisson-Gamma mixture. Formally, the conditional mean of the Poisson is specified as:

$$\widetilde{\lambda_i} = \exp(X_i\beta + \varepsilon_i) = \exp(X_i\beta)\exp(\varepsilon_i) = \lambda_i\delta_i$$

where ε_i and X_i are uncorrelated. Under this specification, the variation in $\widetilde{\lambda_i}$ depends not only on different individual attributes but also on unobserved heterogeneity reflected in the random term. The marginal distribution of y_i is obtained by integrating out δ_i:

$$f(y_i \mid X_i) = \int \frac{\exp(-\lambda_i\delta_i)(\lambda_i\delta_i)^{y_i}}{\Gamma(y_i+1)} g(\delta_i)dv_i$$

where $g(\delta_i)$ is the density of δ_i. Further, it is assumed that δ_i follows a Gamma distribution $G(\alpha_i, \gamma_i)$ with the shape parameter α_i and the scale parameter γ_i being equal so that $E(\delta_i) = 1$ and $Var(\delta_i) = 1/\alpha_i$. The marginal density is:

$$f(y_i \mid X_i) = \frac{\Gamma(y_i+\alpha_i)}{\Gamma(y_i+1)\Gamma(\alpha_i)} \left(\frac{\alpha_i}{\alpha_i+\lambda_i}\right)^{\alpha_i} \left(\frac{\lambda_i}{\alpha_i+\lambda_i}\right)^{y_i}$$

with conditional variance

$$Var(y_i \mid X_i) = \lambda_i(1 + \frac{\lambda_i}{\alpha_i}) > \lambda_i = E(y_i \mid X_i)$$

Since the conditional variance exceeds the conditional mean the model allows for overdispersion. The NB model is identified when assumptions about the mean and variance are made. The mean is identified by imposing $E(\delta_i) = 1$. The NB2 is taken for restricting the variance of δ to be constant: $Var(\delta_i) = 1/\alpha_i = \alpha > 0$ which means that α_i must be the same across individuals. In this case:

$$Var(y_i \mid X_i) = \lambda_i(1 + \alpha\lambda_i)$$

The NB1 is obtained by setting $Var(\delta_i) = \alpha/\lambda_i$ which implies

$$Var(y_i \mid X_i) = \lambda_i(1 + \alpha)$$

The dispersion parameter α explains overdispersion and is estimated by the model. Clearly, if $\alpha = 0$ the model reduces to the PRM. Moreover, the increased variance of the NB model allows for higher proportion of low counts and handles better a higher percentage of zeros than the PRM. Yet, in applications with high proportion of zeros an explicit modeling of the process that generates zeros is required. This purpose is well served by hurdle models.

The term "hurdle" was first introduced by Cragg (1971). Since then, hurdle models have gained popularity because of their ability to model two-part decision making processes. The first part refers to the decision of initiating or not an event while the second part refers to the number of times the event occurs given that it has been initiated. Such a dual decision making structure is frequently met in health literature. The decision to utilize medical services can be thought of as consisting of two separate stages. At the first stage, the patient decides independently whether to visit a physician or not based on factors such as health status, insurance coverage, access to health care and others. At the second stage, the physician decides the intensity of treatment (subsequent visits or referrals to a specialist). This stage is sometimes referred to as the frequency decision. An appropriate modeling would require the contact decision and the frequency decision treated as

different stochastic processes. In the hurdle model this is achieved by assuming two different data generating processes for the zero and positive counts.

Let y_i, $i = 1, ..., N$ be the discrete response variable and X_i a vector of explanatory variables. Zero counts are generated with probability $P(y_i = 0 \mid X_i, \theta_1)$ where θ_1 is a parameter vector. With probability $1 - P(y_i = 0 \mid X_i, \theta_1)$ the hurdle is crossed and positive counts are generated according to a truncated at zero distribution denoted as $P(y_i \mid y_i > 0, X_i, \theta_2)$ where θ_2 is a parameter vector. Therefore, the first stage (contact decision) is modeled by a binary model governing whether the hurdle is crossed and the second stage (frequency decision) is modeled by a truncated at zero model generating strictly positive counts. The parameter sets θ_1 and θ_2 are different so that the explanatory variables have different impact at each stage. The likelihood function for N independent observations is given by:

$$L = \prod_{i=1}^{N} \left\{ I(y_i = 0) P(y_i = 0 \mid X_i, \theta_1) + I(y_i > 0) [1 - P(y_i = 0 \mid X_i, \theta_1)] \frac{P(y_i \mid X_i, \theta_2)}{P(y_i > 0 \mid X_i, \theta_2)} \right\} \tag{1.1}$$

where $I(\omega \in \Omega)$ is an indicator function taking a value of one if $\omega \in \Omega$ and zero otherwise.

The likelihood can be expressed as the product of two likelihood functions that are parametrically independent and therefore can be maximized separately:

$$\begin{aligned} L &= \prod_{i=1}^{N} \{ I(y_i = 0) P(y_i = 0 \mid X_i, \theta_1) + I(y_i > 0) [1 - P(y_i = 0 \mid X_i, \theta_1)] \} \\ &\quad \times \prod_{i=1}^{N} \left\{ I(y_i = 0) + I(y_i > 0) \frac{P(y_i \mid X_i, \theta_2)}{P(y_i > 0 \mid X_i, \theta_2)} \right\} \end{aligned} \tag{1.2}$$

The first product is the likelihood for the binary choice model (zero counts versus positive counts) and the second product is the likelihood for the truncated at zero model.

The conditional mean and variance depend on the probability the hurdle is crossed and the

moments of the truncated at zero distribution:

$$\begin{aligned} E(y_i \mid X_i) &= P(y_i > 0 \mid X_i, \theta_1) E_T(y_i \mid y_i > 0, X_i) \quad (1.3) \\ Var(y_i \mid X_i) &= P(y_i > 0 \mid X_i, \theta_1) Var_T(y_i \mid y_i > 0, X_i) \\ +P(y_i &= 0 \mid X_i, \theta_1)[E_T(y_i \mid y_i > 0, X_i)]^2 \quad (1.4) \end{aligned}$$

Various specifications for the binary and the truncated part have been proposed in the literature. Mullahy (1986) proposed a Poisson and a Geometric hurdle model. For the Poisson specification:

$$\begin{aligned} P(y_i = 0 \mid X_i, \beta_1) &= \exp[-\exp(X_i\beta_1)], \quad (1.5) \\ P(y_i \mid y_i > 0, X_i, \beta_2) &= \frac{\exp(y_i X_i \beta_2)}{y_i!\{\exp[\exp(X_i\beta_2)] - 1\}} \quad (1.6) \end{aligned}$$

and the likelihood function is:

$$\begin{aligned} L &= \prod_{i=1}^{N} \{I(y_i = 0) \exp[-\exp(X_i\beta_1)] + I(y_i > 0)[1 - \exp(-\exp(X_i\beta_1))]\} \\ &\quad \times \prod_{i=1}^{N} \left\{ I(y_i = 0) + I(y_i > 0) \frac{\exp(y_i X_i \beta_2)}{y_i!\{\exp[\exp(X_i\beta_2)] - 1\}} \right\} \quad (1.7) \end{aligned}$$

Obviously, when $\beta_1 = \beta_2$ the likelihood reduces to the likelihood of the parent Poisson.

For the geometric specification,

$$P(y_i = 0 \mid X_i, \beta_1) = \frac{1}{1 + \exp(X_i\beta_1)}, \quad (1.8)$$

$$P(y_i \mid y_i > 0, X_i, \beta_2) = \frac{\exp[(y_i - 1)X_i\beta_2]}{[1 + \exp(X_i\beta_2)]^{y_i}} \quad (1.9)$$

i.e. the first stage is described by a standard logit model while the second stage by a truncated at zero geometric model.

Pohlmeier and Ulrich (1994) assumed that the underlying distribution for both stages is NB1.

Their hurdle model estimated the demand for health care in West Germany as a two part decision-making process using data from the German Socioeconomic Panel. They claimed that the choice of a hurdle specification is justified by the institutional setting of the German's health insurance system. The insured receives on request a sickness voucher from the sickness fund each quarter. It is to her discretion to visit a physician (of her choice) but once he has submitted the voucher, medical services are provided by this physician only. Hence, at the first stage the patient makes the decision whether to visit a physician or not while at the second stage it is the physician who determines the length of treatment.

Gerdtham (1997) employed a hurdle model to test the null hypothesis of no horizontal inequity in the delivery of health care in Sweden. Having defined equity as equal treatment for equal need he focused on two measures of treatment: physician visits and hospital care weeks. The use of a hurdle model in his application is justified by the large proportion of zeros (38.8% for physician visits and 91.3% for hospital care weeks) as well as on theoretical grounds. For example, the two part structure makes it possible to assess whether income has its effect on utilization through the contact decision or the frequency decision. His hurdle specification is a logit model for the binary choice and a truncated at zero NB2 model for the number of physician visits or hospital care weeks.

Gurmu (1998) proposed generalized hurdle models. For the first stage he suggested a generalization of the logit model that allows for asymmetry in the two tails of the cumulative probability curve $P(y_i > 0 \mid X_i, \theta_1)$. For the second part he proposed a truncated at zero NB2 or a truncated at zero Double Poisson (DP) model that allows for overdispersed as well as underdispersed count data.

In another paper, Gurmu (1997) developed a semiparametric estimation method for hurdle models. Parametric estimation that allows for unobserved heterogeneity requires knowledge of the distribution of unobservables. For example, the NB hurdle assumes that the unobserved heterogeneity is Gamma distributed. Misspecification of the distribution of unobserved heterogeneity leads to inconsistent estimates. The advantage of the semiparametric estimation is that it avoids a distributional assumption for the unobserved heterogeneity. The method is applied to study the impact of managed care programmes for Medicaid eligibles on the number of doctor and health centre visits.

Winkelmann (2004) combined a probit model for the binary choice part with a truncated Poisson-log-normal mixture for strictly positive outcomes. The Poisson-log-normal mixture can be derived introducing the unobserved heterogeneity in the conditional mean and assuming that it is normally distributed. The above hurdle specification was used to investigate the impact of the German health care reform of 1997 on the number of doctor visits.

The interpretation of the hurdle model as an intuitive way to describe a two-part decision process is appealing. However, there is a problem mentioned by several authors. In cases where the observation period is relatively long (one year) the probability of observing multiple illness spells and therefore multiple first contacts increases. For example, it is reasonable to think that in an one-year period patients with more than one trip to a hospital consist of a high percentage among the population. Although, the first part of the hurdle model can still be thought as describing the contact decision, the second part should be interpreted either as the frequency decision determined by physicians or as the patient's decision to initiate subsequent trips to the hospital. If there is no reason to believe that the decision to contact a physician for the first time is different than the decision for subsequent visits, the two-part structure can be questioned. It is worthy to mention that quarterly data (as in Pohlmeier and Ulrich (1994)) reduce significantly the possibility of multiple illness spells thereby increasing the appeal of the structural interpretation of the hurdle model.

In light of the above criticism, Deb and Trivedi (2002) contrasted the hurdle model (referred in their paper as the two-part model TPM) with a latent class model (LCM) and they found strong evidence in favor of the LCM. Instead of distinguishing between users and non-users of health care, the LCM distinguishing between frequent and infrequent users. This is achieved by assuming that the count variable is drawn from an additive mixture of S distinct subpopulations so that the density for observation $i = 1, ..., n$ is given by:

$$f(y_i \mid \theta) = \sum_{j=1}^{S} \pi_j f_j(y_i \mid \theta_j), \;\; \sum_{j=1}^{S} \pi_j = 1, \;\; 0 < \pi_j < 1$$

where the subpopulation densities f_j are usually the same distributions but with different parameterizations θ_j for each component. The slope coefficients of the index function $\pi_1, ..., \pi_S$ are the mixing probabilities and are estimated along with parameters θ_j.

A natural interpretation of the mixture components is possible: Deb and Trivedi hypothesized that unobserved differences in individual's long-term health status (not captured by self-perceived health status and chronic conditions) split the population into subpopulations. For instance, the model with two latent classes can be thought as dividing the population into the "healthy" group characterized by low utilization of medical services and the "ill" group with high utilization. Therefore, in addition to unobserved heterogeneity within each component, the LCM assumes that unobserved heterogeneity splits the population into latent classes.

Santos Silva and Windmeijer (2001) proposed a two-part model that allows for multiple illness spells. The total number of visits to a physician resulting from S spells of illness is expressed as:

$$V = \sum_{j=1}^{S} R_j$$

where R_j is the number of visits in the jth spell. Assuming that the conditional expectation of R_j does not depend on j and that S and R_j are conditionally independent, the total number of visits V has a stopped-sum distribution. An example of a stopped-sum distribution is the negative binomial which is derived assuming a Poisson distribution for the number of spells S and a logarithmic distribution for the number of visits in the jth spell R_j.

1.4 Model Specification

We observe N independent observations y_i $(i = 1, ..., N)$. For each individual i we define a latent variable H_i (thereafter referred to as hurdle) which can be thought of as the individual's propensity to initiate a treatment episode. When the zero hurdle is crossed ($H_i \geqslant 0$), positive counts are generated according to a truncated at zero Poisson log-normal mixture with conditional mean $exp(\mu_i)$ which is specified below. Otherwise, zero counts are generated. The counts are measures of health care utilization (outpatient physician visits in the application). Therefore:

$$\begin{cases} y_i \mid y_i > 0 \sim P\left(exp(\mu_i)\right) & \text{if} \quad H_i \geq 0 \\ y_i = 0 & \text{if} \quad H_i < 0 \end{cases} . \tag{1.10}$$

The insurance (selection) equation is defined by specifying latent variable Z_i, reflecting the propensity to purchase managed care type of health insurance, and a binary treatment variable d_i taking a value of 1 if the individual chooses the treated state (chooses to enroll in a managed care plan) and zero otherwise:

$$Z_i = W_i\alpha + u_i, \quad u_i \sim N(0,1), \tag{1.11}$$

$$d_i = \begin{cases} 1, & Z_i \geq 0 \\ 0, & Z_i < 0 \end{cases}, \tag{1.12}$$

where W_i is a vector of exogenous explanatory variables, Z_i is linked to the treatment d_i through the observability condition (1.12) and random variable u_i captures the effect of all unobserved factors that affect the insurance decision. For identification purposes its variance must be fixed such that u_i is assumed to be standard normally distributed.

We model endogeneity by specifying the hurdle and the conditional mean as

$$H_i = X_i\beta + \rho_1 d_i + \pi_1 u_i + \eta_i, \quad \eta_i \sim N(0,1), \tag{1.13}$$

$$\mu_i = X_i\gamma + \rho_2 d_i + \pi_2 u_i + \varepsilon_i, \ \ \varepsilon_i \sim N(0,\sigma), \tag{1.14}$$

where η_i and ε_i are independent of u_i, such that $cov(u_i,\eta_i) = 0$ and $cov(u_i,\varepsilon_i) = 0$. H_i and μ_i are linear in the set of explanatory variables X_i and the treatment d_i. In addition, we assume that $cov(\varepsilon_i,\eta_i) = 0$.

Under this specification, the unobserved parts of H_i and μ_i, captured by random variables $\pi_1 u_i + \eta_i$ and $\pi_2 u_i + \varepsilon_i$, are correlated with the insurance equation through random variable u_i. This is the mechanism through which endogeneity of insurance to utilization is modeled.

We estimate the Endogenous Hurdle Poisson-lognormal model choosing prior distributions as follows. Since no prior information is available about parameters $\alpha, \beta, \gamma, \rho_1, \rho_2$, we center their priors at zero values with a relatively large variance such that:

$$\alpha \sim N(0_p, 10I_p), \beta \sim N(0_k, 10I_k), \gamma \sim N(0_k, 10I_k).$$

We choose priors for π_1 and π_2:

$$\pi_1 \sim N(0, 1/2) \text{ and } \pi_2 \sim N(0, 1/2),$$

as well as for ρ_1 and ρ_2:

$$\rho_1 \sim N(0, 1/2) \text{ and } \rho_2 \sim N(0, 1/2)$$

Finally, we select the following prior for σ:

$$\sigma^{-1} \sim G(\frac{\eta_0}{2}, (\frac{c_0}{2})^{-1}) \text{ where } \eta_0 = 5 \text{ and } c_0 = 3.$$

1.5 Estimation

We use MCMC methods to estimate the posterior distributions of the parameters of our model as well as various treatment effects.

Let $\widetilde{\Delta} = (\alpha, \beta, \gamma, \rho_1, \rho_2, \pi_1, \pi_2, \sigma)$ denote the parameter set, and $\Delta_i = (\widetilde{\Delta}, X_i, W_i)$. Then the joint density of observed and latent data for observation i is given by:

$$\begin{aligned}
&f(Y_i, d_i, Z_i, H_i, \mu_i \mid \Delta_i) = \\
&\quad \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}(Z_i - W_i\alpha)^2) \\
&\times [I(d_i = 1)I(Z_i \geq 0) + I(d_i = 0)I(Z_i \leq 0)] \\
&\times [I(y_i = 0)I(H_i \leq 0) + \frac{\exp(-\exp(\mu_i))\exp(\mu_i y_i)}{y_i!(1 - \exp(-\exp(\mu_i)))} I(H_i \geq 0)I(y_i > 0)] \\
&\times \varphi(H_i \ ; \ X_i\beta + \rho_1 d_i + \pi_1(Z_i - W_i\alpha), 1) \\
&\times \varphi(\mu_i \ ; \ X_i\gamma + \rho_2 d_i + \pi_2(Z_i - W_i\alpha), \sigma)
\end{aligned} \tag{1.15}$$

where $\varphi\left(\bullet \ ; \mu, \sigma^2\right)$ is the $N\left(\mu, \sigma^2\right)$ pdf, and $I\left(x \in X\right)$ is an indicator function taking a value of 1 if the random variable x belongs to the set X and zero otherwise. The posterior density is proportional to the product of the prior density and the distribution of observables and latent variables over all N observations.

We organize the MCMC algorithm in the following steps by blocking the parameters as Z_i, H_i,

μ_i, α, $(\beta'$, ρ_1, $\pi_1)'$, $(\gamma'$, ρ_2, $\pi_2)'$, σ.

1. From the full conditional density for Z_i and our square completion we find that:

$$Z_i \overset{i.i.d.}{\sim} TN_{(0,\infty)}\left(\overline{Z_i}, \overline{H_i}^{-1}\right) \quad \text{if} \quad d_i = 1, \tag{1.16}$$

$$Z_i \overset{i.i.d.}{\sim} TN_{(-\infty,0)}\left(\overline{Z_i}, \overline{H_i}^{-1}\right) \quad \text{if} \quad d_i = 0, \tag{1.17}$$

where

$$\overline{Z_i} = W_i\alpha + \overline{H_i}^{-1}\left[\pi_1\left(H_i - X_i\beta - \rho_1 d_i\right) + \pi_2\sigma^{-1}\left(\mu_i - X_i\gamma - \rho_2 d_i\right)\right], \tag{1.18}$$

$$\overline{H_i} = 1 + \pi_1^2 + \pi_2^2\sigma^{-1}. \tag{1.19}$$

2. The full conditional of H_i is

$$H_i \overset{i.i.d.}{\sim} TN_{(-\infty,0)}\left(X_i\beta + \rho_1 d_i + \pi_1(Z_i - W_i\alpha), 1\right) \quad \text{if} \quad y_i = 0, \tag{1.20}$$

$$H_i \overset{i.i.d.}{\sim} TN_{(0,\infty)}\left(X_i\beta + \rho_1 d_i + \pi_1(Z_i - W_i\alpha), 1\right) \quad \text{if} \quad y_i > 0. \tag{1.21}$$

3. If $H_i < 0$ one has to draw μ_i from $N\left(X_i\gamma + \rho_2 d_i + \pi_2(Z_i - W_i\alpha), \sigma\right)$. If $H_i > 0$ the full conditional density for μ_i is:

$$\pi\left(\mu_i \mid y_i, d_i, Z_i, \Delta_i\right) \propto \frac{\exp(-\exp(\mu_i))\exp(\mu_i y_i)}{y_i!\left(1 - \exp(-\exp(\mu_i))\right)} \times \varphi\left(\mu_i \,;\, X_i\gamma + \rho_2 d_i + \pi_2(Z_i - W_i\alpha), \sigma\right). \tag{1.22}$$

In order to sample μ_i we utilize the Metropolis-Hasting algorithm which uses a proposal density independent across draws. Geweke (1989) argued that the proposal density should have tails at least as fat as the posterior. For that reason we select the t-distribution with a small number of degrees of freedom as the candidate-generating density. Following Chib et al. (1998), we tailor the proposal density to the target density centered at its modal value $\overset{\wedge}{\mu}_i = \arg\max \ln \pi\left(\mu_i \mid y_i, d_i, Z_i, \Delta_i\right)$. To obtain the mode we implement the Newton-Raphson algorithm using the gradient vector g_{μ_i} and the Hessian H_{μ_i} of $\ln \pi\left(\mu_i \mid y_i, d_i, Z_i, \Delta_i\right)$. Then

the proposal t-distribution is $q(\mu_i) = f_T(\mu_i \mid \hat{\mu}_i, V_{\hat{\mu}_i}, v)$ where $\hat{\mu}_i$is the location parameter, $V_{\hat{\mu}_i} = -\left(H_{\hat{\mu}_i}\right)^{-1}$ is the negative inverse of the Hessian evaluated at the mode, and v are degrees of freedom. The chain moves to a value $\mu_i^\dagger$ drawn from the proposal density, with probability

$$\alpha\left(\mu_i, \mu_i^\dagger\right) = \min\left\{\frac{\pi\left(\mu_i^\dagger \mid y_i, d_i, Z_i, \Delta_i\right) q(\mu_i)}{\pi(\mu_i \mid y_i, d_i, Z_i, \Delta_i)\, q\left(\mu_i^\dagger\right)}, 1\right\}, \tag{1.23}$$

and stay at the current value μ_i with probability $1 - \alpha\left(\mu_i, \mu_i^\dagger\right)$.

4. Let the prior distribution of α be $N\left(\underline{\alpha}, \underline{H}_\alpha^{-1}\right)$. The full conditional density of α is $\alpha \sim N\left(\overline{\alpha}, \overline{H}_\alpha^{-1}\right)$ where

$$\overline{H}_\alpha = \underline{H}_\alpha + \sum_{i=1}^{N} W_i'\left(1 + \pi_1^2 + \pi_2^2\sigma^{-1}\right) W_i \tag{1.24}$$

$$\overline{\alpha} = \overline{H}_a^{-1} A \tag{1.25}$$

where A is defined as follows:

$$A = \underline{H}_\alpha\alpha + \sum_{i=1}^{N} W_i'(1 + \pi_1^2 + \pi_2^2\sigma^{-1})Z_i - \pi_1(H_i - X_i\beta - \rho_1 d_i) - \pi_2\sigma^{-1}(\mu_i - X_i\gamma - \rho_2 d_i) \tag{1.26}$$

5. We specify priors $\beta \sim N\left(\underline{\beta}, \underline{H}_\beta^{-1}\right)$, $\rho_1 \sim N\left(\underline{\rho_1}, \underline{H}_{\rho_1}^{-1}\right)$, $\pi_1 \sim N\left(\underline{\pi_1}, \underline{H}_{\pi_1}^{-1}\right)$

and let $\delta_1 = (\beta',\ \rho_1,\ \pi_1)'$ and $C_i = (X_i,\ d_i,\ \ (Z_i - W_i\alpha))$

Then the full conditional density of δ_1 is $\delta_1 \sim N\left(\overline{\delta_1}, \overline{H}_{\delta_1}^{-1}\right)$ where

$$\overline{H}_{\delta_1} = \begin{bmatrix} \underline{H}_\beta & \mathbf{0} & 0 \\ \mathbf{0} & \underline{H}_{\rho_1} & 0 \\ 0 & 0 & \underline{H}_{\pi_1} \end{bmatrix} + \sum_{i=1}^{N} C_i' C_i \tag{1.27}$$

$$\overline{\delta_1} = \overline{H}_{\delta_1}^{-1} \left(\underline{H}_{\delta_1} \underline{\delta_1} + \sum_{i=1}^{N} C_i' H_i \right) \tag{1.28}$$

6. Similarly, we specify priors

$$\gamma \sim N\left(\underline{\gamma}, \underline{H}_{\gamma}^{-1}\right), \rho_2 \sim N\left(\underline{\rho_2}, \underline{H}_{\rho_2}^{-1}\right), \pi_2 \sim N\left(\underline{\pi_2}, \underline{H}_{\pi_2}^{-1}\right)$$

Let $\delta_2 = (\gamma', \rho_2, \pi_2)'$. Then the full conditional density of δ_2 is

$$\delta_2 \sim N\left(\overline{\delta_2}, \overline{H}_{\delta_2}^{-1}\right)$$

where

$$\overline{H}_{\delta_2} = \begin{bmatrix} \underline{H}_{\gamma} & \mathbf{0} & 0 \\ \mathbf{0} & \underline{H}_{\rho_2} & 0 \\ 0 & 0 & \underline{H}_{\pi_2} \end{bmatrix} + \sum_{i=1}^{N} C_i' \sigma^{-1} C_i \tag{1.29}$$

$$\overline{\delta_2} = \overline{H}_{\delta_2}^{-1} \left(\underline{H}_{\delta_2} \underline{\delta_2} + \sum_{i=1}^{N} \mu_i \sigma^{-1} C_{i\prime} \right) \tag{1.30}$$

7. Finally, we specify the prior $\sigma_{\varepsilon}^{-1} \sim G\left(\frac{n}{2}, \left(\frac{c_1}{2}\right)^{-1}\right)$ so that the full conditional of σ^{-1} is

$$G\left(\frac{n+N}{2}, \left\{ \frac{c_1}{2} + \frac{\sum_{i=1}^{N} \left[\mu_i - (X_i\gamma + \rho_2 d_i + \pi_2(Z_i - W_i\alpha))\right]^2}{2} \right\}^{-1} \right) \tag{1.31}$$

1.6 Hypothesis testing

To test the null hypothesis of no endogeneity of the treatment variable ($H_0 : \pi_1 = \pi_2 = 0$) we need to compare the unrestricted model M_1 and model M_0 under the constraints $\pi_1 = \pi_2 = 0$. For model comparison with equal prior weight attached to each model the posterior odds ratio becomes the Bayes Factor defined as:

$$B_{01} = \frac{\pi\left(y \mid M_0\right)}{\pi\left(y \mid M_1\right)}$$

where $\pi(y \mid M_i)$ is the marginal likelihood for model M_i. Since the models are nested we can apply the Savage-Dickey ratio approach to calculate the Bayes factor (see Verdinelli and Wasserman, 1995). Suppose that the parameter set of a model is $\theta = (\omega, \psi)$ and we want to test the null hypothesis $\omega = \omega_0$. Dickey (1971) showed that if the prior for ψ in the restricted and unrestricted models is the same at the point $\omega = \omega_0$ then the Bayes factor can be calculated as

$$B_{01} = \frac{\pi(\omega_0 \mid y)}{\pi(\omega_0)}$$

Therefore, under weak assumptions the Bayes factor is calculated as:

$$B_{01} = \frac{\pi(\pi_1^*, \pi_2^* \mid y)}{\pi(\pi_1^*, \pi_2^*)}$$

where $\pi(\pi_1^*, \pi_2^* \mid y)$ is the joint posterior density and $\pi(\pi_1^*, \pi_2^*)$ the joint prior density evaluated at $\pi_1^* = \pi_2^* = 0$.

The priors for π_1 and π_2 must be specified carefully. In Bayesian estimation the effect of priors diminishes as the sample size becomes fairly large. However, as shown in Kass and Raftery (1995), Bayesian testing differs from estimation in that the effect of the prior density may not vanish as the sample size increases. As a result, the Bayes factor is sensitive to the choice of priors. Improper priors (that can be used for estimation and prediction) will always choose the null hypothesis and therefore are not acceptable for hypothesis testing. Even informative priors with a large spread on π_i under H_1 may provide overwhelming support to the model with fewer parameters regardless of the data, thereby favoring the restricted model. For that reason we choose informative priors.

1.7 Treatment effects

We are interested in evaluating treatment effects of the insurance choice. The literature on treatment effects provides various treatment parameters (see Heckman et al., 2001). We calculate two of them: the average treatment effect (ATE) and the average treatment effect on the treated (ATET).

Denote Y_{MC} and Y_{FFS} two potential utilization variables in the treated and untreated states respectively. The treatment effect (TE) for each individual i in the sample is the expected gain, in

terms of utilization, from enrollment in a managed care plan relative to being in a fee-for-service plan:

$$\hat{TE} = \hat{E}[Y_{MC,i} - Y_{FFS,i}|X_i].$$

The average treatment effect is obtained by averaging the treatment effect across all individuals:

$$\hat{ATE} = \frac{1}{N}\sum_{i=1}^{N} \hat{E}[Y_{MC,i} - Y_{FFS,i}|X_i]$$

We define the treated group as the individuals covered by any managed care plan (HMO or PPO), i.e. those with $d_i = 1$. Then, the treatment effect for the treated (TET) for treated individual i is defined as:

$$\hat{TET} = \hat{E}[Y_{MC,i} - Y_{FFS,i} \mid X_i, d_i = 1] = \hat{E}[Y_{MC,i} \mid X_i, d_i = 1] - \hat{E}[Y_{FFS,i} \mid X_i, d_i = 1]. \quad (1.32)$$

Accordingly, the average treatment effect on the treated (ATET) is defined as the mean effect of treatment on those who actually enrolled in a managed care plan:

$$\hat{ATET} = \frac{1}{N_T}\sum_{i=1}^{N_T} \hat{E}[Y_{MC,i} - Y_{FFS,i} \mid X_i, d_i = 1], \quad (1.33)$$

where N_T is the number of such enrollees.

We estimate the posterior distributions of the ATE and ATET evaluating expressions (1.32) and (1.33) at the posterior distributions of the parameters of the model and report their posterior means and standard deviations.

1.8 Numerical Example

In this section we provide a numerical example to check the performance of the estimation method. We also estimate a restricted model that ignores the endogeneity of the treatment variable to demonstrate the signs and magnitudes of self-selection biases. We begin by generating artificial data (5,000 observations) such that about 50% of the counts are positive (they cross the hurdle) and the remaining are zeros. Also, 50% of the observations have the treatment ($d = 1$) and for the

rest $d = 0$. We generate the data as follows:

1. Generate:

$$\begin{aligned} X_i &= (1, x_i), x_i \sim N(0,1), \\ W_i &= (1, w_i), w_i \sim N(0,1), \\ u_i &\sim N(0,1), \\ e_{1i} &= \pi_1 u_i + \eta_i, \eta_i \sim N(0,1), \\ e_{2i} &= \pi_2 u_i + \varepsilon_i, \varepsilon_i \sim N(0,\sigma), \end{aligned}$$

   and specify: $\alpha = (0,1)$, $\beta = (-0.25, 1)$, $\gamma = (1,2)$,

   $\pi_1 = \pi_2 = -0.5$, $\rho_1 = \rho_2 = 0.5$, $\sigma = 0.5$.

   As the negative signs assigned to the correlation parameters π_1 and π_2 reveal, data are generated assuming favorable selection bias in both stages of the hurdle model. On the other hand, the positive signs of $\rho_1$and ρ_2 imply that the pure treatment effects (or incentive effects) are positive.

2. Generate $Z_i = W_i\alpha + u_i$ such that $d_i = \begin{cases} 1, \text{ if } Z_i \geq 0 \\ 0, \text{ if } Z_i < 0 \end{cases}$,

   $H_i = X_i\beta + \rho_1 d_i + e_{1i}$ such that $h_i = \begin{cases} 1, \text{ if } H_i \geq 0 \\ 0, \text{ if } H_i < 0 \end{cases}$,

   and $\mu_i = \exp(X_i\gamma + \rho_2 d_i + e_{2i})$.

3. Finally, generate y_i such that $\begin{cases} y_i = 0, \text{ if } h_i = 0 \\ y_i \sim \text{truncated } P(\mu_i), \text{ if } h_i > 0 \end{cases}$.

We run our MCMC algorithm for 5,000 replications and we collect every 10th replication after discarding the first 1000 replications of the burn-in phase. In addition, we generate 500 observations according to the proposed model in order to check the performance of the estimation method for different sample sizes.

Table A.1[1] presents estimates of the posterior means and standard deviations for the unrestricted models with 500 and 5000 observations, and the restricted ($\pi_1 = \pi_2 = 0$) model. The estimates ρ_1 and ρ_2 of the restricted model are far from their true values implying a self-selection effect. However, when the endogeneity of the treatment variable is taken into account the posterior means of the coefficients are very close to the true values.

Formally, we compare the two models utilizing the Savage-Dickey density ratio approach. The Bayes factor is found to be 1.35×10^{-7} which provides decisive evidence against the null hypothesis $H_0 : \pi_1 = \pi_2 = 0$. The autocorrelation functions (ACF) of the coefficients die off after 5 lags for all parameters.

We calculate the Average Treatment Effect and the Average Treatment Effect for the Treated to be: $ATE = 0.44$ (0.26) and $ATET = 0.41$ (0.28). Due to favorable selection bias the unobserved factors that induce individuals to be in the treated state, affect negatively the propensity to cross the hurdle and the conditional mean of the count variable. Therefore, those in the treated group will expect to gain less in terms of utilization from choosing the treatment than those without the treatment.

1.9 Application

We investigate the effect of managed care on outpatient physician visits. Our sample consists of the privately insured US population aged 21-64 and is obtained from the Medical Expenditure Panel Survey. MEPS is a nationally representative survey conducted by the Agency for Healthcare Research and Quality (AHRQ) and it provides data on health care use, expenditures, sources of payment, and insurance coverage for the US civilian non-institutionalized population. It uses an overlapping panel design according to which data are collected from each household for two calendar years while a new sample of households is surveyed each subsequent year. In other words, in each survey year (after the first), one sample of persons is in its second year of responses while another one is surveyed for the first time. We use data from all available surveys (1996-2002). However, we exclude observations of those in their first year of responses (and therefore all 1996 observations)

[1] All tables of this chapter are placed in Appendix A.

since one of our exclusion restrictions is the lag of the spouse' s private insurance. We further restrict our sample to employed but not self-employed reducing the sample size to 16,513.

A preliminary description of the distribution of the dependent variable is given in Table A.2. Outpatient physician visits present a slightly greater unconditional mean in the population insured by a FFS insurance plan. The percentage of zero counts is 85.07% which justifies the use of a hurdle model.

Definitions and summary statistics of all variables are given in Table A.3. The explanatory variables in the hurdle and mean equations are the same. Specifically, we include demographic variables (BLACK, HISPANIC, FAMSIZE, FEMALE, MARRIED, EDUCYR, AGE), self-perceived health status variables (VEGOOD, GOOD, FAIRPOOR) excluding excellent health status, variables measuring the number of chronic conditions and the number of injuries (NUMCHRON, NUMINJ), geographical variables (NOREAST, MIDWEST, SOUTH, MSA) which capture differences in the local healthcare markets, year dummies excluding the dummy for 1997, an economic variable (INCOME), and an insurance variable (MANAGED). In addition the variable AGE2 is included to capture non-linear dependence of utilization and age.

The explanatory variables in the insurance equation are the same with the addition of three exclusion restrictions: spouse's age (SPOUAGE), the square of spouse's age (SPOUAGE2), and SPOUPRIV which takes a value of one if the spouse of the individual had private insurance coverage in the previous year and zero otherwise. The purpose of the exclusion restrictions is to identify the correlation between the hurdle and mean equations and the insurance equation. Good instruments are those which are uncorrelated with the error terms in the hurdle and mean equations (exogeneity) and in addition they are highly correlated with the endogenous variable (insurance status). Both spouse's age and the lag of spouse's private insurance do not affect utilization, therefore satisfying exogeneity. We use the lag of spouse's private health insurance because the current private health insurance of the spouse may not be exogenous to utilization. On the other hand both variables have predictive power over the individual's insurance decision.

1.10 Results

We estimate our model running the Markov chain for 5000 replications and discarding the first 1000 replications of the burn-in phase during which the chain converges to the stationary distribution.

In Table A.4 we report the posterior means and standard deviations of the parameters. Health status variables have strong positive impacts on the decision to utilize services (the hurdle part), the conditional mean of outpatient visits (truncated part) and the insurance decision. In contrast, chronic conditions and injuries do not affect insurance. Family size has a strong negative effect on the decision to visit a physician but no effect on the frequency of visits and the insurance decision. Being female, black or Hispanic has strong positive impact on the decision to subscribe to managed care plans and no effect on the number of outpatient visits. However, it is less likely for blacks and Hispanics to make an outpatient physician visit at least once within a year. Education has strong positive impacts on both the hurdle part and the conditional mean but affects negatively the enrollment to a managed care plan. It is also interesting that income is weak in all cases.

The posterior means and standard deviations for the correlation parameters π_1 and π_2 are -0.586(0.354) and 0.022(0.155) respectively while for the coefficients of managed care ρ_1 and ρ_2 the posterior means and standard deviations are 0.894 (0.555) and 0.015 (0.253). That is, only π_1 and ρ_1 are strong. This is an indication of selection bias in the hurdle part. Unobserved factors that affect the insurance decision, do also affect the decision whether to make zero or positive outpatient visits. An interpretation of the negative sign of π_1 is that individuals who self-select themselves into HMO or PPO plans are on average healthier than those who are randomly assigned to these plans. Since they are healthier, they will initiate fewer treatment episodes on average.

The incentive effect is captured by ρ_1. Since we have controlled for selection bias on both observed and unobserved factors, ρ_1 can be interpreted as the pure effect of the insurance variable on the first stage utilization decision. The positive sign implies that managed care insurance plans influence positively their patient's decision whether to visit an outpatient center or not. This result is quite expected: managed care plans have attempted to control resource utilization through administrative efforts that encourage outpatient services and discourage the use of inpatient procedures. For instance, more stringent utilization review has been employed to ensure that surgical

procedures are not performed in the hospital when they can be performed in outpatient centers.

The separation of the incentive effect from the selection effect has strong policy implications. Ignoring self-selectivity on unobservables could have led to the misleading conclusion that managed care plans affect negatively the decision to make an outpatient visit.

As explained in section 6, a formal test for the endogeneity of the treatment variable is based on the calculation of the Bayes factor. The estimated Bayes factor is 1.223 (0.341), which provides no evidence in favor of one of the competing models. This is expected since the correlation parameter π_2 is weak thereby failing to identify a selection effect in the mean equation. The calculation of the Bayes factor with a different set of priors was also inconclusive.

Finally, we are interested in evaluating treatment effects from subscribing in MCOs. The average treatment effect for the treated is ATET = 0.094 (0.061) and the average treatment effect is ATE = 0.129 (0.085). Therefore, those in the treated group will expect to gain less in terms of utilization from choosing the treatment than those without the treatment.

Next we divide treated individuals into three groups according to the self perceived health status (excellent, very good, good and fair or poor health) and report in Table A.5 the mean and standard deviation of the ATET for each group. Clearly, the ATET is greater for those with fair or poor health implying that selection is stronger for this group.

1.11 Conclusions

Empirical research has provided mixed evidence on the effect of managed care insurance plans on the demand for medical services. A well known problem in this literature is that managed care plan members may differ from conventional insurance enrollees in terms of their health status and other observed and unobserved attributes. While most of the studies employ rich sets of control variables to account for differences in observed characteristics, unobserved differences among enrollees in different types of insurance plans are ignored in most cases.

The endogenous Hurdle Poisson log-Normal model developed in this paper accounts for selection on unobservables by explicitly modeling unobserved factors that may affect both the insurance decision and the utilization of outpatient care. In addition, it handles the large proportion of zero

counts (85% zero outpatient physician visits) in this particular application.

The model was applied to a sample of privately insured individuals aged 21-64 all whom are employed but not self-employed. Estimation of the model provided evidence of favorable selection in the hurdle part. A possible explanation is that managed care plans attract relatively healthier individuals who expect to make zero outpatient visits. After controlling for selection bias, the impact of a managed care plan on the decision to make an outpatient visit was found positive. Ignoring self-selectivity on unobservables could have led to the misleading conclusion that managed care plans reduce the number of positive outpatient visits. Calculation of the Average Treatment Effects (ATE) and the Average Treatment Effects for the Treated (ATET) revealed that managed care enrollees gain less in terms of utilization from choosing a managed care plan than the average individual. In addition, selection bias is found to be larger for the group of individuals with poor health.

Chapter 2

Bayesian Estimation of an Endogenous Panel Count Data Model with Correlated Random Effects: an Application to Universal Health Insurance Coverage

2.1 Introduction

The objective of this chapter is to develop and estimate a panel count data model with correlated random effects, which accounts for endogeneity of a binary choice variable. The estimation approach is Bayesian and parametric based on Markov Chain Monte Carlo (MCMC) methodology. Hierarchical priors are used for the unobserved individual effects so that the model emerges as the Bayesian analogue of a random effects approach in classical statistics. The empirical part of the essay estimates the effect of nearly universal insurance coverage – the endogenous treatment variable - on medical utilization measured as the number of doctor visits. The age limit for Medicare eligibility at age 65 serves as an exclusion restriction.

The main motivation for using panel data methods is the ability to model the intertemporal behavior of individuals and control for individual heterogeneity without observing it. In linear panel data models, the unobserved effects can be differenced out. However, differencing is not applicable in count data models where a natural non-linearity is produced by the non-negative discrete nature

of the data.[1] The most widely used estimation approach for fixed effects count data models is the conditional maximum likelihood (CML) proposed by Hausman et al. (1984). The motivation for the CML is to solve the incidental parameters problem which is present in most non linear panel data models.[2] The likelihood function is constructed conditioning on a statistic that contains all the information about the individual specific parameters. For the Poisson fixed effects model specifically, conditioning on the sufficient statistic – the total number of counts for each individual – removes the individual specific parameters from the likelihood.[3] However, the Poisson fixed effects model is one of the few non-linear models for which maximizing the unconditional likelihood gives the same (consistent) estimates as CML estimation. Unlike the Poisson model, Negative Binomial (NB) models are affected by the incidental parameters problem. Hausman et al. (1984) estimated the fixed effects NB1 using CML (see also Greene, 2007).[4]

In addition to fixed effects, Hausman et al. (1984) proposed a Poisson model with gamma distributed random effects and a NB2 model with beta distributed random effects. Finally, Chib et al. (1998), utilized Markov Chain Monte Carlo techniques to estimate a Poisson model with Gaussian random effects.

The estimators discussed above require exogenous explanatory variables. However, in many applications, the explanatory variables are correlated with the idiosyncratic shocks leading to inconsistent estimates. Endogeneity and selectivity in count data cross sectional regression models have been addressed in several ways including method of moments estimators, switching regression models, matching estimators, and Bayesian methods.

One strand of the literature concerns generalized method of moments (GMM) techniques. Windmeijer and Santos Silva (1997) proposed a model in which the count variable is specified with an additive error term and is simultaneously determined with the binary endogenous variable. Use of

[1] Differencing in the linear regression sense is not an option. However, moment based methods that use a differencing transformation to eliminate the (multiplicative) fixed effect can be applied (see Cameron and Trivedi, 1998).

[2] The incidental parameters problem (Neyman and Scott, 1948; Lancaster, 2000) arises in certain applications of maximum likelihood estimation. In the context of longitudinal data, when the number of individuals increases (while the number of time periods remains constant and relatively small) the number of parameters in a fixed effects model increases at the same rate as the sample size leading to inconsistent estimates for all parameters (not only the individual specific parameters). The logistic fixed effects model is the most well known model affected by the incidental parameters model. CML fixed effects logit estimation was proposed by Chamberlain (1984).

[3] Cameron and Trivedi (1998) provide the conditional log-likelihood which is similar to that of the multinomial logit model.

[4] However, Allison and Waterman (2002) criticized the model for not controlling for all stable covariates.

the additive moment condition and instruments that are correlated with the endogenous regressor but not with the unobservables give consistent GMM estimates. The model has also been specified with a multiplicative error in the count specification (Mullahy, 1997). Although the additive and multiplicative models are observationally equivalent when only the first order moment condition is utilized, Mullahy (1997) and Windmeijer and Santos Silva (1997) showed that the same set of instruments will not, in general, be orthogonal to both error types. Therefore, the two sets of moment conditions cannot be valid at the same time. Terza (1998) discussed a FIML approach which is efficient but computationally burdensome. He also developed a two-stage method of moments (TSM) estimator for a partially parametric version of the model. The TSM estimator is a nonlinear least-squares analogue to Heckman's two stage estimator for linear models with endogenous dummies and avoids the numerical integral approximations required by the FIML estimation. Finally, he introduced a nonlinear weighted least-squares (NWLS) estimator as an alternative to the FIML for the fully parametric model. The NWLS estimator is computationally efficient relative to FIML estimation and statistically efficient relative to the partially parametric TSM estimator.

Van Ophem (2000) presented a method for modeling endogenous selectivity in count data which is similar in spirit with the switching regression model. The endogenous dummy variable defines two regimes which are distinguished by different data generating processes. The regime choice is allowed to be correlated with the observed count in each of the regimes.

Propensity score matching that has been used extensively to deal with selectivity in applications where a social program or "treatment" is evaluated (see Rosenbaum and Rubin, 1983, 1985; Rubin and Thomas, 1996; Heckman et al., 1997) can be applied without additional concerns in cases where the outcome of the treatment is a count. However, an important limitation of PSM is that it accounts only for selection on observables.

Another line of research on endogeneity in count data models involves Bayesian methods. Recent studies have analyzed the effect of endogenous multinomial choice indicators on an ordinal dependent variable (Munkin and Trivedi, 2008), a self selection model with endogenous binary treatment and multiple outcomes one of which is a count (Munkin and Trivedi, 2003), a Roy-type model of selectivity with count outcomes (Deb et al., 2006b), and a two-part model with endogenous multinomial treatment (Deb et al., 2006a).

For panel data models, several methods have been suggested to deal with endogeneity when the outcome equation is linear (see for example Wooldridge, 1995; Kyriazidou, 1997; Vella and Verbeek, 1999; Charlier et al., 2001; Kyriazidou, 2001; Semykina and Wooldridge, 2005). Less attention has been paid to endogeneity in panel count data: in the case of predetermined regressors,[5] Chamberlain (1992, 1993) and Wooldridge (1997) proposed a quasi-differenced GMM estimator which has been applied by Cincera (1997), Crepon and Duguet (1997) and Montalvo (1997), and extended to a dynamic linear feedback model for count data by Blundell et al. (2002). Windmeijer (2000, 2006) showed that the moment condition proposed by Wooldridge (1991, 1997) can also be used to consistently estimate the parameters in fixed effects count data models with endogenous regressors.

The present essay contributes to the literature on endogenous panel count data models by proposing a model which assumes a count variable generated according to a Poisson process and a binary choice variable which is modeled to be endogenous to the count. Unobserved heterogeneity across individuals enters in the linear specifications of both the Poisson mean and the latent variable associated with the binary choice by the way of correlated individual specific random effects.

The proposed model is applied to the hotly debated issue of universal health insurance coverage. Advocates of nearly universal coverage argue that the benefits from imposing individual or employer mandates or providing subsidies to those who are presently uninsured and would like to buy insurance outweigh the costs. The lower access to and use of health care services by the uninsured reduces their health status and productivity. Universal access to health services will result in improved health for the uninsured and increased productivity. In addition, part of current health care costs are uncompensated care for the uninsured. Hospitals are required by federal or state law to provide emergency care treatment to the uninsured regardless of their ability to pay. These costs are shifted to taxpayers supporting public hospitals and private insurers who in turn pass them on to employers by charging higher premiums. Universal coverage is likely to reduce cost shifting especially if the utilization gap between insured and uninsured is small. Another argument in favor of universal coverage is that marketing and underwriting costs, estimated to 1 or 2 percent of health insurance premiums would be eliminated if insurers were required to accept all applicants, regardless of health status.

[5] Regressors that are correlated with past shocks.

On the other hand, universal coverage will raise the use of health care services by the previously uninsured. The larger the current access gap between insured and uninsured, the greater the additional resource cost. Furthermore, administrative costs may as well increase if individual mandates are introduced. In order to be effective mandates require enforcement which may be tough and costly especially if subsidies are inefficient and benefits are inappropriate.

Apart from measuring administrative costs (and benefits), costs associated with additional health care used by the uninsured, savings from cost shifting reduction, and improvements in health status and productivity a cost benefit analysis would also take into account changes in utilization per se. It is often stated that universal coverage will reduce unnecessary use of ER by uninsured individuals who have nowhere else to turn to for primary care. On the other hand, a severe overcrowding in primary care physician offices throughout the state of Massachusetts has been reported since the passage of a universal health plan that helped more than 340,000 uninsured residents in the state get the coverage they need. Therefore, it is of great importance to estimate changes in utilization of services resulting from expanding insurance protection to Americans who now lack coverage. Taking into account utilization estimates along with population growth and the growth rate of active physicians one can predict the pressure on physicians and hospital capacity for treatment (Long and Marquis, 1994).

Selection bias and endogeneity of insurance are inherent concerns for modeling health care utilization. In studies with observational data individuals are not randomly assigned to have insurance coverage but they select themselves to the "insured" state based on health status, risk preferences and socioeconomic factors. Therefore, observed differences in utilization rates between the insured population and the uninsured may reflect both lower out of pocket costs for the covered individuals (incentive effects) and selection effects. Universal coverage is expected to change drastically the utilization incentives of the presently uninsured. Failing to separate incentive from selection effects the researcher runs the risk of deriving misleading estimates of the impact of universal insurance. The proposed model estimates the pure incentive effect by modelling the endogeneity of insurance status. To identify the effect of the treatment (insurance) on the outcome (utilization) and make the estimates more robust we use Medicare eligibility at 65 as an exclusion restriction. The age limit at 65 is a natural instrument since insurance coverage rates increase sharply as individuals

pass their 65th birthday and any changes in utilization rates at 65 are attributed to postponed use of medical services in anticipation of Medicare coverage (see Card et al., 2004).

The chapter is organized in the following manner. In section 2.2 we formulate the model and specify prior distributions. The MCMC algorithm is developed in section 2.3. In section 2.4 we discuss computation of treatment effects and Bayes factor for hypothesis testing. A numerical example with artificial data tests the performance of the MCMC estimation method in section 2.5. Section 2.6 presents the empirical application and section 2.7 concludes.

2.2 Model Specification

We assume T periods and N individuals. The insurance equation is specified as follows:

$$\begin{aligned} D_{it} &= W_{it}\alpha + \theta_{1i} + \varepsilon_{it}, \\ \varepsilon_{it} &\sim N(0,1) \end{aligned} \tag{2.1}$$

where D_{it} is a latent variable that reflects individual's i propensity to be insured at time t, W_{it} is an $1 \times p_a$ vector of exogenous explanatory variables, $\theta_{1i} \sim N(0, \sigma_{11})$ is the individual specific random effect, α is $p_a \times 1$ parameter vector and $\varepsilon_{it} \sim N(0,1)$ is the the idiosyncratic error that captures the unexplained by W_{it} and θ_{1i} part of the insurance equation. What we observe is:

$$d_{it} = \begin{cases} 1, & D_{it} \geq 0 \\ 0, & D_{it} < 0 \end{cases} \tag{2.2}$$

where d_{it} $(i = 1, ..., N,\ t = 1, ...T)$ is a binary treatment variable taking a value of one if individual i has any type of health insurance at time t and zero otherwise,

Utilization is modeled through a Poisson model

$$y_{it} \dot{\sim} P(\exp(\mu_{it})) \tag{2.3}$$

where the conditional mean μ_{it} is specified as a linear function of explanatory variables X_{it} and

the treatment d_{it}. Specifically:

$$\mu_{it} = X_{it}\beta + \rho d_{it} + \theta_{2i} + \delta\varepsilon_{it} + u_{it}, \;\; u_{it} \sim N(0, \sigma_u) \tag{2.4}$$

X_{it} is an $1 \times p_\beta$ vector of exogenous regressors, $\theta_{2i} \sim N(0, \sigma_{22})$ the individual specific random effect, β is $p_\beta \times 1$, ρ is 1×1, and δ is 1×1 parameter vectors, and u_{it} is an error term. Under this specification, the idiosyncratic errors of the insurance and utilization equations are correlated. This is the mechanism through which endogeneity of insurance to utilization is modeled. In addition we allow for the individual specific random effects θ_{1i} and θ_{2i} to be correlated with covariance matrix

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}. \tag{2.5}$$

That is, the vector of the random effects $\theta = (\theta_{1i}, \theta_{1i})'$ follows a bivariate Normal distribution, $\theta \sim BN(0, \Sigma)$.

We estimate the model choosing prior distributions as follows. Since no prior information is available about parameters α, β we center their priors at zero values with a relatively large variance such that:

$$\alpha \sim N(0_p, 10I_p), \;\; \beta \sim N(0_k, 10I_k)$$

as well as for ρ_1 and ρ_2:

$$\rho_1 \sim N(0, 1/2) \;\; \text{and } \rho_2 \sim N(0, 1/2)$$

Finally, we select conjugate priors for σ_u and Σ, specifically a gamma prior for σ_u

$$\sigma_u^{-1} \sim G(\frac{\eta_0}{2}, (\frac{c_0}{2})^{-1}) \text{ where } \eta_0 = 5 \text{ and } c_0 = 3.$$

and an Inverse Wishart prior for Σ

$$\Sigma \sim W^{-1}(\underline{v}, \underline{V}) \text{ where } \underline{v} = 5 \text{ and } \underline{V} = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}$$

2.3 Estimation

We use MCMC methods to estimate our model. Let $\widetilde{\Delta} = (\alpha, \beta, \delta, \rho, \Sigma, \sigma_u)$ be the set of parameters, and $\Delta_{it} = (\widetilde{\Delta}, X_{it}, W_{it})$. Then the joint density (likelihood) of observed and latent variables for observation i is:

$$\begin{aligned} & f(y_{it}, d_{it}, D_{it}, \mu_{it}, \theta_{1i}, \theta_{2i} \mid \Delta_{it}) = \\ & \prod_{t=1}^{T} \frac{e^{-e^{\mu_i t}} e^{\mu_{it} y_{it}}}{y_{it}!} \\ & \times \prod_{t=1}^{T} [d_{it} I(D_{it} \geqslant 0) + (1 - d_{it}) I(D_{it} < 0)] \\ & \times \prod_{t=1}^{T} \varphi_1(\mu_{it} \ ; \ X_{it}\beta + \rho d_{it} + \delta(D_{it} - W_{it}\alpha - \theta_{1i}) + \theta_{2i}, \sigma_u) \\ & \times \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi}} \exp\{-0.5(D_{it} - W_{it}\alpha - \theta_{1i})^2 \\ & \times \varphi_2(\theta; 0, \Sigma) \end{aligned} \tag{2.6}$$

where $\varphi_1(\bullet; \mu, \sigma^2)$ denotes the Normal density $N(\mu, \sigma^2)$, $\varphi_2(\bullet; \mu, \Sigma)$ denotes the bivariate Normal density $BN(0, \Sigma)$, and $I(x \in X)$ is an indicator function taking a value of one if the random variable x belongs to the set X and zero otherwise. The posterior density is the product of the likelihood $f(Y_{it}, d_{it}, D_{it}, \mu_{it}, \theta_{1i}, \theta_{2i} \mid \Delta_{it})$ for all N observations times the prior.

We organize the MCMC algorithm in the following steps by blocking the parameters as D_{it}, μ_{it}, θ_{1i}, θ_{2i}, α, $(\beta'$, ρ, $\delta)'$, σ_2.

1. From the full conditional density for D_{it} and our square completion we find that:

$$D_{it} \overset{i.i.d.}{\sim} TN_{(0,\infty)}\left(\overline{D_{it}}, \overline{V_{D_{it}}^{-1}}\right) \text{ if } d_{it} = 1 \tag{2.7}$$

and

$$D_{it} \overset{i.i.d.}{\sim} TN_{(-\infty,0)}\left(\overline{D_{it}}, \overline{V_{D_{it}}^{-1}}\right) \text{ if } d_{it} = 0 \tag{2.8}$$

where

$$\overline{D_{it}} = W_{it}a + \theta_{1i} + \overline{V_{D_{it}}^{-1}}\left[\frac{\delta}{\sigma_u}(\mu_{it} - X_{it}\beta - \rho d_{it} - \theta_{2i})\right], \tag{2.9}$$

$$\overline{V_{D_{it}}} = \frac{\delta^2}{\sigma_u} + 1 \tag{2.10}$$

We draw values from these truncated Normal distributions utilizing an algorithm proposed by Geweke (1991).

2. From the full conditional density for θ_{1i} and our square completion we get:

$$\theta_{1i} \sim N(\overline{\theta_{1i}}, \overline{V_{\theta_{1i}}^{-1}}) \tag{2.11}$$

where

$$\overline{\theta_{1i}} = \overline{V_{\theta_{1i}}^{-1}}\left[(1 + \frac{\delta^2}{\sigma_u})\sum_{t=1}^{T}(D_{it} - W_{it}a) - \frac{\delta}{\sigma_u}\sum_{t=1}^{T}(\mu_{it} - X_{it}\beta - \rho d_{it} - \theta_{2i}) + \frac{\sigma_{12}\sigma_{22}^{-1}\theta_{2i}}{\sigma_{11} - \sigma_{12}\sigma_{22}^{-1}\sigma_{12}}\right], \tag{2.12}$$

$$\overline{V}_{\theta_{1i}} = 4 + \frac{4\delta^2}{\sigma_u} + \frac{1}{\sigma_{11} - \sigma_{12}\sigma_{22}^{-1}\sigma_{12}} \tag{2.13}$$

3. From the full conditional density for θ_{2i} and our square completion we get:

$$\theta_{2i} \sim N(\overline{\theta_{2i}}, \overline{V_{\theta_{2i}}^{-1}}) \tag{2.14}$$

where

$$\overline{\theta_{2i}} = \overline{V_{\theta_{2i}}^{-1}}\left\{\frac{1}{\sigma_u}\sum_{t=1}^{T}\left[\mu_{it} - X_{it}\beta - \rho d_{it} - \delta(D_{it} - W_{it}\alpha - \theta_{1i})\right] + \frac{\sigma_{12}\sigma_{11}^{-1}\theta_{1i}}{\sigma_{22} - \sigma_{12}\sigma_{11}^{-1}\sigma_{12}}\right\}, \tag{2.15}$$

$$\overline{V_{\theta_{2i}}} = \frac{4}{\sigma_u} + \frac{1}{\sigma_{22} - \sigma_{12}\sigma_{11}^{-1}\sigma_{12}} \tag{2.16}$$

4. The full conditional density for μ_{it} is:

$$\pi\left(\mu_{it} \mid y_{it}, d_{it}, Z_{it}, \Delta_{it}\right) \propto \frac{e^{-e^{\mu_{it}} t} e^{\mu_{it} y_{it}}}{y_{it}!} \times \varphi\left(\mu_{it} \; ; \; X_{it}\beta + \rho d_{it} + \delta(D_{it} - W_{it}\alpha - \theta_{1i}) + \theta_{2i}, \sigma_u\right) \tag{2.17}$$

In order to sample μ_{it} we utilize the independence chain Metropolis-Hasting algorithm which uses a proposal density independent across draws. Geweke (1989) argued that the proposal density should have tails at least as fat as the posterior. For that reason we select the t-distribution with a small number of degrees of freedom as the candidate-generating density. Following Chib et al. (1998), we tailor the proposal density to the target density centered at its modal value $\overset{\wedge}{\mu}_{it} = \arg\max \ln \pi\left(\mu_{it} \mid y_{it,} d_{it}, Z_i, \Delta_i\right)$. To obtain the mode we implement the Newton-Raphson algorithm using the gradient vector g_{μ_i} and the Hessian H_{μ_i} of $\ln \pi\left(\mu_i \mid y_{i,} d_i, Z_i, \Delta_i\right)$. Then the proposal t-distribution is $q\left(\mu_i\right) = f_T(\mu_i \mid \overset{\wedge}{\mu}_i, V_{\overset{\wedge}{\mu}_i}, v)$ where $\overset{\wedge}{\mu}_i$is the location parameter, $V_{\overset{\wedge}{\mu}_i} = -\left(H_{\overset{\wedge}{\mu}_i}\right)^{-1}$ is the negative inverse of the Hessian evaluated at the mode, and v are degrees of freedom. The chain moves to a value $\mu_i^{\dagger}$ drawn from the proposal density, with probability

$$\alpha\left(\mu_i, \mu_i^{\dagger}\right) = \min\left\{\frac{\pi\left(\mu_i^{\dagger} \mid y_{i,} d_i, Z_i, \Delta_i\right) q\left(\mu_i\right)}{\pi\left(\mu_i \mid y_{i,} d_i, Z_i, \Delta_i\right) q\left(\mu_i^{\dagger}\right)}, 1\right\} \tag{2.18}$$

and stay at the current value μ_i with probability $1 - \alpha\left(\mu_i, \mu_i^{\dagger}\right)$.

5. Let the prior distribution of α be $N(\underline{\alpha}, \underline{H}_a^{-1})$. The full conditional density of α is $\alpha \sim N(\overline{\alpha}, \overline{H}_a^{-1})$ where

$$\overline{H}_a = \underline{H}_a + (1 + \frac{\delta^2}{\sigma_u}) \sum_{i=1}^{N} \sum_{t=1}^{T} W_{it}' W_{it} \tag{2.19}$$

and

$$\overline{\alpha} = \overline{H}_a^{-1} A \tag{2.20}$$

where A is defined as follows:

$$A = \underline{H}_a\underline{\alpha} + (1+\frac{\delta^2}{\sigma_u})\sum_{i=1}^{N}\sum_{t=1}^{T}[W'_{it}(D_{it}-\theta_{1i})] - \frac{\delta}{\sigma_u}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[W'_{it}(\mu_{it}-X_{it}\beta-\rho d_{it}-\theta_{2i})\right] \quad (2.21)$$

6. We specify priors $\beta \sim N(\underline{\beta}, \underline{H}_{\beta}^{-1})$, $\rho \sim N(\underline{\rho}, \underline{H}_{\rho}^{-1})$, $\delta \sim N(\underline{\delta}, \underline{H}_{\delta}^{-1})$ and let $\gamma = [\beta' \ \rho \ \delta]'$ and $C_{it} = [X_{it} \ d_{it} \ (D_{it} - W_{it}\alpha - \theta_{1i})]$

   Then the full conditional density of γ is $\gamma \sim N(\overline{\gamma}, \overline{H_{\gamma}}^{-1})$ where

$$\overline{H_{\gamma}} = \underline{H}_{\gamma} + \frac{1}{\sigma_u}\sum_{i=1}^{N}\sum_{t=1}^{T}C'_{it}C_{it} \quad (2.22)$$

   and

$$\overline{\gamma} = \overline{H_{\gamma}}^{-1}\left[\underline{H}_{\gamma}\underline{\gamma} + \frac{1}{\sigma_u}\sum_{i=1}^{N}\sum_{t=1}^{T}C'_{it}(\mu_{it}-\theta_{2i})\right] \quad (2.23)$$

7. Given the Gamma prior $\sigma_u^{-1} \sim G(\frac{n_0}{2}, (\frac{c_0}{2})^{-1})$ and letting

$$J_{it} = \mu_{it} - [(X_{it}\beta + \rho d_{it} + \delta D_{it} - W_{it}\alpha - \theta_{1i}) + \theta_{2i}] \quad (2.24)$$

   we find the full conditional density of σ_u^{-1} to be

$$\sigma_u^{-1} \sim G\left(\frac{n_0}{2} + 2N, \left[\frac{\sum_{i=1}^{N}J'_{it}J_{it}}{2} + \frac{c_0}{2}\right]^{-1}\right). \quad (2.25)$$

8. Finally, given the Inverse Wishart prior $\Sigma \sim W^{-1}(\underline{v}, \underline{V})$, the full conditional density of Σ is

$$\Sigma \sim W^{-1}\left(N + \underline{v}, \theta\theta' + \underline{V}\right) \quad (2.26)$$

2.4 Bayes factor and Treatment effects

To test the null hypothesis of no endogeneity of the treatment variable ($H_0 : \delta = 0$) we need to compare the unrestricted model M_1 and the restricted model M_0 with $\delta = 0$. For model comparison

with equal prior weight attached to each model the posterior odds ratio becomes the Bayes Factor defined as:

$$B_{01} = \frac{\pi\left(y \mid M_0\right)}{\pi\left(y \mid M_1\right)}$$

where $\pi\left(y \mid M_i\right)$ is the marginal likelihood for model M_i. Since the models are nested we can apply the Savage-Dickey ratio approach to calculate the Bayes factor (see Verdinelli and Wasserman, 1995). Suppose that $\theta = (\omega, \psi)$ is the parameter set and we want to test the null hypothesis $\omega = \omega_0$. Dickey (1971) showed that if the prior for ψ in the restricted and unrestricted models is the same at the point $\omega = \omega_0$ then the Bayes factor can be calculated as

$$B_{01} = \frac{\pi\left(\omega_0 \mid y\right)}{\pi\left(\omega_0\right)}$$

Therefore, under weak assumptions the Bayes factor is calculated as:

$$B_{01} = \frac{\pi\left(\delta^* \mid y\right)}{\pi\left(\delta^*\right)}$$

where $\pi\left(\delta^* \mid y\right)$ is the posterior density and $\pi\left(\delta^*\right)$ the prior density evaluated at $\delta^* = 0$.

The prior for δ must be specified carefully. In Bayesian estimation the effect of priors diminishes as the sample size becomes fairly large. However, as shown in Kass and Raftery (1995), Bayesian testing differs from estimation in that the effect of the prior density may not vanish as the sample size increases. As a result, the Bayes factor is sensitive to the choice of priors. Improper priors (that can be used for estimation and prediction) will always choose the null hypothesis and therefore are not acceptable for hypothesis testing. Even informative priors with a large spread on π_i under H_1 may provide overwhelming support to the model with fewer parameters regardless of the data, thereby favoring the restricted model. For that reason we choose informative priors.

In our application, we are interested in measuring the impact of the transition from the uninsured to the insured state on utilization of medical services. Let y_{it}^I denote the outcome with treatment (insurance coverage) and y_{it}^U the outcome without treatment (lack of insurance) for individual i in period t. The observed outcome can be written as a function of y_{it}^I and its counterfactual

y_{it}^U

$$y_{it} = d_{it} y_{it}^I + (1 - d_{it}) y_{it}^U \tag{2.27}$$

since we can not observe the same individual in both states — insured and uninsured — at the same time.

The treatment effect for the treated (TET), defined as the insured's expected gain (in terms of utilization) from having insurance coverage is given by:

$$E[y_{it}^I - y_{it}^U \mid d_{it} = 1] = \exp(X_{it}\beta + \rho + \theta_{2i} + \delta\varepsilon_{it} + u_{it}) - \exp(X_{it}\beta + \theta_{2i} + \delta\varepsilon_{it} + u_{it}) \tag{2.28}$$

where t=1,...,T and individual i belongs to the treated group.

2.5 Numerical example

In this example we generate artificial data based on the proposed model to test our MCMC algorithm. Specifically, we generate a balanced panel consisting of 4 periods and 3000 observations for each period. The data are generated such that the binary treatment variable takes the value 1 in about 90% of cases resembling our real data. In addition we generate a panel of 500 observations for each of the 4 periods to compare estimates from different sample sizes. The following steps illustrate the data generation process.

1. Generate: $X_{it} = (1, x_{it})$, $x_{it} \sim N(0, 1)$

   $W_{it} = (1, w_{it})$, $w_{it} \sim N(0, 1)$

   and specify: $\alpha = (2, 1)$, $\beta = (1, 1)$, $\rho = -0.5$, $\delta = 0.5$, $\sigma_{11} = \sigma_{22} = 0.2$, $\sigma_{12} = 0.1$ and $\sigma_u = 0.6$

2. Generate the vector of random effects from a bivariate Normal:

$$\theta = \begin{pmatrix} \theta_{1i} \\ \theta_{2i} \end{pmatrix} \sim BN \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix} \right)$$

3. Generate $\varepsilon_{it} \sim N(0,1)$ and

   $D_{it} = W_{it}\alpha + \theta_{1i} + \varepsilon_{it}$ such that $d_{it} = \begin{cases} 1, \text{ if } D_{it} \geq 0 \\ 0, \text{ if } D_{it} < 0 \end{cases}$,

4. Generate $v_{it} \sim N(0,1)$, $u_{it} = \delta\varepsilon_{it} + \sqrt{\sigma_u}v_{it}$ and

   $\mu_{it} = X_{it}\beta + \rho d_{it} + \theta_{2i} + u_{it}$

5. Finally, generate y_{it} such that $y_{it} \sim Poisson\left(\exp(\mu_{it})\right)$

We run our MCMC algorithm for 5,000 replications and we collect every 10th replication after discarding the first 1000 replications of the burn-in phase. Table B.1[6] presents estimates of the posterior means and standard deviations for the unrestricted models with 500 and 5000 observations. As the sample size increases the posterior means are moving closer to the true values and the standard deviations become smaller. With 5000 observations the posterior means are very close to the true parameters. The autocorrelation functions of the coefficients die off after 12 lags for all parameters except from σ_{11} which dies off after 20 lags.

We also estimate the model ignoring the endogeneity of the treatment variable. To implement this, in the estimation phase we fix the draws of δ to zero. Estimates of the posterior means and standard deviations for the restricted ($\delta = 0$) model are also presented in table B.1. The estimates of β_1 and ρ (0.40 and 0.14 respectively) are far from their true values 1 and -0.5. The inconsistency of the treatment coefficient ρ reveals the problem of ignoring the endogeneity issue. While the data have been generated such that the effect of treatment on outcome is negative, the estimated posterior mean of ρ is found positive because it compounds the pure effect of treatment and selection effects. The Bayes factor is found to be 1.35×10^{-286} which provides decisive evidence against the null hypothesis $H_0 : \delta = 0$.

[6] All tables of this chapter are placed in Appendix B.

2.6 Application

Health and Retirement Study (HRS) is a national longitudinal survey of individuals over age 50 and their spouses. Its goal is to provide panel data to support research on retirement and health insurance related issues. The survey provides information about demographics, income, assets, health, cognition, family structure, health care utilization and costs, insurance, housing, job status, and expectations. HRS is a not only rich but also extraordinarily complex. It consists of 4 cohorts presented along with interview dates in table B.2.[7] An initial HRS cohort including individuals born between 1931 and 1941 was first interviewed in 1992 and subsequently every two years. The second cohort (AHEAD) consists of individuals born in 1923 or before. They were first interviewed in 1993 and subsequently in 1995, 1998, 2000, and 2002. In addition to respondents, the survey interviewed the spouses (or partners) of married respondents regardless of age. Some spouses of the initial HRS entry cohort respondents were age 70 or older and were subsequently included in the AHEAD study as 1992 interviews. The HRS and AHEAD cohorts were merged, with a single interview schedule, in 1998 and at the same time two new cohorts were added. The children of depression (CODA) cohort are individuals born 1924 to 1930. Finally the war baby (WB) cohort are those born 1942 to 1947. The last two cohorts were first interviewed in 1998 and subsequently every two years.

We use the RAND HRS files - a user friendly version of a subset of the HRS - created by the RAND center for the study of Aging. The RAND HRS data contain cleaned and processed variables that are comparable across 6 survey waves. Wave 1 is from 1992. Wave 2 is from 1994 data for the HRS cohort, and 1993 for the AHEAD cohort. Wave 3 is from 1996 data for the HRS cohort and 1995 for the AHEAD cohort. Waves 4, 5 , and 6 are from 1998, 2000, and 2002 respectively.

In our application we use only the initial HRS cohort of individuals born between 1931 and 1941 ignoring the AHEAD, CODA, and WB cohorts for two reasons. First, we want to avoid the complexity of the ground rules for following baseline respondents and their spouses in subsequent interview waves. Second, and most important, we want to construct a balanced panel. Furthermore, we use 4 waves (waves 3-6) of the initial HRS cohort. The first 2 waves are excluded since none of

[7] Table B.2 is a modification of a table that appears in RAND HRS documentation, version E, August 2005.

the individuals has reached the threshold of 65 years. In addition, the instructions accompanying the question about respondents doctor utilization are significantly different in the first two waves. The resulting panel consists of 5295 individuals present in all 4 waves, thereafter referred to as periods 1-4.

Table B.3 gives definitions and summary statistics for the variables in each of the 4 waves. The dependent variable DRVIS is defined as the number of times the respondent has seen or talked to a medical doctor including emergency room or clinic visits in the reference period. The number of doctor visits increases over time from 7.07 in the 1996 wave to 8.33 in the 2002 wave. The vector of explanatory variables used in the mean equation includes the socioeconomic variables AGE, GENDER, EDUCYR, BLACK, HISPANIC, MARRIED, the self-perceived health status variables EXCHLTH, VGHLTH, GOODHLTH (FAIRPOORHLTH is the reference category), the economic variables INC and WEALTH, the year dummies YEAR96, YEAR98, YEAR00 corresponding to waves 1 to 3 (wave 4 is the reference category), and the binary variable INS indicating whether individual has any insurance coverage. Since health status indicators are self perceived rather than measured based on diagnoses coding (dcg scores) it is likely to leave part of health status variation unexplained. The unexplained part included in the error terms of the insurance and mean equations may be responsible for selection on unobservables which motivates the modeling strategy. Income is the sum of all income in household (respondent's and spouse's earned income plus capital income plus income from employer pension plus income from social security retirement plus income from unemployment benefits and government transfers). Wealth is constructed as the net value of total wealth excluding second residency. In order to avoid problems caused by extreme values, income and wealth are transformed to the categorical variables INC and WEALTH as shown in table B.4. The same explanatory variables with the addition of the exclusion restriction THRES indicating whether an individual is over 65 or older are included in the insurance equation.

All U.S. citizens and legal residents 65 years of age and older who have an employment record (or are married to someone with an employment record) that entitles them to Social Security benefits are eligible for Medicare. This accounts for 95% of the elderly population. In addition, people with disabilities who are younger than age 65 and who have received Social Security Disability Insurance (SSDI) or Railroad Retirement disability payments for 24 months, people with end-stage renal

disease (ESRD) requiring dialysis or a transplant, and people with amyotrophic lateral sclerosis (ALS) are covered by Medicare. Figure E.1 [8] graphs Medicare coverage by age for all individuals in the sample (line with squares), whites with at least 12 years of education (line with stars) and less educated blacks or Hispanics (line with circles). For all three groups there are discontinuous changes in Medicare coverage at age 65 with less educated minorities having the largest Medicare coverage rate just before the age of 65.

Medicare eligibility at age 65 is a powerful instrumental variable for insurance coverage. Figure E.2 shows insurance coverage rate by age for the overall population and the two sub-groups of educated whites and less educated blacks or Hispanics. Coverage rate for the overall population represented by the line with squares rises at age 65 from 88% to 98%. The line with stars shows that educated whites have higher coverage rate (93%) just before age 65 than the overall population but the gap narrows once they become eligible for Medicare. The discrete jump in coverage rate at age 65 for less educated minorities is notably larger. Only 68% are insured prior to age 65 while 93% have insurance coverage just after their 65th birthday.

Only 5% of individuals are 65 or older in period 1. The percentage of those crossed the Medicare eligibility threshold increases to 20% in period 2, and reaches 57% in the last period. Accordingly the number of insured rises over time from 89% in 1996 to 96% in 2002. Table B.5 shows utilization by insurance status. The gap in doctor visits between insured and uninsured widens from 0.98 visits in period 1 to 3.60 in period 4. Table B.6 presents differences in utilization and insurance by group. DOC is a binary variable that shows whether the individual had at least one doctor visit.

We estimate the model running the Markov chain for 5000 replications and discarding the first 1000 replications of the burn-in phase. Table B.7 reports the posterior means and standard deviations of the parameters. As expected, the health status variables EXCHLTH, VGHLTH, and GOODHLTH have strong negative effects on the utilization decision compared to the reference category which includes individuals in fair or poor health. However, their impact on insurance is weak. Being female has a strong positive impact on the number of doctor visits. Blacks and Hispanics are less likely to have insurance coverage. The opposite is true for married individuals who also utilize more doctor visits. Education has strong positive effects on both insurance and

[8] All figures are placed in Appendix E.

utilization. Individuals in households with higher income have a higher number of doctor visits but it is household wealth that affects positively the decision to buy insurance.

The posterior means of the treatment parameter ρ and the correlation parameter δ, 0.328(0.044) and -0.042(0.021) respectively, are strong. Unobserved factors that affect positively whether an individual has insurance coverage have also a strong negative impact on the number of doctor visits implying favorable selection bias. An interpretation could be that insured individuals are on average unobservably healthier or less risky than if they were randomly assigned to the insured state and therefore, they utilize fewer doctor visits on average. Since selection bias on both observables and unobservables has been taken into account, the treatment parameter ρ can be seen as the pure incentive effect of insurance on utilization. Ignoring self selectivity on unobservables would have resulted in a misleadingly smaller estimate for the incentive effect since the treatment parameter would compound selection and incentive effects.

We are interested in comparing the model that permits endogeneity of insurance and a model that treats insurance as exogenous. The Bayes factor calculated by the Savage-Dickey ratio is 2.4×10^{-6} which provides decisive evidence against the null hypothesis $H_0 : \delta = 0$.

Average treatment effects for the treated (ATET) are calculated for all four periods as explained in section 4. They are reported in table B.8 as ATET1-ATET4 for periods 1-4 respectively. Insured individuals in period 1 utilize 1.92 more doctor visits than if they had not been covered. Their gain in utilization from the treatment (insurance) rises from 1.92 visits in period 1 to 2.04 visits in period 2, 2.12 visits in period 3 and 2.27 doctor visits in period 4. Average treatment effects (ATE) are presented in table B.9 and are slightly lower than the ATET for all waves. Compared to the unconditional differences in utilization between the insured and the uninsured, average treatment effects differ substantially in periods 1 and 4. The actual utilization difference in wave 1 is 0.98 visits, remarkably lower than the ATE (1.92 visits) and the ATET (1.88 visits). This gap implies favorable selection bias. Incentive effects measured by ATE and ATET are positive and result in utilization increase of approximately 2 doctor visits. However, due to favorable selection observed differences in utilization are lower than those estimated by the treatment effects. The results are reversed in wave 4. The difference in unconditional means of the insured and uninsured is 3.60 doctor visits while the ATE estimates a rise of only 2.23 doctor visits when individuals switch

from the uninsured to the insured state. This difference can be explained by the existence of adverse selection effects. In waves 2 and 3 actual differences in utilization and treatment effects are relatively close. It is interesting that the size of the selection effect is large only in waves 1 and 4 and that the direction of selection bias is the opposite in the two waves. How can this be explained by the data? Table B.10 presents the means of health variables by insurance status. In wave 1 insured individuals have higher means of "excellent health" and "very good health" and lower means of "good health" and "fair or poor health". Since insured individuals are healthier they utilize less doctor visits than the uninsured. Surprisingly, in wave 4 the mean of "excellent health" is higher for the uninsured which may be the source of adverse selection. Therefore, after removing adverse selection on observables and favorable selection on unobservables reflected in the negative correlation coefficient δ, the remaining utilization difference of 2.23 doctor visits is the incentive effect in wave 4.

At the theoretical level, universal coverage may reduce the use of the emergency department by the previously uninsured and cause overcrowding of primary care physician offices. It would therefore be more appropriate to estimate the effect of insurance on various measures of utilization such as primary care physician visits, specialist visits, hospital admissions, and outpatient visits. Unfortunately, HRS defines doctor visits as the number of times an individual has seen a medical doctor including clinic and ER visits. Another limitation is that health insurance is heterogeneous in the sense that Medicare provides different benefits than other insurance plans. Therefore, differences in utilization may be observed because individuals covered by other forms of insurance switch to Medicare at age 65 and not because uninsured individuals obtain coverage. Besides the limitations, the application offers an example of the model's ability to deal with endogenous treatment in panel data applications with count dependent variables.

2.7 Conclusions

The chapter proposed a panel count data model with correlated random effects that accounts for endogeneity of a binary choice variable. The model was applied to analyze the effect of health insurance coverage on utilization of doctor visits allowing for potential endogeneity of the insurance

status. The sample consists of individuals over age 50 and their spouses followed for 4 waves from 1996 t0 2002. The incentive effects of health insurance are separated from selection bias on both observables and unobservables and are found to be positive and strong. Insured individuals utilize 1.92 more doctor visits in wave 1, 2.04 in wave 2, 2.12 in wave 3 and 2.27 in wave 4. Compared to the observed utilization rates that compound incentive and selection effects, these numbers show a more consistent pattern in utilization differences between insured and uninsured population. Depending on data availability, the model can be used to estimate the impact of insurance coverage on other measures of health care utilization such as specialist visits, hospital admissions and outpatient services. Therefore, it adds an empirical tool in the current discussion on universal health insurance coverage as it can predict overcrowding in primary care physician offices or shifting of services from the costly emergency department to the outpatient setting.

Chapter 3

Demand for Cigarettes: A Mixed Binary-Ordered Probit Aproach

This chapter is a lightly revised version of a paper by the same name published in the journal *Applied Economics* in 2008 by Panagiotis Kasteridis, Murat Munkin, and Steven Yen:

Kasteridis P., M. Munkin, and S. Yen, 2008. Demand for cigarettes: a mixed binary-ordered probit approach. *Applied Economics*, in press. DOI: 10.1080/00036840701704402.

My use of "we" in this chapter refers to my co-authors and myself. My primary contributions to this paper include (1) most of the estimation, and (2) part of the writing.

3.1 Introduction

This paper proposes an approach to analyze the demand for cigarettes utilizing a rich dataset that identifies non-smokers, potential smokers, quitters and actual smokers in the sample. This information brings additional gains in efficiency than that accomplished in the double-hurdle approach (e.g., Jones, 1989a) which has been the primary econometric tool in cigarette demand modeling. Following a general approach suggested in Jones (1989a), the proposed model consists of an equation that explains the decision to be a non-smoker, an equation that explains the quitting decision for those who started smoking in the past, and an equation that models the level (number) of cigarettes consumed. Five competing specifications are considered for the level equation, some of which have been attempted in the literature.

There is a long-standing interest in the empirical analysis of cigarette smoking by individuals because the health effect of cigarette smoking is an important public-health issue. Many studies are

based on micro survey data, which allow for investigation of the roles of detailed socio-demographic characteristics. In modeling cigarette demand with microdata, it has become a standard approach to assume that cigarette consumption is subject to two decisions: whether to smoke and how much to smoke (Fry and Pashardes, 1994; Garcia and Labeaga, 1996; Jones, 1989a, 1989b, 1995; Labeaga, 1999; Mullahy, 1985). These models vary in specifications of the two censoring mechanisms. The double-hurdle model, a bivariate generalization of the Tobit model (Tobin, 1958), assumes that zero observations are attributed to both nonparticipation and censoring (e.g., Atkinson et al., 1984; Jones, 1989a). Specifically, with two separate processes to govern participation and consumption in the double-hurdle model, zero observations are generated by those who are non-smokers and those who are potential smokers but choose not to consume. In many datasets potential smokers are usually not identifiable. However, availability of such information can lead to efficiency gains and simplifications of functional forms. Jones (1989a) introduces a trivariate model which features three stochastic processes accommodating starting, quitting and the level of cigarette smoking. Not only do we know individuals in our dataset who never smoked but also we can identify those who started smoking in the past and quit. Those who identify themselves as non-smokers and quitters have zero cigarette consumption. It is interesting to note that a few individuals among those who are smokers choose not to smoke, perhaps, trying to quit smoking. All this information is utilized in our approach.

Another important issue this paper addresses relates to the distribution of the level variable. The number of cigarettes is a count variable. Many of the double-hurdle model applications have been based on the bivariate normal distribution. While the normal distribution may be appropriate for applications based on cigarette expenditure data (Atkinson et al., 1984; Jones and Labeaga, 2003) or weekly consumption data (Jones, 1989a), it is unlikely to accommodate other forms of reported consumption. It is plausible to assume the Poisson or Negative Binomial distribution for the count variable. However, a close examination of the dependent variable in our dataset shows that the underlying distribution is neither normal nor Poisson related and it is difficult to expect a good fit from such models. Since most of the observed consumption values are reported as a fraction of a pack of cigarettes it is reasonable to group observations in categories and utilize the ordered outcome approach. We consider two competing specifications along that line: the ordered

probit and sequential probit models.

In this paper, we accommodate our particular form of data by following the trivariate model of Jones (1989a) and trying alternative distributions for the level equation. Specifically, the model features a starting equation, a quitting equation, and an equation that explains the level of consumption with five alternative specifications. First we analyze a joint sample that includes both male and female individuals. We find that the ordered probit model specification is preferred to the others based on the Akaike information criterion (Akaike, 1973). To facilitate interpretations of results we propose to calculate the marginal effects of variables for the preferred ordered probit model. The calculated effect with respect to gender suggests that further analysis applied to the male and female subsamples is desirable. We also check sensitivity of the calculated marginal effects to alternative categorization of the level variable.

The rest of the paper is organized as follows. Section 2 develops the model and presents five competing specifications of the level equation. Section 3 describes the data and variables used in the study. Section 4 considers an application and presents the results. Section 5 concludes the paper. The computational appendix presents the likelihood function for the fully-specified (dependent) trivariate model and formulae for the marginal effects.

3.2 Econometric Specification

Assume we observe N independent observations where the dependent variable of interest, y_i ($i = 0, 1, ..., N$), is a count variable that displays a large proportion of zeros. Each individual i belongs to one of three groups. The first group includes individuals who had not started smoking by the time the survey was conducted. The second group are those who had smoked in the past but decided to quit and consider themselves quitters. The last group are current smokers. The binary decisions to start and quit smoking are both modeled with probit models. The starting equation is characterized by a latent equation

$$s_i^* = X_i\alpha_1 + \varepsilon_{1i}, \tag{3.1}$$

where latent variable s_i^* measures the difference in utility derived by individual i from starting and not starting smoking, X_i is a vector of exogenous variables, α_1 is a conformable parameter vector, and the error terms ε_{1i} are independently and identically distributed as standard normal, that is, $\varepsilon_{1i} \sim N(0,1)$. The observed binary variable for starting (S_i) relates to the latent variable (s_i^*) such that

$$S_i = \begin{cases} 1 \text{ iff } s_i^* > 0 \\ 0 \text{ otherwise} \end{cases}, \tag{3.2}$$

taking a value of 1 if the individual ever started smoking. Then the probability of starting is

$$\Pr(S_i = 1) = \Phi\left(X_i \alpha_1\right), \tag{3.3}$$

where $\Phi(\cdot)$ is the cumulative distribution function (cdf) of the standard normal. Likewise, quitting is characterized by latent variable q_i^*, which measures the difference in utility derived from smoking and quitting states, and specified as

$$q_i^* = X_i \alpha_2 + \varepsilon_{2i}, \tag{3.4}$$

where $\varepsilon_{2i} \sim N(0,1)$. The observed binary variable for quitting is defined as

$$Q_i = \begin{cases} 1 \text{ iff } q_i^* > 0 \\ 0 \text{ otherwise} \end{cases}, \tag{3.5}$$

where $Q_i = 1$ if a person continues smoking and $Q_i = 0$ if she quits, and

$$\Pr(Q_i = 1) = \Phi\left(X_i \alpha_2\right). \tag{3.6}$$

Let y_i be our dependent variable, measuring the level of smoking. Variable y_i takes the value of zero for either non-smokers ($S_i = 0$) or quitters ($Q_i = 0 | S_i = 1$). Non-zero values can only be

observed conditional on ($S_i = 1$ and $Q_i = 1$). The joint likelihood of observing the sample is

$$\begin{aligned} L(y|\theta) &= \prod_{S_i=0} [1 - \Pr(S_i = 1)] \prod_{S_i=1,Q_i=0} \Pr(S_i = 1) [1 - \Pr(Q_i = 1|S_i = 1)] \\ &\quad \times \prod_{S_i=1,Q_i=1} \Pr(S_i = 1) \Pr(Q_i = 1|S_i = 1) \Pr(y_i|S_i = 1, Q_i = 1) \end{aligned} \tag{3.7}$$

where θ is a vector containing all parameters in the model, and the products are taken over sample observations satisfying $(S_i = 0)$, $(S_i = 1, Q_i = 0)$ and $(S_i = 1, Q_i = 1)$ conditions respectively.

For the conditional density $\Pr(y_i|S_i = 1, Q_i = 1)$, we consider five different specifications: a Gaussian model that truncates the error term to ensure non-negativity, two negative binomial models, an ordered probit and a sequential ordered probit model. The Gaussian specification which is the driving mechanism in the double-hurdle model (Jones, 1989a), and the single-hurdle model (Yen, 2005) assumes that y_i follows a truncated normal distribution with mean $X_i\beta$ and variance σ^2 such that

$$y_i = X_i\beta + u_i, \tag{3.8}$$

$$u_i > -X_i\beta \tag{3.9}$$

and

$$\Pr(y_i|S_i = 1, Q_i = 1) = \sigma^{-1} \frac{\phi\left((y_i - X_i\beta)/\sigma\right)}{\Phi(X_i\beta/\sigma)}, \tag{3.10}$$

where $\phi(\cdot)$ is the probability density function (pdf) of the standard normal.

We also consider two forms of the negative binomial model, namely NB1 and NB2, with probability mass functions of the forms

$$\Pr(y_i|S_i = 1, Q_i = 1) = \frac{\Gamma(y_i + \upsilon^{-1}\lambda_i)}{\Gamma(y_i + 1)\Gamma(\upsilon^{-1}\lambda_i)} \left(\frac{\upsilon^{-1}}{\upsilon^{-1} + 1}\right)^{\upsilon^{-1}\lambda_i} \left(\frac{1}{1 + \upsilon^{-1}}\right)^{y_i}, \tag{3.11}$$

$$\Pr(y_i|S_i = 1, Q_i = 1) = \frac{\Gamma(y_i + \upsilon^{-1})}{\Gamma(y_i + 1)\Gamma(\upsilon^{-1})} \left(\frac{\upsilon^{-1}}{\upsilon^{-1} + \lambda_i}\right)^{\upsilon^{-1}} \left(\frac{\lambda_i}{\upsilon^{-1} + \lambda_i}\right)^{y_i}, \tag{3.12}$$

where $\lambda_i = \exp(X_i\beta)$, $\Gamma(\cdot)$ is the Gamma function, and υ is the overdispersion parameter.

Finally, we specify two ordered outcome models. Construction of the ordered dependent variable

is discussed in section 3. Assume that the dependent variable y_i, measuring the level of smoking conditional on $S_i = 1$ and $Q_i = 1$, takes integer values from 0 to M. Our first ordered outcome model specification assumes that y_i follows the ordered probit model. The latent variable Z_i which measures the propensity to smoke at different levels is assumed to be linear in X_i through the structural equation

$$Z_i = X_i\beta + u_i, \tag{3.13}$$

where the error term $u_i \sim N(0,1)$. Variable y_i relates to Z_i according to the observability condition

$$y_i = \begin{cases} 1 \text{ iff } Z_i \leq \tau_1 \\ 2 \text{ iff } \tau_1 < Z_i \leq \tau_2 \\ 3 \text{ iff } \tau_2 < Z_i \leq \tau_3 \quad , \\ . \\ M \text{ iff } \tau_{M-1} < Z_i \end{cases} \tag{3.14}$$

where $\tau = (\tau_1, ..., \tau_{M-1})$ are threshold parameters and τ_1 is restricted to zero for identification. To complete the likelihood function specify

$$\Pr(y_i = m | S_i = 1, Q_i = 1) = \Phi(\tau_m - X_i\beta) - \Phi(\tau_{m-1} - X_i\beta), \tag{3.15}$$

so that

$$\Pr(y_i | S_i = 1, Q_i = 1) = \prod_{m=1}^{M} [\Pr(y_i = m | S_i = 1, Q_i = 1)]^{d_{im}}, \tag{3.16}$$

where d_{im} is a binary indicator such that $d_{im} = 1$ iff $y_i \in m$th category and 0 otherwise.

Our last model specification choice is the sequential probit model which assumes that the ordered variable y_i can take the value m only after the levels $1, ..., m-1$ have been reached. Then the conditional probability of reaching level m given the levels $1, ..., m-1$ have been reached is

$$\Pr(y_i = m \mid y_i \geq m, \beta, \tau) = \Phi(\tau_m - X_i\beta), \tag{3.17}$$

where X_i is a vector of covariates, β is a parameter vector, $\tau = (\tau_1, ..., \tau_{M-1})$ are threshold parameters. Then, the unconditional probabilities are

$$\Pr(y_i = m|\beta, \tau) = \Phi(\tau_m - X_i\beta) \prod_{k=1}^{m-1} [1 - \Phi(\tau_k - X_i\beta)],\ m = 1, ..., M-1. \tag{3.18}$$

$$\Pr(y_i = M \mid \beta, \tau) = \prod_{k=1}^{M-1} \left[1 - \Phi\left(\tau_k - X_i\beta\right)\right],\ m = M \tag{3.19}$$

To highlight a potential advantage of this model over the ordered probit model consider the latent representation of the sequential probit model which assumes a latent variable structure that generates the count outcome. Define latent variables

$$\xi_{mi} = X_i\beta + u_{mi}, \tag{3.20}$$

where $u_{mi} \sim N(0, 1)$. The latent variables ξ_{mi} represent propensities to continue to the next level. Applied to the number of cigarettes smoked, level m can be reached only after an individual makes it to level $m-1$ and ξ_{mi} defines propensities to smoke additional cigarettes to move to the next level of consumption. Counts are generated according to

$$y_i = \begin{cases} 1 & \text{iff } \xi_{1i} \leq \tau_1 \\ 2 & \text{iff } \xi_{1i} > \tau_1, \xi_{2i} \leq \tau_2 \\ 3 & \text{iff } \xi_{1i} > \tau_1, \xi_{2i} > \tau_2, \xi_{3i} \leq \tau_3 \\ . & . \\ M-1 & \text{iff } \xi_{1i} > \tau_1, \xi_{2i} > \tau_2, ..., \xi_{M-2,i} > \tau_{M-2}, \xi_{M-1,i} \leq \tau_{M-1} \\ M & \text{iff } \xi_{1i} > \tau_1, \xi_{2i} > \tau_2, ..., \xi_{M-2,i} > \tau_{M-2}, \xi_{M-1,i} > \tau_{M-1} \end{cases}. \tag{3.21}$$

Then

$$\Pr(y_i|S_i = 1, Q_i = 1) = \prod_{m=1}^{M-1} \left[\Pr(y_i = m \mid \beta, \tau)\right]^{d_{im}} \left[\Pr(y_i = M \mid \beta, \tau)\right]^{d_{iM}}. \tag{3.22}$$

As can be seen from Equation 21, the propensities to move to the next level are not ordered since the threshold parameters do not follow the order condition similar to that of the ordered probit

model. It is not clear *a priori* whether propensities to smoke more cigarettes increase with the level of smoking but allowing for a flexible propensity structure seems to be justified. It remains an empirical matter whether the sequential model would perform better than the ordered probit.

We restrict the error covariances across equations to be zeros. This is done because our data do not provide variables that would affect one equation but not the others. This is a limitation to our study. The assumption of independence cannot be tested and the fully-specified trivariate model cannot be estimated. We however present its likelihood function for the ordered probit case which can be used in future applications with more conducive datasets (see Appendix D). For the current application with independent error terms, the likelihood function (Equation 7) suggests that the starting equation can be estimated separately with the whole sample, the quitting equation with the starter sample ($S = 1$), and the level equation with the smoker sample ($S = 1, Q = 1$).

3.3 Data and Variables

Data are compiled from the 1994–96 Continuing Survey of Food Intakes by Individuals (CSFII), collected by the Agricultural Research Service of the US Department of Agriculture (2000) (see US Dept Agr, 2000) A nationally representative survey, the CSFII 1994–96 were stratified, multistage area probability samples targeting individuals of all ages. With an overall response rate of 76.1 percent, the three-year data initially included 20607 individuals of all ages, of whom 10721 age 12 or over were asked about lifestyle and cigarette smoking. We focus on individuals age 15 or over as few of those age 12–14 reported smoking any cigarettes. After deleting observations with missing values on important variables, a total of 9587 individuals (4923 men and 4664 women) remain for analysis.

In the CSFII, each individual was asked (i) whether she/he had smoked 100 or more cigarettes in the entire life, and if yes, (ii) whether she/he currently smoked at the time of the survey. Responses to these questions allow identification of starters and quitters, respectively. In addition, among the current smokers, each was asked the question: 'On average, how many cigarettes do you smoke per day?' The reported number is used as the quantity variable.

The quantities of cigarettes smoked are reported in the form of 'number per day', which feature

a pile-up of counts at 10 cigarettes (0.5 pack), 20 (1 pack), 30 and 40 cigarettes, and so forth. It is unlikely that such pile-ups of counts can be adequately accommodated by the normal distribution. The histogram of the number of cigarettes smoked by the smokers only (Figure E.3) has spikes at 5, 10, 15, 20, 30 cigarettes, which are self-reported consumption levels.

Since the number of cigarettes is self-reported and perhaps due to convenience of (or errors in) reporting there are disproportionately larger shares of outcomes measured in packs of 0.25, 0.5, 0.75, 1 and 1.5. Self-reported measurement errors are likely present in the sample as individuals might have rounded the number of cigarettes reported smoked to the closest integer in multiples of 5, such as 5, 10 and 20. It seems therefore reasonable to categorize the reported quantities in the following categories: 0–5 cigarettes, 6–10, 11–15, 16–20, 21–30, and over 31. The histogram of the constructed quantity variable is presented in Figure E.4 for smokers in the sample. Another interesting feature of the sample is that there are a few individuals who identified themselves as smokers who never quit but whose cigarette consumption is zero.

The CSFII also includes detailed demographic information on each individual. Variables commonly used in the cigarette demand literature are included in the starting, quitting and consumption equations (Blaylock and Blisard, 1992a, 1992b; Jones, 1989a, 1989b, 1994, 1995). The explanatory variables are income, body mass, education and age, along with dummy variables indicating urbanization (city, suburban), region (Northeast, Midwest, South), years of survey (Year95, Year96), race (Black), ethnicity (Hispanic), self-evaluated health, and whether the individual was a white-collar worker, had been diagnosed with cancer, high blood pressure or heart problems. Also included are lifestyle variables indicating whether the individuals had consumed alcohol in the past three months, had exercised regularly (no exercise and intensively), or was on any special diet (see Table 1). Price information is not available in the survey and so is included in the constant terms. However, the regional, urbanization and year dummy variables are expected to capture some price variations. In view of the literature in which socio-demographic variables are used as proxies for missing prices such as wage rate (e.g., Wales and Woodland, 1980) and the fact that, due to the insignificant role of transportation cost, cigarette prices are likely to be dominated by state level taxes, our use of regional, urbanization and time dummy variables serves as a remedial measure

for the omission of the price variable[1]. Use of these variables in the literature and in the current study is elaborated in the empirical section below.

Of the final sample (N=9587), 4743 individuals (49.47%) ever started smoking, 2430 (25.35%) had smoked in the past but had quit. Among the 2313 (24.12%) current smokers, the average number of cigarettes is 18.64 per day. Detailed definitions and sample statistics for all variables by gender and for the pooled sample are presented in Table C.1.[2]

3.4 Application and Results

The application section is organized as follows. First we concentrate our analysis on the pooled sample including both males and females. We estimate five competing models and perform model selection based on the Akakie Information Criterion (AIC). The preferred model is the ordered probit for which we calculate marginal effects and partial changes. The estimated partial effect with respect to gender invites further analysis with the separate male and female subsamples, for which the ordered probit model is found to be the preferred model as well. As a robustness check we calculate marginal effects for the ordered probit model for a different segmentation of the observations, that is, with alternative definitions of the ordered dependent variable.

ML estimates of the starting and quitting equations for the pooled sample are presented in Table C.2. These results are the same for all five competing models since the first two equations are independent of the conditional part and therefore parameters from the three equations are separable. The significant determinants of the starting variable at the 5% level of significance are education (negative), age, Black (negative), Hispanic (negative), healthy (negative), white-collar (negative), male, alcohol and no exercise. Quitting ($Q = 0$) is affected positively by the individual's income, body mass, education, age, the geographic variables city, Northeast and suburban, and the variables white-collar and diet. On the other hand, alcohol consumption has a negative impact on quitting. Blacks are less likely to quit smoking than others, while Hispanics and males are more successful in quitting.

[1]State-level cigarette taxes would be a good proxy for price. However, for confidentiality reasons the CSFII sampling units were not identifiable by state.

[2]All tables of this chapter are placed in Appendix C.

The results of the starting equation should be interpreted as the effects of the covariates on the decision to be a non-smoker, which is an up-to-date decision taken throughout the entire life up to the point of the survey. Thus, all variables affecting level are included in the starting equation because they affect the current choice to be a non-smoker. The same reasoning applies to the current decision to be and remain a quitter for those who once started smoking.

We find that body mass does not affect the decision to be a non-smoker but individuals with a higher body mass are more likely to quit smoking. This result is consistent with the findings of Jones (1994, 1995) and Blaylock and Blisard (1992b). Education improves the individual's cognitive skills regarding the risks associated with smoking, thereby discouraging starting and motivating quitting. Similar result was reported by Blaylock and Blisard (1992a). It is also in line with Hsieh (1998) who reports that the probability of quitting smoking increases with years of formal education and Ault et al. (2004) who find that those who attend high school have a higher probability of smoking and a lower probability of quitting than others. As expected, consumption of alcohol, another addictive good, decreases the probability of being a non-smoker and the probability of quitting. Individuals who exercise only rarely have a smaller probability to be a non-smoker but the lack of exercise has no effect on quitting.

Jones (1995) and Yen (2005) found that age has a negative impact on participation arguing that most smokers start smoking as a teenager or young adult. We find that age is negatively correlated with the non-smoking decision. However, older individuals are more likely to quit. Individuals who are black are more likely to be non-smokers but less successful in quitting. The empirical literature on the effects of self-perceived health status is mixed. We find that poor health is negatively related to non-smoking. As to the role of gender, males are more prone to starting smoking but are more successful in quitting.

Table C.3 presents ML results for the conditional portion of the five competing models outlined above. For the NB1, NB2, ordered probit, and sequential ordered probit models, body mass index, age, Midwest, South and sex are positive and significant while city, Black, ethnicity (Hispanic), and white-collar are negative and significant. Note that education is not significant for any of the competing models.

The ordered probit model fits the data the best according to the log-likelihood values and the

respective AIC[3]. The threshold parameters $\tau_2, ..., \tau_5$ are all significant at the 1% level, justifying the use of all six categories over combining some categories. It is known for the ordered probit model that the signs of the coefficients may not relate directly to the directions of the effects of variables on the probabilities of categories. In addition, it is useful to relate each category to the quantity level it stands for. To accomplish this goal we calculate marginal effects of each variable on the ordered probit probabilities

$$\Pr(y_i = m|X_i) = \left[\Phi\left(\tau_m - X_i\beta\right) - \Phi\left(\tau_{m-1} - X_i\beta\right)\right]\Phi(X_i\alpha_1)\Phi(X_i\alpha_2), \tag{3.23}$$

and the conditional mean of the dependent variable

$$E(y_i|X_i) = \sum_{m=1}^{M} \overline{y}_m \Pr(y_i = m|X_i), \tag{3.24}$$

where $\overline{y}_m$ is the category mean for the mth category. Since the ordered model involves combining observations into categories our results are likely to depend on how the categories are defined (an issue investigated later) and how averaging takes place. To address this issue we also use category medians and modes as weights in calculating the marginal effects but since they produce similar results the corresponding marginal effects are not reported. The marginal effects on the conditional mean (Equation 3.24) is the sum of the marginal effects on the probabilities (Equation 23), weighted by the category means $\overline{y}_m$. The effects of each binary explanatory variable are derived by simulating a finite change (i.e., from 0 to 1) in the variable, holding all other variables constant. Analytical expressions for the marginal effects are enclosed in the appendix.

The marginal effects of explanatory variables are presented in Table C.4. It is interesting to note that significance and signs of the marginal effects vary with categories and the weighted average. Income has a negative and significant effect on the cigarette consumption level, with higher income individuals smoking fewer cigarettes than others. The roles of income in cigarette smoking are largely inconclusive in the literature. Blaylock and Blisard (1992a) find that cigarettes are

[3]Note that the ordered probit specification for conditional level can be extended to the multinomial probit (or logit). The lack of variations among some of the explanatory variables however prevented this pursuit without consolidating the ordered dependent variable.

an inferior good. Tansel (1993) reports a low but positive income elasticity of demand, which is consistent with the addictive nature of cigarettes, whilst Goel and Nelson (2005) find that income does not affect smoking prevalence among adults.

Body mass, education and age all have significant and overall negative effects. Body mass and age both have positive effects on the probability of consuming in the lowest (0–5) category and negative effects on the probabilities of the higher categories, whereas education has negative effects on the probabilities of all but the highest (> 30) categories. The persistently negative effects of body mass, education and age translate into the significant and negative effects on the level of cigarette consumption.

Also presented in Table C.4 are the discrete effects of binary explanatory variables. Positive effects on level are seen in Midwest, South and male, which are due to the positive (negative) effects of these variables on the probabilities of consuming in the higher (lower) categories. Thus, men consume more cigarettes than women, and individuals residing in the Midwest and the South consume more cigarettes than those residing in the West. Consumption of alcohol and lack of exercise are also positive and significant. Opposite effects are seen in the other variables, which include residing in the city or a suburban area (relative to rural area), residing in the Northeast, being on a special diet and being Black, Hispanic, healthy and white-collar worker. Individuals with these characteristics consume fewer cigarettes than others. Overall, the effects of most variables on the probabilities and conditional level, though statistically significant, are fairly small. The more notable effects of variables are seen in ethnicity, gender, race, job status, use of alcohol and diet, and lack of exercise, with individuals of the Hispanic origin smoking 3.43 fewer cigarettes (per day) than non-Hispanics, men smoking 1.58 more cigarettes than women, alcohol consumers smoking 2.21 more cigarettes than non-consumers, and Blacks smoking 1.17 fewer cigarettes than non-Blacks. The effects of other variables are all very small, with individuals 10 years older smoking only 0.77 fewer cigarette than their younger counterparts, and with all other binary variables having the effects of less than 1 cigarette per day on average.

The estimated marginal effects with respect to gender in the pooled men-women sample invite further analysis for men and women separately. Such separate analysis can provide further insight into gender differences in cigarette consumption. We perform a formal test to determine whether

male and female samples can be pooled or should be used separately in modelling cigarette demand. The results, reported in Table C.5, suggest the hypothesis of equal parameters between genders is rejected for all competing models, favoring estimation of each model for men and women separately.

Table C.6 presents estimation results for men and women, using the preferred ordered probit specification for the conditional level. White-collar men are more likely to be non-smokers and more successful in quitting. In contrast, having a white-collar job has no effect on the starting or quitting variables among women. This result for women contradicts the usual argument that antismoking messages have the greatest effect on women in better jobs. Men on a special diet are more likely to be non-smokers and they smoke fewer cigarettes on average; these effects of special diet are not seen in women. Other differences between genders are observed in the effects of city, Hispanic, healthy and alcohol on quitting, and the effects of age and Northeast on the level of smoking. Table 6 presents the mean marginal effects and average partial changes. The signs and significance of the effects are similar to those for the pooled sample. However, the marginal effects of body mass and age on the level, as well as the discrete effects of city, suburban, Midwest, South, no exercise, Hispanic, and white-collar are smaller for women than for men. Yen (2005) calculated elasticities with respect to the same variables using hurdle models and found smaller effects for women as well. On the other hand our marginal effects with respect to Black and healthy are larger for women, whilst Yen documents the opposite. Unconditional on smoking, a black woman smokes 1.41 fewer cigarettes than other women while a black man smokes 0.97 fewer cigarettes than other men. In general, our marginal effects are smaller than the those reported by Yen for both men and women.

Finally, we investigate sensitivity of the marginal effects to the choice of categories. Our original categorization of quantity, labeled "Categorization 1" in Table C.7, was based on the assumption that self-reported numbers of cigarettes smoked are rounded upward to the closest fraction of a cigarette pack during reporting. Self-reported consumption levels may be subject to non-systematic measurement errors. While our paper does not address the measurement errors per se, as a robustness check of the marginal effects we categorized quantity with an alternative set of cut-offs (Categorization 2): 0–4 cigarettes, 5–9, 10–14, 15–19, 20–29, and over 30. The marginal effects calculated at the conditional means based on this alternative categorization are presented in Table

6. The results show that the calculated mean effects are similar between the two categorization schemes.

3.5 Conclusions

We address the issue of zero observations in modeling cigarette smoking. Our dataset contains information which allows investigation of three key elements of cigarette smoking: starting, quitting and the level of consumption. Despite a lack of exclusion restrictions, a shortcoming, which prevents estimation of the fully specified model with dependent covariances, the special feature of the data allows us to construct a statistical model that accommodate skewness of distribution and pile-ups of counts in the number of cigarettes smoked. We model starting and quitting as binary variables, and use alternative specifications to accommodate the level of smoking. The ordered probit model provides the best fit to the data. Unlike other ordered probit specifications based entirely on the category information, the approach we follow allows a way to relate each category to the level associated with the category, thereby allowing the calculation of marginal effects of variables on the level of cigarettes smoked.

The empirical analysis was carried out separately for men and women, and we find strong evidence of gender differences in cigarette consumption in terms of parameter estimates and marginal effects of explanatory variables. Estimation of our preferred model, with the ordered probit specification for the conditional level of smoking, requires categorization of the quantity variable and we find our results are robust to alternative categorizations of the variable. Although our data do not allow estimation of the dependent trivariate model due to a lack of exclusion restrictions, we present the likelihood function which can be used in future applications when the data become available. Whilst we apply the statistical model to cigarette smoking, the model can be useful in other applications, such as consumption of soft drink or vegetables which are likely to be reported in cans or servings, in which dependent variable may be censored and may feature skewness in distribution and pile-ups at certain values.

Bibliography

Akaike H., 1973. Information theory and an extension of the maximum likelihood principle. In Petrov, B. N. and Csaki, F. (eds), 2nd International Symposium on Information Theory, Budapest: Akademiai Kiado, 267–81.

Albert, H.J. and S. Chib, 1993. Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association* 88 (422), 669-679.

Allison, P. and R.Waterman, 2002. Fixed effects Negative Binomial regression models. *Sociological Methodology* 32, 247-265.

Atkinson, A.B., J. Gomulka and N. Stern, 1984. Household expenditure on tobacco 1970–1980: evidence from the Family Expenditure Survey, London: London School of Economics.

Ault, R.W., R.B. Ekelund, J.D. Jackson and R.P. Saba, 2004. Smokeless tobacco, smoking cessation and harm reduction: an economic analysis. *Applied Economics* 36, 17–29.

Blaylock, J.R. and W.N. Blisard, 1992a. US cigarette consumption: the case of low income women. *American Journal of Agricultural Economics* 74, 698–705.

Blaylock, J.R. and W.N. Blisard, 1992b. Self-evaluated health status and smoking behavior. *Applied Economics* 24, 429–35.

Blundell, R., R. Griffith and F. Windmeijer, 2002. Individual effects and dynamics in count data models. *Journal of Econometrics* 108, 113-131.

Bruce, Welch, et al., 1987. The RAND health insurance study: a summary critique. *Medical Care* 25 (2), 148-156.

Cameron, A.C. and P.K. Trivedi, 1998. Regression analysis of count data. Econometric Society Monographs No 30.

Card, D., C. Dobkin and N. Maestas, 2004. The impact of nearly universal insurance coverage on health care utilization and health. NBER working paper series. Working paper 10365.

Chamberlain, G., 1984. Panel data. In:Griliches, Z., Intriligator, M., (Eds), Handbook of Econometrics. North Holland, Amsterdam.

Chamberlain, G., 1992. Comment: sequential moment restrictions in panel data. *Journal of Business and Economic Statistics* 10, 20-26.

Chamberlain, G., 1993. Feedback in panel data models. *Mimeo, Econometric Society Summer Meeting* Boston, June.

Charlier, E., B. Melenberg and A. Van Soest, 2001. An analysis of housing expenditure using semiparametric models and panel data. *Journal of Econometrics* 101, 71-107.

Chib, S., E. Greenberg and R. Winkelmann, 1998. Posterior simulation and Bayes factors in panel count data models. *Journal of Econometrics* 86, 33-54.

Chib, S. and B.H. Hamilton, 2000. Bayesian analysis of cross section and clustered data treatment models. *Journal of Econometrics* 97, 25-50.

Cincera, M., 1997. Patents, RD, and technological spillovers at the firm level: some evidence from econometric count models for panel data. *Journal of Applied Econometrics* 12, 265-280.

Cragg, G.J., 1971. Some statistical models for limited dependent variables with application to the demand for durable goods. *Econometrica* 19 (5), 829-844.

Crepon, B. and E. Duguet, 1997. Estimating the innovation function from patent numbers: GMM on count panel data. *Journal of Applied Econometrics* 12, 243-263.

Deb,P. and P.K. Trivedi, 2002. The structure of demand for health care: latent class vs two-part models. *Journal of Health Economics* 21, 601-625.

Deb, P., M.K. Munkin and P.K. Trivedi, 2006a. Bayesian analysis of the two-part model with endogeneity: application to health care expenditure. *Journal of Applied Econometrics* 21, 1081-1099.

Deb, P., M.K. Munkin and P.K. Trivedi, 2006b. Private insurance, selection, and health care use: a Bayesian analysis of a Roy-type model. *Journal of Business and Economic Statistics* 24 (4), 403-415.

Dickey, J., 1971. The weighted likelihood ratio, linear hypotheses on Normal location parameters. *The annals of Statistics* 42, 204-223.

Fry, V. and P. Pashardes, 1994. Abstention and aggregation in consumer demand: zero tobacco expenditures. *Oxford Economic Papers* 46, 502–18.

Garcia, J. and J.M. Labeaga, 1996. Alternative approaches to modelling zero expenditure: an application to Spanish demand for tobacco. *Oxford Bulletin of Economics and Statistics* 58, 489–506.

Gerdtham, Ulf-G., 1997. Equity in health care utilization: further tests based on hurdle models and Swedish micro data. *Health Economics* 6, 303-319.

Geweke, J., 1989. Bayesian inference in econometric models using Monte Carlo integration. *Econometrica* 57 (6), 1317-1339.

Geweke, J., 1991. Efficient simulation from the Multivariate Normal and Student-t distributions subject to linear constraints. In E. M. Keramidas (ed.), Computing Science and Statistics: Proceedings of the Twenty-Third Symposium on the Interface, 571-578.

Glied S., 2000. Managed care. In Handbook of Health Economics, vol. 1, Culyer AJ, Newhouse JP (eds). Elsevier: New York, 2000, 707-753.

Goel, R.K. and M.A. Nelson, 2005. Tobacco policy and tobacco use: differences across tobacco types, gender and age. *Applied Economics* 37, 765–71.

Greene, W., 2007. Fixed and random effects for count data. New York University, Leonard N. Stern School of Business, Department of Economics Working Papers, number 07-16.

Gurmu, S., 1997. Semi-parametric estimation of hurdle regression models with an application to medicaid utilization. *Journal of Applied Econometrics*. Special Issue: Econometric models of event counts (May-June., 1997) 12 (3), 225-242.

Gurmu, S., 1998. Generalized hurdle count data regression models. *Economics Letters* 58, 263-268.

Hausman, J., B. Hall and G. Griliches, 1984. Econometric models for count data and an application to the patents-R&D relationship. *Econometrica* 52, 909-938.

Heckman, J., H. Ichimura and P.E. Todd, 1997. Matching as an econometric estimator: evidence from evaluating a job training programme. *The review of Economics and Statistics* 64 (4), 605-654.

Heckman, J., L.J. Tobias and E. Vytlacil, 2001. Four parameters of interest in the evaluation of social programs. *Southern Economic Journal* 68 (2), 210-223.

Hsieh, C.R., 1998. Health risk and the decision to quit smoking. *Applied Economics* 30, 795–804.

Jones, A.M., 1989a. A double-hurdle model of cigarette consumption. *Journal of Applied Econometrics* 4, 23–39.

Jones, A.M., 1989b. The UK demand for cigarettes 1954–86, a double-hurdle approach. *Journal of Health Economics* 8, 133–41.

Jones, A.M., 1994. Health, addiction, social interaction and the decision to quit smoking. *Journal of Health Economics* 13, 93–110.

Jones, A.M., 1995. A microeconometric analysis of smoking in the UK Health and Lifestyle Survey, Discussion Paper 139, The University of York, September.

Jones, A.M. and J.M. Labeaga, 2003. Individual heterogeneity and censoring in panel data estimates of tobacco expenditure. *Journal of Applied Econometrics* 18, 157–77.

Kass, E.R. and A.E. Raftery, 1995. Bayes factors. *Journal of the American Statistical Association* 90 (430), 773-795.

Kasteridis, P., M. Munkin and S. Yen, 2008. Demand for cigarettes: a mixed binary-ordered probit approach. *Applied Economics*, in press DOI: 10.1080./00036840701704402.

Kyriazidou E., 1997. Estimation of a panel data sample selection model. *Econometrica* 65 (6), 1335-1364.

Kyriazidou, E., 2001. Estimation of dynamic panel data sample selection models. *The Review of Economic Studies* 68 (3), 543-572.

Labeaga, J.M., 1999. A double-hurdle rational addiction model with heterogeneity: estimating the demand for tobacco. *Journal of Econometrics* 93, 49–72.

Lancaster, T., 2000. The incidental parameter problem since 1948. *Journal of Econometrics* 95, 391-413.

Li, K., 1998. Bayesian inference in a simultaneous equation model with limited dependent variables. *Journal of Econometrics* 85, 387-400.

Long, S.H. and S. Marquis, 1994. Universal health insurance and uninsured people: effects on use and cost. Report for Congress. NTIS order #PB95-109724.

Luft H., 1981. Health Maintenance Organizations: Dimensions of Performance. John Willey & Sons: New York, 1981.

Manning W.G., A. Leibowitz, G.A. Goldberg et al., 1984. A controlled trial of the effect of a prepaid group practice on use of services. *New Engl J Med* 310, 1505-1510.

Miller, R.H. and H.S. Luft, 1994. Managed care plan performance since 1980. *Journal of American Medical Association* 271, 1512-1519.

Miller, R.H. and H.S. Luft, 1997. Does managed care lead to better or worse quality of care? *Health Afairs* 16, 7-25.

Miller, R.H. and H.S. Luft, 2002. HMO plan performance update: An analysis of the literature, 1997-2001. *Health Affairs* 21 (4), 63-86.

Montalvo, J.G., 1997. GMM estimation of count panel data models with fixed effects and predetermined instruments. *Journal of Business and Economic Statistics* 15 (1), 82-89.

Mullahy, J., 1985. Cigarette smoking, habits, health concerns, and heterogeneous unobservables in a microeconometric analysis of consumer demand, Ph.D. dissertation, University of Virginia, 1985.

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

Mullahy, J., 1997. Instrumental-variable estimation of count data models: applications to models of cigarette smoking behavior. *The Review of Economic and Statistics* 79, 586-593.

Mullahy, J., 1998. Much ado about two: reconsidering retransformation and the two-part model in health econometrics. *Journal of Health Economics* 17, 247-281.

Munkin, M.K. and P.K. Trivedi, 2003. Bayesian analysis of self-selection model with multiple outcomes using simulation-based estimation: an application to the demand for healthcare. *Journal of Econometrics* 114, 197-220.

Munkin, M.K. and P.K. Trivedi, 2008. Bayesian analysis of the ordered probit model with endogenous selection. *Journal of Econometrics* 143, 334-348.

Neyman, J. and E.L. Scott, 1948. Consistent estimation from partially consistent observations. *Econometrica* 16, 1-32.

Pohlmeier, W. and V. Ulrich, 1994. An econometric model of the two-part decision-making process in the demand for health care. *The Journal of Human Resources* XXX2, 339-361.

Reschovsky, J.D., 2000. Do HMOs make a difference? Data and methods. *Inquiry* 36, 378-389.

Reschovsky, J.D. and P. Kemper, 2000. Do HMOs make a difference? Introduction. *Inquiry* 36, 374-377.

Rosenbaum, P. and D. Rubin, 1983. The central role of the propensity score in observational studies for causal effects. *Biometrika* 70(1):41-55.

Rosenbaum, P. and D. Rubin, 1985. Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician* 39 (1), 33-38.

Rubin, D.B. and N. Thomas, 1996. Matching using estimated propensity scores: relating theory to practice. *Biometrics* 52, 249-264.

Santos Silva, J.M.C. and F.A.G. Windmeijer, 2001. Two-part multiple spell models for health care demand. *Journal of Econometrics* 104, 25-50.

Semykina, A. and M. Wooldridge, 2005. Estimating panel data models in the presence of endogeneity and selection: theory and application. Working paper.

Tanner, A.M. and W.H., Wong, 1987. The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association* 82, (398), 528-540.

Tansel, A., 1993. Cigarette demand, health scares and education in Turkey. *Applied Economics* 25, 521–29.

Terza, J.V., 1998. Estimating count data models with endogenous switching: sample selection and endogenous treatment effects. *Journal of Econometrics* 84, 129-154.

Tobin, J., 1958. Estimation of relationships for limited dependent variables. *Econometrica* 26, 24–36.

US Department of Agriculture, 2000. Continuing Survey of Food Intakes by Individuals 1994–96 and 1998, CD-ROM, Agricultural Research Service, Washington, DC.

Van Ophem, H., 2000. Modeling selectivity in count-data models. *Journal of Business & Economic Statistics* 18 (4), 503-511.

Vella, F. and M. Verbeek, 1999. Two-step estimation of panel data models with censored endogenous variables and selection bias. *Journal of Econometrics* 90, 239-263.

Verdinelli, I. and L. Wasserman, 1995. Computing Bayes factors using a generalization of the Savage-Dickey density ratio. *Journal of the American Statistical Association* 90 (430), 614-618.

Wales, T.J. and A.D. Woodland, 1980. Sample selectivity and the estimation of labor supply functions. *International Economic Review* 21, 437–68.

Winkelmann, R., 2004. Health care reform and the number of doctor visits – an econometric analysis. *Journal of Applied Econometrics* 19, 455-472.

Windmeijer, F.A.G. and J.M.C., Santos Silva, 1997. Endogeneity in count data models: an application to demand for health care. *Journal of Applied Econometrics* 12, 281-294.

Windmeijer, F., 2000. Moment conditions for fixed effects count data models with endogenous regressors. *Economics Letters* 68, 21-24.

Windmeijer, F., 2006. GMM for panel count data models. Discussion paper No. 06/591.

Wooldridge, J.M., 1991. Multiplicative panel data models without the strict exogeneity assumption. MIT Working paper No. 574.

Wooldridge, J.M., 1995. Selection corrections for panel data models under conditional mean independence assumptions. *Journal of Econometrics* 68, 115-132.

Wooldridge, J.M., 1997. Multiplicative panel data models without the strict exogeneity assumption. *Econometric Theory* 13, 667-678.

Yen, S.T., 2005. Zero observations and gender differences in cigarette consumption. *Applied Economics* 37, 1839–49.

Appendices

Appendix A

Tables — Chapter 1

Table A.1: Numerical Example

Coefficient	True Value	Unrestricted N=500	Unrestricted N=5000	ACF(2)	ACF(5)	Restricted N=5000
beta1	-0.25	-0.30	-0.25	0.54	0.05	0.04
		(0.13)	(0.05)			(0.06)
beta2	1	0.95	1.06	0.37	-0.02	0.93
		(0.09)	(0.03)			(0.06)
gamma1	1	1.36	1.06	0.49	0.09	1.42
		(0.14)	(0.05)			(0.07)
gamma2	2	1.88	1.94	0.46	0.12	1.99
		(0.07)	(0.02)			(0.05)
alpha1	0	0.01	0.05	0.06	-0.02	-0.02
		(0.06)	(0.02)			(0.05)
alpha2	1	0.99	0.96	0.26	-0.01	1.01
		(0.08)	(0.03)			(0.06)
pi1	-0.5	-0.35	-0.50	0.60	0.06	
		(0.16)	(0.06)			
rho1	0.5	0.26	0.46	0.57	0.03	-0.05
		(0.21)	(0.08)			(0.09)
pi2	-0.5	-0.43	-0.46	0.42	0.01	
		(0.12)	(0.04)			
rho2	0.5	0.27	0.50	0.33	0.01	-0.04
		(0.17)	(0.06)			(0.08)
sigma	0.5	0.54	0.53	0.30	-0.02	0.67
		(0.08)	(0.03)			(0.05)

Note: Standard deviations in parentheses.

Table A.2: Utilization by Insurance Status

Outpatient visits (85.07% zeroes)		
	MC	FFS
Mean	0.406	0.415
St Dev	2.528	2.454
Cells	%	%
0	85.63	84.12
1	8.58	9.28
2	2.57	3.15
3	1.14	1.15
4	0.47	0.66
≥ 5	1.61	1.64

Table A.3: Summary of the Data

Variable	Definition	Mean	St. Dev.
OPPVIS	Number of outpatient physician visits	0.19	1.36
FAMSIZE	Family size	3	1.43
AGE	Age in years (divided by 10)	4.41	1.06
EDUCYR	Completed years of education	13.55	2.57
MARRIED	Equals 1 if the person is married	0.74	0.43
FEMALE	Equals 1 if the person is female	0.56	0.5
BLACK	Equals 1 if the person is African American	0.1	0.31
HISPANIC	Equals 1 if the person is hispanic	0.14	0.34
NOREAST	Equals 1 if the person lives in Northeastern US	0.18	0.39
MIDWEST	Equals 1 if the person lives in Midwest US	0.24	0.43
SOUTH	Equals 1 if the person lives in South US	0.36	0.48
MSA	Equals 1 if the person is found in a metropolitan area	0.8	0.4
VEGOOD	Equals1 if self perceived health is very good	0.36	0.48
GOOD	Equals1 if self perceived health is good	0.28	0.45
FAIRPOOR	Equals1 if self perceived health is fair or poor	0.1	0.3
NUMCHRON	number of chronic conditions	0.82	1.37
NUMINJ	number of injuries	0.26	0.66
INCOME	income in $1000	34.9	28.25
MANAGED	Equals 1 if the person is covered by an HMO or a PPO plan	0.36	0.48
SPOUAGE	Spouse's age (divided by 10)	2.75	2.39
SPOUPRIV	Equals 1if spouse had private insurance last year	0.58	0.49
YEAR DUMMIES	dummies for 98, 99, 00, 01, 02		

Table A.4: MCMC Estimates

Variable	Hurdle	Mean	Insurance	Variable	Hurdle	Mean	Insurance
CONST	-1.043	-0.301	0.001	VEGOOD	0.067	0.194	0.026
	(2.267)	(2.277)	(2.238)		(0.036)	(0.084)	(0.024)
FAMSIZE	-0.041	0.006	0.007	GOOD	0.162	0.350	0.055
	(0.013)	(0.024)	(0.009)		(0.039)	(0.08)	(0.027)
AGE	-0.016	0.065	-0.125	FAIRPOOR	0.381	0.552	0.105
	(0.014)	(0.221)	(0.118)		(0.058)	(0.102)	(0.046)
AGE2	0.019	-0.005	0.008	NUMCHRON	0.160	0.091	-0.001
	(0.014)	(0.025)	(0.013)		(0.016)	(0.019)	(0.009)
EDUCYR	0.011	0.028	-0.008	NUMINJ	0.195	0.063	-0.003
	(0.006)	(0.012)	(0.004)		(0.023)	(0.026)	(0.016)
MARRIED	-1.524	-0.355	0.032	INCOME	0.000	-0.001	-0.001
	(2.224)	(2.273)	(2.226)		(0.001)	(0.001)	(0.001)
FEMALE	0.412	-0.039	0.037	SPOUAGE			0.214
	(0.036)	(0.055)	(0.002)				(0.010)
BLACK	-0.376	0.049	0.168	SPOUAGE2			-0.026
	(0.080)	(0.118)	(0.036)				(0.011)
HISPANIC	-0.218	-0.099	0.151	SPOUPRIV(-1)			0.090
	(0.063)	(0.096)	(0.032)				(0.033)
NOREAST	0.359	0.092	0.023	π_1	-0.586		
	(0.050)	(0.093)	(0.032)		(0.354)		
MIDWEST	0.542	0.116	-0.400	ρ_1	0.894		
	(0.125)	(0.095)	(0.029)		(0.555)		
SOUTH	0.259	-0.110	-0.264	π_2		0.022	
	(0.083)	(0.090)	(0.028)			(0.155)	
MSA	-0.310	-0.077	0.502	ρ_2		0.015	
	(0.127)	(0.080)	(0.024)			(0.253)	

Note: Standard deviations in parentheses.

Table A.5: Average Treatment Effects for the Treated

	excellent	very good	good	fair or poor
mean	0.064	0.086	0.116	0.187
st. dev.	0.040	0.059	0.073	0.112

Appendix B

Tables — Chapter 2

Table B.1: Numerical example

Parameter	True value	Unrestricted (N=500)	Unrestricted (N=5000)	Restricted (N=5000)
α_1	2	2.16	2.01	2.07
		(0.13)	(0.04)	(0.05)
α_2	1	1.09	1.02	1.03
		(0.09)	(0.03)	(0.03)
β_1	1	1.10	1.00	0.40
		(0.18)	(0.08)	(0.05)
β_2	1	1.01	1.01	1.01
		(0.03)	(0.01)	(0.01)
ρ	-0.5	-0.57	-0.49	0.14
		(0.19)	(0.08)	(0.05)
δ	0.5	0.56	0.52	
		(0.11)	(0.05)	
σ_{11}	0.2	0.44	0.22	0.28
		(0.13)	(0.04)	(0.04)
σ_{22}	0.2	0.19	0.22	0.19
		(0.03)	(0.02)	(0.01)
σ_{12}	0.1	0.15	0.09	0.12
		(0.04)	(0.02)	(0.02)
σ_u	0.6	0.61	0.59	0.82
		(0.11)	(0.05)	(0.02)

Note: Standard deviations in parentheses.

Table B.2: Source of data for entry cohorts in RAND HRS data file by wave.

Wave	Entry Cohort			
	HRS (1931-1941)	AHEAD (-1923)	CODA (1924-1930)	WB (1942-1947)
1	1992	1992 (HRS/AHEAD overlaps only)	N/A	N/A
2	1994	1993	N/A	N/A
3	1996	1995	N/A	N/A
4	1998	1998	1998	1998
5	2000	2000	2000	2000
6	2002	2002	2002	2002

Table B.3: Summary statistics.

Variable	Definition	Mean (St. Dev.)			
		1996	1998	2000	2002
DRVIS	Number of doctor visits	7.07 (8.39)	7.40 (8.66)	7.77 (8.61)	8.33 (9.18)
AGE	Age in years (divided by 10)	5.94 (0.32)	6.12 (0.32)	6.32 (0.32)	6.53 (0.32)
EDUCYR	Completed years of education	12.33 (3.11)	12.33 (3.11)	12.33 (3.11)	12.33 (3.11)
GENDER	=1 if person is male	0.45 (0.50)	0.45 (0.50)	0.45 (0.50)	0.45 (0.50)
BLACK	=1 if race is Black	0.14 (0.35)	0.14 (0.35)	0.14 (0.35)	0.14 (0.35)
HISPANIC	=1 if person is Hispanic	0.08 (0.28)	0.08 (0.28)	0.08 (0.28)	0.08 (0.28)
MARRIED	=1 if person is married	0.74 (0.44)	0.72 (0.45)	0.71 (0.46)	0.69 (0.46)
EXCHLTH	=1 if in excellent health	0.20 (0.40)	0.14 (0.34)	0.14 (0.35)	0.13 (0.33)
VGHLTH	=1 if in very good health	0.34 (0.47)	0.30 (0.46)	0.33 (0.47)	0.32 (0.47)
GOODHLTH	=1 if in good health	0.28 (0.45)	0.33 (0.47)	0.32 (0.46)	0.33 (0.47)
INC	Household income (categorical)	1.21 (0.71)	1.21 (0.71)	1.21 (0.76)	1.19 (0.74)
WEALTH	Total household wealth (categorical)	4.13 (1.72)	4.21 (1.77)	4.35 (1.83)	4.41 (1.85)
INS	=1 if individual has any insurance coverage	0.89 (0.32)	0.90 (0.30)	0.93 (0.25)	0.96 (0.21)
THRES	=1 if age$\geq$ 65	0.05 (0.21)	0.20 (0.40)	0.37 (0.48)	0.57 (0.50)

Table B.4: Categorical income and wealth.

Household income	Categorical income (INC)	Total household wealth	Categorical wealth (WEALTH)
$<$ \$10,000	1	$<$ \$0	1
[\$10,000, \$20,000)	2	[\$0, \$300)	2
[\$20,000, \$30,000)	3	[\$300, \$1000)	3
[\$30,000, \$40,000)	4	[\$1000, \$2000)	4
[\$40,000, \$50,000)	5	[\$2000, \$3000)	5
[\$50,000, \$60,000)	6	[\$3000, \$5000)	6
[\$60,000, \$70,000)	7	$\geq$ \$5,000	7
[\$70,000, \$80,000)	8		
[\$80,000, \$90,000)	9		
[\$90,000, \$100,000)	10		
[\$100,000, \$120,000)	11		
[\$120,000, \$150,000)	12		
[\$150,000, \$200,000)	13		
$\geq$ \$200,000	14		

Table B.5: Utilization of Dr visits by insurance.

	PERIOD 1		PERIOD 2		PERIOD 3		PERIOD 4	
	INS	UNINS	INS	UNINS	INS	UNINS	INS	UNINS
Mean	7.19	6.21	7.62	5.35	7.90	6.01	8.49	4.89
St. Dev.	8.27	9.22	8.70	8.02	8.58	8.86	9.23	7.17
% of zeroes	7.10	21.75	7.11	21.37	5.96	22.10	6.07	24.79

Table B.6: Differences in utilization and insurance by group.

	OBS	DRVIS	DOC	INS
Population	21180	7.64	0.92	0.92
		(8.73)	(0.27)	(0.27)
Population$\geq$ 65	6272	8.17	0.93	0.99
		(8.95)	(0.25)	(0.09)
Population$<$ 65	14908	7.42	0.92	0.89
		(8.62)	(0.27)	(0.31)
Educated white	13596	7.31	0.93	0.95
		(8.11)	(0.25)	(0.21)
Educated white$\geq$ 65	3983	8.00	0.95	0.996
		(8.46)	(0.23)	(0.06)
Educated white$<$ 65	9613	7.03	0.93	0.94
		(7.94)	(0.25)	(0.25)
Minorities	2256	8.33	0.88	0.76
		(10.02)	(0.33)	(0.42)
Minorities$\geq$ 65	686	8.86	0.88	0.96
		(10.54)	(0.33)	(0.19)
Minorities$<$ 65	1570	8.11	0.88	0.68
		(9.78)	(0.33)	(0.47)

Note: Standard deviations in parentheses.

Table B.7: MCMC estimates.

	Insurance	Mean
AGE	0.149	0.211
	(0.023)	(0.010)
GENDER	0.103	-0.224
	(0.064)	(0.021)
MARRIED	0.186	0.044
	(0.066)	(0.022)
BLACK	-0.350	0.048
	(0.087)	(0.033)
HISPANIC	-0.843	-0.001
	(0.103)	(0.040)
EDUCYR	0.106	0.042
	(0.012)	(0.004)
EXCHLTH	0.104	-0.920
	(0.100)	(0.026)
VGHLTH	0.134	-0.671
	(0.075)	(0.020)
GOODHLTH	0.079	-0.379
	(0.061)	(0.019)
INC	0.088	0.024
	(0.052)	(0.010)
WEALTH	0.189	-0.005
	(0.020)	(0.006)
YEAR96	-0.911	0.009
	(0.060)	(0.015)
YEAR98	-0.879	-0.046
	(0.058)	(0.015)
YEAR00	-0.737	0.003
	(0.067)	(0.014)
THRES	1.790	
	(0.122)	
ρ		0.328
		(0.044)
δ		-0.042
		(0.021)
σ_{11}	2.248	
	(0.133)	
σ_{22}		0.419
		(0.013)
σ_u		0.336
		(0.006)
σ_{12}	0.197	
	(0.028)	

Note: Standard deviations in parentheses.

Table B.8: Average treatment effects for the treated.

ATET1	ATET2	ATET3	ATET4
1.92	2.04	2.12	2.27
(2.06)	(2.19)	(2.17)	(2.33)

Note: Standard deviations in parentheses.

Table B.9: Average treatment effects.

ATE1	ATE2	ATE3	ATE4
1.88	1.98	2.07	2.23
(2.08)	(2.18)	(2.16)	(2.31)

Note: Standard deviations in parentheses.

Table B.10: Health variables by insurance status.

Mean	PERIOD 1		PERIOD 2		PERIOD 3		PERIOD 4	
	INS	UNINS	INS	UNINS	INS	UNINS	INS	UNINS
Excellent health	0.21	0.13	0.14	0.10	0.14	0.10	0.12	0.16
Very good health	0.35	0.26	0.31	0.22	0.34	0.25	0.32	0.25
Good health	0.28	0.32	0.33	0.32	0.32	0.33	0.33	0.34
Fair or poor health	0.16	0.29	0.37	0.20	0.32	22.10	0.23	0.25

Appendix C

Tables — Chapter 3

Table C.1: Summary statistics

Variable	Definition	Mean		
		Pooled (N=9587)	Male (N=4923)	Female (N=4664)
Cigarettes	Number per day (full sample)	4.50 (9.98)	5.19 (11.04)	3.76 (8.68)
	Number per day (consuming sample)	18.64 (12.23)	20.29 (13.03)	16.66 (10.89)
	Proportion of smokers	24.12%	25.59%	22.58%
	Proportion ever started smoking	49.47%	56.65%	41.90%
	Proportion quitters	25.35%	31.06%	19.32%
Income	Per capita income in thousand USD	15.29 (12.54)	15.84 (13.02)	14.70 (11.98)
Body mass	Quetelet's body mass index[a]	26.06 (5.25)	26.26 (4.55)	25.84 (5.90)
Education	Years of formal education	12.50 (3.09)	12.56 (3.18)	12.43 (2.99)
Age	Age in years	47.19 (18.95)	47.42 (18.97)	46.95 (18.93)
Dummy variables (yes = 1, no = 0)				
Male	Gender is male	0.51		
City	Resides in central city	0.29	0.28	0.31
Suburban	Resides in suburban area	0.45	0.46	0.43
Rural	Resides in rural area (reference)	0.26	0.26	0.26
Northeast	Resides in the North or Northeast	0.18	0.18	0.18
Midwest	Resides in the Midwest	0.24	0.24	0.25
South	Resides in the South	0.37	0.36	0.37
West	Resides in the West (reference)	0.21	0.22	0.20
Black	Race is Black	0.12	0.10	0.13
Hispanic	of Hispanic origin	0.04	0.04	0.04
Healthy	Self-evaluated health is fair or better	0.83	0.84	0.82
White-collar	A white-collar worker	0.23	0.25	0.21
Cancer	Has been diagnosed of cancer	0.06	0.06	0.06
BP-heart	Has had blood pressure/heart problems	0.27	0.28	0.27
Alcohol	Consumed alcohol in past 3 months	0.62	0.67	0.56
Intensive exercise	Exercises 2–4 times per week or more	0.50	0.57	0.42
Moderate exercise	Exercises 1–4 times per month (reference)	0.12	0.12	0.14
No exercise	Exercises rarely or never	0.38	0.31	0.44
Diet	On a special diet	0.17	0.14	0.20
Year94	Survey conducted in 1994 (reference)	0.33	0.33	0.34
Year95	Survey conducted in 1995	0.34	0.34	0.34
Year96	Survey conducted in 1996	0.33	0.33	0.32

Note: Standard deviations in parentheses. [a]Quetelet's body mass index=(weight in kg)/(height in metres).

Table C.2: ML estimates for binary part: pooled sample

Variable	Starting (N=9587)	Not quitting[c] (N=4743)
Constant	-0.501[a]	3.163[a]
	(0.116)	(0.189)
Income	-0.0002	-0.005[a]
	(0.0012)	(0.002)
Body mass	0.001	-0.027[a]
	(0.003)	(0.004)
Education	-0.031[a]	-0.065[a]
	(0.005)	(0.008)
Age	0.012[a]	-0.028[a]
	(0.001)	(0.001)
City	-0.022	-0.157[a]
	(0.037)	(0.055)
Suburban	-0.044	-0.181[a]
	(0.037)	(0.049)
Northeast	-0.004	-0.208[a]
	(0.044)	(0.064)
Midwest	-0.014	-0.011
	(0.041)	(0.059)
South	-0.014	0.016
	(0.038)	(0.056)
Black	-0.125[a]	0.289[a]
	(0.044)	(0.068)
Hispanic	-0.471[a]	-0.287[a]
	(0.076)	(0.133)
Healthy	-0.223[a]	-0.086[b]
	(0.039)	(0.052)
White-collar	-0.103[a]	-0.109[a]
	(0.036)	(0.054)
Cancer	0.107[b]	-0.108
	(0.060)	(0.075)
BP-heart	-0.026	-0.047
	(0.035)	(0.048)
Male	0.360[a]	-0.158[a]
	(0.027)	(0.041)
Alcohol	0.568[a]	0.137[a]
	(0.030)	(0.046)
Intensive exercise	-0.028	-0.074
	(0.042)	(0.066)
No exercise	0.144[a]	0.106
	(0.044)	(0.068)
Diet	-0.025	-0.260[a]
	(0.038)	(0.054)
Year95	-0.044	0.112[a]
	(0.033)	(0.048)
Year96	-0.005	0.069
	(0.032)	(0.048)

Note: Standard errors in parentheses.
[a,b] Denote significance at the 5% and 10% levels respectively.
[c] We label the second column as "not quitting" to be consistent with definition of the quitting equation in the text (Equation 5).

Table C.3: ML estimates for conditional part: pooled sample

	Level (N=2313)				
Variable	Gaussian	NB1	NB2	OP	SOP
Constant	0.322	2.476^a	2.384^a	0.648^a	0.683^a
	(3.247)	(0.123)	(0.125)	(0.208)	(0.161)
Income	0.058	0.001	0.001	0.002	0.002
	(0.040)	(0.001)	(0.001)	(0.002)	(0.002)
Body mass	0.343^a	0.011^a	0.013^a	0.018^a	0.014^a
	(0.082)	(0.003)	(0.003)	(0.004)	(0.003)
Education	-0.214	-0.005	-0.003	-0.014	-0.011
	(0.170)	(0.006)	(0.006)	(0.010)	(0.007)
Age	0.088^a	0.002^a	0.004^a	0.004^a	0.004^a
	(0.029)	(0.001)	(0.001)	(0.002)	(0.001)
City	-4.419^a	-0.125^a	-0.134^a	-0.233^a	-0.182^a
	(1.095)	(0.036)	(0.035)	(0.060)	(0.045)
Suburban	-1.377	-0.024	-0.027	-0.042	-0.030
	(0.934)	(0.032)	(0.031)	(0.055)	(0.041)
Northeast	-1.602	-0.017	-0.066	-0.083	-0.098
	(1.485)	(0.050)	(0.048)	(0.078)	(0.060)
Midwest	3.788^a	0.174^a	0.129^a	0.261^a	0.152^a
	(1.272)	(0.044)	(0.041)	(0.068)	(0.050)
South	4.586^a	0.208^a	0.152^a	0.331^a	0.218^a
	(1.162)	(0.041)	(0.038)	(0.064)	(0.048)
Black	-12.677^a	-0.366^a	-0.415^a	-0.652^a	-0.526^a
	(1.445)	(0.041)	(0.043)	(0.068)	(0.057)
Hispanic	-20.782^a	-0.635^a	-0.707^a	-1.140^a	-0.909^a
	(4.247)	(0.100)	(0.122)	(0.170)	(0.163)
Healthy	-1.290	0.001	-0.040	-0.059	-0.070
	(1.048)	(0.037)	(0.035)	(0.063)	(0.047)
White-collar	-3.051^a	-0.098^a	-0.101^a	-0.165^a	-0.117^a
	(1.137)	(0.038)	(0.038)	(0.063)	(0.047)
Cancer	0.316	0.021	0.004	-0.025	-0.027
	(1.780)	(0.054)	(0.060)	(0.095)	(0.072)
BP-heart	-0.703	-0.038	-0.019	-0.055	-0.029
	(1.012)	(0.036)	(0.034)	(0.060)	(0.045)
Male	6.337^a	0.182^a	0.219^a	0.345^a	0.278^a
	(0.830)	(0.027)	(0.027)	(0.046)	(0.035)
Alcohol	-0.337	-0.027	-0.020	-0.044	-0.021
	(0.886)	(0.030)	(0.031)	(0.052)	(0.041)
Intensive exercise	-2.096	-0.039	-0.049	-0.095	-0.022
	(1.317)	(0.047)	(0.044)	(0.074)	(0.055)
No exercise	1.594	0.094^a	0.058	0.108	0.110^a
	(1.329)	(0.047)	(0.044)	(0.075)	(0.056)
Diet	-1.760	-0.070	-0.068	-0.101	-0.059
	(1.303)	(0.046)	(0.043)	(0.073)	(0.055)
Year95	0.211	0.017	0.016	-0.030	-0.014
	(0.887)	(0.032)	(0.031)	(0.054)	(0.041)
Year96	-0.681	-0.017	-0.009	-0.055	-0.044
	(0.032)	(0.033)	(0.032)	(0.053)	(0.041)
σ	14.562^a				
	(0.581)				
α		6.492^a	0.361^a		
		(0.263)	(0.015)		
τ_2				0.701^a	0.365^a

continued on next page

Table C3: Continued

	Level (N=2313)				
Variable	Gaussian	NB1	NB2	OP	SOP
				(0.030)	(0.046)
τ_3				0.995[a]	0.134[a]
				(0.034)	(0.052)
τ_4				1.944[a]	1.421[a]
				(0.043)	(0.050)
τ_5				2.446[a]	1.267[a]
				(0.048)	(0.066)
Log-likelihood[e]	-17600.97	-17621.43	-17632.10	-12712.14	-39458.79
AIC	3.686	3.691	3.693	2.667	8.247

Note: Standard errors in parentheses.
[a,b] Denote significance at the 5% and 10% levels respectively.
[d]OP and SOP stand for Ordered Probit and Sequential Ordered Probit respectively.
[e]The log-likelihood values and AIC reported at the bottom of each model correspond to the trivariate models.

Table C.4: Mean marginal effects and average discrete changes: pooled sample

Variable	Mean	Probabilities					
		$m = 1$	$m = 2$	$m = 3$	$m = 4$	$m = 5$	$m = 6$
		Marginal effects of continuous variables					
Income	-0.033[a]	0.00030	-0.00023[a]	-0.00010[a]	-0.00025[a]	-0.00005	-0.00054
	(0.018)	(0.00044)	(0.00010)	(0.00004)	(0.00012)	(0.00007)	(0.00046)
Body mass	-0.209[a]	0.00278[a]	-0.00131[a]	-0.00055[a]	-0.00110[a]	-0.00011	-0.00399[a]
	(0.036)	(0.00085)	(0.00021)	(0.00009)	(0.00027)	(0.00014)	(0.00088)
Education	-0.208[a]	-0.00507[a]	-0.00269[a]	-0.00159[a]	-0.00564[a]	-0.00230[a]	0.00077
	(0.080)	(0.00189)	(0.00046)	(0.00021)	(0.00058)	(0.00031)	(0.00196)
Age	-0.077[a]	0.00057[b]	-0.00051[a]	-0.00025[a]	-0.00070[a]	-0.00021[a]	-0.00117[a]
	(0.014)	(0.00032)	(0.00007)	(0.00004)	(0.00009)	(0.00005)	(0.00035)
		Average effects of discrete variables					
City	-1.094[a]	0.00759[a]	0.00114	-0.00171	-0.01448[a]	-0.00900[a]	-0.01286[a]
	(0.250)	(0.00356)	(0.00271)	(0.00120)	(0.00387)	(0.00200)	(0.00278)
Suburban	-0.826[a]	-0.00312	-0.00577[a]	-0.00355[a]	-0.01319[a]	-0.00563[a]	-0.00669[a]
	(0.240)	(0.00315)	(0.00258)	(0.00114)	(0.00352)	(0.00186)	(0.00288)
Northeast	-0.841[a]	-0.00086	-0.00426	0.00310[a]	-0.01303[a]	-0.00604[a]	-0.00746[a]
	(0.294)	(0.00402)	(0.00322)	(0.00143)	(0.00457)	(0.00225)	(0.00317)
Midwest	0.552[b]	-0.01242[a]	-0.00871[a]	-0.00242[b]	0.00224	0.00551[a]	0.01157[a]
	(0.311)	(0.00318)	(0.00294)	(0.00137)	(0.00439)	(0.00232)	(0.00380)
South	0.822[a]	-0.01562[a]	-0.01002[a]	-0.00239[b]	0.00535	0.00792[a]	0.01528[a]
	(0.285)	(0.00327)	(0.00278)	(0.00126)	(0.00406)	(0.00222)	(0.00349)
Black	-1.166[a]	0.04418[a]	0.02072[a]	0.00342[a]	-0.01394[a]	-0.01358[a]	-0.02010[a]
	(0.246)	(0.00649)	(0.00370)	(0.00151)	(0.00468)	(0.00201)	(0.00206)
Hispanic	-3.425[a]	0.02796[a]	-0.01457[a]	-0.01430[a]	-0.06060[a]	-0.02591[a]	-0.02823[a]
	(0.249)	(0.01106)	(0.00480)	(0.00226)	(0.00555)	(0.00199)	(0.00179)
Healthy	-1.227[a]	-0.00600	-0.00947[a]	-0.00555[a]	-0.01983[a]	-0.00821[a]	-0.00954[a]
	(0.317)	(0.00387)	(0.00320)	(0.00133)	(0.00439)	(0.00256)	(0.00399)
White-collar	-1.031[a]	0.00224	-0.00270	-0.00290[a]	-0.01493[a]	-0.00782[a]	-0.01040[a]
	(0.236)	(0.00348)	(0.00262)	(0.00113)	(0.00363)	(0.00189)	(0.00265)
Cancer	-0.052	0.00192	0.00101	0.00020	-0.00052	-0.00061	-0.00096
	(0.429)	(0.00572)	(0.00462)	(0.00202)	(0.00637)	(0.00326)	(0.00497)
BP-heart	-0.365	0.00091	-0.00081	-0.00095	-0.00514	-0.00277	-0.00381
	(0.248)	(0.00353)	(0.00268)	(0.00112)	(0.00359)	(0.00196)	(0.00300)
Male	1.579[a]	-0.00990[a]	-0.00176	0.00221[a]	0.02001[a]	0.01288[a]	0.01908[a]
	(0.197)	(0.00257)	(0.00217)	(0.00095)	(0.00304)	(0.00164)	(0.00240)
Alcohol	2.216[a]	0.02148[a]	0.02586[a]	0.01332[a]	0.04003[a]	0.01349[a]	0.01251[a]
	(0.200)	(0.00250)	(0.00228)	(0.00127)	(0.00319)	(0.00165)	(0.00241)
Intensive exercise	-0.533[b]	0.00207	-0.00045	-0.00110	-0.00710	-0.00416[b]	-0.00599
	(0.322)	(0.00425)	(0.00351)	(0.00155)	(0.00485)	(0.00248)	(0.00373)
No exercise	1.126[a]	0.00111	0.00517	0.00385[a]	0.01669[a]	0.00809[a]	0.01054[a]
	(0.371)	(0.00442)	(0.00378)	(0.00176)	(0.00547)	(0.00284)	(0.00429)
Diet	-1.114[a]	-0.00194	-0.00631[a]	-0.00434[a]	-0.01747[a]	-0.00788[a]	-0.00959[a]
	(0.262)	(0.00376)	(0.00277)	(0.00118)	(0.00387)	(0.00211)	(0.00305)

Note: Standard errors in parentheses.

[a,b] Denotes significance at the 5% and 10% levels respectively.

Table C.5: Likelihood-ratio tests for gender differences and AIC for male and female subsamples

Model	Pooled	Male		Female		LR	df	p-value
	Log-likelihood	Log-likelihood	AIC	Log-likelihood	AIC			
Gaussian	-17600.97 (70)	-9463.85 (67)	3.872	-8016.70 (67)	3.466	240.84	64	< 0.001
NB1	-17621.43 (70)	-9487.46 (67)	3.882	-8011.67 (67)	3.464	244.60	64	< 0.001
NB2	-17632.10 (70)	-9494.60 (67)	3.884	-8021.60 (67)	3.469	231.80	64	< 0.001
OP[a]	-12712.14 (73)	-6715.02 (70)	2.756	-5870.96 (70)	2.548	252.32	67	< 0.001
SOP	-39458.79 (73)	-21139.79 (70)	10.039	-18198.90 (70)	7.834	240.20	67	< 0.001

[a]OP and SOP stand for Ordered Probit and Sequential Ordered Probit respectively.

Table C.6: ML estimation of the ordered probit parts for male and female subsamples

	Male			Female		
	Starting	Not quitting	Level	Starting	Not quitting	Level
Variable	(N=4923)	(N=2789)	(N=1260)	(N=4664)	(N=1954)	(N=1053)
Constant	-0.334^b	3.238^a	0.518^b	-0.529^a	3.054^a	0.560^b
	(0.173)	(0.269)	(0.302)	(0.158)	(0.280)	(0.304)
Income	-0.001	-0.004^a	0.002	(0.002)	-0.007^a	0.003
	(0.002)	(0.002)	(0.003)	(0.002)	(0.003)	(0.003)
Body mass	0.005	-0.032^a	0.019^a	0.003	-0.026^a	0.023^a
	(0.004)	(0.006)	(0.007)	(0.003)	(0.005)	(0.007)
Education	-0.042^a	-0.055^a	-0.013	-0.017^a	-0.083^a	-0.002
	(0.008)	(0.010)	(0.013)	(0.008)	(0.014)	(0.015)
Age	0.019^a	-0.032^a	0.007^a	0.005^a	-0.021^a	0.003
	(0.001)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)
City	-0.061	-0.101	-0.261^a	0.049	-0.252^a	-0.155^b
	(0.053)	(0.073)	(0.081)	(0.053)	(0.085)	(0.089)
Suburban	-0.074	-0.132^a	-0.042	-0.012	-0.263^a	0.020
	(0.047)	(0.064)	(0.076)	(0.049)	(0.077)	(0.082)
Northeast	-0.010	-0.149^b	0.043	0.0004	-0.285^a	-0.247^a
	(0.062)	(0.084)	(0.105)	(0.063)	(0.097)	(0.116)
Midwest	-0.003	-0.019	0.298^a	0.001	0.025	0.253^a
	(0.057)	(0.079)	(0.094)	(0.059)	(0.091)	(0.099)
South	0.017	-0.020	0.380^a	-0.045	0.081	0.328^a
	(0.052)	(0.075)	(0.087)	(0.055)	(0.087)	(0.094)
Black	-0.117^b	0.316^a	-0.594^a	-0.146^a	0.236^a	-0.773^a
	(0.065)	(0.090)	(0.089)	(0.062)	(0.103)	(0.111)
Hispanic	-0.298^a	-0.394^a	-1.154^a	-0.713^a	-0.113	-1.101^a
	(0.100)	(0.162)	(0.205)	(0.125)	(0.248)	(0.291)
Healthy	-0.183^a	-0.150^a	0.076	-0.271^a	-0.025	-0.122
	(0.058)	(0.070)	(0.090)	(0.055)	(0.078)	(0.088)
White-collar	-0.131^a	-0.207^a	-0.139^b	-0.082	0.065	-0.269^a
	(0.050)	(0.072)	(0.084)	(0.053)	(0.083)	(0.095)
Cancer	-0.104	-0.152	-0.055	0.212^a	-0.055	-0.017
	(0.088)	(0.108)	(0.155)	(0.082)	(0.109)	(0.123)
BP-heart	0.033	-0.064	-0.039	-0.079	-0.009	-0.064
	(0.050)	(0.063)	(0.083)	(0.051)	(0.075)	(0.089)
Alcohol	0.493^a	0.202^a	-0.074	0.610^a	0.076	0.006
	(0.043)	(0.062)	(0.075)	(0.043)	(0.071)	(0.075)
Intensive exercise	-0.014	-0.042	-0.055	-0.00001	-0.140	-0.146
	(0.059)	(0.091)	(0.105)	(0.059)	(0.098)	(0.107)
No exercise	0.098	0.134	0.130	0.223^a	0.050	0.117
	(0.066)	(0.095)	(0.112)	(0.061)	(0.099)	(0.103)
Diet	-0.129^a	-0.372^a	-0.239^a	0.033	-0.155^a	-0.024
	(0.057)	(0.078)	(0.120)	(0.050)	(0.077)	(0.094)
Year95	-0.115^a	0.107^b	-0.006	0.055	0.121^b	-0.022
	(0.047)	(0.063)	(0.073)	(0.046)	(0.073)	(0.082)
Year96	-0.059	0.035	-0.069	0.057	0.120	0.032
	(0.046)	(0.064)	(0.072)	(0.047)	(0.075)	(0.081)
τ_2			0.572^a			0.841^a
			(0.039)			(0.047)
τ_3			0.854^a			1.140^a
			(0.044)			(0.051)
τ_4			1.778^a			2.148^a
			(0.055)			(0.065)

continued on next page

Table C5: Continued

	Male			Female		
Variable	Starting (N=4923)	Not quitting (N=2789)	Level (N=1260)	Starting (N=4664)	Not quitting (N=1954)	Level (N=1053)
τ_5			2.259[a]			2.699[a]
			(0.061)			(0.078)
Log-likelihood[c]			-6715.02			-5870.96
AIC			2.756			2.548

Note: Standard errors in parentheses.

[a,b] Denote significance at the 5% and 10% levels respectively.

[c]We report log-likelihood values and AIC for the trivariate models.

Table C.7: Mean marginal effects and average discrete changes

	Categorization 1		Categorization 2		
	Male	Female	Pooled	Male	Female
Mean Marginal Effects					
Income	-0.031	-0.033	-0.029^{b}	-0.029	-0.035^{b}
	(0.024)	(0.024)	(0.016)	(0.024)	(0.021)
Body mass	-0.247^{a}	-0.211^{a}	-0.197^{a}	-0.228^{a}	-0.179^{a}
	(0.056)	(0.049)	(0.034)	(0.054)	(0.046)
Education	-0.254^{a}	-0.244^{b}	-0.251^{a}	-0.319^{a}	-0.199^{a}
	(0.107)	(0.130)	(0.068)	(0.095)	(0.093)
Age	-0.110^{a}	-0.052^{a}	-0.076^{a}	-0.101^{a}	-0.047^{a}
	(0.021)	(0.019)	(0.013)	(0.019)	(0.015)
Average discrete changes					
City	-1.267^{a}	-0.744^{a}	-0.946^{a}	-1.113^{a}	-0.690^{a}
	(0.383)	(0.295)	(0.242)	(0.374)	(0.293)
Suburban	-0.862^{a}	-0.620^{a}	-0.811^{a}	-0.841^{a}	-0.634^{a}
	(0.359)	(0.280)	(0.241)	(0.335)	(0.290)
Northeast	-0.524	-1.108^{a}	-0.714^{a}	-0.324	-0.954^{a}
	(0.450)	(0.338)	(0.290)	(0.461)	(0.346)
Midwest	0.752^{b}	0.596	0.659^{a}	0.859^{b}	0.675^{b}
	(0.450)	(0.383)	(0.301)	(0.477)	(0.380)
South	1.007^{a}	0.750^{a}	0.913^{a}	1.121^{a}	0.777^{a}
	(0.434)	(0.368)	(0.279)	(0.421)	(0.356)
Black	-0.971^{a}	-1.407^{a}	-1.215^{a}	-1.060^{a}	-1.485^{a}
	(0.386)	(0.277)	(0.242)	(0.415)	(0.281)
Hispanic	-3.839^{a}	-3.006^{a}	-3.462^{a}	-3.805^{a}	-3.097^{a}
	(0.401)	(0.312)	(0.242)	(0.386)	(0.287)
Healthy	-0.974^{a}	-1.264^{a}	-1.147^{a}	-0.975^{a}	-1.281^{a}
	(0.453)	(0.376)	(0.296)	(0.430)	(0.371)
White-collar	-1.557^{a}	-0.630^{a}	-1.035^{a}	-1.494^{a}	-0.651^{a}
	(0.359)	(0.297)	(0.227)	(0.355)	(0.293)
Cancer	-1.012^{b}	0.492	-0.060	-0.950	0.474
	(0.543)	(0.490)	(0.406)	(0.598)	(0.496)
BP-heart	-0.239	-0.417	-0.330	-0.281	-0.298
	(0.370)	(0.300)	(0.239)	(0.360)	(0.302)
Alcohol	2.232^{a}	2.177^{a}	2.237^{a}	2.248^{a}	2.212^{a}
	(0.307)	(0.258)	(0.191)	(0.283)	(0.259)
Intensive exercise	-0.377	-0.650^{b}	-0.523^{b}	-0.315	-0.653^{b}
	(0.471)	(0.360)	(0.317)	(0.491)	(0.372)
No exercise	1.196^{a}	1.091^{a}	1.141^{a}	1.261^{a}	1.085^{a}
	(0.552)	(0.370)	(0.335)	(0.516)	(0.376)
Diet	-2.229^{a}	-0.307	-1.095^{a}	-2.228^{a}	-0.290
	(0.379)	(0.306)	(0.241)	(0.363)	(0.304)
Male			1.527^{a}		
			(0.189)		

Note: Standard errors in parentheses.

^{a,b} Denotes significance at the 5% and 10% levels respectively.

Appendix D

The trivariate model

Likelihood Function of the Dependent Trivariate Model

The dependent trivariate model is not estimated due to data limitation but here we present the likelihood function for the unrestricted (dependent) binary-binary-ordinal specification which can be used in future applications when data allow. Assume the error terms of the starting, quitting and level equations are distributed as standard trivariate normal with correlation matrix R, that is, $[\varepsilon_1', \varepsilon_2', u]' \sim N(0, R)$, where

$$R = \begin{bmatrix} 1 & \rho_{21} & \rho_{31} \\ \rho_{21} & 1 & \rho_{32} \\ \rho_{31} & \rho_{32} & 1 \end{bmatrix}.$$

Denote θ as a vector containing all parameters in the model, $\Phi_k(...; B)$ as the k-variate standard normal cdf ($k = 2, 3$) with correlation matrix (or scalar) B, and define diagonal matrix $W = diag[-1, -1, 1]$. Also, for generality we use vectors X_{1i}, X_{2i} and X_{3i} in the starting, quitting and level equations, respectively. Then the joint likelihood of observing the sample is

$$\begin{aligned} L(y|\theta) = \prod_{S_i=0} & [1 - \Phi(X_{1i}\alpha)] \prod_{S_i=1, Q_i=0} \Phi_2(X_{1i}\alpha_1, -X_{2i}\alpha_2, -\rho_{21}) \\ \times \prod_{S_i=1, Q_i=1} & \prod_{m=1}^{M} \left\{ \left[\Phi_3(X_{1i}\alpha_1, X_{2i}\alpha_2, \tau_m - X_{3i}\beta; WRW') \right.\right. \\ & \left.\left. - \Phi_3(X_{1i}\alpha_1, X_{2i}\alpha_2, \tau_{m-1} - X_{3i}\beta; WRW') \right] \right\}^{d_{im}}, \end{aligned}$$

where binary indicator d_{im} is defined in the text (Equation 16). It can be verified that by restricting all estimable correlation coefficients to zero, that is, $\rho_{21} = \rho_{31} = \rho_{32} = 0$, the likelihood function above reduces to that for the independent model presented in the text (Equations (3.3), (3.6) and (3.16)).

Marginal Effects for the Trivariate Ordered Probit with Independence
Differentiating the conditional mean of the dependent variable (Equation 24) and averaging the derivatives over the sample, the mean (average) marginal effect of the conditional mean with respect to the explanatory (continuous) variable X_{ik} is

$$\begin{aligned} MME &= \frac{1}{N}\sum_{i=1}^{N}\frac{\partial E\left(y_i \mid X_i\right)}{\partial X_{ik}} \\ &= \frac{1}{N}\sum_{i=1}^{N}\left\{\sum_{m=0}^{M}\frac{\partial\left[\overline{y}_m \Pr\left(y_i = m \mid X_i\right)\right]}{\partial X_{ik}}\right\} \\ &= \frac{1}{N}\sum_{i=1}^{N}\left\{\sum_{m=1}^{M}\overline{y}_m\frac{\partial\left[\Pr\left(y_i = m \mid S_i = 1, Q_i = 1\right)\Pr(S_i = 1)\Pr(Q_i = 1)\right]}{\partial X_{ik}}\right\}. \end{aligned}$$

Note that for $m = 0$, y_i is zero so that $\overline{y}_m = 0$.This simplifies our calculations since the term $\overline{y}_m \Pr\left(y_i = m \mid X_i\right)$ vanishes for $m = 0$. One can calculate the partial derivative with respect to X_{ik} as

$$\begin{aligned} &\frac{\partial\left[\Pr\left(y_i = m \mid S_i = 1, Q_i = 1\right)\Pr(S_i = 1)\Pr(Q_i = 1)\right]}{\partial X_{ik}} = \Pr(y_i = m|X_i) \\ &\times\left[\frac{\beta_k[\phi\left(\tau_{m-1} - X_i\beta\right) - \phi\left(\tau_m - X_i\beta\right)]}{\Phi\left(\tau_m - X_i\beta\right) - \Phi\left(\tau_{m-1} - X_i\beta\right)} + \frac{\alpha_{1k}\phi(X_i\alpha_1)}{\Phi(X_i\alpha_1)} + \frac{\alpha_{2k}\phi(X_i\alpha_2)}{\Phi(X_i\alpha_2)}\right], \end{aligned}$$

where $\Pr(y_i = m|X_i)$ is defined in the text (Equation 23).

Appendix E

Figures — Chapter 2

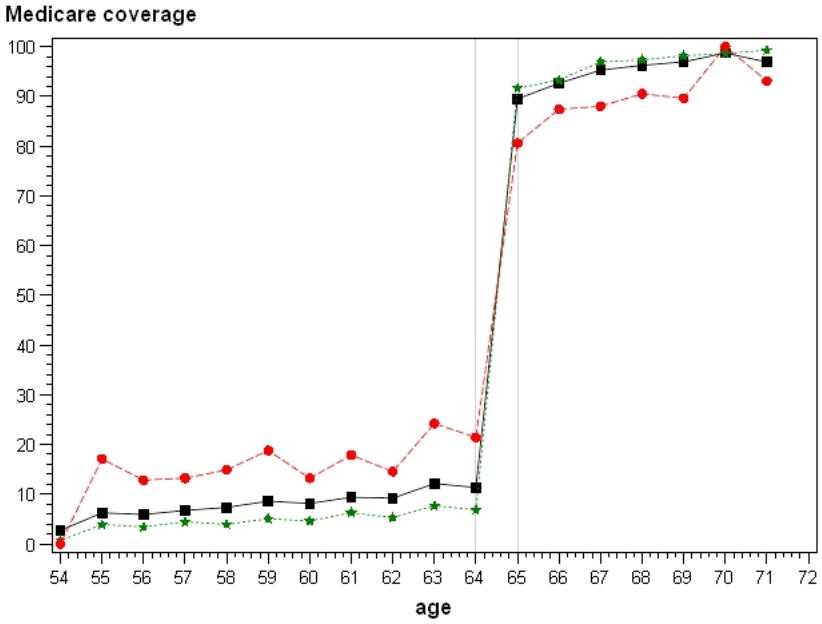

Figure E.1: Medicare coverage by age

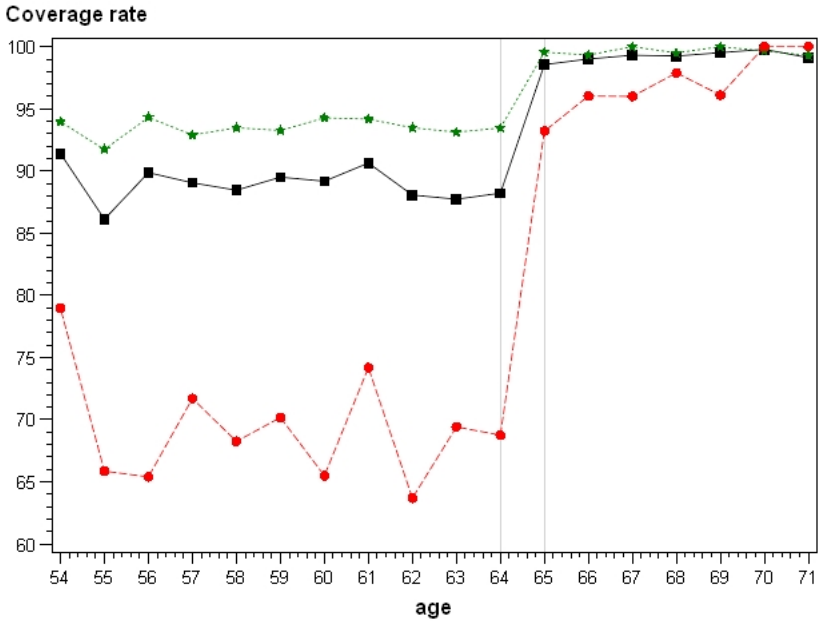

Figure E.2: Insurance coverage by age

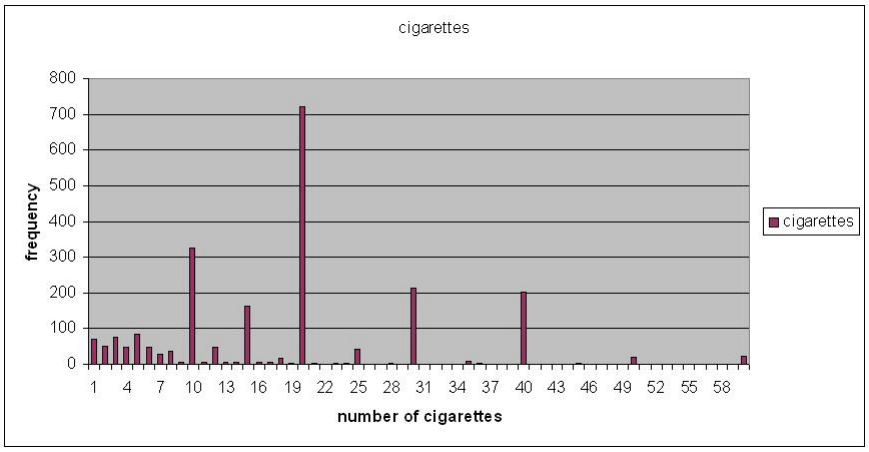

Figure E.3: Histogram of cigarettes smoked

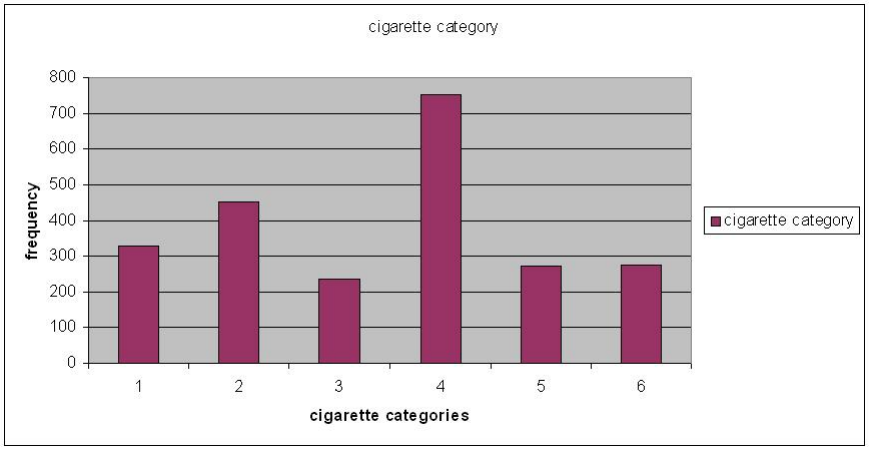

Figure E.4: Histogram of cigarette categories

Vita

Panagiotis Kasteridis was born in Thessaloniki, Greece on December 13, 1972. He received a Bachelor of Science degree in Mathematics from Aristotle University of Thessaloniki in 1997. After military service he entered the University of Macedonia, Greece and graduated with a Master of Science degree in economics in 2001. The same year he entered the graduate program in economics at The University of Tennessee, Knoxville and in 2003 he received a Master of Arts degree in economics. He specialized in econometrics and in 2007 he accepted a position as a research scientist with BlueCross BlueShield of Tennessee, where he is currently employed. He was awarded the Doctor of Philosophy degree with a major in economics in May, 2009.